Deep learning in neuroradiology

Resources and payment methods: Canada vs US

Reperfusion failure causes in stent-retriever thrombectomy
INDICATIONS FOR USE:

The LVIS® and LVIS® Jr. devices are indicated for use with neurovascular embolization coils in patients ≥ 18 years of age for the treatment of wide-neck (neck width ≥ 4 mm or dome to neck ratio < 2) saccular intracranial aneurysms arising from a parent vessel with a diameter ≥ 2.0 mm and ≤ 4.5 mm.

Rx Only: Federal (USA) law restricts this device to sale by or on the order of a physician.

The HydroCoil® Embolic System (HES) and MicroPlex® Coil System (MCS) are intended for the endovascular embolization of intracranial aneurysms and other neurovascular abnormalities such as arteriovenous malformations and arteriovenous fistulae. The HES and MCS are also intended for vascular occlusion of blood vessels within the neurovascular system to permanently obstruct blood flow to an aneurysm or other vascular malformation and for arterial and venous embolizations in the peripheral vasculature.

The device should only be used by physicians who have undergone pre-clinical training in all aspects of HES/MCS procedures as prescribed by MicroVention.
A Complete Coil Portfolio
Hydrogel Embolic Coils & Platinum Coils

MicroVention’s comprehensive portfolio features clinically proven Hydrogel coils, which can be used exclusively or in combination with our trusted Platinum coils to treat a wide range of aneurysms and neurovascular lesions.

Breakthrough Hydrogel Technology

- Less Recurrence
- Less Retreatment
- More Progressive Occlusion

*Compared to platinum coils with comparable safety*¹

REFERENCES:

1. Taschner et al. Second-generation Hydrogel Coils for the Endovascular Treatment of Intracranial Aneurysm: A Randomized Controlled Trial. 2018;43:00-00. DOI:10.1161/STRUKAHA.117.018707

For more information or a product demonstration, contact your local MicroVention representative:
MAGIC catheters are designed for general intravascular use. They may be used for the controlled, selective regional infusion of therapeutic agents or embolic materials into vessels.

1. Magic Catheters IFU – Ind 19
MKTG-068 Rev. A
18 Technology Drive #169, Irvine Ca 92618
P 949.788.1443 | F 949.788.1444
Now you have **24 hours** to make a lifetime of difference in stroke patients like Nora.

The Trevo Retriever is the only device cleared to **reduce disability in stroke patients up to 24 hours** from time last seen well.

For more information, visit strykerneurovascular.com/trevo24hours

---

*Trevo® XP*  
PROVUE RETRIEVER
You are invited! Join us next year in Boston to present the best scientific research in neuroradiology.

Submit your abstract online at 2019.asnr.org
Deadline: 5:00pm (PST), Wednesday, November 7, 2018

Submit an abstract for the ASNR 57th Annual Meeting (May 18-23, 2019, Boston, MA) in one of the following presentation categories:

- Scientific Paper .............................................. Oral
- Scientific Poster .......................................... Print
- Electronic Scientific Poster .......................... Electronic
- Electronic Educational Exhibit .................. Electronic
- Educational Exhibit Poster ......................... Print
- Excerpta ..................................................... Oral

This activity has been approved for AMA PRA Category 1 Credits™.

Abstract Submissions Information and Criteria

1. Abstracts should describe the learning objectives of the presentation.
2. The American Journal of Neuroradiology (AJNR) encourages presenters to submit manuscripts based on their work to AJNR before considering other journals.
3. Presenters of accepted abstracts must register for the ASNR 57th Annual Meeting.
4. Published or previously presented works should NOT be submitted.
5. Format abstract text using headings required for submission category.
6. Submit each abstract in one category only.
7. Submission topic areas include: Adult Brain, Spine, Head and Neck, Pediatrics, Functional Imaging, Interventional, Health Policy, and AI/Informatics.
8. Maximum length: 300 words including graphics. Charts, graphs and tables may be included as an image, for all categories.
9. Submission site allows uploading of files into the system.
10. Changes can be made to the abstract until the deadline.

Complete instructions for submissions are available online: 2019.asnr.org
CALL FOR AJNR EDITORIAL FELLOWSHIP CANDIDATES

2019 Candidate Information and Requirements

GOALS
- Increase interest in editorial and publication-related activities in younger individuals.
- Increase understanding and participation in the AJNR review process.
- Incorporate into AJNR's Editorial Board younger individuals who have previous experience in the review and publication process.
- Fill a specific need in neuroradiology not offered by other similar fellowships.
- Increase the relationship between “new” generation of neuroradiologists and more established individuals.
- Increase visibility of AJNR among younger neuroradiologists.

ACTIVITIES OF THE FELLOWSHIP
- Serve as Editorial Fellow for one year. This individual will be listed on the masthead as such.
- Review at least one manuscript per month for 12 months. Evaluate all review articles submitted to AJNR.
- Learn how electronic manuscript review systems work.
- Be involved in the final decision of selected manuscripts together with the Editor-in-Chief.
- Participate in all monthly Senior Editor telephone conference calls.
- Participate in all meetings of the Editors during the annual meetings of ASNR and RSNA and the Radiology Editors Forum as per candidate's availability. The Foundation of the ASNR will provide $2000 funding for this activity.
- Evaluate progress and adjust program to specific needs in annual meeting or telephone conference with the Editor-in-Chief.
- Embark on an editorial scientific or bibliometric project that will lead to the submission of an article to AJNR or another appropriate journal as determined by the Editor-in-Chief. This project will be presented by the Editorial Fellow at the ASNR annual meeting.
- Serve as liaison between AJNR and ASNR's Young Professionals Network. Participate in meetings and telephone calls with this group. Design one electronic survey/year, polling the group regarding readership attitudes and wishes.
- Recruit trainees as reviewers as determined by the Editor-in-Chief.
- Organize and host a Fellows’ Journal Club podcast.
- Serve as Guest Editor for an issue of AJNR's News Digest with a timely topic.

QUALIFICATIONS
- Be a fellow in neuroradiology from North America, including Canada (this may be extended to include other countries).
- Be a junior faculty neuroradiology member (< 3 years) in either an academic or private environment.
- Be an “in-training” or member of ASNR in any other category.

APPLICATION
- Include a short letter of intent with statement of goals and desired research project. CV must be included.
- Include a letter of recommendation from the Division Chief or fellowship program director. A statement of protected time to perform the functions outlined is desirable.
- Applications will be evaluated by AJNR’s Senior Editors prior to the ASNR meeting. The name of the selected individual will be announced at the meeting.
- Applications should be received by March 1, 2019 and sent to Ms. Karen Halm, AJNR Managing Editor, electronically at khalm@asnr.org.

ASNR and AJNR are pleased once again to join efforts with other imaging-related journals that have training programs on editorial aspects of publishing for trainees or junior staff (3–5 years after training), including Radiology (Olmsted fellowship), AJR (Figley and Rogers fellowships), JACR (Bruce J. Hillman fellowship), and Radiologia.
AXS Infinity LS® Plus Long Sheath

See package insert for complete indications, contraindications, warnings, and instructions for use. INDICATIONS FOR USE

The AXS Infinity LS® Plus Long Sheath is indicated for the introduction of interventional devices into the peripheral, coronary, and neurovasculature.

RX ONLY

CONTRAINDICATIONS

There are no known contraindications.

POTENTIAL ADVERSE EVENTS

• Acute vessel occlusion
• Embolization
• Hemorrhage or hematomas
• Ischemia
• Malfunction of device
• Numbness or paresthesia
• Thrombus

WARNINGS

Contents supplied STERILE and unexpanded (EX_EN_US). Do not use in sterile barrier is damaged. If damage is found, call your Stryker Neurovascular representative.

For single use only. Do not re-use, re-process, or re-stereilize. Re-use, re-processing or re-stereilization may compromise the structural integrity of the device and/or lead to device failure which, in turn, may result in patient injury, death or illness. Re-use, re-processing or re-stereilization may also create a risk of contamination of the device and/or cause patient infection or cross-infection. Exceeding, including, but not limited to, the transmission of infection/disinfection among/some/to patient. Contamination of the device may lead to injury, death or patient illness.

After use, dispose of product and packaging in accordance with hospital, administrative and/or local government policy.

The AXS Infinity LS® Plus Long Sheath has not been evaluated for more than one (1) clot removal procedure.

RF Introducer is difficult to withdraw from the vessel, do not torque RF Introducer: Advance Microcatheter slowly, gently pull Retriever back into Microcatheter, and remove Retriever and Microcatheter as a unit. Further resistance may be encountered when withdrawing the Retriever into the Microcatheter, consider operating the Retriever using the AXS Vascular GDC guidewire extender REFINISH 22008b to ensure that Microcatheter can be advanced over a larger diameter catheter such as a GADI® Catheter. Jointly withdraw the Retriever into the larger diameter catheter.

• Administration of angiography and fluoroscopy may result in exposure to ionizing radiation.

• Acute vessel occlusion
• Embolization
• Hemorrhage or hematomas
• Ischemia
• Malfunction of device
• Numbness or paresthesia
• Thrombus

© 2018 Stryker

AXS Vecta® 71 Aspiration Catheter

See package insert for complete indications, contraindications, warnings, and instructions for use. INDICATIONS FOR USE

The AXS Vecta® 71 Aspiration Catheter (such as Merci® Balloon Guide Catheter and FlowGate® Balloon Guide Catheters) are recommended for use during thrombus removal procedures.

WRAP ONLY

• Ischemia
• Intracranial hemorrhage
• Hematoma or hemorrhage at the puncture site
• Distal embolization
• Death
• Acute vessel occlusion
• Thrombus

• To prevent thrombus formation and contrast media crystal formation, maintain a constant infusion of appropriate flush solution.

• Exposure to temperatures above 54°C (130°F) may damage device and accessories. Do not expose to outside temperatures above 54°C (130°F) or in-automobiles in direct sunlight or in heat. Handle with care. Do not autoclave.

• Do not use if labeling is illegible or inaccessible.

AXS Vecta® 71 Aspiration Catheter

Scout is a trademark of InNeuroCo, Inc.

AXSInfinity LS, AXS Vecta, Stryker. All other trademarks are trademarks of their respective owners or holders.

The AXS Vecta 71 Aspiration Catheter is a single use, single lumen, flexible, variable stiffness catheter. It has an inside diameter of 0.035" (1 mm), an outside diameter of 0.064" (1.63 mm) and an Luer lock end. The AXS Vecta 71 Aspiration Catheter shaft has a lumen coating at the distal and to reduce friction during use. The Scout Intro is an introducer and a sheath that can be used for the AXS Vecta 71 Aspiration Catheter. The Scout is an introducer and a sheath that can be used for the AXS Vecta 71 Aspiration Catheter. The inner lumen of the Scout Intro is compatible with guidewires and microcatheters of an outer diameter of less than 0.044". Each package includes a guidewire, a Scout Intro sheath, a hemostatic valve, and two post-assembly dimensions. The AXS Vecta 71 Aspiration Catheter and AXS Vecta 71 Aspiration Catheters are available in 3 different lengths, the device configurations including the length of the Scout package with each catheter and the recommended microcatheter length is presented in the table below.

<table>
<thead>
<tr>
<th>AXS Vecta 71 Aspiration Catheter</th>
<th>AXS Vecta 71 Aspiration Catheter</th>
<th>AXS Vecta 71 Aspiration Catheter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length (mm)</td>
<td>105</td>
<td>165</td>
</tr>
<tr>
<td>Recommended microcatheter length</td>
<td>0.044 max</td>
<td>0.044 max</td>
</tr>
<tr>
<td>Scout assembly</td>
<td>190</td>
<td>190</td>
</tr>
<tr>
<td>AXS Vecta 71 Aspiration Catheter</td>
<td>Recommended microcatheter length</td>
<td>0.044 max</td>
</tr>
<tr>
<td>105</td>
<td>165</td>
<td></td>
</tr>
<tr>
<td>Scout Intro length</td>
<td>0.044 max</td>
<td></td>
</tr>
<tr>
<td>Scout sheath length</td>
<td>0.044 max</td>
<td></td>
</tr>
<tr>
<td>Scout sheath outer diameter</td>
<td>0.044 max</td>
<td></td>
</tr>
</tbody>
</table>

POTENTIAL ADVERSE EVENTS

• Acute vessel occlusion
• Embolization
• Allergic reaction and anaphylaxis from contrast media
• Arteriovenous fistula
• Death
• Distal embolization
• Embolism
• False aneurysm formation
• Hematoma or hematomas at the puncture site
• Inability to complete re-aspiration thrombectomy
• Ischemia
• Kidney damage from contrast media
• Neuropathic deficit or stroke
• Neurologic deficit
• Retreiver...
1st Annual Meeting of the
AMERICAN SOCIETY OF
Pediatric Neuroradiology

January 18-20, 2019
Royal Sonesta Hotel - New Orleans • New Orleans, Louisiana

Please contact Educational Symposia, at 813-806-1000 or ASPNR@edusymp.com or visit www.ASPNR.org for additional information.
**WARNING: NEPHROGENIC SYSTEMIC FIBROSIS (NSF)**

Gadolinium-based contrast agents (GBCAs) increase the risk for NSF among patients with impaired elimination of the drugs. Avoid use of GBCAs in these patients unless the diagnostic information is essential and not available with non-contrast MRI or other modalities. NSF may result in fatal or debilitating fibrosis affecting the skin, muscle and internal organs.

- The risk for NSF appears highest among patients with:
  - Chronic, severe kidney disease (GFR < 30 mL/min/1.73 m²), or
  - Acute kidney injury.
- Screen patients for acute kidney injury and other conditions that may reduce renal function. For patients at risk for chronically reduced renal function (e.g., age > 60 years, hypertension, diabetes), estimate the glomerular filtration rate (GFR) through laboratory testing.
- For patients at highest risk for the drug, do not exceed the recommended DOTAREM dose and allow a sufficient period of time for elimination of the drug from the body prior to any re-administration.

**INDICATIONS AND USAGE**

DOTAREM® (gadoterate meglumine) injection is a prescription gadolinium-based contrast agent indicated for intravenous use with magnetic resonance imaging (MRI) in brain (intracranial), spine and associated tissues in adult and pediatric patients (including term neonates) to detect and visualize areas with disruption of the blood brain barrier (BBB) and/or abnormal vascularity.

**CONTRAINDICATIONS**

History of clinically important hypersensitivity reactions to DOTAREM.

**WARNINGS AND PRECAUTIONS**

- Hypersensitivity Reactions: Anaphylactic and anaphylactoid reactions have been reported with DOTAREM, including cardiovascular, respiratory, and/or cutaneous manifestations. Some patients experienced circulatory collapse and died. In most cases, initial symptoms occurred within minutes of DOTAREM administration and resolved with prompt emergency treatment.
- Before DOTAREM administration, assess all patients for any history of a reaction to contrast media, bronchial asthma and/or allergic disorders. These patients may have an increased risk for a hypersensitivity reaction to DOTAREM.
- Administer DOTAREM only in situations where trained personnel and therapies are promptly available for the treatment of hypersensitivity reactions, including personnel trained in resuscitation.
- Gadolinium Retention: Gadolinium is retained for months or years in several organs. The highest concentrations have been identified in the brain, followed by bone, skin, kidney, liver and spleen. The duration of retention also varies by tissue, and is longest in bone. Linear GBCAs cause more retention than macrocyclic GBCAs.
- Consequences of gadolinium retention in the brain have not been established. Adverse events involving multiple organ systems have been reported in patients with normal renal function without an established causal link to gadolinium retention.
- Acute Kidney Injury: In patients with chronically reduced renal function, acute kidney injury requiring dialysis has occurred with the use of GBCAs. The risk of acute kidney injury may increase with increasing dose of the contrast agent; administer the lowest dose necessary for adequate imaging.
- Extravasation and Injection Site Reactions: Ensure catheter and venous patency before the injection of DOTAREM. Extravasation into tissues during DOTAREM administration may result in tissue irritation.

**ADVERSE REACTIONS**

- The most common adverse reactions associated with DOTAREM in clinical trials were nausea, headache, injection site pain, injection site coldness and rash.
- Serious adverse reactions in the Postmarketing experience have been reported with DOTAREM. These serious adverse reactions include but are not limited to: arrhythmia, cardiac arrest, respiratory arrest, pharyngeal edema, laryngospasm, bronchospasm, coma and convulsion.

**USE IN SPECIFIC POPULATIONS**

- **Pregnancy:** GBCAs cross the human placenta and result in fetal exposure and gadolinium retention. Use only if imaging is essential during pregnancy and cannot be delayed.
- **Lactation:** There are no data on the presence of gadoterate in human milk, the effects on the breastfed infant, or the effects on milk production. However, published lactation data on other GBCAs indicate that 0.01 to 0.04% of the maternal gadolinium dose is present in breast milk.
- **Pediatric Use:** The safety and efficacy of DOTAREM at a single dose of 0.1 mmol/kg has been established in pediatric patients from birth (term neonates ≥ 37 weeks gestational age) to 17 years of age based on clinical data. The safety of DOTAREM has not been established in preterm neonates. No cases of NSF associated with DOTAREM or any other GBCA have been identified in pediatric patients age 6 years and younger.

You are encouraged to report negative side effects of prescription drugs to the FDA. Visit www.fda.gov/medwatch or call 1-800-FDA-1088.

Please see the full Prescribing Information, including the patient Medication Guide, for additional important safety information.

**References:**
2. Maravilla K et al. Comparison of Gadoterate Meglumine and Gadobutrol in the Diagnosis of Primary Brain Tumors: A Double-Blind Randomized Controlled Intraindividual Crossover Study (the REMIND Study). 2017 June 29. doi: 10.3174/ajnr.A5316. [Epub ahead of print].

DOTAREM® is a registered trademark of Guerbet LLC, and is available by prescription only.

Gadavist® is a registered trademark of the Bayer group of companies, and is available by prescription only.
DOTAREM®
(gadoterate meglumine) Injection

Comparison of Dotarem® to Gadavist® for Overall Visualization and Characterization in the MRI Diagnosis of Primary Brain Tumors

Despite the difference in relaxivity between these 2 GBCAs

There Is No Measurable Difference in Clinical Benefit Observed Between Dotarem® and Gadavist®

as demonstrated by the REMIND Study, a multicenter, double-blind, randomized, controlled intraindividual crossover study.

This study “demonstrates the non-inferiority of gadoterate meglumine [Dotarem®] versus gadobutrol [Gadavist®] for overall visualization and characterization of primary brain tumors”. Additionally, there was no preference of the readers for either contrast agent, in most cases, regarding border delineation, internal morphology, and the qualitative degree of contrast enhancement, despite quantitative mean lesion percentage enhancement being higher with gadobutrol.

- For all readers in the REMIND Study, more than 90% of patients presented with good or excellent overall lesion visualization and characterization with either Dotarem® or Gadavist®.
- The REMIND Study also demonstrated a low incidence of immediate reported AEs with Dotarem® and with Gadavist®, as shown in multiple previous studies.
- Dotarem® is the only imaging contrast with macrocyclic and ionic structure for high thermodynamic and kinetic stability.
- Dotarem® is not only trusted for high molecular stability; the REMIND Study demonstrates that it is as effective as Gadavist® for MRI diagnosis of primary brain tumors.

Please see Important Safety information on opposite page. For more information on Dotarem®, please see Full Prescribing Information including Boxed Warning and Medication Guide.
Redefine aspiration.

AXS Vecta™ 71
ASPIRATION CATHETER

Packaged with the Scout Introducer, a 0.044in lumen nitinol cross coil catheter that replaces the need for a 3MAX or other delivery catheter.

AXS Vecta 71 Aspiration Catheter

Big 0.071in ID aspiration lumen to ingest more clot

Deliver through a 0.088in ID long sheath or the new 0.091in AXS Infinity LS™ Plus Long Sheath
1775 PERSPECTIVES
B. Thomas

1776 REVIEW ARTICLE
Deep Learning in Neuroradiology
G. Zaharchuk, et al.
ADULT BRAIN

1785 PRACTICE PERSPECTIVES
Comparison of Advanced Imaging Resources, Radiology Workforce, and Payment Methodologies between the United States and Canada
H.A. Valand, et al.

1791 GENERAL CONTENTS
Clinical Value of Hybrid TOF-PET/MR Imaging–Based Multiparametric Imaging in Localizing Seizure Focus in Patients with MRI-Negative Temporal Lobe Epilepsy
K. Shang, et al.
ADULT BRAIN

1799 Longitudinal Persistence of Meningeal Enhancement on Postcontrast 7T 3D-FLAIR MRI in Multiple Sclerosis
S.N. Jonas, et al.
ADULT BRAIN

1806 Automated Integration of Multimodal MRI for the Probabilistic Detection of the Central Vein Sign in White Matter Lesions
J.D. Dworkin, et al.
ADULT BRAIN

1814 Predicting Genotype and Survival in Glioma Using Standard Clinical MR Imaging Apparent Diffusion Coefficient Images: A Pilot Study from The Cancer Genome Atlas
C.-C. Wu, et al.
ADULT BRAIN

1821 Near-Term Decrease in Brain Volume following Mild Traumatic Injury Is Detectable in the Context of Preinjury Volumetric Stability: Neurobiologic Insights from Analysis of Historical Imaging Examinations
A.E. Goldman-Yassen, et al.
ADULT BRAIN

1827 3D Black-Blood Luminal Angiography Derived from High-Resolution MR Vessel Wall Imaging in Detecting MCA Stenosis: A Preliminary Study
X. Bai, et al.
ADULT BRAIN

1833 Clinical Evaluation of Highly Accelerated Compressed Sensing Time-of-Flight MR Angiography for Intracranial Arterial Stenosis
S.s. Lu, et al.
ADULT BRAIN

1839 Cerebrovascular Reactivity during Prolonged Breath-Hold in Experienced Freedivers
V.C. Keil, et al.
ADULT BRAIN
<table>
<thead>
<tr>
<th>Page</th>
<th>Title</th>
<th>Author(s)</th>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>1854</td>
<td>Thrombus Permeability on Dynamic CTA Predicts Good Outcome after Reperfusion Therapy</td>
<td>Z. Chen, et al.</td>
<td>INTERVENTIONAL</td>
</tr>
<tr>
<td>1860</td>
<td>Identification of Hostile Hemodynamics and Geometries of Cerebral Aneurysms: A Case-Control Study</td>
<td>B.J. Chung, et al.</td>
<td>INTERVENTIONAL</td>
</tr>
<tr>
<td>1867</td>
<td>Integrating 3D Rotational Angiography into Gamma Knife Planning</td>
<td>H. Hasegawa, et al.</td>
<td>INTERVENTIONAL</td>
</tr>
<tr>
<td>1871</td>
<td>Quantification of Blood Velocity with 4D Digital Subtraction Angiography Using the Shifted Least-Squares Method</td>
<td>Y. Wu, et al.</td>
<td>INTERVENTIONAL</td>
</tr>
<tr>
<td>1878</td>
<td>Correlation between Human Papillomavirus Status and Quantitative MR Imaging Parameters including Diffusion-Weighted Imaging and Texture Features in Oropharyngeal Carcinoma</td>
<td>M. Ravanelli, et al.</td>
<td>HEAD &amp; NECK</td>
</tr>
<tr>
<td>1884</td>
<td>Negative Predictive Value of NI-RADS Category 2 in the First Posttreatment FDG-PET/CT in Head and Neck Squamous Cell Carcinoma</td>
<td>P. Wangaryattawanich, et al.</td>
<td>HEAD &amp; NECK</td>
</tr>
<tr>
<td>1903</td>
<td>MR Imaging of the Facial Nerve through the Temporal Bone at 3T with a Noncontrast Ultrashort Echo Time Sequence</td>
<td>J.P. Guenette, et al.</td>
<td>HEAD &amp; NECK</td>
</tr>
<tr>
<td>1912</td>
<td>Multidelay Arterial Spin-Labeling MRI in Neonates and Infants: Cerebral Perfusion Changes during Brain Maturation</td>
<td>H.G. Kim, et al.</td>
<td>PEDIATRICS</td>
</tr>
<tr>
<td>1919</td>
<td>The Impact of Persistent Leukoencephalopathy on Brain White Matter Microstructure in Long-Term Survivors of Acute Lymphoblastic Leukemia Treated with Chemotherapy Only</td>
<td>N.D. Sabin, et al.</td>
<td>PEDIATRICS</td>
</tr>
<tr>
<td>1926</td>
<td>Apparent Diffusion Coefficient Levels and Neurodevelopmental Outcome in Fetuses with Brain MR Imaging White Matter Hyperintense Signal</td>
<td>E. Katorza, et al.</td>
<td>PEDIATRICS</td>
</tr>
<tr>
<td>1932</td>
<td>Longitudinal Findings of MRI and PET in West Syndrome with Subtle Focal Cortical Dysplasia</td>
<td>Y. Sakaguchi, et al.</td>
<td>PEDIATRICS</td>
</tr>
<tr>
<td>1938</td>
<td>Imaging of Clival Hypoplasia in CHARGE Syndrome and Hypothesis for Development: A Case-Control Study</td>
<td>C.M. de Geus, et al.</td>
<td>PEDIATRICS</td>
</tr>
<tr>
<td>1943</td>
<td>Multiple Brain Developmental Venous Anomalies as a Marker for Constitutional Mismatch Repair Deficiency Syndrome</td>
<td>S.I. Shiran, et al.</td>
<td>PEDIATRICS</td>
</tr>
<tr>
<td>1953</td>
<td>Dilated Vein of the Filum Terminale on MRI: A Marker for Deep Lumbar and Sacral Dural and Epidural Arteriovenous Fistulas</td>
<td>W. Brinjikji, et al.</td>
<td>SPINE</td>
</tr>
<tr>
<td>1957</td>
<td>Spinal Instrumentation Rescue with Cement Augmentation</td>
<td>A. Cianfoni, et al.</td>
<td>SPINE</td>
</tr>
</tbody>
</table>
ONLINE FEATURES

LETTERS

E108 Reply  S. Karimi
E109 Vertebroplasty: Expectation or Evidence-Based Interventional Radiology?  A. Brallion, et al.
E110 Reply  R.V. Chandra, et al.

BOOK REVIEWS R.M. Quencer, Section Editor
Please visit www.ajnrblog.org to read and comment on Book Reviews.

The goal is to classify MR images into 4 specific diagnoses. Multiple different images form the training set. For each new case, the image is broken down into its constituent voxels, each one of which acts as an input into the network. This example has 3 hidden layers with 7 neurons in each layer. Final output is the probabilities of the 4 classification states. At the bottom is a zoomed-in view of an individual neuron in the second hidden layer, which receives input from the previous layer, performs a standard matrix multiplication, passes this through a nonlinear function, and outputs a single value to all the neurons of the next layer.
Title: Purple Toronto. Oil on canvas 22" × 15". This is a view of the harborfront from the central island, recollections from my fellowship days.

Bejoy Thomas, MD, DNB, PDCC, Trivandrum, Kerala, India
Deep learning is a form of artificial intelligence, roughly modeled on the structure of neurons in the brain, which has shown tremendous promise in solving many problems in computer vision, natural language processing, and robotics. It has recently become the dominant form of machine learning, due to a convergence of theoretic advances, openly available computer software, and hardware with sufficient computational power. The current excitement in the field of deep learning stems from new data suggesting its excellent performance in a wide variety of tasks. One benchmark of machine learning performance is the ImageNet Challenge. In this annual competition, teams compete to classify millions of images into discrete categories (tens of different kinds of dogs, fish, cars, and so forth). A watershed year was 2012, when the first neural network–based entry bested the competition and prior years’ results by a wide margin. Since then, every winning entry has used a deep learning framework, with performance now exceeding that of humans.

Deep learning has the potential to revolutionize entire industries, including medical imaging. Given the centrality of neuroimaging in the diagnosis and treatment of neurologic diseases, deep learning will likely affect neuroradiologists first and most profoundly. This article will introduce deep learning methods, overview their current successes, and speculate on the future evolution of these methods, focusing on their application to neuroradiology.

What is Deep Learning?

It is useful to consider where deep learning fits into the broader context of artificial intelligence (Fig 1). One definition suggested for artificial intelligence is any computer method that performs tasks normally requiring human intelligence. Machine learning is one type of artificial intelligence that develops algorithms to enable computers to learn from existing data without explicit programming. Examples are classification algorithms such as clustering, logistic regression, and support vector machines.

Machine learning methods can be further divided into supervised and unsupervised learning. In supervised learning, some “ground truth” exists, which is used to train the algorithms. One example is a collection of brain CT scans that a neuroradiologist has classified into different groups (ie, hemorrhage versus no hemorrhage). In contrast, for unsupervised learning, no criterion standard images or classifications are used—the computer itself must determine the classes. One example is clustering, in which images are placed in multiple groups based on similarity metrics.
without knowing a priori what is driving the separation. While unsupervised learning holds great promise for medical imaging, this review focuses on supervised learning.

Viewed in this context, deep learning is a supervised machine learning method that uses a specific architecture, namely some form of neural network. The power of these techniques is in their scalability, which is largely based on their ability to automatically extract relevant features. In the past, constructing an image-classification algorithm took years of effort on the part of domain experts and experienced artificial intelligence researchers. Deep learning allows such a classifier to be created automatically from a labeled dataset in days. These neural networks are loosely inspired from how the brain is structured, with hidden layers representing interneurons (Fig 2). While modern neural networks share these similarities to the brain, whether more fidelity to known brain structures would improve performance is an actively debated question. For example, in computer vision applications, many of the features to which hidden layers are sensitive (such as edges in different orientations) have correlates in the mammalian visual cortex.

For neuroimaging, a simple deep learning model may accept image data as a vector composed of voxel intensities, with each voxel serving as an input “neuron.” While the examples below assume the use of individual images, more generally, the input can consist of entire imaging series, multiple series, or even multiple modalities. Next, one must determine how many layers (how deep) and how many neurons per layer (how wide) to include; this is known as the network architecture (Fig 3). Each neuron stores a numeric value, and each connection between neurons represents a weight. Weights connect the neurons in different layers and represent the strength of connections between the neurons. A “fully connected” layer in which all neurons in one layer are connected to all neurons in the next can be interpreted and implemented as a matrix multiplication. Finally, it is customary to include a nonlinear “activation function” at the output of the neuron. This introduces nonlinearity into the equations so that complex functions can be represented that would not otherwise be possible. Historically, sigmoids and hyperbolic tangents have been used on the basis of insights from neuroscience; however, researchers have since found that the rectified linear unit is both simpler to implement and more effective. The rectified linear unit function outputs the value of the neuron for positive values and zero for negative values.

The choice of network architecture for a specific application is not always obvious, though some typical configurations and assumptions exist. The number of neurons in the hidden layers tends to be larger than that in either the input or the output layers. The final layer encodes the desired outcomes or labeled states. For example, if one wishes to classify an image as “hemorrhage” or “no hemorrhage,” 2 final layer neurons are appropriate. Commonly, the value stored by each final neuron is interpreted as the probability that the training example corresponds to a specific class. The goal of training is to optimize the network weights so that when a new sample image is input, the probabilities...
FIG 2. Parallels between artificial and biologic neural networks. Hidden layers of artificial neural networks can be thought to be analogous to brain interneurons.

FIG 3. Example of a simple deep network architecture. The goal of this network is to classify MR images into 4 specific diagnoses (normal, tumor, stroke, hemorrhage). Multiple different images form the training set. For each new case, the image is broken down into its constituent voxels, each one of which acts as an input into the network. This example has 3 hidden layers with 7 neurons in each layer, and the final output is the probabilities of the 4 classification states. All layers are fully connected. At the bottom is a zoomed-in view of an individual neuron in the second hidden layer, which receives input from the previous layer, performs a standard matrix multiplication (including a bias term), passes this through a nonlinear function (the rectified linear unit function in this example), and outputs a single value to all the neurons of the next layer.
measured at the output are heavily skewed to the correct class. For example, if we input an image with hemorrhage, we would like the model to output a high probability for hemorrhage and low probability for other classes. How is this accomplished?

**Training Simple Neural Network Deep Learning Models**

Neural networks are ideally trained using large numbers of cases that are divided into several groups. Usually the largest fraction is used for model training (50%–60%), with another 10%–20% for validation and 20%–40% for testing. The training cases are used to set the model parameters; large training sets are important because even relatively shallow networks may have 100,000s of free parameters (weights). The training dataset is looped through multiple times (epochs) until the accuracy of the model converges. At first, predictions will be poor. However, the beauty of this setup is that you can compare the output of the model with the ground truth via the use of a “cost function,” a single number that quantifies how far off the model is. Back-propagation, a technique whereby the strength of connections between neurons (weights) can be adjusted on the basis of the value of the cost function, is then used to reinforce correct predictions and penalize incorrect predictions. This procedure is repeated using separate training examples and multiple iterations, thereby optimizing the weights and effectively training the model. Once the model is trained, there are several “hyperparameters” to optimize, including the learning rate and number of epochs. Finally, the testing set is used to assess the model accuracy on data that has not been used for training. This assessment will yield an error rate similar to or higher than that for the training set, and this helps to gauge how well the final model will perform on real world data. While training the model is often time-intensive, the application of the final trained model to new data is usually computationally fast.

Choosing the right cost function is important. For classification, the value of the cost function should be low when the model predicts the correct class and high when its predictions are off. A popular cost function for classification is “cross-entropy loss,” an extension of logistic regression to multiple classes that can be implemented using the softmax function. For image prediction, common cost functions include the root-mean-square error between the predicted and reference images and measures of similarity, such as the structural similarity index metric. One promising approach is replacing the cost function itself with a network whose goal is to make it optimally difficult to distinguish reference images from predicted images, an approach known as the “generative adversarial network” approach. Generative adversarial networks strive to eliminate systematic differences between the predicted and reference images, which is highly desirable in the radiology setting.

**From Simple to Convolutional Neural Networks**

Fully connected neural networks are computationally expensive because the number of weights is very large, especially with images of typical matrix sizes (256 × 256 = 65,536 voxels). With even just 1 slice, >4 billion weights are required to implement a fully connected layer. Thus, much research in image-based deep learning has moved to using more computationally efficient structures, specifically convolutional neural networks (CNNs).

CNNs are well-suited for imaging. Instead of full connections, a small “kernel” of weights is applied at each image position to determine the value of the neuron of the next layer (Fig 4). The approach mimics the mathematical operation of convolution. Be-
tween 2 layers, the only weights required are those for the kernel, which is then rastered across the image to obtain the next layer. This method has several advantages. First, it markedly reduces the number of weights. Second, it allows spatial invariance: Image features may occur in different locations, and a CNN allows their identification independent of their precise location. Often, CNNs pool adjacent voxels or slide the kernel across these images at spaced intervals (a hyperparameter known as the “stride length”), so that the dimensions in each subsequent layer are smaller than those in the last one. For each layer, multiple different kernels can be trained, creating multiple “channels” in each layer; such a structure allows the network to learn many location-invariant features, such as edges, textures, and other nonlinear representations of the data. With pooling or increased stride lengths, it is possible to incorporate ever larger features into the hidden layers of the network. Indeed, the ability of CNNs to extract relevant imaging features in a location-invariant way parallels the structure of the visual system of the brain; Hubel and Wiesel5 showed in the 1960s that different regions of the cat brain responded strongly to features such as edges oriented in different directions.

For classification, ≥1 fully connected layer is typically added to reach the final output layer. For image prediction, upsampling layers are used to “re-form” the smaller dimension hidden layers back into the original size of the input image. Such an architecture is called an “encoder-decoder” because it represents the image in terms of increasing abstraction (encoding) in the hidden layers and then uses them to recreate (decode) the image.

**Overfitting and Data Augmentation**

As described above, typical deep learning models have millions of weights. Analogous to the idea that you need more equations than variables to solve algebraic equations, if a deep network is trained on too few examples, it is possible to perfectly represent the transform between the input and output states. However, this approach will not generalize to new cases, a problem known as “overfitting.” The best solution to overfitting is collecting more training examples, though other solutions such as regularization and drop-out can also be used.7,9 Another potential solution is data augmentation.

Data augmentation is a method of increasing the amount of training data. Because most image data should be recognizable whether offset in the x-y plane, rotated, flipped, or slightly stretched or skewed, it is conventional to perform such image manipulations to augment the training data. While these image alterations do not add more data, they have been shown to improve the robustness of the models, possibly by preventing the model from learning features that occur only in a specific orientation.

**Broad Classes of Applications**

Deep learning can address many aspects of neuroradiology. The overall flow of work in neuroradiology is a useful framework in which to consider these applications. This starts with referring clinicians ordering studies and then moves to image acquisition. Next, the images are put before radiologists, and tasks surrounding detection and segmentation of lesions and differential diagnosis arise. Each link in this chain can potentially benefit from a deep learning approach.

**Imaging Logistics**

After a study is ordered, it needs to be triaged to a specific neuroimaging protocol. This process often involves the precious time of radiologists and relies on their knowledge of imaging protocols and attention to clinicians’ specific requests, which are often encapsulated in the order history as free text. Deep learning methods to interpret natural language are already mature, making automation of the protocolling process quite feasible. In theory, this problem is just about classification, with the different protocols being the classes to predict and the input being the order itself and patient metadata. The protocolling application is ideal for deep learning because of the immense amount of training data that already exists; all prior studies that have been protocollated by humans can be used for training.

Another promising application is to triage image review in the order of suspected acuity. For example, if models can be trained to identify critical findings on images, it is possible to prioritize radiologic review of these studies, even if they were not initially ordered as “stat” studies. For large organizations, such triage offers the potential to reduce the time between acquisition and interpretation for critical cases, with presumably positive effects on patient outcome.

**Image Acquisition and Improvement**

Deep learning methods can be used to perform image reconstruction and improve image quality. Deep learning frameworks are capable of “learning” standard MR imaging reconstruction techniques, such as Cartesian and non-Cartesian acquisition schemes.10 Combining deep learning to k-space undersampling with model-based/compressed sensing reconstruction schemes holds the potential to revolutionize imaging science by optimizing how image data are collected.11-13

Also, one could apply deep learning methods to improve image quality. If images at low-resolution and high-resolution are available, it is possible to use a deep network for super-resolution.14 If paired image sets of low and high quality are available, learning the optimal nonlinear transformation between them can be considered. This has already been applied to CT imaging and has been demonstrated to be of value on a dataset consisting of normal-dose and simulated low-dose CT.15 Another approach uses paired MR images of the same anatomy, which are acquired at different field strengths. A study using 3T input data and 7T output data showed that a deep network can be trained to create simulated “7T-like” images from 3T data.16 Often obtaining a certain imaging sequence can be very time-consuming; an example is DTI, in which the need for multiple angular directions lengthens the examination beyond what many patients can tolerate. A deep learning approach can reduce imaging duration 12-fold by predicting final parameter maps (fractional anisotropy, mean diffusivity, and so forth) from relatively few angular directions.17 By acquiring paired arterial spin-labeling (ASL) CBF images with 2 and 30 minutes of acquisition time, our group has trained a deep network to boost the SNR of ASL significantly (Fig 5).18
Image Transformation

An extension of this is to create images with different contrast or with features of different modalities. For example, using the National Alliance for Medical Imaging Computing data base (http://www.insight-journal.org/midas/community/view/17), Velumalapalli et al19 used a deep network to predict T1 images from T2 images, and vice versa. Another application is to PET/MR imaging; unlike PET/CT, in which CT is used to calculate an attenuation map, MR images do not directly yield attenuation images. However, if there is information about soft tissue, air, and bone in the MR images, these sequences can be used as input to a deep network. Crucially here, the image to predict is no longer another MR image, but rather a coregistered CT scan of the same subject. Proof of principle was recently demonstrated for brain MR imaging attenuation correction, with performance superior to that of competing techniques.20 Another study demonstrated a similar use of MR imaging to create synthetic CT for radiation therapy.21 In clinical trials, situations arise in which patients may not be able to undergo a certain diagnostic technique, such as patients with MR imaging–incompatible implants. Alternatively, they may lack images at specific time points. While statistical techniques can be used to account for such missing data, if enough patients drawn from the same population have completed all imaging examinations, it is possible to train a deep learning network to recreate these data. Li et al22 demonstrated this using the Alzheimer’s Disease Neuroimaging Initiative (ADNI; http://www.adni-info.org/). They trained a CNN on patients with both FDG-PET and T1-weighted MR imaging and then used this network on a test set to predict the expected PET images from patients’ MR imaging studies,22 showing that the CNN method outperformed more traditional methods.

Lesion Detection and Segmentation

Detecting and segmenting lesions is an onerous task for humans but is well-suited to machine learning. While related, they are really 2 different tasks. The former starts with an unlabeled image and marks potential abnormalities. The goal of the latter is to circumscribe the regions encompassing the abnormal structures. Identifying and delineating the margins of a lesion is important because neuroradiologists are often tasked with monitoring the change in size or activity of known lesions across time or in response to treatment. Deep learning also has advantages for segmenting normal brain structures because existing methods are time-consuming and may not generalize to younger or older subjects.23-25 Furthermore, many research projects rely on manual delineation of image lesions. One can train a deep network to take images as input and hand-drawn manually segmented masks as output. Indeed, such an approach has shown great early success. While there are many examples in different neurologic disease conditions, we will reference 3 representative areas: detecting microhemorrhages, identifying infarcts and predicting final infarct volumes in patients with stroke, and segmenting brain tumors. Dou et al26 described a process to detect brain microhemorrhages by training a CNN on an annotated dataset of susceptibility-weighted images. They proposed a cascaded, 2-step approach, in which candidate lesions are first identified by the CNN and then only these lesions are input to a discriminatory CNN (ie, true microhemorrhage or mimic). With this approach, they achieved a sensitivity of >93%, with an average of about 3 false-positive identifications per subject.

Automatic identification and outlining of infarcted brain tissue would be useful in the acute stroke setting. Chen et al27 used DWI as input to a 2-stage deep learning algorithm and were able to detect 94% of all acute infarcts. Using the Dice coefficient as a marker of accuracy, they showed a mean score of 0.67 in a large cohort of patients 2 days after stroke. Another study using a 3-layer-deep CNN followed by 2 fully connected layers showed similar results and outperformed several other machine learning methods.28 Another intriguing application is to predict final infarct volume from early DWI and/or PWI in patients with acute stroke. Currently, the diffusion-perfusion mismatch approach is
the dominant paradigm, which states that DWI lesions represent irreversibly damaged tissue, while PWI identifies tissue at risk of infarction. Rather than using such hand-crafted features, a deep neural network can be trained using initial DWI and PWI maps as input and final infarct size measured several days later as the output. Using such a framework, Nielsen et al demonstrated that a deep learning architecture outperforms traditional state-of-the-art lesion prediction methods in acute stroke. They also showed that a 37-layer architecture outperformed a shallower 3-layer architecture, highlighting the importance of the depth feature of the network. One exciting application of this approach is to train separate networks based on different treatments. In stroke, one could train networks in patients who received stroke treatment and those who did not. New predictions using these 2 different models could give insight into whether treatment would lead to a reduced infarct volume (Fig 6).

Automated segmentation of brain tumors, not just their enhancing margins, but also other features such as regions of enhancement and necrosis, would be useful for a wide range of indications, including diagnosis, presurgical planning, and follow-up. The Brain Tumor Image Segmentation dataset is a publicly available dataset of brain tumor images with expert manual segmentations that has been a useful proving ground for new segmentation algorithms. The highest performance on the Brain Tumor Image Segmentation dataset in 2016 was achieved using a fully convolutional residual neural network, built on the structure that won the 2015 ImageNet Challenge, with Dice coefficients for complete tumor, core tumor, and enhancing tumor between 0.72 and 0.87. In a separate study, Korfiatis et al trained a deep autoencoder-decoder to segment T2-FLAIR lesions on the Brain Tumor Image Segmentation dataset, which included manually drawn outlines in 186 patients. They then applied their model to a separate group of 135 patients with tumor in which they had 3 expert segmentations and measured a Dice coefficient of 0.88 based on a method that incorporated the individual tracings of the 3 readers. They did note significant variability among their readers’ segmentations, pointing out the importance of how the segmentation criterion standard is implemented.

**Deep Learning for Image-Based Diagnosis**

The “Holy Grail” of machine learning in radiology is the so-called “end-to-end” solution, in which images are used as input and the output is a draft radiology report encompassing all the salient features of the image that an expert radiologist would include. As improbable as this might seem, progress is being made on this front with deep learning. Such approaches require tremendous amounts of annotated training data, which exist currently in many heterogeneous forms. Structured reporting, use of standard lexicons (such as RadLex; http://www.rsna.org/RadLex.aspx), and the standardization of electronic medical record platforms are all steps that are enabling the formation of the required large imaging/diagnosis datasets. This section will deal with some early applications of deep learning to neuroradiologic diagnosis.

Gao et al classified 285 noncontrast brain CT examinations with a deep network into 1 of 3 categories (normal aging, lesion [such as tumor], or Alzheimer disease [AD]). With an average classification accuracy of 88%, the approach performed marginally better than other approaches that relied on hand-crafted features. Another study showed that a multimodal stacked deep polynomial network using the ADNI dataset could classify patients into different binary groups (ie, AD versus healthy control [NC], or mild cognitive impairment converters [MCI-C] versus nonconverters [MCI-nonconverters]). Instead of using images as input, they used the volumes of 93 structures segmented from T1-weighted images along with the $[^{18}F]$ FDG-PET signal intensity in these same regions. For distinguishing patients with AD

![FIG 6. An example of the predicted risk of final infarct for 2 patients with acute ischemic stroke using 2 neural networks trained, respectively, on patients with and without rtPA administration. Patient A is a 76-year-old woman with an admission NIHSS score of 10 scanned 1.5 hours after symptom onset. The +rtPA network estimates a negligible permanent lesion, consistent with the acute DWI suggesting little permanent tissue damage at follow-up. The -rtPA network indicates that without treatment, a considerable volume of the acute ischemic region will progress to a permanent lesion. Patient B is a 72-year-old man also with an admission NIHSS score of 10, scanned 2 hours after onset. In this case, the 2 networks indicate little expected impact of treatment, likely due to the progression at the time of imaging of the ischemic event as seen on the DWI. CMRO2 indicates cerebral metabolic rate of oxygen; Tmax, time-to-maximum. Figure courtesy of Kim Mouridsen and Anne Nielsen/Aarhus University, Combat Stroke.](https://example.com/fig6.png)
from NCs, they showed an impressive area under the curve of 0.97. For the more challenging task of predicting MCI converters from nonconverters, they still showed areas under the curve of >0.80. Suk et al35 showed that similar features could be combined with CSF data in the ADNI dataset using a deep-weighted sparse multitask learning framework to improve classification, showing 95% accuracy to distinguish patients with AD from NCs. However, when trying to predict among 3 classes (AD, NC, and MCI), the accuracy dropped to 63%, which further dropped to 54% when trying to classify among 4 groups (AD, NC, MCI-C, and MCI-nonconverter). This latter point speaks to the challenges of moving beyond simple binary classifications into the kinds of tasks in which many diagnostic groups are possible, a situation familiar to neuroradiologists.

Another application is to use deep networks to identify the presence of hemorrhage on noncontrast brain CT. While general machine learning techniques have been applied to these cases with great success,36,37 only recently have deep learning methods been evaluated. Phong et al38 demonstrated a particularly interesting approach, using several “pretrained” deep networks as starting points for their training. Specifically, they used the optimal weights from the GoogLeNet (http://deeplearning.net/tag/googlenet/) or Inception-ResNet (https://keras.rstudio.com/reference/application_inception_resnet_v2.html) that were trained on nonmedical images and then used their data to train a final fully connected layer. This approach (called “transfer learning”) is attractive, given that it allows networks to be trained with less data than if they were trained from scratch.39 Trained on 80 cases and tested on 20 cases, they showed classification accuracies of >98%.

Plis et al40 examined both structural and functional MR imaging as input to deep networks for predicting various neurologic diseases. For structural imaging, they showed that they could distinguish patients with schizophrenia and Huntington disease from healthy subjects. For fMRI, they showed that deep networks performed similar to independent component analysis for identifying functional networks but tended to preserve edge details better. Similar results using a temporal-autoencoding neural network to predict the next time point in a resting-state fMRI time-series were applied to the Human Connectome Project data and showed a similar ability to identify task-specific networks.41

**Impact of Deep Learning on Neuroradiology Practice**

One concern is that if these approaches are successful, some work that radiologists have traditionally performed may become obsolete. Recently, a framework for inputting medical images (in this case, pathology slides) and outputting diagnostic text-based reports has been reported.52 While this technology is still very rudimentary, it is not difficult to imagine training a similar network with CT scans and their reports. While deep learning does hold much promise to automate tasks that radiologists find unpleasant, ways of checking and verifying results will be needed. Another concern with deep learning is that we have little insight into the inner workings of the models; they work well for prediction, but precisely how they accomplish this is unclear. This contrasts with much prior radiology research, which relied heavily on domain knowledge and the building of realistic models. Understanding how and why deep networks perform so well is an active area of research in the artificial intelligence community.

Also, the application of deep learning is still limited by requirements for large amounts of annotated training datasets and the challenge of keeping models current as source data and practice patterns change. Among applications that are amenable to disruption, it is still unclear which applications are supported by sufficient clinical need to drive widespread adoption. Thus, it is important that radiologists remain engaged with artificial intelligence scientists to both understand the capabilities of existing methods and direct future research in an intelligent way. Although this technology is still developing, state-of-the-art deep learning models show little evidence that they could replace all functions of radiologists, though this issue is contentious. For the foreseeable future, they will likely serve as powerful image-processing and decision-support tools that will augment the accuracy and efficiency of radiologists.

**Outlook**

The wealth of applications of deep learning will likely lead to increased use of this technology. How fast this will happen depends on a few key factors: Deep learning can learn more from larger datasets, usually by adding additional hidden layers. As larger labeled datasets become available, the power of deep learning approaches will increase. The importance of data-sharing initiatives such as ADNI and the Cancer Imaging Archive (https://public.cancerimagingarchive.net/ncia/legalRules.jsf) cannot be overstated. Second, the computer hardware required to run these methods continues to improve and become less expensive. The availability of open-source software frameworks, such as Caffe, Tensorflow, PyTorch, and Keras, is greatly facilitating progress. However, before these methods become a routine part of clinical practice, vendors will need to provide “turn-key” systems that integrate well into current workflow patterns. As more neuroradiology researchers and practitioners become comfortable with these methods, through exposure at medical imaging conferences and conversations with colleagues, applications will branch out from the current “low-hanging fruit” to address more complex and specialized questions. Therefore, we can expect further advances in this field with accompanying benefits in many areas of radiology.

**REFERENCES**

3. Mwangi B, Soares JC, Hasan KM. Visualization and unsupervised
predictive clustering of high-dimensional multimodal neuroimaging data. J Neurosci Methods 2014;236:19–25 CrossRef Medline
Comparison of Advanced Imaging Resources, Radiology Workforce, and Payment Methodologies between the United States and Canada


ABSTRACT

SUMMARY: The purpose of this Practice Perspectives was to review the United States and Canadian approaches to health care access and payment for advanced imaging. The historical background, governmental role, workforce, coding, payment, radiologic challenges, cost, resource intensity, and overall outcomes in longevity are reviewed.

ABBREVIATION: BC — British Columbia

Although Canada and the United States share a 5,425-mile border, common heritage, culture, and language, the health care system of each country, payment methodology, and radiologist workforce differ and are detailed below.

Health Systems

Canada. In the post–World War II era, Canada and the United States faced similar hospital and bed shortages. During the 1950s, Premier Tommy Douglas designed a hospitalization plan to address these shortages in rural and low-income regions in the Canadian province of Saskatchewan. The 1984 Canada Health Act mandated that “medically necessary” comprehensive services be provided in the provinces, largely at no cost. Today, Canadians in all 10 provinces and 3 territories enjoy “reasonable access to medically necessary hospital and physician services without paying out of pocket.” Administration of health services, including imaging, remains the responsibility of the provinces and territories under the 5 principles of the Canada Health Act: public administration, comprehensiveness, universality, portability, and accessibility.

All provinces and territories have a publicly funded health plan that covers all residents living in each jurisdiction for a minimum of 6–8 months annually, depending on the province or territory. The resident health card permits access to a full range of medical services, including diagnostic imaging at no or minimal out-of-pocket cost. Charging additional fees for medically necessary services is strictly forbidden by the Canada Health Act. Portability of health care services exists for Canadians who travel or work in other provinces, whereby their provincial health system compensates the delivering province at an agreed interprovincial rate.

Six of 10 Canadian provinces prohibit private insurance from covering services that are provided under the provincial health care plan, and 3 of the remaining 4 provinces allow private insurance coverage of these services but at a considerable disadvantage to physicians who decide to opt out from public payers. For example, it is illegal for a physician in Nova Scotia to charge fees that exceed the provincial public rate. Saskatchewan and New Brunswick physicians are not reimbursed if they opt-out of the provincial plan. Only Newfoundland and Labrador authorize private insurance coverage of medical services reimbursable by the public plan without any economic disincentives to opt-out of the provincial health plan. Many Canadians have supplemental insurance for uncovered services, such as prescription drugs, ambulance services, vision, and dental care.

Unlike the United States, payment of Canadian physicians is generally restricted to specialty, as defined by the Royal College of Physicians and Surgeons of Canada and recorded by the provincial medical licensing authorities—that is, in Canada, only a radiologist can provide professional radiology services. Professional fees of physicians are negotiated between the province and provincial medical association. Radiologist and MR imaging services are largely hospital-based performed in a hospital setting. Professional fees of Canadian radiologists vary between...
the Provinces and Territories of Canada as they are independently negotiated with the medical associations, Ministry of Health and providers.19

More than 90% of CT and MR imaging services are delivered in publicly funded hospitals in Canada. CT and MR imaging fees in Canadian provinces are either tied to the provincial fee schedule or negotiated on a regional basis. The remaining CT and MR imaging services are provided at private imaging clinics. According to the Canadian Agency for Drugs and Technologies in Health 2017, there are approximately 561 CT and 366 MR imaging scanners in Canada.20 Additionally, there are approximately 39 true private payer MR imaging and CT facilities across Canada, located in a subset of provinces in urban areas: British Columbia (BC) (n = 17), Alberta (n = 10), Quebec (n = 31), and Nova Scotia (n = 1).21 In Ontario, there are approximately 184 CT scanners and 120 MR imaging scanners, all in hospitals.20

**United States.** The United States has a multipayer system dependent on the private marketplace or government subsidy. Nongovernment plans include fee for service and prepay plans. There are 6 government-based plans: Medicare, Medicaid, Children’s Health Insurance Program, Tricare, Indian Health Service, and the Veterans Health Administration. Medicare plans are regionized into 7 Medicare Administrative Contractors, which standardize covered benefits and rates. There are 274 different managed Medicaid plans, each with distinct coverage policies and reimbursement.22 In 1965, President Lyndon B. Johnson signed the bill that led to Medicare and Medicaid. Medicare is federally funded by tax revenue benefiting largely the elderly (65 years or older) and individuals with disabilities. Medicaid is a federal and state cost-funded program for those with limited financial resources.23 There are 3 basic types of private health insurance plans in the United States: Health Maintenance Organizations, Preferred Provider Organizations, and Point of Service Plans, which combine Health Maintenance and Preferred Provider Organization features.24

**Spending, Work Force, Coding, and Payment**

In 2015, the United States spent $3.2 trillion on total health care, roughly $9994 per capita. Canada spent $170 billion (US dollars), approximately $4734 (in US dollars) per capita, a little more than half that of the United States. Despite these expenditures, the average life expectancy in the United States was 78.7 years, whereas it was 82.1 years in Canada in 2015 (Table 1).25–27

### Table 1: Comparing vital and economic statistics of Canada versus US in 2015

<table>
<thead>
<tr>
<th></th>
<th>Canada</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>35.8 million</td>
<td>320.9 million</td>
</tr>
<tr>
<td>Uninsured (% of total population)</td>
<td>~0%b</td>
<td>10%</td>
</tr>
<tr>
<td>Average life expectancy (yr)</td>
<td>82.1</td>
<td>78.7</td>
</tr>
<tr>
<td>Per capita health care spending</td>
<td>$4734</td>
<td>$9994</td>
</tr>
<tr>
<td>Total health care GDP</td>
<td>$170 billion</td>
<td>$3.2 trillion</td>
</tr>
<tr>
<td>Total health care spending (% of GDP)</td>
<td>$16 trillion</td>
<td>$18.1 trillion</td>
</tr>
</tbody>
</table>

**Note:** a GDP indicates gross domestic product.

b According to the Wellesley Institute report, approximately 200,000–500,000 uninsured individuals reside in Canada. They include students, workers from overseas, undocumented refugees, and newly landed immigrants.26

In 2013, the radiology workforce in the United States was 8.14 radiologists per 100,000 with 25,730 radiologists serving a total United States population of 320.9 million.28 In 2013, the Canadian per capita radiology workforce was 6.82 with 2396 radiologists, serving a Canadian population of 35.16 million.29,30

The Canadian procedure codes, including imaging, are unique and individualized to each province with updates as needed. Canadian fee codes for CT and MR imaging are organized along body systems and contrast (without contrast, with contrast, and pre and post contrast). With the exceptions of cardiac CT and CT colonography, Canadian radiologists are paid on a fee-for-service basis. (Table 2).

In the United States, the imaging procedures are standardized by the 5-digit Current Procedural Terminology code, a methodology published and owned by the American Medical Association with standing meetings by the Current Procedural Terminology Editorial Panel 3 times a year.31,32 Medicare payments are based on a relative value unit formula weighted by the work of physicians, professional liability insurance, and practice expense with a geographic variation modifier as defined by the resource-based Relative Value Scale Update Committee, which also has standing meetings 3 times a year to evaluate updates and revisions, and, when approved by the United States Congress, with a conversion factor converting the relative value units into dollars.33,34 Medicaid reimbursement and private insurance payments use the Medicare rates as a reference, but each plan individually determines the rate.35

### Table 2: Examples of differing procedure codes and professional reimbursement for an unenhanced CT of the head in the 10 provinces of Canada

<table>
<thead>
<tr>
<th>Province</th>
<th>Fee Code</th>
<th>Professional Fee (CAD$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>British Columbia</td>
<td>08690</td>
<td>44.88</td>
</tr>
<tr>
<td>Alberta</td>
<td>7113</td>
<td>49.15</td>
</tr>
<tr>
<td>Saskatchewan</td>
<td>4423</td>
<td>6.14</td>
</tr>
<tr>
<td>Manitoba</td>
<td>7113</td>
<td>49.15</td>
</tr>
<tr>
<td>Ontario</td>
<td>X400</td>
<td>43.25</td>
</tr>
<tr>
<td>Quebec</td>
<td>08259</td>
<td>33.40</td>
</tr>
<tr>
<td>New Brunswick</td>
<td>3166</td>
<td>67.00</td>
</tr>
<tr>
<td>Nova Scotia</td>
<td>1105</td>
<td>42.33</td>
</tr>
<tr>
<td>Prince Edward Island</td>
<td>8925</td>
<td>83.16</td>
</tr>
<tr>
<td>Newfoundland and Labrador</td>
<td>73800</td>
<td>55.53</td>
</tr>
</tbody>
</table>

a Benefits for noninvasive diagnostic procedures performed in hospitals and urgent care clinics are payable through hospitals or health authorities.14

b Not listed in the Payment Schedule for Insured Services Provided by a Physician, October 1, 2017, published by the Government of Saskatchewan.

In the United States, CT and MR Imaging (Advanced Imaging) Wait Times. More CT and MR imaging scanners have been approved in the past decade, but an undersupply continues to exist in every province. The combination of an inadequate number of scanners and underfunding for the operation of these scanners has led to prolonged wait times for both modalities, but especially MR imaging. This issue is occurring with an increasing demand to meet longstanding indications as well as new clinical applications (eg, breast MR imaging and CT colonography) driven by a growing and aging population. The Fraser Institute survey, based on physician recollection and self-reported methodology for “medically necessary
Table 3: Comparison of CT scanner use in Canada versus US (2016)\textsuperscript{62}

<table>
<thead>
<tr>
<th>Scans</th>
<th>Canada</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total CT use (hospital + ambulatory + other)</td>
<td>5.68</td>
<td>82</td>
</tr>
<tr>
<td>Total No. (MM)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Per 1000 people</td>
<td>156.6</td>
<td>253.8</td>
</tr>
<tr>
<td>Per scanner</td>
<td>10,561.9</td>
<td>6062.8</td>
</tr>
</tbody>
</table>

CT use

- CT in hospital setting
  - Total No. (MM) | 5.61 | 65.7 |
  - Per 1000 people | 154.4 | 203.3 |
  - Per scanner | 8973.8\textsuperscript{a} | 7373.7 |

- CT in ambulatory care setting
  - Total No. | 79.261 | 16.3 MM |
  - Per 1000 people | 2.2 | 50.4 |
  - Per scanner | 2735.3\textsuperscript{a} | 3532 |

Note: MM indicates million.
\textsuperscript{a} Data were obtained in 2012. Remainder are from 2015.

Table 4: Comparison of MRI scanner use in Canada versus US (2016)\textsuperscript{62}

<table>
<thead>
<tr>
<th>Scans</th>
<th>Canada</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total MRI use (hospital + ambulatory + other)</td>
<td>2.03</td>
<td>39</td>
</tr>
<tr>
<td>Total No. (MM)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Per 1000 people</td>
<td>55.9</td>
<td>120.7</td>
</tr>
<tr>
<td>Per scanner</td>
<td>5946.9</td>
<td>3287</td>
</tr>
</tbody>
</table>

MRI use

- MRI in hospital setting
  - Total No. (MM) | 1.8  | 16.5 |
  - Per 1000 people | 49.4  | 51.1 |
  - Per scanner | 6485.6\textsuperscript{a} | 3064.1 |

- MRI in ambulatory care setting
  - Total No. | 234,308 | 22.5 MM |
  - Per 1000 people | 6.5  | 69.6 |
  - Per scanner | 3170.9\textsuperscript{a} | 3471.2 |

\textsuperscript{a} Data were obtained in 2012. Remainder are from 2015.

Enrolled doctors, however, may operate clinics for non-medically necessary services such as cosmetic surgery.

The Uninsured. All Canadians have provincial government health insurance. There are approximately 28.2 million uninsured United States citizens (Table 1),\textsuperscript{45} 9% of the United States population according to the 2016 National Center for Health Statistics report.\textsuperscript{46} In Canada, the Aboriginal population has access difficulties; “medically necessary” services for Aboriginals are the direct responsibility of the provinces, though the federal government does provide some services for mental health, chronic conditions, and prescription drugs.\textsuperscript{47} The confusing and bureaucratic “patchwork” of coverage and the often rural location of Aboriginals have led to challenging care for this community.

Professional Liability Insurance Cost.

Professional liability insurance costs differ between Canada and the United States. California legislated a $250,000 cap on noneconomic damages and fostered the low professional liability insurance rate for radiology. State-filed malpractice premiums of diagnostic radiologists in California range from $5729 to $21,095, with an average of $10,934; low-risk physicians receive a discount up to 50% on state-filed rates.\textsuperscript{48} Greater than 90% of Canadian physicians are members of the Canadian Medical Protective Association, a mutual legal defense association that aims to decrease medical-legal risk for its members and improve the safety of health care. In BC, the annual fees (premiums) for diagnostic radiologists are near the median with other Canadian physicians, CaD$2597 (47 US dollars), to radiologists. In the United States, the malpractice claim rate of radiologists is approximately 7%,\textsuperscript{50} compared with a Canadian malpractice claim rate of 1.9% in 2013 per the Canadian Medical Protective Association report of 46 claims against radiologists.\textsuperscript{51} Nearly all 2396 radiologists in Canada were covered by the Canadian Medical Protective Association.

Advanced Imaging Equipment Funding and Use

Funding of Radiology Equipment. In Canada, each provincial ministry of health funds most hospital operating budgets, either Professional liability insurance costs differ between Canada and the United States. California legislated a $250,000 cap on noneconomic damages and fostered the low professional liability insurance rate for radiology. State-filed malpractice premiums of diagnostic radiologists in California range from $5729 to $21,095, with an average of $10,934; low-risk physicians receive a discount up to 50% on state-filed rates. Greater than 90% of Canadian physicians are members of the Canadian Medical Protective Association, a mutual legal defense association that aims to decrease medical-legal risk for its members and improve the safety of health care. In BC, the annual fees (premiums) for diagnostic radiologists are near the median with other Canadian physicians, CaD$2597 (47 US dollars), to radiologists. In the United States, the malpractice claim rate of radiologists is approximately 7%, compared with a Canadian malpractice claim rate of 1.9% in 2013 per the Canadian Medical Protective Association report of 46 claims against radiologists. Nearly all 2396 radiologists in Canada were covered by the Canadian Medical Protective Association.
directly or indirectly through Regional Health Authorities. Hospital imaging equipment purchases involve some combination of an approval of the health authority, ministry of health capital equipment funding, or funds raised through the charitable foundation of the hospital. The ministry/health authority controls operation and approval of capital purchases. In the United States, except for the Veterans Administration and large government-owned hospitals, manufacturers and providers negotiate imaging equipment costs directly with the hospitals and facilities.52

Advanced Imaging Use and Units. The variation in the volume of advanced imaging procedures in Canada and the United States is marked (Tables 3 and 4). The Organization for Economic Cooperation and Development 2016 reported 157 CT procedures per 1000 individuals in Canada and 254 CT procedures per 1000 individuals in the United States. The number of MR imaging scans in Canada was 56 per 1000, while in the United States, the number was 121 per 1000 individuals. As of 2015, the 9.48 MR imaging units per million Canadian inhabitants was dwarfed by the United States ratio of 38.96 MR imaging units per million. During the same year, Canada had 15.01 CT units per million individuals, compared with 40.98 units per million in the United States.53 The proportionally lower number of CT and MR imaging units in Canada is accompanied by the need for greater numbers of procedures per unit per day. For CT, 10,600 examinations were performed each year in Canada, while this number was 3300 in the United States.55 According to the Canadian Agency for Drugs and Technologies in Health 2017 report, there were approximately 1 million neurologic CT procedures, 19.8% of 5.68 million total CT procedures.20 The United States performs 29.5 million neurologic CT procedures, 36% of 82 million (Table 5). Correcting for the population difference between the United States and Canada, the CT use/1000 Canadians is 31 (19.8% of 156.6 all CT procedures/1000 Canadians) versus 96.4 neurologic CT procedures/1000 Americans (Table 6). In comparison with Canada, the United States performs ~3 times as many neurologic CT examinations/1000 individuals. Similarly, the neurologic MR imaging use/1000 individuals in Canada is 14.2 neurologic MR imaging/1000 Canadians versus 65.2 neurologic MR imaging/1000 Americans (4.6 times higher) (Tables 6 and 7).

CONCLUSIONS

The American and Canadian systems both strive to provide health care to their citizens. The system in United States uses standardized coding within a dual system of government and private payer health insurance. The Canadian universal coverage is largely funded from tax revenue collected by the federal legalization payments and provincial government but is independently administered by provinces and territories using different coding and payment policies. The Canadian health system, including advanced imaging, is highly regulated and accounts for the per capita expenditure for health care in Canada being more than half that of the United States. As the United States struggles with increasing health care expenditures, Canada is struggling with the demand for greater access. The United States population is 10 times that of Canada, with health care costs nearly twice as high per beneficiary. The explanation of less expensive care and longer outpatient procedure wait times is complex, though perhaps due to a less adverse malpractice climate in Canada compared with the United States. Delivery in a nonprofit Canadian Government–controlled system may be an oversimplification; perhaps differing populations, poverty rates, and medical conditions served despite similar culture and origins may account for the apparent discordance of health care cost and access.


REFERENCES


63. IMV Medical Information Division. IMV 2016 CT Benchmark Report. Des Plaines; IMV; Accessed March 14, 2018

64. IMV Medical Information Division. IMV 2016 MR Benchmark Report. Des Plaines; IMV; Accessed March 4, 2018
Clinical Value of Hybrid TOF-PET/MR Imaging–Based Multiparametric Imaging in Localizing Seizure Focus in Patients with MRI-Negative Temporal Lobe Epilepsy

K. Shang, J. Wang, X. Fan, B. Cui, J. Ma, H. Yang, Y. Zhou, G. Zhao, and J. Lu

ABSTRACT

BACKGROUND AND PURPOSE: Temporal lobe epilepsy is the most common type of epilepsy. Early surgical treatment is superior to prolonged medical therapy in refractory temporal lobe epilepsy. Successful surgical operations depend on the correct localization of the epileptogenic zone. This study aimed to evaluate the clinical value of hybrid TOF-PET/MR imaging–based multiparametric imaging in localizing the epileptogenic zone in patients with MR imaging-negative for temporal lobe epilepsy.

MATERIALS AND METHODS: Twenty patients with MR imaging-negative temporal lobe epilepsy who underwent preoperative evaluation and 10 healthy controls were scanned using PET/MR imaging with simultaneous acquisition of PET and arterial spin-labeling. On the basis of the standardized uptake value and cerebral blood flow, receiver operating characteristic analysis and a logistic regression model were used to evaluate the predictive value for the localization. Statistical analyses were performed using statistical parametric mapping. The values of the standardized uptake value and cerebral blood flow, as well as the asymmetries of metabolism and perfusion, were compared between the 2 groups. Histopathologic findings were used as the criterion standard.

RESULTS: Complete concordance was noted in lateralization and localization among the PET, arterial spin-labeling, and histopathologic findings in 12/20 patients based on visual assessment. Concordance with histopathologic findings was also obtained for the remaining 8 patients based on the complementary PET and arterial spin-labeling information. Receiver operating characteristic analysis showed that the sensitivity and specificity of PET, arterial spin-labeling, and combined PET and arterial spin-labeling were 100% and 81.8%, 83.3% and 54.5%, and 100% and 90.9%, respectively. When we compared the metabolic abnormalities in patients with those in healthy controls, hypometabolism was detected in the middle temporal gyrus (P < .001). Metabolism and perfusion asymmetries were also located in the temporal lobe (P < .001).

CONCLUSIONS: PET/MR imaging–based multiparametric imaging involving arterial spin-labeling may increase the clinical value of localizing the epileptogenic zone by providing concordant and complementary information in patients with MR imaging-negative temporal lobe epilepsy.

ABBREVIATIONS: AI = asymmetry index; ASL = arterial spin-labeling; EZ = epileptogenic zone; FCD = focal cortical dysplasia; HS = hippocampal sclerosis; SPM = statistical parametric mapping; SUV = standardized uptake value; SUVr = standardized uptake value ratio; TLE = temporal lobe epilepsy

Epilepsy is a common chronic neurologic disorder characterized by recurrent spontaneous seizures. It has an incidence of 50 per 100,000 persons per year. Temporal lobe epilepsy (TLE) is the most common type of epilepsy. A published randomized trial reported that early surgical treatment is superior to prolonged medical therapy in refractory TLE. Successful operations depend on the correct localization of the epileptogenic zone (EZ). MR imaging is a powerful tool in identifying the lesions causing epilepsy, such as hippocampal sclerosis (HS). However, approximately 16% of patients with TLE have a normal MR imaging appearance. Histopathologic studies have shown that many focal cortical dysplasias (FCDs) are small or subtle and are difficult to identify visually using MR imaging. FCDs are identified as the most common histo-
pathologic abnormality in MR imaging-negative TLE. Therefore, additional functional imaging is needed for localization. The increase in neuronal activity triggered by an epileptic seizure is associated with an increase in neuronal metabolism and regional blood flow. Therefore, analyses of cerebral perfusion and the metabolic status are widely used in presurgical evaluation to identify the EZ.

[18F] FDG-PET reveals areas of interictal cerebral hypometabolism associated with epileptic activity and epileptogenic lesions and is being used for presurgical evaluation of the EZ. PET has been shown beneficial in patients with MR imaging negative epilepsy or nonconcordant electroencephalography and neuroimaging findings. Moreover, FDG-PET-positive, MR imaging-negative TLE patients had excellent surgical outcomes. Arterial spin-labeling (ASL) MR imaging is a noncontrast perfusion technique increasingly used to evaluate the brain cerebral blood flow (CBF). It uses the magnetically labeled arterial blood as an endogenous contrast agent. Previous studies have shown increased perfusion during the perictal period, decreased perfusion during the postictal period, and hemispheric hemodynamic asymmetry in patients with epilepsy. Thus, ASL perfusion MR imaging might help to confirm the location and extent of the EZ in the presurgical work-up of epilepsy.

Previous studies were obtained with separate PET, MR imaging, and coregistered PET/MR imaging. Hybrid PET/MR imaging provides simultaneous acquisition in the same physiologic or pathophysiologic states and may be advantageous in various neurologic disorders. A few studies with PET/MR imaging that evaluated refractory focal epilepsy have been reported. Ding et al showed specific patterns of metabolic abnormality and asymmetry in patients with epilepsy, which may help to understand the etiopathogenesis. Shin et al evaluated the improved accuracy of hybrid PET/MR imaging compared with separate MR imaging and PET/CT in localizing the EZ. Boscolo Galazzo et al reported a high level of correlation between PET and ASL and found that the statistical approach could improve the diagnostic value. However, as far as we know, no study has revealed the relationship between metabolism and perfusion with the EZ confirmed by histopathology in hybrid PET/MR imaging. This pilot study aimed to investigate the clinical value of localizing the EZ using hybrid TOF-PET/MR imaging-based multiparametric imaging in patients with MR imaging-negative TLE.

**MATERIALS AND METHODS**

**Subjects**

Twenty patients (13 males and 7 females; 26.5 ± 9.48 years of age) who were diagnosed with medically refractory TLE and who underwent standard anterior temporal lobectomy with en bloc resection between September 2015 and April 2017 at Xuanwu Hospital were retrospectively evaluated. The inclusion criteria were as follows: 1) TLE confirmed by histopathology, 2) no structural abnormalities revealed on MR images, and 3) no seizure for at least 72 hours before PET/MR imaging. The exclusion criteria were as follows: 1) abnormalities observed in MR images, and 2) serious medical or nonepilepsy neurologic disorders. All patients received the same standardized preoperative assessment protocol, including a detailed clinical and medical history, neuropsychologic examination, long-term ictal and interictal scalp electroencephalography surveillance, and anatomic and functional neuroimaging studies, which included a brain MR imaging, PET/CT, and hybrid PET/MR imaging. Patient demographics, clinical characteristics, surgical approach, and histopathologic findings are summarized in Table 1.

For comparison, 10 healthy controls were recruited (6 men and 4 women, 44.1 ± 8.0 years of age). All healthy controls were right-handed with normal brain MR imaging findings. No familial or personal history of neurologic or psychiatric diseases was noted. The clinical study was approved by the ethics committee of Xuanwu Hospital, and written informed consent was obtained from all participants before the study.

**Table 1: Study population demographics**

<table>
<thead>
<tr>
<th>Patient No.</th>
<th>Sex</th>
<th>Age (yr)</th>
<th>Age at Onset (yr)</th>
<th>Epilepsy Duration (yr)</th>
<th>Seizure Frequency</th>
<th>Side of Operation</th>
<th>Histopathology</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>26</td>
<td>14</td>
<td>12</td>
<td>4–5 Times/mo</td>
<td>Left</td>
<td>Left temporal region (FCD type I)</td>
</tr>
<tr>
<td>2</td>
<td>M</td>
<td>28</td>
<td>7</td>
<td>21</td>
<td>1–2 Times/day</td>
<td>Left</td>
<td>Left temporal region (FCD type I)</td>
</tr>
<tr>
<td>3</td>
<td>M</td>
<td>31</td>
<td>26</td>
<td>5</td>
<td>2–4 Times/mo</td>
<td>Left</td>
<td>Left temporal region (FCD type I)</td>
</tr>
<tr>
<td>4</td>
<td>M</td>
<td>23</td>
<td>14</td>
<td>9</td>
<td>2–5 Times/mo</td>
<td>Left</td>
<td>Left temporal region (FCD type I)</td>
</tr>
<tr>
<td>5</td>
<td>M</td>
<td>29</td>
<td>15</td>
<td>14</td>
<td>6–8 Times/mo</td>
<td>Left</td>
<td>Left temporal region (FCD type IIIa-HS)</td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>24</td>
<td>12</td>
<td>12</td>
<td>2–3 Times/day</td>
<td>Left</td>
<td>Left temporal region (FCD type IIIa-HS)</td>
</tr>
<tr>
<td>7</td>
<td>F</td>
<td>38</td>
<td>17</td>
<td>21</td>
<td>3–5 Times/mo</td>
<td>Left</td>
<td>Left temporal region (FCD type IIIa-HS)</td>
</tr>
<tr>
<td>8</td>
<td>F</td>
<td>27</td>
<td>2</td>
<td>25</td>
<td>2–3 Times/mo</td>
<td>Right</td>
<td>Right temporal region (FCD type IIIa-HS)</td>
</tr>
<tr>
<td>9</td>
<td>M</td>
<td>23</td>
<td>16</td>
<td>7</td>
<td>3–4 Times/mo</td>
<td>Right</td>
<td>Right temporal region (FCD type IIIa-HS)</td>
</tr>
<tr>
<td>10</td>
<td>F</td>
<td>14</td>
<td>8</td>
<td>6</td>
<td>2–3 Times/day</td>
<td>Left</td>
<td>Left temporal region (FCD type II)</td>
</tr>
<tr>
<td>11</td>
<td>M</td>
<td>16</td>
<td>10</td>
<td>6</td>
<td>6–9 Times/mo</td>
<td>Left</td>
<td>Left temporal region (FCD type II)</td>
</tr>
<tr>
<td>12</td>
<td>M</td>
<td>26</td>
<td>14</td>
<td>12</td>
<td>3–4 Times/mo</td>
<td>Right</td>
<td>Right temporal region (FCD type II)</td>
</tr>
<tr>
<td>13</td>
<td>M</td>
<td>29</td>
<td>5</td>
<td>24</td>
<td>1 Time/day</td>
<td>Right</td>
<td>Right temporal region (FCD type II)</td>
</tr>
<tr>
<td>14</td>
<td>F</td>
<td>16</td>
<td>13</td>
<td>3</td>
<td>1 Time/mo</td>
<td>Right</td>
<td>Right temporal region (FCD type II)</td>
</tr>
<tr>
<td>15</td>
<td>M</td>
<td>24</td>
<td>11</td>
<td>13</td>
<td>2–4 Times/mo</td>
<td>Right</td>
<td>Right temporal region (FCD type II)</td>
</tr>
<tr>
<td>16</td>
<td>M</td>
<td>35</td>
<td>12</td>
<td>23</td>
<td>3–5 Times/mo</td>
<td>Right</td>
<td>Right temporal region (FCD type II)</td>
</tr>
<tr>
<td>17</td>
<td>M</td>
<td>19</td>
<td>14</td>
<td>5</td>
<td>3 Times/mo</td>
<td>Right</td>
<td>Right temporal region (FCD type II)</td>
</tr>
<tr>
<td>18</td>
<td>M</td>
<td>22</td>
<td>15</td>
<td>7</td>
<td>7–8 Times/mo</td>
<td>Right</td>
<td>Right temporal region (FCD type II)</td>
</tr>
<tr>
<td>19</td>
<td>M</td>
<td>17</td>
<td>1</td>
<td>16</td>
<td>1–2 Times/mo</td>
<td>Left</td>
<td>Left temporal region (FCD type Ic)</td>
</tr>
<tr>
<td>20</td>
<td>F</td>
<td>52</td>
<td>12</td>
<td>40</td>
<td>2–3 Times/mo</td>
<td>Left</td>
<td>Left temporal region (FCD type Ic)</td>
</tr>
</tbody>
</table>
Surgical Approach and Outcome
All patients underwent standard anterior temporal lobectomy, including 11 who had an operation on the left side and 9 who had an operation on the right side. Nine patients were evaluated after a postoperative follow-up period of at least 1 year. The Engel classification was used in the evaluation of surgical outcomes (class I: free of disabling seizures; class II: rare disabling seizures [almost seizure-free]; class III: worthwhile improvement; and class IV: no worthwhile improvement).23 There were 7 patients with Engel class I and 2 with Engel class II results.

Histopathologic Examination and Findings
The laboratory protocols used for specimen preparation were similar in all samples as described previously.24 Tissue sections were fixed overnight in 10% buffered formalin and then oriented and cut perpendicularly to the cortical surface. Following routine paraffin embedding, 4- or 8-μm-thick sections were stained with hematoxylin-eosin and Luxol fast blue. The selected sections were also tested for immunohistochemical reactions. Histopathologic diagnoses were made by an experienced neuropathologist. There were 15 cases of FCD type I and 5 of FCD type III–HS.

PET/MR Imaging Acquisition
Each subject was scanned with an imaging protocol consisting of injecting FDG (mean, 264.0 ± 46.8 MBq), with the scan being initiated 40 minutes after the injection. All studies were performed on an integrated simultaneous Signa PET/MR imaging system (GE Healthcare, Milwaukee, Wisconsin). The PET bed position included a simultaneous 18-second 2-point Dixon scan for MR imaging–based attenuation correction as well as additional diagnostic MR images. The attenuation map for the head frame was created with atlas-based methods, in which a coregistered MR imaging–CT atlas dataset was used to derive a pseudo-CT image from the patient’s MR image.

The ordered subsets expectation maximization algorithm was used for PET image reconstruction, and the detailed parameters were the following: 8 iterations, 32 subsets, and full width at half maximum of 3.0 mm. The PET images were reconstructed to a matrix of 192 × 192, and the slice thickness was 2.44 mm. Imaging parameters of the ASL perfusion MR images were as follows: TR = 4852 ms, TE = 20.7 ms, TI = 2025 ms, FOV = 24 × 24 cm², gap = 0 mm, matrix size = 512 × 8, postlabel delay = 2 seconds. CBF maps were generated on-line from the console using an AW4.6 Workstation (GE Healthcare). We also obtained routine anatomic acquisitions: axial T2-weighted fast spin-echo (TR = 9600 ms, TE = 149 ms, matrix size = 256 × 256, slice thickness = 3.0 mm, gap = 1.0 mm); axial T1-weighted fast spin-echo (TR = 3286 ms, TE = 24 ms, matrix size = 288 × 256, slice thickness = 3.0 mm, gap = 1.0 mm).

Visual Inspection
All PET and CBF maps were evaluated in a blinded manner by at least 2 experienced neuroradiologists independently, without any knowledge of the patients, operations, histopathology, or follow-up results. They were requested to identify the brain lobes with abnormalities of metabolism and perfusion. Any disagreement between the 2 observers was resolved by consulting a third reader to reach a final consensus.

Statistical Parametric Mapping Analysis of PET and ASL Data
All standardized uptake value (SUV) and CBF images were processed and analyzed using statistical parametric mapping (SPM8; http://www.fil.ion.ucl.ac.uk/spm/software/spm12) and Matlab 7.14 (MathWorks, Natick, Massachusetts) with a dedicated in-house code developed for this study. PET images were spatially normalized to the Montreal Neurological Institute template space. The images were then smoothed using an isotropic Gaussian kernel with a full width at half maximum of 8 mm for all directions. PET data in the Montreal Neurological Institute space were transformed into maps representing the standardized uptake value ratio (SUVr), which were normalized by scaling to a common value (50) for all scans. Similarly, each CBF map was registered to a specific template in the Montreal Neurological Institute space with a resolution of 2 × 2 × 2 mm³ using a nonlinear registration and then smoothed using an 8-mm full width at half maximum Gaussian kernel. According to the histopathologic findings, FDG-PET SUV images and ASL MR imaging perfusion CBF maps of the right hemisphere needed to be left-right flipped for further data analysis to ensure a homogeneous group with all patients having the EZ on the same side.

Analysis of the Asymmetry Index of FDG-PET and ASL MR Imaging
An asymmetry index (AI), which was used for detecting left-right asymmetries, was calculated for the FDG-PET SUV and ASL MR imaging CBF based on the following equation: AI = 2 (Left − Right) / (Left + Right).25 The normalization of the SUV and CBF images allowed identification of left-right asymmetries in the cerebral hemispheres; voxelwise AIs for both the SUV and CBF maps of each patient were then calculated. After we calculated the mean (μm) and SD (SDm) of the overall AI map, a voxelwise AI z score map (Zm) was derived as follows: Zm = (VAlm − μm) / SDm. The metabolic and perfusion asymmetries were evaluated by referring to the normal distribution values of the healthy controls.

Statistical Analysis
A κ test was used to determine the degree of concordance between the 2 readers. Correlation analysis between regional SUVr and CBF was used, and the Spearman rank correlation coefficient was obtained. On the basis of the values of the SUVr and CBF, receiver operating characteristic analysis was used to assess the predictive value of each parameter. The highest area under the curve and sensitivity and specificity were obtained. A logistic regression model with stepwise regression was used to select the optimal model for the location of the EZ. Statistical analyses were performed using SPSS 21.0 (IBM, Armonk, New York). A P value < .05 was considered statistically significant.

For statistical parametric mapping (SPM) analysis, the 2-sample t test was used for comparing the difference in SUV,
CBF, and AI in the whole brain between patients and healthy controls. Hypometabolism, hypoperfusion, and AI were regarded as statistically significant using an uncorrected \( P \) value of < .001 with a minimum cluster size of 50 voxels.

**RESULTS**

The FDG-PET, ASL perfusion MR imaging, and histopathologic findings are summarized in Table 2. The concordance between the 2 neuroradiologists in the interpretation of the PET and ASL CBF maps was good (\( \kappa = 0.86, 0.74 \)).

**Comparison of Histopathologic Findings with PET and ASL**

Hypometabolism was observed by visual assessment of PET images in 19/20 patients. Concordant lateralization with the EZ was confirmed by histopathology in 18/20 patients, and 15/18 patients showed focal hypometabolism concordant with the localization of the EZ. The remaining 3/18 patients had at least 1 hypometabolic area, which showed partial concordance with histopathologic findings.

In ASL perfusion MR imaging, the readers agreed on the presence of hypoperfusion in 14 of 20 patients. In the 14 patients, 11 were in complete agreement with the localization of the EZ. In the other 3 patients, ASL and histopathologic findings had partial concordance with 1 common lobe identified.

The best performance was FDG-PET SUVr with an area under the curve of 0.932, a sensitivity of 100%, and a specificity of 81.8%, followed by ASL perfusion MR imaging CBF with an area under the curve of 0.636, a sensitivity of 83.3%, and a specificity of 54.5% (Fig 1).

**Evaluation of Combined FDG-PET and ASL MR Imaging in the Location of EZ**

A complete concordance in lateralization and localization among the PET, ASL, and histopathologic results was found in 12/20 patients. A representative image set is shown in Fig 2. Concordance of histopathologic results with either FDG-PET or ASL MR imaging was also obtained in the remaining 8 patients.

Six of 8 patients (patients 3, 5, 8, 14, 19, and 20) showed normal perfusion in ASL images; however, concordant hypometabolic areas in PET images were observed. A representative patient with normal finding in ASL image and hypometabolism in PET images is shown in Fig 3. In addition, patient 1 showed a PET image with findings that seemed normal on visual inspection; however, focal hypoperfusion was identified as the EZ by histopathology. Bilateral brain regions (right temporal and left frontal lobe) with hypometabolism were seen in patient 13. However, only the right temporal lobe with hypoperfusion was identified as the EZ (Fig 4).

Scatterplots of the mean SUVr and CBF values from FDG-PET and ASL MR imaging in a series of ROIs with hypometabolism and hypoperfusion are presented in Fig 5. The correlation between PET and ASL across 20 patients was good (\( r = 0.587, P < \)

---

**Table 2: Summary of findings of PET and ASL in hybrid PET/MR imaging and histopathology**

<table>
<thead>
<tr>
<th>Patient No.</th>
<th>PET Findings in PET/MR Imaging</th>
<th>ASL Findings in PET/MR Imaging</th>
<th>Histopathology</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Normal</td>
<td>Left temporal</td>
<td>Left temporal region (FCD type I)</td>
</tr>
<tr>
<td>2</td>
<td>Left parietal and temporal</td>
<td>Left frontal, parietal, and temporal</td>
<td>Left temporal region (FCD type I)</td>
</tr>
<tr>
<td>3</td>
<td>Left temporal</td>
<td>Negative</td>
<td>Left temporal region (FCD type Ib)</td>
</tr>
<tr>
<td>4</td>
<td>Left temporal</td>
<td>Left temporal</td>
<td>Left temporal region (FCD type Ib)</td>
</tr>
<tr>
<td>5</td>
<td>Left temporal</td>
<td>Negative</td>
<td>Left temporal region (FCD type Ila-HS)</td>
</tr>
<tr>
<td>6</td>
<td>Left temporal</td>
<td>Left temporal</td>
<td>Left temporal region (FCD type Ila-HS)</td>
</tr>
<tr>
<td>7</td>
<td>Left frontal, parietal, and temporal</td>
<td>Left frontal, parietal, and temporal</td>
<td>Left temporal region (FCD type Ila-HS)</td>
</tr>
<tr>
<td>8</td>
<td>Right temporal</td>
<td>Negative</td>
<td>Right temporal region (FCD type Ila-HS)</td>
</tr>
<tr>
<td>9</td>
<td>Right temporal</td>
<td>Right temporal</td>
<td>Right temporal region (FCD type I)</td>
</tr>
<tr>
<td>10</td>
<td>Left temporal</td>
<td>Left temporal</td>
<td>Left temporal region (FCD type I)</td>
</tr>
<tr>
<td>11</td>
<td>Left frontal, parietal, and temporal</td>
<td>Left frontal, parietal, and temporal</td>
<td>Left temporal region (FCD type I)</td>
</tr>
<tr>
<td>12</td>
<td>Right temporal</td>
<td>Right temporal</td>
<td>Right temporal region (FCD type I)</td>
</tr>
<tr>
<td>13</td>
<td>Left frontal and right temporal</td>
<td>Right temporal</td>
<td>Right temporal region (FCD type I)</td>
</tr>
<tr>
<td>14</td>
<td>Right temporal</td>
<td>Negative</td>
<td>Right temporal region (FCD type Ila-HS)</td>
</tr>
<tr>
<td>15</td>
<td>Right temporal</td>
<td>Left temporal</td>
<td>Right temporal region (FCD type lc)</td>
</tr>
<tr>
<td>16</td>
<td>Right temporal</td>
<td>Right temporal</td>
<td>Right temporal region (FCD type lc)</td>
</tr>
<tr>
<td>17</td>
<td>Right temporal</td>
<td>Right temporal</td>
<td>Right temporal region (FCD type lc)</td>
</tr>
<tr>
<td>18</td>
<td>Right temporal</td>
<td>Right temporal</td>
<td>Right temporal region (FCD type lc)</td>
</tr>
<tr>
<td>19</td>
<td>Right temporal</td>
<td>Negative</td>
<td>Left temporal region (FCD type lc)</td>
</tr>
<tr>
<td>20</td>
<td>Left temporal</td>
<td>Negative</td>
<td>Left temporal region (FCD type lc)</td>
</tr>
</tbody>
</table>
Significantly improved logistic regression models were identified for combining FDG-PET SUVr and ASL perfusion MR imaging CBF to predicate localization with a sensitivity of 100% and a specificity of 90.9% (Fig 1).

**SPM Analyses of the Patient Group**

Figure 6 illustrates the results from the statistical SPM analysis between patients and healthy controls in FDG-PET SUV images. The interictal metabolic activities in patients were compared with those in healthy controls using voxelwise normalization methods: Hypometabolism of the middle temporal gyrus was mainly detected, which reached perfect agreement with the histopathologic findings ($P < .001$). As discussed, metabolic abnormalities were mainly identified in the temporal lobe region by visual inspection, which was concordant with the results of SPM analysis.

A more focal area of functional alteration using SPM analysis was detected. Similarly, statistical analysis automatically identified the areas of decreased perfusion in all patients compared with the healthy controls. However, no areas with a statistically significant hypoperfusion were detected in the patient group.

The $Z_{AI}$ map of the assessment of metabolic and perfusion abnormalities is shown in Fig 7. The metabolism asymmetry mainly located in the middle temporal gyrus was found to be statistically significant compared with that in the control group, showing agreement with the histopathologic findings ($P < .001$). When we compared the patient group with the control group, a statistically significant asymmetry of perfusion was located in the superior temporal gyrus and insula, in line with the histopathologic results ($P < .001$).
DISCUSSION

This study demonstrated the value of hybrid PET/MR imaging in localizing the EZ in patients with MR imaging-negative TLE using the simultaneous acquisition of PET and ASL. The metabolic abnormalities seen in the PET images and abnormal perfusion in ASL images corresponded well with the EZ confirmed by histopathology. Furthermore, PET and ASL could provide concordant and complementary information in localizing the EZ.

The value of localization using FDG-PET based on visual inspection varied from 36% to 73% in extratemporal epilepsy and reached up to 90% exclusively in temporal lobe cases. The maximum value for clinical management of FDG-PET can be achieved in patients with MR imaging-negative TLE because it can correctly localize lesions in 80% of the patients. For FCD, FDG-PET is more sensitive than MR imaging. Our study found that the sensitivity and specificity of FDG-PET were 100% and 81.8%, respectively. There were 3 patients who showed a widespread area of hypometabolism extending beyond the EZ and 2 patients who showed incorrect lateralization and localization, in agreement with the findings of previous studies.

Many previous ASL studies have reported interictal hypoperfusion at the site of seizure in patients with TLE and found that interictal ASL might help in localizing the EZ in FCD. The recommended postlabel delay was used in this study to ensure that the brain parenchyma was sufficiently perfused, which could help to avoid false reduction of relative CBF as estimated with ASL. The results of the ASL perfusion MR imaging study showed that the sensitivity and specificity were 83.3% and 54.5%, respectively. The hypoperfusion regions extending beyond the EZ in 3 patients were likely due to confounding factors, such as age, leukoaraiosis, and so forth.

Previous studies reported that areas of hypoperfusion on the ASL MR imaging were well-associated with the hypometabolic areas on the PET. Our study with simultaneous acquisition of FDG-PET and ASL MR imaging also showed good overall correlation between them. Combined FDG-PET and ASL MR imaging was studied to find the optimal model for evaluating the value of EZ localization. Receiver operating characteristic analysis and logistic regression models are widely used to evaluate and optimize the performance of clinical diagnostic tests, especially for the predictive value of neurologic diseases. A combined PET and ASL model was identified with a higher area under the curve compared with PET or ASL only. Moreover, the specificity of combined FDG-PET and ASL MR imaging was improved, which suggested that multiparametric imaging is potentially valuable for EZ localization. In terms of individual patients, there were 6 patients with hypometabolism in PET but normal findings in ASL and 2 cases showing that ASL can also provide complementary values. These findings suggest that the combined use of PET and ASL can yield excellent performance in lateralization and localization.

Multiparametric imaging is now widely practiced in presurgi-
ASL had excellent performance. However, the nonsynchronized and nonsimultaneous acquisition of different functional parameters might lead to potential biases in the EZ localization. The hybrid PET/MR imaging can provide simultaneous acquisition of PET and ASL in the same physiologic or pathophysiologic states.

Although visual analysis is widely used in clinical practice, it is subjective and relies on the knowledge and experience of physicians. SPM has proved to be an objective and useful method for quantitative analysis in epilepsy with FDG-PET and ASL-MR imaging. Metabolic abnormalities were mainly found in the middle temporal gyrus when comparing patients with healthy controls. However, no statistically significant hypoperfusion was observed in the qualitative assessment of CBF maps, potentially due to the limiting size of the patient cohort. In addition, the AI is also suggested for analysis. Previous studies used voxelwise or region-wise asymmetries to evaluate metabolism and perfusion abnormalities between the 2 hemispheres. Our results show that the asymmetries of metabolism and perfusion were mainly located in the temporal lobe. These findings suggest that PET and ASL had excellent performance.

There were several limitations to this study. First, the number of the patients and healthy controls was relatively small, and the participants were not age-matched because the participants had to be exposed to radioactive material. Future studies with larger cohorts are needed for further confirmation of the findings. Second, visual assessment combined with SPM analysis should be used with each patient in localizing the EZ. Third, the spectrum of histopathologic changes observed in tissue specimens from the patients with MR imaging negative for epilepsy could be divided into 4 categories as follows: FCD, HS, abnormalities in white matter, and microscopically nonlesional (e.g., astrogliosis). The most common histopathologic abnormality identified is FCD. Increased numbers of heterotopic white matter neurons can be encountered in epileptic brain tissue, which are usually not visible at the MR imaging level. Five percent to ten percent of patients with histopathologically confirmed TLE with HS do not present with MR imaging abnormalities. Eight percent of all specimens were microscopically reported nonlesional or negative for epilepsy. Because the pathologic findings in MR imaging-negative TLE were all FCDs in our study, multicenter studies involving other pathologic types will be needed to make the results more comprehensive and of statistical significance in the future.

**CONCLUSIONS**

Hybrid PET/MR imaging plays an increasingly important role in the study of refractory focal epilepsy, especially with MR imaging-negative TLE. Complementary information with simultaneous acquisition of metabolism and perfusion was shown to be useful in localizing the EZ for MR imaging-negative TLE.

**ACKNOWLEDGMENTS**

We thank Judy Buchanan, Division of Nuclear Medicine, Johns Hopkins University, and Abby Barre, Laboratory for the Study of the Brain Basis of Individual Differences, Massachusetts General Hospital, for their helpful editing and suggestions.

**REFERENCES**


30. Rheims S, Jung J, Ryvlin P. Combination of PET and magnetoencephalography in the presurgical assessment of MRI-negative epilepsy. *Front Neurol* 2013;4:188 CrossRef Medline


Longitudinal Persistence of Meningeal Enhancement on Postcontrast 7T 3D-FLAIR MRI in Multiple Sclerosis

S.N. Jonas, I. Izbudak, A.A. Frazier, and D.M. Harrison

ABSTRACT

BACKGROUND AND PURPOSE: Preliminary research has demonstrated that postgadolinium 3D-FLAIR MR imaging at 7T may be a valuable tool for detecting abnormal meningeal enhancement and inflammation in MS; however, researchers have not systematically investigated its longitudinal persistence. We hypothesized that persistence of meningeal enhancement in MS varies on the basis of pattern of enhancement as well as demographic and clinical factors such as treatment status, disease phenotype, and disability score.

MATERIALS AND METHODS: Thirty-one subjects with MS were prospectively scanned before and after intravenous contrast administration at 2 time points, approximately 1 year apart. Fifteen subjects in the cohort were scanned at another time approximately 1 year later. Foci of enhancement were categorized into 4 subtypes: subarachnoid spread/fill, subarachnoid nodular, vessel wall, and dural foci. We reviewed follow-up scans to determine whether foci changed between time points and then compared persistence with demographic and clinical variables.

RESULTS: Persistence ranged from 71% to 100% at 1 year and 73% to 100% at 2 years, depending on the enhancement pattern. Subarachnoid spread/fill and subarachnoid nodular subtypes persisted less often than vessel wall and dural foci. Persistence was not significantly different between those on/off treatment and those with progressive/nonprogressive disease phenotypes. The number of persisting foci was significantly different in subjects with/without increasing Expanded Disability Status Scale scores (median, 12 versus 7.5, P = .04).

CONCLUSIONS: Longitudinal persistence of meningeal enhancement on 3D-FLAIR at 7T in MS varies by pattern of enhancement and correlates with worsening disability; however, it is not significantly different in those on/off treatment or in those with progressive/nonprogressive disease phenotypes.

ABBREVIATIONS: EDSS = Expanded Disability Status Scale; MP2RAGE = magnetization-prepared rapid acquisition of 2 gradient echoes

MS is a chronic inflammatory demyelinating disorder classically affecting white matter within the brain and spinal cord. In the past few decades, an additional pathophysiologic mechanism—meningeal inflammation—has been elucidated in MS, which is now believed to directly contribute to cortical demyelination, cortical neuroaxonal loss, microglial activation, and oligodendrocyte dysfunction. Visualization of meningeal inflammation on MR imaging has become an active, and somewhat controversial, area of recent investigation. Landmark studies have demonstrated that gadolinium-enhanced 3D-FLAIR sequences, which have long been useful for identifying meningeal infection and carcinomatosis, can also be used to image meningeal disease in the MS population. At a magnetic field strength of 3T, Absinta et al found that approximately 25% of patients with MS demonstrated leptomeningeal enhancement on gadolinium-enhanced FLAIR.

Protocols for imaging meningeal enhancement were improved by Zivadinov et al, who showed the benefit of both pre- and postcontrast acquisitions and generating subtraction images when assessing meningeal enhancement, because these techniques decrease false-positives and reduce interpretation time. Recent preliminary research has also suggested that 7T MR imaging may be more sensitive than 3T for detecting meningeal enhancement. Although no direct 3T-versus-7T comparisons have...
been made in the same study population, up to 90% of patients with MS undergoing contrast-enhanced brain MR imaging at 7T demonstrated at least 1 enhancing focus.6 This result closely approximates the 89% of patients with MS reported to show some element of leptomeningeal inflammation at postmortem examination.10,11 Given this radiologic-pathologic concordance, it is conceivable that 7T 3D-FLAIR may soon provide a noninvasive in vivo method of detecting and accurately quantifying the extent of meningeal inflammation in patients with MS.

Meningeal enhancement was noted to be a persistent phenomenon in prior 3T studies7; however, at 7T, where sensitivity for meningeal enhancement in MS appears to be significantly higher, it remains unknown whether smaller, more subtle foci of enhancement wax and wane in a predictable pattern across time or whether they remain longitudinally stable. Also unknown is the degree to which enhancement persistence with time is associated with previously described enhancement shape and morphology, including subarachnoid spread/fill and subarachnoid nodular patterns.7,9 Because prior studies have shown that the prevalence of meningeal enhancement varies with enhancement morphology,7,9 in this study, we hypothesized that persistence of meningeal enhancement in MS would vary on the basis of the morphology of enhancement as well as demographic and clinical factors such as treatment status, disease phenotype, and disability scores. Greater understanding of the imaging and clinical characteristics of meningeal enhancement is necessary if these features are to aid in the diagnosis of and prognosis for patients with MS.

MATERIALS AND METHODS

Standard Protocol Approval and Informed Consent

The institutional review boards at the authors’ institutions approved this Health Insurance Portability and Accountability Act–compliant, prospective study. Written, informed consent was obtained from all participants.

Participants

Thirty-one volunteers, 26–61 years of age, with diagnoses of relapsing-remitting MS, secondary-progressive MS, or primary-progressive MS according to the 2010 revised McDonald Criteria were recruited.12 Exclusion criteria included contraindications to contrast-enhanced MR imaging.

MR Imaging Protocol

Study participants were prospectively scanned at 2 time points approximately 1 year apart on a 7T Achieva scanner (Philips Healthcare, Best, the Netherlands) with a volume-transmit/32-channel head coil (Nova Medical, Wilmington, Massachusetts). Fifteen patients in the cohort were scanned at an additional third visit approximately 1 year later. Scans were obtained between September 9, 2014, and August 21, 2017. Dielectric padding was used for improved image homogeneity.13 Scanning parameters are listed in Table 1. Images were acquired before the administration of contrast and again after the intravenous administration of 0.1 mmol/kg of gadoteridol (ProHance; Bracco Diagnostics, Princeton, New Jersey). Magnetization-prepared rapid acquisition of 2 gradient echoes (MP2RAGE) images were initiated approximately 3 minutes after contrast administration, and magnetization-prepared FLAIR images were initiated approximately 20 minutes after contrast administration.

Table 1: MRI sequence parameters

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Resolution (mm)</th>
<th>TR</th>
<th>TI</th>
<th>TE</th>
<th>Parallel Imaging</th>
<th>Flip Angle</th>
<th>Time (min:sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MP2RAGE</td>
<td>0.7 × 0.7 × 0.7</td>
<td>TR&lt;sub&gt;volume&lt;/sub&gt; = 8.25 s, TR&lt;sub&gt;TE&lt;/sub&gt; = 6.9 ms, T&lt;sub&gt;1&lt;/sub&gt;&lt;sub&gt;e&lt;/sub&gt; = 1 s, T&lt;sub&gt;2&lt;/sub&gt;&lt;sub&gt;e&lt;/sub&gt; = 3.3 s</td>
<td>1.97 ms</td>
<td>SENSE = 2 × 2</td>
<td>FA&lt;sub&gt;1&lt;/sub&gt; = 7°</td>
<td>FA&lt;sub&gt;2&lt;/sub&gt; = 5°</td>
<td>9.46</td>
</tr>
<tr>
<td>MPFLAIR</td>
<td>0.7 × 0.7 × 0.7</td>
<td>8000 ms</td>
<td>2077 ms</td>
<td>400 ms</td>
<td>SENSE = 2 × 3</td>
<td>90°</td>
<td>10.48</td>
</tr>
</tbody>
</table>

Note:—SENSE indicates sensitivity encoding; MPFLAIR, magnetization-prepared FLAIR; FA, flip angle; TFE, turbo field echo.

Image Processing and Analysis

MP2RAGE images were processed to create a T1-weighted image and a T1 map.14 Images were then manipulated using Medical Image Processing, Analysis, and Visualization (Version 7.2; http:// mipav.cit.nih.gov). Magnetization-prepared FLAIR images underwent N4 inhomogeneity correction before analysis.15 Pre- and postcontrast magnetization-prepared FLAIR images were registered to the precontrast T1 map. A magnetization-prepared FLAIR subtraction image was created by direct subtraction of the registered pre- and postcontrast images.

The magnetization-prepared FLAIR subtraction image, alongside the pre- and postcontrast magnetization-prepared FLAIR images, was reviewed by 2 independent raters (a postgraduate year 4 radiology resident and an academic MS neurologist) who were blinded to subject identity, disease state, and treatment regimen. Hyperintensities noted on the subtraction image were located on anatomic images and demarcated, if present, in the meningeal space on postcontrast images only. All foci were localized in 3 orthogonal planes before notation. When needed, coregistered MP2RAGE T1-weighted images were used for confirmation of anatomic locations. The pattern of enhancement was categorized on the basis of location and morphology and stratified into 1 of 4 subtypes. Subarachnoid spread/fill foci were characterized by the presence of contrast in the subarachnoid space distributed in an amorphous manner (Fig 1A). Subarachnoid nodular foci were characterized by small, round areas of contrast, usually 1–2 voxels (0.7–1.4 mm) and were adherent to the pial surface (Fig 1B). Vessel wall enhancement was characterized by contrast outlining the outer margin of a large meningeal vessel with a signal void in the lumen of the vessel, often resulting in a characteristic tram-track appearance (Fig 1C). Dural foci were characterized by discrete regions of enhancement clearly situated along the dural surface without extension into the subarachnoid space (Fig 1D).

Following both independent reviews, a consensus review was performed under the supervision of an expert third rater (an academic neuroradiologist with 12 years of experience). After consensus review, follow-up images underwent linear registration (with 9 df) to baseline images. Consensus regions of contrast enhancement on baseline images were reviewed for their presence or absence on follow-up scans. The total number of foci per subject that persisted between scans was compared among different morphologies of meningeal enhancement and correlated with previous descriptions.
with demographic and clinical data. Additionally, the proportion of baseline foci per subject that persisted to follow-up scans was also compared with morphologic, demographic, and clinical factors.

**Disability Measures**

The Kurtzke Expanded Disability Status Scale (EDSS) was used to characterize disability. EDSS progression was defined as an increase of the EDSS score at follow-up of \( \geq 1.0 \) if the baseline EDSS score was \( \geq 5.0 \) or an increase of \( \geq 0.5 \) if the baseline EDSS score was \( \geq 5.0 \). The Modified Fatigue Impact Scale was used to assess MS-related fatigue. The Symbol Digit Modalities Test was used to assess cognitive functioning. These tests were administered at each study visit.

**Statistical Analysis**

Statistical analysis was performed in Stata 10.1 IC (StataCorp, College Station, Texas). Nonparametric testing was used due to the non-normal distribution of data. We performed group comparisons for demographic and clinical variables using the Wilcoxon rank sum statistic. We computed the Spearman rank correlation for correlation testing. All statistical tests were performed with a significance threshold of \( P < 0.05 \). Due to the small sample size and exploratory nature of this study, adjustment for multiple comparisons was not performed.

**RESULTS**

We recruited 31 patients with MS; most had the relapsing-remitting MS phenotype (\( n = 21, 68\% \)), though 7 subjects had secondary-progressive MS (23%) and 3 subjects had primary-progressive MS (10%) (Table 2). No subject had a comorbid neuroinflammatory disorder. Most subjects were on disease-modifying therapy (\( n = 25, 81\% \)). This was a relatively stable and moderately disabled patient population with a median of 0 (range, 0–3) relapses in the year before enrollment and a median Expanded Disability Status Scale score of 3 (range, 1–6.5). Enhancing white matter lesions were seen in 3 subjects on review of
T1-weighted images, with 2 subjects having 1 enhancing lesion and 1 subject having 2 enhancing lesions.

At baseline, a total of 284 enhancing foci were identified across all 31 subjects. Table 3 lists the anatomic distribution of these foci within the brain. Most (>98%) of the observed foci were located supratentorially. Figure 2 shows the percentage of enhancing meningeal foci identified at baseline that persisted at later time points. Figures 3 and 4 provide examples of persisting and resolving enhancing meningeal foci from each group. Table 4 and Online Tables 1–3 compare the persistence of meningeal enhancement with demographic and clinical variables. We found no significant difference in the total number or proportion of longitudinally persistent enhancing meningeal foci between those on or off treatment or between those with progressive phenotypes (primary-progressive MS and secondary-progressive MS) versus a relapsing phenotype (relapsing-remitting MS). However, we did find significantly more (P = .04) persistent foci in Expanded Disability Status Scale progressors (median, 12; range, 1–15) compared with those who were not progressors (median, 7.5; range, 1–24). We also observed a nonsignificant trend toward a negative association (ρ = −0.31, P = .09) between the proportion of persisting foci overall and the interval change in Symbol Digit Modalities Test scores at 1 year (On-line Table 3). Surprisingly and counterintuitively, we observed a positive correlation (ρ = 0.45, P = .01) between the proportion of enhancing meningeal foci that persisted at 1 year and baseline Symbol Digit Modalities Test scores (On-line Table 3). This association was driven by the correlation (ρ = 0.48, P = .01) between the proportion of subarachnoid spread/fill subtype that persisted at 1 year and baseline Symbol Digit Modalities Test scores. We also observed 15 foci of meningeal enhancement that developed in the interval between baseline and follow-up scans. The morphologies of these 15 foci were as follows: 6 subarachnoid spread/fill, 4 subarachnoid nodular, 2 vessel wall, and 3 dural foci.

DISCUSSION

In this study, we catalogued 2 enhancement patterns described in prior analyses (subarachnoid spread/fill and subarachnoid nodular)10–12 in addition to describing 2 new patterns of meningeal enhancement for the first time: vessel wall enhancement and dural foci. Previous studies without precontrast comparison sequences excluded from consideration regions of postcontrast hyperintensity in/near the dural sinuses, large subarachnoid veins, and the basal meninges to reduce false-positives because these structures often manifest precontrast T1 or FLAIR hyperintensity.7,9 Using techniques similar to those in the recent investigation by Zivadinov et al,8 we coregistered and subtracted pre- and postcontrast magnetization-prepared FLAIR sequences in all cases. Given this protocol, we did not have to exclude any structures a priori, and we were confident in our ability to differentiate true vessel wall and dural foci enhancement from intrinsically increased signal. Which anatomic/pathologic substrates are represented by vessel wall and dural foci is unknown, but most interesting, both closely match what was recently described for visualization of meningeal lymphatics by FLAIR MR imaging.20–22 Thus, these findings may represent gadolinium absorption by lymphatic structures after leakage into the CSF. The accumulation of gadolinium signal alongside the outer wall of vessels in the vessel wall pattern is also very reminiscent of the expected location and direction of drainage along the recently described glymphatic system.23 Alternatively, it is also possible that vessel wall and dural foci could represent the reaccumulation, under hydrostatic pressure, of CSF-leaked gadolinium back into the venous system. They could also feasibly represent the actual sites of blood-brain and blood-CSF barrier disturbance secondary to ongoing inflammation.24 Because age- and sex-matched healthy controls were not used in this study, the specificity of dural and vessel wall enhancements to MS is unknown. Future

Table 2: Cohort baseline characteristics

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>N/A</th>
<th>Median (Range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at enrollment (yr)</td>
<td></td>
<td>49 (26–61)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td>11/31 Men (35%), 20/31 women (65%)</td>
</tr>
<tr>
<td>Disease subtype at enrollment</td>
<td></td>
<td>21/31 (68%) RR, 7/31 (23%) SP, 3/31 (10%) PP</td>
</tr>
<tr>
<td>Disease duration at enrollment (mo)</td>
<td></td>
<td>109 (8–461)</td>
</tr>
<tr>
<td>Patients with new relapses in past 30 days</td>
<td></td>
<td>0 (0–3)</td>
</tr>
<tr>
<td>No. of relapses in past year per subject</td>
<td></td>
<td>43 (0–78)</td>
</tr>
<tr>
<td>Modified Fatigue Impact Scale score at enrollment</td>
<td></td>
<td>50 (35–81)</td>
</tr>
<tr>
<td>Symbol Digit Modality Test score at enrollment</td>
<td></td>
<td>3 (1–6.5)</td>
</tr>
<tr>
<td>Expanded Disability Status Scale score at enrollment</td>
<td></td>
<td>12 (1–15)</td>
</tr>
<tr>
<td>Immunomodulatory treatment status at baseline</td>
<td></td>
<td>10/31 (33%) Dimethyl fumarate, 6/25 (24%) Glatiramer, 3/25 (12%) Natalizumab, 1/25 (4%) Fingolimod, 1/25 (4%) Teriflunomide, 10/31 (32%) On treatment, 6/31 (19%) Not on treatment</td>
</tr>
</tbody>
</table>

Note: RR indicates relapsing-remitting MS; SP, secondary-progressive MS; PP, primary-progressive MS.

*Median values are shown with the range of observed values in parentheses.

Table 3: Anatomic distribution within the brain of enhancing meningeal foci at baseline

<table>
<thead>
<tr>
<th>Brain Region</th>
<th>No. of Foci at Baseline</th>
<th>Percentage of Foci at Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right frontal</td>
<td>60</td>
<td>21%</td>
</tr>
<tr>
<td>Left frontal</td>
<td>64</td>
<td>22.5%</td>
</tr>
<tr>
<td>Right parietal</td>
<td>44</td>
<td>15.5%</td>
</tr>
<tr>
<td>Left parietal</td>
<td>44</td>
<td>15.5%</td>
</tr>
<tr>
<td>Right occipital</td>
<td>20</td>
<td>7.0%</td>
</tr>
<tr>
<td>Left occipital</td>
<td>24</td>
<td>8.4%</td>
</tr>
<tr>
<td>Right temporal</td>
<td>16</td>
<td>5.6%</td>
</tr>
<tr>
<td>Left temporal</td>
<td>8</td>
<td>2.8%</td>
</tr>
<tr>
<td>Right cerebellum</td>
<td>2</td>
<td>0.7%</td>
</tr>
<tr>
<td>Left cerebellum</td>
<td>2</td>
<td>0.7%</td>
</tr>
</tbody>
</table>

In this study, we coregistered and subtracted pre- and postcontrast magnetization-prepared FLAIR sequences in all cases. Given this protocol, we did not have to exclude any structures a priori, and we were confident in our ability to differentiate true vessel wall and dural foci enhancement from intrinsically increased signal.
work is needed to determine whether such findings are specific to MS, neuroinflammatory disease in general, or are seen in all patients.

We found no significant difference in the total number or proportion of persisting meningeal enhancement per subject between those on treatment and those off treatment. Lack of a significant difference between subjects on/off treatment may, in part, be explained by the low statistical power of our study because a relatively small number of untreated subjects were included. However, the lack of difference is not surprising because prior studies have also failed to show differences in meningeal enhancement between those who are and are not taking disease-modifying medications. Our data reinforce the notion that current immunomodulatory medications may not adequately control meningeal inflammation. Of note, none of our subjects received a course of corticosteroids during the study. However, 10 of 31 subjects in this cohort switched between disease-modifying therapies from baseline to follow-up scans, including 2 subjects who changed to rituximab and 1 subject who switched to alemtuzumab—both monoclonal antibodies that impact B-cell function. Despite such changes, most foci remained stable. Given the sample size of this report, we would not want to comment on the persistence (or lack thereof) of foci with individual therapy changes because conclusions from 1 or 2 examples would not be generalizable. Future comparative studies are needed to determine whether changes in any specific disease-modifying therapies or monoclonal regimens alter the longitudinal persistence of meningeal enhancement.

Surprisingly, we did not detect a significant difference in the persistence of meningeal enhancement between MS subjects with progressive phenotypes (primary-progressive MS and secondary-progressive MS) and those with re-
This finding runs counter to the previously proposed theory that meningeal enhancement may be a substrate specific to progressive MS, with the associated cortical demyelination and volume loss representing a distinctly late marker of disease. Indeed, previous studies have shown that the presence of leptomeningeal enhancement at 3T was 1.7-fold higher in progressive MS compared with relapsing-remitting MS, and postmortem findings of meningeal inflammation were more profound in those with secondary-progressive MS. However, our 7T data showed no difference in the frequency of longitudinal persistence of meningeal enhancement between patients with MS with progressive and relapsing phenotypes.

Although the persistence of enhancing foci was not related to clinical phenotype, we did detect a significant relationship between the persistence of foci (especially subarachnoid spread/fill foci) at 1 year with disability progression (by the Expanded Disability Status Scale) during the same time period. If postcontrast meningeal enhancement on magnetization-prepared FLAIR is indeed representative of meningeal inflammation, this finding may indicate that persistent rather than transient meningeal inflammation is required to affect prognosis in patients with MS. This notion is supported by postmortem data showing that the development of structures that support ongoing inflammation, such as ectopic lymphoid follicular tissue, in the meninges of patients with MS is associated with earlier onset of disease, shorter diagnosis-death interval, and more severe cortical pathology. If this relationship can be confirmed in larger studies, perhaps the elimination of persistently enhancing meningeal foci can become a target outcome for patients with MS.

While we found a significant relationship between changes in Expanded Disability Status Scale scores and the persistence of enhancement, we were unable to detect any similar association between imaging findings and MS-related fatigue or cognitive deficits (On-line Tables). This difference may be due to the small sample size of our study or the short follow-up duration. Still to be investigated is the rate at which foci of meningeal enhancement develop with time, whether the rate of development is associated with enhancement morphology, and whether the rate of development is associated with demographic and clinical parameters. This study focused on subjects with MS only; future work is needed to elucidate the rate of longitudinal persistence of meningeal enhancement in other neuroinflammatory diseases. Finally, our study is limited by possible false discovery because we did not perform multiple-comparison correction, given the exploratory nature of this investigation. Therefore, our results will require replication before widespread acceptance of these conclusions. Despite these limitations, these preliminary results provide

![FIG 4. Examples of resolving foci of meningeal enhancement on delayed postcontrast FLAIR at 7T.](image)

<table>
<thead>
<tr>
<th>Table 4: Wilcoxon rank sum test for longitudinal persistence of meningeal enhancement versus demographic and clinical factors*</th>
</tr>
</thead>
<tbody>
<tr>
<td>On Treatment</td>
</tr>
<tr>
<td>Total No. of overall foci persisting at 1 yr per subject</td>
</tr>
<tr>
<td>Total No. of subarachnoid spread/fill foci persisting at 1 yr per subject</td>
</tr>
<tr>
<td>Total No. of subarachnoid nodular foci persisting at 1 yr per subject</td>
</tr>
<tr>
<td>Total No. of vessel wall foci persisting at 1 yr per subject</td>
</tr>
<tr>
<td>Total No. of dural foci persisting at 1 yr per subject</td>
</tr>
</tbody>
</table>

*Median values are listed with the range of observed values in parentheses.

* Criteria for EDSS progressor status is listed in the “Materials and Methods” section.

* *P < .05.
important new insight into the longitudinal activity of meningeal enhancement in MS.

CONCLUSIONS

Here we describe the results of a prospective, systematic investigation into the longitudinal persistence of meningeal enhancement in MS using 7T 3D-FLAIR. Given our pre- and postcontrast techniques, we are able to include, for the first time, vessel wall and dural foci subtypes, which persist most frequently; their appearance very closely matches recent descriptions of meningeal lymphatics or the glial lymphatics system.19-22 Longitudinal persistence of meningeal enhancement is not significantly different between those on or off immunomodulatory treatment, and there is not a significant difference in the rates of longitudinal persistence between those with progressive clinical phenotypes (primary-progressive MS and secondary-progressive MS) and those without a progressive clinical phenotype (relapsing-remitting MS). However, there is a significantly increased number of persistent foci in subjects who have worsening Expanded Disability Status Scale scores at 1 year compared with those who do not, suggesting that persistently enhancing meningo foci may be an in vivo imaging marker for ongoing meningeal inflammation causing clinical progression.

Disclosures: Samuel N. Jonas—UNRELATED: Support for Travel to Meetings for the Study or Other Purposes: 2018 Radiological Society of North America Resident Travel Grant. Comments: $500, Provision of Writing Assistance, Medicines, Equipment, or Administrative Support: Department of Radiology Pagination Editor. Izlem Izbudak—UNRELATED: Consultancy: Alexian; Comments: Neuroimyelitis Optica Relapse Adjudication Committee MRI reads; Grants/Grants Pending: Siemens, Biogen, Comments: Siemens DTI of the spinal cord prospective research study, Biogen MS PATHS MRI scan-rescan study.* Aletta A. Frazier—UNRELATED: Consultancy: American Institute for Radiologic Pathology, Comments: Section Chief, Cardiovascular Radiology; Associate Chief, Chest Radiology, American Institute for Radiologic Pathology, a program of the American College of Radiology, Silver Spring, Maryland; Employment: Clinical Professor of Radiology, University of Maryland School of Medicine, Comments: Clinical Professor of Diagnostic Radiology and Nuclear Medicine, University of Maryland School of Medicine, Baltimore, Maryland. Daniel M. Harrison—RELATED: Grant: EMD-Serono, Comments: research grant*; UNRELATED: Consultancy: EMD-Serono, Genentech, Genzyme, Biogen, Comments: honorsaria for 1-time advisory board meetings; Payment for Lectures Including Service on Speakers Bureaus: Houston Methodist Hospital; Other: American College of Physicians, Comments: textbook chapter writing and multiple-choice question development for Medical Knowledge Self Assessment Program (MKSAP). *Money paid to the institution.

REFERENCES


5. Kowarik MC, Cepok S, Sellner J, et al. CXCL13 is the major determin-
Automated Integration of Multimodal MRI for the Probabilistic Detection of the Central Vein Sign in White Matter Lesions


ABSTRACT

BACKGROUND AND PURPOSE: The central vein sign is a promising MR imaging diagnostic biomarker for multiple sclerosis. Recent studies have demonstrated that patients with MS have higher proportions of white matter lesions with the central vein sign compared with those with diseases that mimic MS on MR imaging. However, the clinical application of the central vein sign as a biomarker is limited by interrater differences in the adjudication of the central vein sign as well as the time burden required for the determination of the central vein sign for each lesion in a patient’s full MR imaging scan. In this study, we present an automated technique for the detection of the central vein sign in white matter lesions.

MATERIALS AND METHODS: Using multimodal MR imaging, the proposed method derives a central vein sign probability, πi, for each lesion, as well as a patient-level central vein sign biomarker, ψi. The method is probabilistic in nature, allows site-specific lesion segmentation methods, and is potentially robust to intersite variability. The proposed algorithm was tested on imaging acquired at the University of Vermont in 16 participants who have MS and 15 participants who do not.

RESULTS: By means of the proposed automated technique, participants with MS were found to have significantly higher values of ψ than those without MS (ψMS = 0.55 ± 0.18; ψnon-MS = 0.31 ± 0.12; P < .001). The algorithm was also found to show strong discriminative ability between patients with and without MS, with an area under the curve of 0.88.

CONCLUSIONS: The current study presents the first fully automated method for detecting the central vein sign in white matter lesions and demonstrates promising performance in a sample of patients with and without MS.

ABBREVIATIONS: CVS = central vein sign; MIMoSA = Method for InterModal Segmentation Analysis

Multiple sclerosis is an inflammatory demyelinating disease of the central nervous system characterized by lesions in the brain and spinal cord. Currently, assessment of MR imaging factors heavily in the diagnosis of MS, with much importance placed on the distribution (dissemination in space) and time course of lesions (dissemination in time)1 in patients presenting with clinical symptoms typical for MS. However, current imaging-based diagnostic criteria favor sensitivity over specificity, making misdiagnosis of MS relatively common.2,3 Misdiagnosis is especially prevalent among disorders that demonstrate white matter lesions similar to those found in MS.4,5

As a means of distinguishing MS lesions from white matter abnormalities arising from other diseases, the identification of a vein traversing the center of a lesion has been proposed as a diagnostic tool because inflammatory demyelination in the MS white matter is perivenular.6,7 The potential for this marker to be used...

Please address correspondence to Jordan D. Dworkin, Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, 218 Blockley Hall, 423 Guardian Drive Philadelphia, PA, 19104; e-mail: jdwor@pennmedicine.upenn.edu. jddwor

Indicates open access to non-subscribers at www.ajnr.org
http://dx.doi.org/10.3174/ajnr.A5765
in the diagnosis of MS has been advanced by recent developments in MR imaging pulse sequences, which have enabled detailed imaging of veins in the brain.\textsuperscript{8–10} Using these sequences, researchers have provided strong evidence that higher proportions of MS lesions show the central vein sign (CVS) compared with lesions resulting from other disease processes commonly mistaken for MS.\textsuperscript{7,11–16} This finding has been demonstrated for neuromyelitis optica spectrum disorder, systemic autoimmune diseases, cerebral small vessel disease, Susac syndrome, and migraine. While further replication in a prospective setting is still necessary, a high proportion of brain MR imaging lesions demonstrating the CVS appear to have potential as a biomarker with high specificity for MS.

Unfortunately, important barriers limit the feasibility of the clinical application of the CVS. Two such limitations are the presence of intra- and interrater variability in the subjective assessment of the CVS and the time required to adjudicate the CVS in every MR imaging lesion per patient. Recent studies have attempted to mitigate the time burden associated with CVS assessment by limiting the number of lesions examined.\textsuperscript{13,17} However, these techniques have the potential to increase variability and have generally not been as successful as the evaluation of the proportion of the CVS in all MR imaging lesions per patient.\textsuperscript{7,18} Most important, in studies that adjudicate all lesions per patient, optimal proportion cutoffs have differed across study sites and disease comparisons.\textsuperscript{7,13,18} This variability highlights the need for thorough comparison and optimization of these cutoffs across samples and diseases, yet the same issues of rater subjectivity and temporal burden make this type of research difficult. Thus, the current study introduces an algorithm for the automatic determination of the CVS in white matter lesions and presents a fully automated patient-level diagnostic biomarker. In this article, we describe the CVS-detection pipeline, present statistical measures of judgment accuracy, and discuss the implications and next steps for this line of research.

**MATERIALS AND METHODS**

**CVS Detection Algorithm**

To adjudicate the CVS for each lesion in a given participant, we perform several steps. We first present the overall summary and then address each step, with associated rationale, in detail below.

To perform the algorithm, we require a T1-weighted volume, T2-weighted FLAIR volume, and T2*-weighted segmented echo-planar imaging volume: 1) A map of the veins present in the T2*-EPI volume is created using a process referred to as “vesselness filtering,” and the vein map is rigidly registered to the T1 volume. 2) White matter lesions are segmented using the T1- and T2-FLAIR volumes. 3) Clear lesion boundaries are then determined using a process that removes ambiguous boundary voxels. 4) Periventricular lesions are removed from candidacy, per guidelines given by the North American Imaging in Multiple Sclerosis Cooperative.\textsuperscript{19} 5) A permutation procedure is performed to determine whether identified veins occur in the center of a given lesion to a greater degree than would be expected by chance. This yields a probability of a CVS for each lesion \( j \) in patient \( i \)’s scan, denoted \( \pi_{ij} \). Lesion-level CVS probabilities are then averaged to obtain a patient-level CVS biomarker, denoted \( \bar{\pi}_i \). 6) Contributions of the lesions to the average can be weighted by the noise in their T2*-EPI intensities to account for scan motion. Figure 1 demonstrates the steps of the algorithm on a sample lesion. Most important, while figures are necessarily presented in 2D space, all methods undertaken for this procedure are conducted in 3D volumetric space and simultaneously consider all 3 planes of the image.

**Vesselness Filtering**

Vein maps in the brain are created to later determine the presence or absence of veins in each lesion. To do this, we applied the Frangi vesselness filter\textsuperscript{20} to the unregistered T2*-EPI volume (for the application to data, this study used the Convert3D Tool; https://sourceforge.net/projects/c3d/), producing a map of scores of \( \geq 0 \), with scores of 0 implying no vesselness qualities. The Frangi filter is a vessel-enhancement algorithm based on the Hessian matrix at each voxel, in which the second-order structure of the image is obtained through convolution with derivatives of Gaussian kernels. The scores are calculated using the eigenvalues of the Hessian matrix, specifically picking up on tubular structures that are darker (or lighter, depending on the implementation) than their surroundings. After being obtained in the unregistered T2*-EPI space, these vesselness maps are then rigidly registered to the T1 space.

**Lesion Segmentation**

To determine the location and shape of white matter lesions, we performed automatic lesion segmentation on coregistered T1- and T2-FLAIR volumes. For the application to data, this study used the Method for InterModal Segmentation Analysis (MIMoSA) model\textsuperscript{21} in the R statistical environment.\textsuperscript{22} The lesion segmentation algorithm produces a map containing the probability that each voxel is part of a lesion. For the results presented in this article, a threshold of 0.30 was applied to this probability map.
to create a binary lesion mask. The threshold of 0.30 was chosen because previous work has found it to be a conservative cutoff that can limit the amount of false-positive lesion tissue. Following the definition of a lesion positive for CVS (CVS+) given by the North American Imaging in Multiple Sclerosis Cooperative, we removed from candidacy lesions detected by the MIMoSA model of $<3$ mm in any plane.

**Lesion-Boundary Determination**

Thresholding of the lesion-probability map often results in pathologically distinct lesions being connected by ambiguous boundary voxels. For these lesions to be properly assessed for the CVS, the proposed algorithm addresses this pseudoconfluence through a recently described technique that removes voxels that are connecting pathologically distinct lesions. The technique works by finding regions in which the texture of the lesion-probability map resembles the center of a lesion. Therefore, the centers that it produces are maintained and used for investigating the CVS for the remainder of this algorithm. Further detail on the implementation of this method can be found in the original publication.

Because the North American Imaging in Multiple Sclerosis guidelines call for the exclusion of confluent lesions, the removal of connecting voxels may represent a deviation from these guidelines in cases of true confluence. However, many lesions that would be judged discrete by expert raters are often merged by automated segmentation methods. This merging can result in drastic and unrealistic degrees of pseudoconfluence in automated lesion masks, sometimes resulting in $\geq50$ distinct lesions being merged into $<10$ lesion components. Thus, relying on automated determinations of confluence in automated lesion masks would likely result in the exclusion of many or most eligible lesions.

**Periventricular Lesion Exclusion**

The density and branching nature of veins near the ventricles makes assessment for the CVS difficult in periventricular lesions, especially in cases in which $>1$ distinct vein traverses the lesion. Thus, the North American Imaging in Multiple Sclerosis Cooperative recommends excluding lesions with $>1$ vein or with branching veins. The proposed algorithm addresses this consideration by excluding periventricular lesions because periventricular lesions typically contain multiple veins. This exclusion is done by performing tissue-class segmentation on the T1 volumes (for the application to data, this study used the FMRIB Automated Segmentation Tool; FAST; http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/ fast) $^{26}$, expanding the CSF region of the brain by 3 mm and eliminating lesions from the lesion-center mask that overlap the expanded CSF region. The choice of a 3-mm expansion was made on the basis of visual inspection of randomly selected T2*-EPI volumes, for which 3 mm appeared to include most of the branching vein structure discussed in the consensus statement, without removing too much of the deep white matter. Notably, although this technique excludes periventricular lesions, it does not exclude other lesions that may have multiple veins. This issue represents a second deviation from the North American Imaging in Multiple Sclerosis recommendations, which could potentially be addressed by future advances in methods for segmenting and counting distinct veins.

**CVS Permutation Procedure**

In lesions that contain central veins, one would expect above-average coherence between the centrality of voxels within the lesion and their vesselness score. The proposed permutation procedure takes advantage of that expectation to examine the degree to which the most veinlike voxels of a lesion are more concentrated in the center of the lesion than one might expect to observe by chance. First, a vein-center coherence score for lesion $j$ in patient $i$’s scan, $C_{ij}$, is calculated by summing the products of the distance-to-nearest-lesion-boundary of each voxel (ie, centrality) score, $d_{ij}$, and its Frangi vesselness score, $f_{ij}$. The coherence formula is given by

$$C_{ij} = \sum_{v} d_{ij} \times f_{ij},$$

where $V$ is the set of all voxels in lesion $j$. Thus, higher values of this score indicate that the highest vesselness values within the lesion tend to occur in the same voxels as the highest centrality values.

A lesion-specific null distribution of coherence scores is created using 1000 random permutations to determine the degree to which this score deviates from chance in cases in which there is no biologic correspondence between vesselness and location within lesions. For each permutation, $p$, the vesselness scores of the voxels in lesion $j$ are randomly resampled without replacement, yielding a randomly ordered set of values, $V_{p}^*$ . A null coherence score is then calculated using the formula,

$$C_{ijp} = \sum_{v \in V_{p}^*} d_{ij} \times f_{ij},$$

This permutation procedure is performed 1000 times, resulting in a sample of 1000 null coherence scores. The lesion-level CVS probability, $\pi_{ij}$, is then calculated as the proportion of chance (null-distributed) CVS scores that are smaller than the observed score, given by

$$\pi_{ij} = \frac{1}{1000} \sum_{p=1}^{1000} 1[C_{ijp} < C_{ij}].$$

To obtain a subject-level CVS biomarker, $\psi_i$, these probabilities are averaged over all lesions observed in patient $i$. The formula for $\psi_i$ is given by

$$\psi_i = \frac{1}{N_i} \sum_{i \in \mathcal{L}_i} \pi_{ij},$$

where $N_i$ is the number of candidate lesions in patient $i$’s scan. The biomarker, $\psi_i$, can be roughly interpreted as the proportion of the patient’s lesions that demonstrate the CVS.

**Optional Noise Weighting**

When one takes the average of the CVS probabilities for a patient’s lesions, some lesions may have more reliable estimates than others. A more stable biomarker can potentially be obtained by weighting each lesion’s contribution to the biomarker by the amount of noise in the voxels of the lesion on the T2*-EPI volume. To estimate the level of noise in a lesion, we first constructed a “noiseless” T2*-EPI by performing anisotropic diffusion on the
A weighted subject-level biomarker, \( \psi^w_i \), is then calculated by summing the products of the CVS probabilities of the lesions with their weights and dividing by the sum of the weights. The weighted biomarker is given by

\[
\psi^w_i = \frac{\sum_{i=1}^{n_i} \pi_{ij} \times w_{ij}}{\sum_{i=1}^{n_i} w_{ij}}
\]

**Implementation and Software**

Accompanying this article, code for the central vein detection algorithm has been made freely available on-line (https://github.com/jdwor/cvs). One file, centralveins_full.R, contains code to run all preprocessing and analysis steps described in the previous section. This file serves to increase the understanding of all steps used in this study and to provide a straightforward tool that can be applied to raw images. A second file, centralveins_simple.R, contains code to be run directly on a probability map and a vein map. This file serves to improve implementations across different sites and scanners, for which researchers and clinicians may have preferred pipelines for preprocessing and lesion segmentation. Following preprocessing and structure segmentation, the centralveins_simple function was found to take an average of 17.7 ± 9.1 minutes and was roughly broken down as a 10-minute baseline with an additional 20 seconds per lesion when run without parallelization. Finally, a third file, helperfunctions.R, provides additional functions used within the previous 2 files.

**Validation**

For this study, data were analyzed for 40 research participants recruited from the University of Vermont neurology clinic as part of a study aiming to improve the diagnostic specificity for MS. Participants were between 20 and 67 years of age, and 37 were women. Ten had MS and no comorbidities known to produce MR imaging white matter abnormalities; 10 had MS and comorbidities known to produce MR imaging white matter abnormalities; 10 had migraine with MR imaging white matter abnormalities; 10 had MS and no comorbidities known to produce MR imaging white matter abnormalities and no other white matter comorbidities; and 10 were previously incorrectly diagnosed with MS and had MR imaging white matter abnormalities and a variety of diagnoses (Table 1).

Whole-brain 3D-T2-FLAIR, T1, and T2*-EPI volumes were acquired on a 3T dStream MR imaging scanner (Philips Health-care, Best, the Netherlands) with a 32-channel dStream head coil. FLAIR and T1 volumes were obtained with 1-mm isotropic resolution, and T2*-EPI volumes were obtained with 0.55-mm isotropic resolution. N4 bias correction was performed on all images, and the T2-FLAIR volume for each participant was interpolated to a voxel size of 1 mm³ and rigidly coregistered to the T1 volume. Extracerebral voxels were removed from the T1 volume using a skull-stripping procedure, and the brain mask was applied to the T2-FLAIR volume.

**Motion-Exclusion Criteria.** Because head motion might occur during the T2*-EPI scan, potentially producing uninterpretable images, each participant’s T2*-EPI scan was manually rated for motion in the relevant white matter regions. Scans were scored from 1 to 5: One indicated “perfect, no artifacts, and excellent signal-to-noise,” 2 indicated “only 1 minor artifact that does not obscure any vessels in supratentorial white matter,” 3 indicated “more than 1 artifact that does not obscure any vessels in supratentorial white matter,” 4 indicated “more than 1 artifact that does obscure some vessels in supratentorial white matter,” and 5 indicated “severe artifacts or bad signal-to-noise that does obscure most vessels in supratentorial white matter.” It was decided a priori that scans that were rated 5 would be removed for the primary analysis because scans with that degree of motion may be unusable in clinical practice as well.

**Performance Assessment.** Because the CVS shows great promise as a diagnostic biomarker, the performance of this algorithm in distinguishing MS and non-MS is of primary interest. To determine whether the automated biomarkers, \( \psi_i \) and \( \psi^w_i \), replicate the findings from previous work that the distribution of manually adjudicated central vein proportion differs between MS and its mimics, we used \( t \) tests to compare the automated CVS values for patients with and without MS. To determine the diagnostic utility of \( \psi_i \) and \( \psi^w_i \), we estimated the area under the curve values of the receiver operating characteristic curves. The presence of a difference in performance between \( \psi_i \) and \( \psi^w_i \) was tested with the DeLong test for comparing the areas under correlated receiver operating characteristic curves using the pROC package in the R statistical environment. Sensitivity and specificity were calcu-
lated using the 40% cutoff, under which inflammatory demyelination is diagnosed if ≥40% of white matter lesions exhibit the CVS, as well as the more recently proposed 50% cutoff. Additionally, locally optimal cutoffs were determined, and their sensitivity and specificity values were compared with those obtained using established cutoffs.

Finally, these cutoffs were compared with the performance of proportion cutoffs applied to manual determinations of the CVS in previous research, as well as the performance of 3 recently proposed clinical decision rules that do not require the assessment of the full set of lesions in a scan. The first such rule, referred to as the rule of 6, states that inflammatory demyelination is diagnosed if there are ≥6 lesions with the CVS or if more than half of lesions show the CVS. The second and third, referred to as select and select*, state that inflammatory demyelination is diagnosed if the CVS is found in at least 2 of 3 lesions preselected on T2-FLAIR and FLAIR* imaging, respectively.

RESULTS
Following manual ratings of scan noise due to motion, 9 participants were excluded and 31 remained for the primary analysis. Of the remaining 31 participants, 16 had MS and 15 did not. Automated CVS detection was performed on these 31 participants using the algorithms and software packages described in the previous section. Two-sample t tests were run to determine whether the automated CVS scores differed between the 16 cases with MS and 15 cases without MS. In both the unweighted (M_{MS} = 0.56 ± 0.17; M_{non-MS} = 0.37 ± 0.12; P < .01) and weighted (M_{MS} = 0.55 ± 0.18; M_{non-MS} = 0.31 ± 0.12; P < .001) variants of the algorithm, the within-patient average CVS probabilities were higher in patients with MS compared with patients without MS. See Fig 2 for breakdowns across all 4 groups.

To determine the diagnostic utility of the automated biomarkers, ψ_i and ψ_i^w, we estimated receiver operating characteristic curves and calculated their areas under the curve. For the unweighted case, ψ_i yielded an area under the curve of 0.84 (Fig 3A). On the basis of the 40% rule, applying a cutoff of 0.40 to ψ_i yielded a sensitivity of 0.94 and a specificity of 0.67. On the basis of the 50% rule, applying a cutoff of 0.50 to this biomarker yielded a sensitivity of 0.56 and a specificity of 0.80. Three locally optimal cutoffs appear to occur at 0.38, at which sensitivity was 1.00 and specificity was 0.67; at 0.44, at which sensitivity was 0.75 and specificity was 0.73; and at 0.50, at which sensitivity was 0.56 and specificity was 0.80 (Table 2).

For the noise-weighted case, ψ_i^w yielded an area under the curve of 0.88 (Fig 3B). Applying a cutoff of 0.40 to ψ_i^w yielded a sensitivity of 0.75 and a specificity of 0.73. Applying a cutoff of 0.50 yielded a sensitivity of 0.56 and a specificity of 0.93. Two locally optimal cutoffs for ψ_i^w appeared to occur at 0.37, at which sensitivity was 0.94 and specificity was 0.73, and at 0.46, at which sensitivity was 0.63 and specificity was 0.93 (Table 2). Although the weighting appeared to produce marginally improved performance, no significant difference was found using the DeLong test (Z = 0.77, P = .22). Robustness analysis on the full sample of 40 participants after reintroducing the motion-obscured scans showed area under the curve values of 0.77 and 0.81 for ψ_i and ψ_i^w, respectively.

Previous studies that used CVS proportions within patients’ full sets of lesions obtained optimal sensitivity/specificity of 1.00/1.00 when comparing cases of MS with undiagnosed cases without MS, patients with microangiopathic lesions, and patients with inflammatory vasculopathies. Prior research on a subset of the current sample was unable to obtain perfect discrimination...
between patients with MS and those with migraine when adjudicating the CVS for all lesions. Compared with cutoffs that used the full set of lesions, decision rules based on a subset of lesions were generally less discriminative between participants with and without MS. The rule of 6 did obtain a sensitivity/specificity of 1.00/1.00 for distinguishing patients with MS and those with small-vessel ischemia, yet in the current sample of MS, migraine, and misdiagnosed patients, the select3 procedure obtained a sensitivity/specificity of 0.81/0.95 and the select3* procedure obtained a sensitivity/specificity of 0.81/0.83.

**DISCUSSION**

Preliminary studies have proposed and validated CVS as a promising biomarker for differentiation of MS from other diseases that cause MR imaging white matter abnormalities. Yet concerns remain regarding the heavy temporal burden on manual adjudication of CVS as well as the subjective differences that may arise in response to variation in the adjudicators’ time constraints and intuition. This study sought to address these issues by introducing an algorithm for automated CVS detection that could, in principle, following further validation, be applied in clinical practice.

In the primary analysis, the algorithm was tested on a cohort of 16 patients with MS (8 with and 8 without other white matter comorbidities) and 15 patients without MS (8 with migraine and 7 misdiagnosed with MS). The fully automated technique replicated previous work that used manual adjudications by demonstrating that proportions of lesions with the CVS differ significantly between MS and its mimics. Additionally, the automated biomarkers, \( \psi \) and \( \psi^w \), were found to have strong diagnostic ability, with areas under the curve of 0.84 and 0.88 and optimal sensitivity/specificity of approximately 0.94/0.70. There is also great promise for this algorithm to perform consistently across study sites and MR imaging scanners because in-house preprocessing and lesion-segmentation methods can be easily substituted and the remaining steps (obtaining vessel-ness scores, finding lesion centers, and calculating CVS probabilities) do not require parameter tuning.

Most important, the automated biomarkers presented in this study did not perform as well as previously obtained proportions of the CVS based on manual ratings of all lesions in patients’ scans. Specifically, the 40% and 50% cutoffs used in prior manually rated studies often achieved perfect discrimination between patients with and without MS, which the automated biomarkers were not able to replicate. However, previous work in a subset of the current sample showed that manual ratings of all lesions did not fully distinguish patients with migraine from those with MS and no white-matter comorbidities. This finding suggests that the patients without MS in the current sample might be more difficult to distinguish from those with MS using the CVS alone than the patients without MS in the studies that did obtain perfect discrimination.

Additionally, although the sensitivity and specificity obtained by these biomarkers were lower than those in the manually obtained CVS proportions, the biomarkers performed comparably with decision rules that use only a subset of lesions in a scan. Thus, while automated adjudication of every lesion in a scan is not yet as accurate as manual adjudication of every lesion in a scan, the proposed automated method shows promise as an alternative to other clinically feasible methods for identifying inflammatory de-

---

**FIG 3.** Receiver operating characteristic curves of MS diagnosis based on patient-level automated CVS biomarker scores. The receiver operating characteristic curve for the unweighted biomarker is shaded blue, and the receiver operating characteristic curve for the weighted biomarker is shaded green. The area under the curve (AUC) values for both curves are displayed in their respective colors.

**Table 2: Diagnostic performance of weighted and unweighted biomarkers**

<table>
<thead>
<tr>
<th>Threshold</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unweighted biomarker (( \psi ))</td>
<td>0.38</td>
<td>1.00 (0.75–1.00)</td>
</tr>
<tr>
<td></td>
<td>0.40</td>
<td>0.94 (0.70–1.00)</td>
</tr>
<tr>
<td></td>
<td>0.44</td>
<td>0.75 (0.50–0.90)</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.56 (0.25–0.65)</td>
</tr>
<tr>
<td></td>
<td>0.37</td>
<td>0.94 (0.70–1.00)</td>
</tr>
<tr>
<td></td>
<td>0.40</td>
<td>0.75 (0.50–0.90)</td>
</tr>
<tr>
<td></td>
<td>0.46</td>
<td>0.63 (0.40–0.80)</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.56 (0.35–0.75)</td>
</tr>
</tbody>
</table>

| Weighted biomarker (\( \psi^w \)) | 0.37 | 0.94 (0.70–1.00) | 0.73 (0.42–0.84) |
| | 0.40 | 0.75 (0.50–0.90) | 0.73 (0.42–0.84) |
| | 0.46 | 0.63 (0.40–0.80) | 0.93 (0.68–1.00) |
| | 0.50 | 0.56 (0.35–0.75) | 0.93 (0.74–1.00) |

*Data in column 2 represent the given cutoff values; data in parentheses are the relevant 95% confidence intervals.*
CONCLUSIONS

Although the potential clinical implications of an automated tool for CVS adjudication call for further study and refinement of such techniques, the current study demonstrates the promising performance of a fully automated method for detecting CVS in white matter lesions. To our knowledge, this is the first automated technique for this challenging aspect of MS diagnosis and represents an important step forward toward a specific MR imaging biomarker for MS lesions.

Disclosures: Jordan D. Dworkin—RELATED: Grant, National Institutes of Health.*  Andrew Solomon—UNRELATED: Consultancy: Teva Pharmaceuticals. Biogen, EMD Serono, Genentech, Comments: consulting and/or advisory boards; Grants/Grants Pending: Biogen, Comments: funding for research.*. Dzung L. Pham—RELATED: Grant: National Multiple Sclerosis Society. Comments: support from NIMSS RG-1507-05453*; UNRELATED: Travel/ Accommodations/Meeting Expenses Unrelated to Activities Listed: Magnetic Resonance Imaging in MS, Comments: gave an invited talk at the Magnetic Resonance Imaging in MS plenary meeting and workshop in April 2018. Although about MS, the talk was unrelated to the work in this article. Daniel Ontaneda—UNRELATED: Consultancy: Biogen Idec, Genzyme, Merck; Expert Testimony: Novartis, Genzyme, Genentech.* Daniel S. Reich—RELATED: Grant: National Multiple Sclerosis Society, Conrad N. Hilton Foundation; UNRELATED: Expert Testimony: Leventhal & Puga; Grants/Grants Pending: National Multiple Sclerosis Society, Conrad N. Hilton Foundation, Adelson Medical Research Foundation.*. Payment for Development of Educational Presentations: At the Limits; Travel/Accommodations/Meeting Expenses Unrelated to Activities Listed: National MS Society, Race to Erase MS Foundation, Adelson Medical Research Foundation, ECTRIMS, LACTRIMS, ACRTRMS, Consortium of MS Centers, MS Society of Canada, Neurological Research Support Society [Istanbul, Alberta MS Network, University of British Columbia, Mayo Clinic, Gordon Research Conferences, Comments: invitations from nonprofit foundations for travel to present scientific lectures or act as federal liaison to an advisory board.* Russell T. Shimosaka—RELATED: Grant, National Institutes of Health, National Multiple Sclerosis Society, Race to Erase MS*; UNRELATED: Board Membership: Genentech/Roche*; Consultancy: Genentech/Roche*; Grants/Grants Pending: National Institutes of Health*, Travel/ Accommodations/Meeting Expenses Unrelated to Activities Listed: European Committee for Treatment and Research in Multiple Sclerosis, Comments: travel grant for conference attendance. *Money paid to the institution.

REFERENCES


Predicting Genotype and Survival in Glioma Using Standard Clinical MR Imaging Apparent Diffusion Coefficient Images: A Pilot Study from The Cancer Genome Atlas


ABSTRACT

BACKGROUND AND PURPOSE: Few studies have shown MR imaging features and ADC correlating with molecular markers and survival in patients with glioma. Our purpose was to correlate MR imaging features and ADC with molecular subtyping and survival in adult diffuse gliomas.

MATERIALS AND METHODS: Presurgical MRIs and ADC maps of 131 patients with diffuse gliomas and available molecular and survival data from The Cancer Genome Atlas were reviewed. MR imaging features, ADC (obtained by ROIs within the lowest ADC area), and mean relative ADC values were evaluated to predict isocitrate dehydrogenase (IDH) mutation, 1p/19q codeletion status, MGMT promoter methylation, and overall survival.

RESULTS: IDH wild-type gliomas tended to exhibit enhancement, necrosis, and edema; >50% enhancing area (P < .001); absence of a cystic area (P = .013); and lower mean relative ADC (median, 1.1 versus 1.6; P < .001) than IDH-mutant gliomas. By means of a cutoff value of 1.08 for mean relative ADC, IDH-mutant and IDH wild-type gliomas with lower mean relative ADC (<1.08) had poorer survival than those with higher mean relative ADC (median survival time, 24.2 months; 95% CI, 0.0–54.9 months versus 62.0 months; P = .003; and median survival time, 10.4 months; 95% CI, 4.4–16.4 months versus 17.7 months; 95% CI, 11.6–23.7 months; P = .041, respectively), regardless of World Health Organization grade. Median survival of those with IDH-mutant glioma with low mean relative ADC was not significantly different from that in those with IDH wild-type glioma. Other MR imaging features were not statistically significant predictors of survival.

CONCLUSIONS: IDH wild-type glioma showed lower ADC values, which also correlated with poor survival in both IDH-mutant and IDH wild-type gliomas, irrespective of histologic grade. A subgroup with IDH-mutant gliomas with lower ADC had dismal survival similar to that of those with IDH wild-type gliomas.

ABBREVIATIONS: IDH = isocitrate dehydrogenase; max = maximum; min = minimum; rADC = relative ADC; rADCmean = mean relative ADC; TCGA = The Cancer Genome Atlas; WHO = World Health Organization

Gliomas are a heterogeneous group of tumors, and the clinical aggressiveness and prognoses are diverse among different histopathologic grades and molecular subtypes. Previous studies have shown that histopathologic classification of diffuse gliomas has high interobserver variation and correlates imperfectly with clinical outcomes.1,2 Nevertheless, molecular markers, particularly isocitrate dehydrogenase (IDH) mutational status, have been demonstrated to be significant and more robust prognostic markers3 and have been incorporated into the classification of diffuse gliomas in the latest update of the World Health Organization (WHO) classification in 2016.4 IDH mutation, a powerful prognostic marker of improved survival in diffuse glioma, is found mainly in lower grade gliomas (WHO grades II and III), but also in glioblastoma (WHO grade IV), though at much lower frequency.5,6 Preoperative and noninvasive determination of molecular subtyping is of great value in the clinical management of patients with glioma. However, studies correlating MR imaging features with IDH-mutation status and patient survival in diffuse gliomas are scarce. Recently, we showed that the “T2-FLAIR mismatch sign,” detectable using conventional MR imaging, is a highly spe-
cific imaging biomarker for the IDH-mutant, 1p/19q noncodeleted molecular subtype in lower grade gliomas. Wang et al demonstrated that the absence of contrast enhancement was associated with longer progression-free and overall survival in patients with IDH1-mutated anaplastic gliomas. MR spectroscopy could detect 2-hydroxyglutarate, a metabolite that accumulates in IDH-mutant gliomas but did not discover a survival difference. Blood volume estimates obtained by MR perfusion have also provided potential markers for noninvasive assessment of IDH status.

ADC can be calculated from DWI, and tumors with more freely mobile water molecules and lesser cellularity have higher ADC values. ADC has been shown to be a valuable imaging marker in the diagnosis of intracranial lesions as well as in grading brain tumors. Therefore, we hypothesized that ADC values obtained from conventional MR imaging could correlate with molecular subtype and patient survival in adult diffuse gliomas.

MATERIALS AND METHODS
This was a retrospective study using data from the publicly available National Institutes of Health/National Cancer Institute-approved databases of The Cancer Genome Atlas (TCGA; https://cancergenome.nih.gov) and The Cancer Imaging Archive (http://www.cancerimagingarchive.net). From all 461 cases with imaging data were reviewed, and only cases of treatment-naive diffuse gliomas (WHO grades II–IV) with available DWI and ADC maps were included. WHO grade, the status of 3 validated molecular prognostic markers (IDH mutation, 1p/19q codeletion, MGMT promoter methylation), and survival data were retrieved from The Cancer Genome Atlas. MR images were reviewed, in consensus, by 2 board-certified neuroradiologists (with 8 and 17 years of experience) who were blinded to pathologic and molecular diagnosis. The order of cases viewed was randomized to avoid bias.

Each tumor was scored for 9 MR imaging features according to the following criteria modified from the Visually Accessible Rembrandt Images MR imaging feature set: T2 signal intensities (higher than gray matter or mixed [the presence equal to or darker than that of gray matter part]); T2 homogeneity (homogeneous or heterogeneous); margin (well-defined or not well-defined [either infiltrative or irregular]); edema (none to minimal or mild to marked); enhancing pattern (non-/minimally enhancing or enhancing); portion of enhancing area (<50% or ≥50%); the presence of cystic areas (presence or absence); and the presence of necrotic areas (presence or absence). We investigated the relationship among 9 different MR imaging features and 3 molecular markers (IDH mutation, 1p/19q codeletion, and MGMT promoter methylation) as well as WHO grade and overall survival.

Diffusion-weighted images were analyzed using OsiriX Imaging Software (http://www.osirix-viewer.com). ADC measurements were generated by manually drawing 3 nonoverlapping ROIs ranging from 40 to 60 mm² within the region of lowest ADC values within the solid component of each tumor on ADC maps. The ADC value was also calculated from contralateral normal-appearing white matter by drawing a single ROI with a size similar to that of a tumoral ROI. We obtained mean, minimum (min), and maximum (max) ADCs of each tumor, respectively, by averaging the 3 ROIs; rADC values (rADCmean, rADCmin, and rADCmax) was calculated by dividing the tumor ADC by the ADC of the contralateral normal-appearing white matter.

Statistical Analysis
The Kolmogorov-Smirnov test was used to determine whether the numeric data (age and relative ADC values) for each group were normally distributed. Independent variables (clinical parameters and MR imaging features) were compared using the χ² test among different molecular groups. Normally distributed continuous variables (eg, age) were compared using the independent t test or ANOVA test, and non-normally distributed continuous variables (rADCmean, rADCmin, and rADCmax) were compared using the Mann-Whitney U test among different molecular groups. The intraobserver reliability of ADC value measuring was tested using intraclass correlation coefficients. The optimal cutoff value of each rADC was obtained from receiver operating characteristic curve analysis when the Youden index reached a maximum. Survival curves were estimated and plotted by the Kaplan-Meier method with log-rank tests to compare Kaplan-Meier curves among groups. Variables were first analyzed by the univariate model. MR imaging features and clinical and molecular parameters (including age, sex, WHO grade, MGMT promoter methylation status, IDH mutation status, and 1p/19q codeletion status) with statistical significance in univariate analysis (P < .05) were entered into a Cox proportional hazards ratio model for multivariate analysis. Statistical significance was defined as P < .05 for all tests. The statistical analyses were performed using the statistical software package SPSS 23.0 (IBM, Armonk, New York) and R statistical and computing software, Version 3.3.2 (http://www.r-project.org).

RESULTS
A total of 131 (59 [45%] IDH wild-type and 72 [55%] IDH-mutant) gliomas were included in this study. Of the 72 IDH-mutant tumors, 26 (36%) were 1p/19q codeleted and 46 (64%) were non-1p/19q codeleted. Patients in the IDH wild-type group (mean, 60 ± 12.2 years) were significantly older than those in the IDH-mutant group (mean, 45.1 ± 13.9 years), and patients with IDH-mutant, 1p/19q codeleted gliomas (mean, 50.7 ± 13.9 years) were older than those with IDH-mutant, non-1p/19q codeleted gliomas (mean, 41.9 ± 12.9 years) (P = .01).

Correlation between Conventional MR Imaging Features and Molecular Subtypes
Among the conventional MR imaging characteristics, IDH wild-type gliomas were more likely to exhibit enhancement (P < .001), >50% enhancing area (P < .001), absence of cystic area (P = .013), the presence of necrosis (P < .001), and the presence of edema (P < .001). Within the IDH-mutant group, there were no MR imaging characteristics to differentiate 1p/19q codeletion status using the features tested (Table 1).

Correlation between rADC Values and Molecular Subtypes
Median rADCmean, rADCmin, and rADCmax values of IDH wild-type gliomas were significantly lower than those of IDH-mutant gliomas (P < .001) (Fig 1 and Table 1). Within the IDH-mutant...
Table 1: MR imaging features and IDH-mutation and 1p/19q codeletion status

<table>
<thead>
<tr>
<th>MR imaging features (No.) (%)</th>
<th>IDH Wild-Type</th>
<th>IDH-Mutant</th>
<th>IDH-Mutant, Non-1p/19q Codeleted</th>
<th>IDH-Mutant, 1p/19q Codeleted</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>T2 signal intensities*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gray matter</td>
<td>13 (22.4%)</td>
<td>30 (43.5%)</td>
<td>21 (48.8%)</td>
<td>9 (34.6%)</td>
<td>.012c</td>
</tr>
<tr>
<td>Mixed</td>
<td>45 (77.6%)</td>
<td>39 (56.5%)</td>
<td>22 (51.2%)</td>
<td>17 (65.4%)</td>
<td>.248</td>
</tr>
<tr>
<td>T2 homogeneity*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homogeneous</td>
<td>6 (10.3%)</td>
<td>14 (20.3%)</td>
<td>10 (23.3%)</td>
<td>4 (15.4%)</td>
<td>.544</td>
</tr>
<tr>
<td>Heterogeneous</td>
<td>52 (89.7%)</td>
<td>55 (79.7%)</td>
<td>33 (76.7%)</td>
<td>22 (84.6%)</td>
<td></td>
</tr>
<tr>
<td>Margin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Well-defined</td>
<td>32 (54.2%)</td>
<td>27 (43.8%)</td>
<td>20 (44.4%)</td>
<td>7 (26.9%)</td>
<td>.143</td>
</tr>
<tr>
<td>Mixed</td>
<td>27 (45.8%)</td>
<td>44 (62.0%)</td>
<td>25 (55.6%)</td>
<td>19 (73.1%)</td>
<td></td>
</tr>
<tr>
<td>Edema</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No-to-minimal</td>
<td>20 (33.9%)</td>
<td>57 (79.2%)</td>
<td>38 (82.6%)</td>
<td>19 (73.1%)</td>
<td>.339</td>
</tr>
<tr>
<td>Mild-to-marked</td>
<td>39 (66.1%)</td>
<td>15 (20.8%)</td>
<td>8 (17.4%)</td>
<td>7 (26.9%)</td>
<td></td>
</tr>
<tr>
<td>Enhancing patternb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None/minimally enhancing</td>
<td>4 (7.0%)</td>
<td>30 (42.3%)</td>
<td>19 (41.3%)</td>
<td>11 (40.0%)</td>
<td>.826</td>
</tr>
<tr>
<td>Enhancing</td>
<td>55 (93.2%)</td>
<td>41 (56.9%)</td>
<td>27 (58.7%)</td>
<td>14 (56.0%)</td>
<td></td>
</tr>
<tr>
<td>Proportion of enhancing areab</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;50%</td>
<td>14 (23.7%)</td>
<td>63 (88.7%)</td>
<td>40 (87.0%)</td>
<td>23 (92.0%)</td>
<td>.704</td>
</tr>
<tr>
<td>≥50%</td>
<td>45 (76.3%)</td>
<td>8 (11.3%)</td>
<td>6 (13.0%)</td>
<td>2 (8.0%)</td>
<td></td>
</tr>
<tr>
<td>Cystic area</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presence</td>
<td>5 (8.5%)</td>
<td>28 (39.4%)</td>
<td>26 (56.5%)</td>
<td>17 (65.4%)</td>
<td>.461</td>
</tr>
<tr>
<td>Absence</td>
<td>54 (91.5%)</td>
<td>43 (59.7%)</td>
<td>20 (43.5%)</td>
<td>9 (34.6%)</td>
<td></td>
</tr>
<tr>
<td>Necrotic areab</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presence</td>
<td>49 (83.1%)</td>
<td>16 (22.5%)</td>
<td>38 (82.6%)</td>
<td>17 (68.0%)</td>
<td>.159</td>
</tr>
<tr>
<td>Absence</td>
<td>10 (16.9%)</td>
<td>55 (77.5%)</td>
<td>8 (17.4%)</td>
<td>8 (32.0%)</td>
<td></td>
</tr>
<tr>
<td>rADCmean (median)</td>
<td>1.1</td>
<td>1.6</td>
<td>1.7</td>
<td>1.5</td>
<td>.071</td>
</tr>
<tr>
<td>rADCmin (median)</td>
<td>0.9</td>
<td>1.4</td>
<td>1.4</td>
<td>1.3</td>
<td>.178</td>
</tr>
<tr>
<td>rADCmax (median)</td>
<td>1.3</td>
<td>1.8</td>
<td>1.9</td>
<td>1.7</td>
<td>.106</td>
</tr>
</tbody>
</table>

* There were 4 cases lacking T2 MR images.

b One case in the IDH-mutant group lacked postcontrast studies.

c Statistically significant (P < .05).

FIG 1. Boxplot representation of rADCmean values by glioma IDH genotype.
The mechanism by which IDH-mutant and IDH wild-type gliomas differ in terms of the rADC value is a simple approach that demonstrates that most IDH-mutant gliomas have significantly longer survival compared with IDH wild-type gliomas, independent of their WHO grade. Additionally, using rADC, we could identify a particularly poor prognosis subset of IDH-mutant gliomas, with outcomes similar to those in patients IDH wild-type disease. Most important, determining the rADC value is a simple approach that requires no specialized software; hence, our findings potentially have immediate clinical impact.

The mechanism by which IDH-mutant and IDH wild-type gliomas differ in terms of the rADC value is a simple approach that demonstrates that most IDH-mutant gliomas have significantly longer survival compared with IDH wild-type gliomas, independent of their WHO grade. Additionally, using rADC, we could identify a particularly poor prognosis subset of IDH-mutant gliomas, with outcomes similar to those in patients IDH wild-type disease. Most important, determining the rADC value is a simple approach that requires no specialized software; hence, our findings potentially have immediate clinical impact.


discussion

Mutations in the IDH genes are among the most important diagnostic and prognostic markers of diffuse gliomas.26 Patients with IDH-mutant gliomas have significantly longer survival compared with those with IDH wild-type gliomas, and management of these 2 molecular subgroups differs significantly. Previous studies have investigated the potential of various conventional and advanced MR imaging characteristics, including perfusion, diffusion tensor, and MR spectroscopy, in identifying genetic subtypes of diffuse gliomas.10,21,22 Here, our results indicate that rADC values correlate with IDH mutation status as well as survival in both IDH-mutant and IDH wild-type diffuse gliomas, independent of their WHO grade. Additionally, using rADC, we could identify a particularly poor prognosis subset of IDH-mutant gliomas, with outcomes similar to those in patients IDH wild-type disease. Most important, determining the rADC value is a simple approach that requires no specialized software; hence, our findings potentially have immediate clinical impact.

The mechanism by which IDH-mutant and IDH wild-type gliomas differ in terms of the rADC value is a simple approach that demonstrates that most IDH-mutant gliomas have significantly longer survival compared with IDH wild-type gliomas, independent of their WHO grade. Additionally, using rADC, we could identify a particularly poor prognosis subset of IDH-mutant gliomas, with outcomes similar to those in patients IDH wild-type disease. Most important, determining the rADC value is a simple approach that requires no specialized software; hence, our findings potentially have immediate clinical impact.

The mechanism by which IDH-mutant and IDH wild-type gliomas differ in terms of the rADC value is a simple approach that demonstrates that most IDH-mutant gliomas have significantly longer survival compared with IDH wild-type gliomas, independent of their WHO grade. Additionally, using rADC, we could identify a particularly poor prognosis subset of IDH-mutant gliomas, with outcomes similar to those in patients IDH wild-type disease. Most important, determining the rADC value is a simple approach that requires no specialized software; hence, our findings potentially have immediate clinical impact.

The mechanism by which IDH-mutant and IDH wild-type gliomas differ in terms of the rADC value is a simple approach that demonstrates that most IDH-mutant gliomas have significantly longer survival compared with IDH wild-type gliomas, independent of their WHO grade. Additionally, using rADC, we could identify a particularly poor prognosis subset of IDH-mutant gliomas, with outcomes similar to those in patients IDH wild-type disease. Most important, determining the rADC value is a simple approach that requires no specialized software; hence, our findings potentially have immediate clinical impact.

The mechanism by which IDH-mutant and IDH wild-type gliomas differ in terms of the rADC value is a simple approach that demonstrates that most IDH-mutant gliomas have significantly longer survival compared with IDH wild-type gliomas, independent of their WHO grade. Additionally, using rADC, we could identify a particularly poor prognosis subset of IDH-mutant gliomas, with outcomes similar to those in patients IDH wild-type disease. Most important, determining the rADC value is a simple approach that requires no specialized software; hence, our findings potentially have immediate clinical impact.

The mechanism by which IDH-mutant and IDH wild-type gliomas differ in terms of the rADC value is a simple approach that demonstrates that most IDH-mutant gliomas have significantly longer survival compared with IDH wild-type gliomas, independent of their WHO grade. Additionally, using rADC, we could identify a particularly poor prognosis subset of IDH-mutant gliomas, with outcomes similar to those in patients IDH wild-type disease. Most important, determining the rADC value is a simple approach that requires no specialized software; hence, our findings potentially have immediate clinical impact.

The mechanism by which IDH-mutant and IDH wild-type gliomas differ in terms of the rADC value is a simple approach that demonstrates that most IDH-mutant gliomas have significantly longer survival compared with IDH wild-type gliomas, independent of their WHO grade. Additionally, using rADC, we could identify a particularly poor prognosis subset of IDH-mutant gliomas, with outcomes similar to those in patients IDH wild-type disease. Most important, determining the rADC value is a simple approach that requires no specialized software; hence, our findings potentially have immediate clinical impact.

The mechanism by which IDH-mutant and IDH wild-type gliomas differ in terms of the rADC value is a simple approach that demonstrates that most IDH-mutant gliomas have significantly longer survival compared with IDH wild-type gliomas, independent of their WHO grade. Additionally, using rADC, we could identify a particularly poor prognosis subset of IDH-mutant gliomas, with outcomes similar to those in patients IDH wild-type disease. Most important, determining the rADC value is a simple approach that requires no specialized software; hence, our findings potentially have immediate clinical impact.

The mechanism by which IDH-mutant and IDH wild-type gliomas differ in terms of the rADC value is a simple approach that demonstrates that most IDH-mutant gliomas have significantly longer survival compared with IDH wild-type gliomas, independent of their WHO grade. Additionally, using rADC, we could identify a particularly poor prognosis subset of IDH-mutant gliomas, with outcomes similar to those in patients IDH wild-type disease. Most important, determining the rADC value is a simple approach that requires no specialized software; hence, our findings potentially have immediate clinical impact.

The mechanism by which IDH-mutant and IDH wild-type gliomas differ in terms of the rADC value is a simple approach that demonstrates that most IDH-mutant gliomas have significantly longer survival compared with IDH wild-type gliomas, independent of their WHO grade. Additionally, using rADC, we could identify a particularly poor prognosis subset of IDH-mutant gliomas, with outcomes similar to those in patients IDH wild-type disease. Most important, determining the rADC value is a simple approach that requires no specialized software; hence, our findings potentially have immediate clinical impact.

The mechanism by which IDH-mutant and IDH wild-type gliomas differ in terms of the rADC value is a simple approach that demonstrates that most IDH-mutant gliomas have significantly longer survival compared with IDH wild-type gliomas, independent of their WHO grade. Additionally, using rADC, we could identify a particularly poor prognosis subset of IDH-mutant gliomas, with outcomes similar to those in patients IDH wild-type disease. Most important, determining the rADC value is a simple approach that requires no specialized software; hence, our findings potentially have immediate clinical impact.

The mechanism by which IDH-mutant and IDH wild-type gliomas differ in terms of the rADC value is a simple approach that demonstrates that most IDH-mutant gliomas have significantly longer survival compared with IDH wild-type gliomas, independent of their WHO grade. Additionally, using rADC, we could identify a particularly poor prognosis subset of IDH-mutant gliomas, with outcomes similar to those in patients IDH wild-type disease. Most important, determining the rADC value is a simple approach that requires no specialized software; hence, our findings potentially have immediate clinical impact.

The mechanism by which IDH-mutant and IDH wild-type gliomas differ in terms of the rADC value is a simple approach that demonstrates that most IDH-mutant gliomas have significantly longer survival compared with IDH wild-type gliomas, independent of their WHO grade. Additionally, using rADC, we could identify a particularly poor prognosis subset of IDH-mutant gliomas, with outcomes similar to those in patients IDH wild-type disease. Most important, determining the rADC value is a simple approach that requires no specialized software; hence, our findings potentially have immediate clinical impact.
Another retrospective study of 112 cases by Tan et al. demonstrated that fractional anisotropy and ADC from diffusion tensor imaging can detect IDH1 mutation in astrocytomas, with the ratio of ADC\textsubscript{min} being the best metric for detecting IDH mutation, regardless of the WHO grade. We found that rADC\textsubscript{mean} can differentiate IDH wild-type from IDH-mutant gliomas with excellent discrimination, regardless of WHO grade. Our study also emphasizes MR imaging and ADC values correlating well with molecular subtype.

The main novel finding of our study is that preoperative rADC values can distinguish favorable and unfavorable prognosis within both IDH-mutant and IDH wild-type glioma subgroups. While IDH-mutant gliomas generally behave less aggressively and have a better prognosis compared with their IDH wild-type counterparts, we identified a small subset (12.5%) of IDH-mutant gliomas with low rADC\textsubscript{mean} values and poor overall survival, which was only slightly better (24 months) than that of IDH wild-type gliomas but was not statistically significant. Concordantly, a study by Liao et al. revealed that a small subgroup (11.7%) of patients with IDH-mutant gliomas across all grades had a dismal prognosis (median survival of 22 months), more similar to IDH wild-type gliomas and glioblastomas in their cohort. These tumors had distinct genetic characteristics, lacking the typical concurrent genetic alterations observed in IDH-mutant gliomas. In addition, a recent study of The Cancer Genome Atlas identified a small subset (5.5%) of IDH-mutant gliomas with markedly worse survival than other IDH-mutant gliomas, and these tumors were associated with relatively decreased global DNA methylation. Together, these data clearly indicate that a subgroup of IDH-mutant gliomas behaves as aggressively as their IDH wild-type counterparts. Although whether the malignant subgroups across these datasets represent the same biology is unknown, our results suggest that rADC values can potentially identify this aggressively behaving IDH-mutant subgroup.

Furthermore, our study highlights how detection of robust imaging-phenotype correlations can be significantly improved by analyzing glioma datasets by molecular subtype rather than by histopathologic classification. We evaluated MR imaging features and prognosis in diffuse gliomas across lower and higher grades in the current study and demonstrated the power of rADC\textsubscript{mean} to differentiate IDH-mutation status and discrete survival subgroups beyond WHO grade. While many previous studies have demonstrated an inverse relationship between ADC and astrocytoma grade, other studies have shown substantial overlap of ADC values between high-versus-low-grade gliomas and no significant differences between grade II versus III or grade III versus IV. These observed variations of ADC in predicting tumor grade are likely due to limitations that make the exact histopathologic classification challenging, with high interobserver variability and molecular constituent and clinical behavior being likely different in tumors with the same histopathologic grade. Similar to a recent meta-analysis by Zulfiqar et al., which showed that low ADC values correlate independently with poor survival in malignant astrocytomas (grades III and IV),
we found an inverse relationship between rADC_{mean} values and prognosis for both IDH-mutant and IDH wild-type tumors independent of WHO grade.

One limitation of our study is its retrospective design, which was necessary to include a relatively large number of patients and to correlate with survival, which is relatively long in patients with IDH mutation. A second limitation is that the studied patients had been scanned on different MR imaging magnet types, and ADC maps were generated by diffusion-weighted imaging or diffusion tensor imaging of all collected data. However, a previous study has verified that ADC datasets from 3-directional DWI and 6-directional diffusion tensor imaging could be analyzed together.40 We calculated the rADC to minimize the differences among absolute ADC values across platforms. Third, the treatment regimen applied to each patient was not available to us in many cases. This issue might have potentially impacted the outcome and survival in each case. However, IDH status has been repeatedly shown to be an independent marker of prognosis in independent datasets. Finally, the ADC value has previously been reported to predict 1p/19q codeletion status, a marker of oligodendroglioma, in lower grade gliomas.35,41 In the study by Johnson et al,41 the ADC values were calculated from sampling both the highest and lowest ADC areas. However, we did not detect a significant correlation between rADC values and 1p/19q codeletion status. Further investigation of the optimal methods of measuring ADC and physiology correlates of ADC values in genetically defined oligodendroglioma is needed.

Our results require independent confirmation, incorporating emerging molecular markers and accounting for different treatment strategies. However, our results expand on and refine the existing correlation between DWI and tumor genetic markers and highlight its potential role as an independent imaging biomarker that can aid in stratification of patients with gliomas, both IDH-mutant and IDH wild-type. Here, we were able to identify a subset of aggressive IDH-mutant gliomas using ADC values easily obtained from common clinical MR images. Ongoing accumulation of tumorigenesis knowledge, together with imaging studies stratified by molecular subgroup rather than histopathologic features, will likely identify additional robust genetic-imaging-clinical phenotype correlations that will improve early detection of clinically meaningful glioma molecular subtypes.

CONCLUSIONS

We demonstrate that ADC values obtained from DWI correlate with IDH-mutation status and overall survival in adult diffuse gliomas. IDH wild-type gliomas showed low ADC values and poor survival compared with IDH-mutant gliomas. Within IDH-mutant gliomas, a small subgroup with lower ADC values had dismal survival, similar to that in IDH wild-type gliomas. ADC values correlated with survival in patients with IDH-mutant and IDH wild-type gliomas regardless of WHO grade. Preoperative ADC estimates may corroborate with molecular subtypes as a prognostic marker and potentially enhance risk stratification, especially within IDH-mutant gliomas.

REFERENCES


2. van den Bent MJ. Interobserver variation of the histopathological diagnosis in clinical trials on glioma: a clinician’s perspective. Acta Neuropathol 2010;120:297–304 CrossRef Medline


Near-Term Decrease in Brain Volume following Mild Traumatic Injury Is Detectible in the Context of Preinjury Volumetric Stability: Neurobiologic Insights from Analysis of Historical Imaging Examinations

A.E. Goldman-Yassen, K.X. Chen, D. Edasery, K. Hsu, K. Ye, and M.L. Lipton

ABSTRACT

BACKGROUND AND PURPOSE: Neurodegeneration after mild traumatic brain injury may manifest as decreasing regional brain volume that evolves from months to years following mild traumatic brain injury and is associated with worse clinical outcomes. We hypothesized that quantitative brain volume derived from CT of the head, performed for clinical indications during routine care, would change with time and provide insights into the putative neuroinflammatory response to mild traumatic brain injury.

MATERIALS AND METHODS: We searched the electronic medical record of our institution for NCCTs of the head performed in patients with mild traumatic brain injury and included those who also underwent NCCTs of the head 1 month to 1 year before and after mild traumatic brain injury for an indication unrelated to trauma. Controls underwent 3 sequential NCCTs of the head with indications unrelated to trauma. The whole-brain and intracranial volume groups were computed using ITK-SNAP. Brain volumes normalized to intracranial volumes were compared across time points using the Wilcoxon signed-rank test.

RESULTS: We identified 48 patients from 2005 to 2015 who underwent NCCTs of the head in the emergency department for mild traumatic brain injury and had NCCTs of the head performed both before and after mild traumatic brain injury. Median normalized brain volumes significantly decreased on the follow-up study post-mild traumatic brain injury (0.86 versus 0.84, \( P < .001 \)) and were similar compared with pre-mild traumatic brain injury studies (0.87 versus 0.86, \( P = .927 \)). There was no significant difference between normalized brain volumes in the 48 controls.

CONCLUSIONS: A decrease in brain volume following mild traumatic brain injury is detectable on CT and is not seen in similar patients with non-mild traumatic brain injury during a similar timeframe. Given the stability of brain volume before mild traumatic brain injury, CT volume loss may represent the subtle effects of neurodegeneration.

ABBREVIATIONS: IQR = interquartile range; mTBI = mild traumatic brain injury; NCCTH = NCCT of the head; TBI = traumatic brain injury

T raumatic brain injury (TBI) affects an estimated 1.7 million patients annually in the United States and is a major cause of morbidity and mortality.\(^1\,^2\) Most, 70%–90% of cases of TBI, are classified as mild TBI (mTBI), which is also known as concussion.\(^3\) Although mTBI-related symptoms resolve in most patients, a minority of patients experience persistent symptoms, which may be disabling.\(^4\)

Conventional CT and MR imaging examinations performed as part of routine clinical care typically reveal no visible indications of trauma in patients with mTBI. Neuroinflammation, proposed as a mechanism underlying persistent dysfunction and brain structural changes following mTBI, however, might manifest as subtle brain swelling on acute timeframe imaging.\(^5\) Although quantitative studies of brain volume have demonstrated both gray and white matter volume loss after mTBI using MR imaging, published studies of brain volume in TBI have an important limitation: They did not have access to pre-mTBI brain imaging and cannot therefore assess change of brain volume from a preinjury baseline.\(^6\)–\(^16\) Moreover, prior studies generally reported change during the long-term after mTBI. Although prior studies have shown a reduction in brain volume after mTBI, in the absence of a baseline, it is unclear whether these changes are the
result of resolving subtle trauma-related edema due to neuroinflammation or atrophy from neurodegeneration.

CT is widely performed as the initial imaging test of choice for patients who present with acute mTBI, but it has not been reported as a method to quantify brain volume changes.17-19 In this study, we aimed to mine existing clinical CT examinations, performed as a part of routine clinical care, to quantify baseline brain volume before mTBI and assess subsequent brain volume change. Because NCCT of the head (NCCTH) is ordered in patients for a variety of reasons unrelated to trauma, we can compare mTBI examinations with CT performed both before and following mTBI in the same patients, to quantify brain volume changes following mTBI and determine whether the change suggests resolution of acute swelling, consistent with resolving neuroinflammation, or loss of volume, consistent with neurodegeneration.

MATERIALS AND METHODS

Study Population and Study Design

After obtaining approval from our institutional review board, including a waiver of informed consent for our Health Insurance Portability and Accountability Act–compliant study, we used Clinical Looking Glass (Streamline Health, Atlanta, Georgia) to search our electronic medical records for patients who underwent NCCTH for the evaluation of mTBI. Specifically, we searched for radiology reports from NCCTHs ordered in the emergency department at Montefiore Medical Center from 2005 to 2015 with clinical indications of “trauma,” “fall,” “concussion,” or “assault.” Subjects were included if they also underwent NCCTH 1–12 months before and after the mTBI scan for a nontraumatic indication (Fig 1). If the patient underwent multiple scans in these intervals, the scan closest to the time of mTBI was used. Controls were identified by searching Clinical Looking Glass for patients who underwent 3 sequential NCCTHs in the emergency department for nontraumatic reasons at time intervals similar to those in the patients with mTBI.

Demographic information and medical comorbidities were obtained for each individual from the electronic medical records at the time of mTBI (cases) or the second CT (controls). Clinical notes associated with NCCTHs performed before mTBI, at the time of mTBI, and after mTBI were reviewed to assess the clinical state of the patient at the time of each scan, including the mechanism of injury, clinical indications, and signs and symptoms.

CT Image Acquisition and Data Storage

During routine clinical care, axial NCCTH images were acquired on LightSpeed VCT and LightSpeed RT16 (GE Healthcare, Milwaukee, Wisconsin) at a kilovolt (peak) ranging from 120 to 140 and milliampere ranging from 280 to 300. For each subject, the hardware and technique remained constant across all time points. We retrieved DICOM files containing 0.625-mm-thick axial slices of the head from the PACS and stored them on a password-protected local hard drive. Subjects were excluded if 0.625-mm-thick images were not available.

Image Processing

Processing and analysis were performed by 3 postgraduate year 3 or postgraduate year 4 radiology residents, supervised by an American Board of Radiology Certificate of Added Qualification–certified neuroradiologist. Before the processing and analysis of all cases, a random subset of 18 cases was independently segmented by 2 residents and interrater reliability was assessed. Once the protocol was considered reliable, the residents then processed and analyzed the remaining cases independently to conserve time and resources. Incomplete datasets were excluded. Next, all images were first visually inspected. Subjects were excluded if there was evidence of acute or active disease, including hemorrhage, mass, acute or chronic infarct, and hydrocephalus. Images degraded by motion or beam-hardening artifacts were also excluded. Patients with age-related abnormalities, such as white matter hypodensities, were not excluded.

We assessed total brain and intracranial volumes using semiautomated threshold segmentation in ITK-SNAP, Version 3.6.0 (www.itksnap.org), adhering to the following standardized protocol for consistency:20 An initial lower threshold of 20 HU and an upper threshold of 75 HU were applied. We chose these thresholds on the basis of maximal differentiation between brain parenchyma and CSF on visual inspection. We then placed 10-mm seeds in the centrum semiovale, corona radiata, basal ganglia, adjacent to the occipital horn of each lateral ventricle, and pons. Seeds were grown to fill the desired volume, followed by inspection of the segmentations and growing of additional seeds to fill parenchyma not included by active contour evolution. Structures spuriously included in the initial segmentation, such as paraspinal soft tissues or the optic nerve, were then manually edited. The inferior extent of the cerebellar tonsils was defined as the caudal limit of the segmentation volume. A representative final segmen-
tation is demonstrated in Fig 2. To calculate total intracranial volume, we started with the previously computed brain segmentation. We then used a lower threshold of −20 HU and an upper threshold of 100 HU to maximize differentiation between the skull and intracranial contents. The segmentation was then grown to fill the intracranial volume, and structures spuriously included, such as paraspinal soft tissues or the optic nerve, were then manually edited. The inferior extent of the cerebellar tonsils was also defined as the caudal limit of the intracranial segmentation volume.

Statistical Analysis
Statistical analysis was performed using STATA, Version 12.1 (StataCorp, College Station, Texas). Each individual’s brain volume was initially normalized to his or her total intracranial volume. Continuous demographic variables were compared using the Mann-Whitney U test, and categoric variables, with the χ² or Fisher’s exact test when appropriate. Interrater reliability was assessed using the intraclass correlation coefficient. Brain volumes were compared within individuals across timepoints using the Wilcoxon signed rank test. Absolute changes in brain volumes were calculated between the initial scan and the index scan and the index scan and follow-up for both the patients and controls. Differences in absolute volume changes between patients and controls were calculated using the Mann-Whitney U test. A 2-tailed P value < .05 was considered statistically significant.

RESULTS
We identified 28,777 patients who underwent NCCTH (“index CT”) in 1 emergency department for evaluation of traumatic head injury. Of these, 751 also underwent NCCTH within 1–12 months both before (“baseline CT”) and after (“follow-up CT”) the index CT (Fig 1). Of the 751 patients, 48 patients had no visible CT abnormalities on any CT or history of head trauma at the time of either the baseline CT or follow-up CT. In addition, 48 matched control subjects were identified who underwent 3 CT examinations, which were similarly named (baseline, index, and follow-up CT).

Patient demographics did not differ between patients with mTBI and controls (Table 1). Similarly, there was no significant difference in the median time between baseline, index, and fol-
Although CT is the recommended initial method for evaluating patients with mild traumatic brain injury, to our knowledge, this is the first report of quantitative analysis of brain volume changes derived from NCCTH in patients with mTBI. While we do not propose the use of CT for routine clinical follow-up, the ubiquity of NCCTH performed during routine clinical practice in the emergency department makes these data ripe for analysis to characterize and gain insight into the underlying disease process. We found a significant decrease in brain volumes from the time of mTBI to follow-up 1–12 months later, with evidence of stable brain volume during 1–12 months before mTBI and across similar timeframes in controls evaluated in the emergency department for reasons other than TBI.

Changes in brain volume following mTBI have been described in many previous studies to correlate with clinical sequelae of long-term brain injury. In 2 studies that reported whole-brain volume changes in patients with TBI with a range of severity, average total brain volume decreases of 7.6 mL were found at 1 year in patients with mTBI and 157.3 mL in patients with mild and moderate TBI. For reference, normal human total brain volume at 32 years of age averages 1273.6 mL in men and 1131.1 mL in women and declines by about 0.2%–0.5% per year in healthy individuals. We showed an absolute decrease of approximately 40 mL after mTBI at a median of 149 days follow-up. Although our study and others showed significant brain volume changes, some studies failed to show any volume differences after mTBI. Most studies that reported changes in regional brain volume showed changes in regions such as the frontal gyri, precuneus, temporal gyri, caudate, cingulum, and hippocampus. The ITK-SNAP protocol we used is robust to discrimination of bone, fluid, and soft tissue on CT but cannot discriminate soft-tissue differences (eg, gray versus white matter) sufficiently to allow regional segmentation. Currently available regional segmentation algorithms such as FreeSurfer (http://surfer.nmr.mgh.harvard.edu) are validated for parcelation of MR imaging. CT images provide substantially lower soft-tissue contrast than MR imaging, and we are not aware of any validated methods for parcellation of brain regions based on CT.

Although decreases in brain volume have been reported after mTBI, it is unclear if these changes reflect resolution of trauma-related edema, which might be at a level not detectable on visual inspection, or development of atrophy in the wake of injury. Animal models have shown both early cortical thickening due to transient edema after trauma followed by reduction of cortical thickness due to atrophy. Our findings of stable brain volume between pretrauma and index mTBI scans with subsequent decreases in volume at follow-up suggest neurodegeneration follow ing mTBI rather than acute edema at the time of mTBI with subsequent resolution. Most volumetric studies of mTBI do not include comparison with preinjury imaging. However, a longitudinal study of ice hockey players included MR imaging at the beginning and end of a season of play as well as additional scans for players who sustained mTBI during that time. Brain volumes at the end of the hockey season declined in all players, regardless of incident mTBI, and no increase in brain volume

### DISCUSSION

#### Table 1: Subject demographics

<table>
<thead>
<tr>
<th></th>
<th>mTBI (n = 48)</th>
<th>Controls (n = 48)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median age (IQR) (yr)</td>
<td>63 (52–79)</td>
<td>71 (50–82)</td>
<td>.202</td>
</tr>
<tr>
<td>Male sex percent (%)</td>
<td>42% (20)</td>
<td>37% (18)</td>
<td>.289</td>
</tr>
<tr>
<td>Median pre- to index interval (IQR) (day)</td>
<td>118 (67–239)</td>
<td>171 (87–274)</td>
<td>.123</td>
</tr>
<tr>
<td>Median index to post interval (IQR) (day)</td>
<td>149 (79–204)</td>
<td>124 (61–219)</td>
<td>.901</td>
</tr>
</tbody>
</table>

* Mann-Whitney U test.
* χ² test.

#### Table 2: No significant differences in comorbidities that may affect brain volume between cases and controls

<table>
<thead>
<tr>
<th></th>
<th>mTBI (n = 48)</th>
<th>Controls (n = 48)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dementia</td>
<td>23% (11)</td>
<td>25% (12)</td>
<td>&gt; .999</td>
</tr>
<tr>
<td>End-stage renal disease</td>
<td>8% (4)</td>
<td>8% (4)</td>
<td>.999</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>48% (23)</td>
<td>38% (18)</td>
<td>.409</td>
</tr>
<tr>
<td>Hypertension</td>
<td>75% (36)</td>
<td>77% (37)</td>
<td>&gt; .999</td>
</tr>
<tr>
<td>Cardiac arrhythmia/valvular disease</td>
<td>19% (9)</td>
<td>19% (9)</td>
<td>&gt; .999</td>
</tr>
<tr>
<td>Seizure disorder</td>
<td>10% (5)</td>
<td>10% (5)</td>
<td>&gt; .999</td>
</tr>
<tr>
<td>Cirrhosis</td>
<td>8% (4)</td>
<td>6% (3)</td>
<td>&gt; .999</td>
</tr>
<tr>
<td>Psychiatric disorders</td>
<td>30% (14)</td>
<td>19% (9)</td>
<td>.339</td>
</tr>
</tbody>
</table>

* Fisher exact test.
was detected at the time of mTBI. 12 The decline in brain volumes regardless of concussion status in these hockey players could represent a consequence of repeat game-related brain trauma in the entire cohort.

In addition to mTBI, many other clinical disorders have been shown to cause longitudinal change of brain volume, including dehydration, schizophrenia, neurodegenerative disorders, and normal aging.32-36 Although our patient group with mTBI did exhibit comorbid medical conditions, the control group, who were of similar age, exhibited a similar prevalence of comorbid disorders but did not demonstrate any significant change in brain volume during similar time periods. This finding makes comorbid disease an unlikely explanation for the effects we found. We considered that administration of IV fluid in the emergency department might result in an increase in brain volume, with a subsequent decrease detectable on follow-up. However, because all 3 CT examinations were performed in the emergency department but the only significant change occurred between the index and follow-up CT, and only in patients with mTBI, we consider hydration status an unlikely explanation of the findings.

Our study has several limitations. First, the retrospective design cannot confirm a causal relationship between incident mTBI and brain volume change. Second, our reliance on clinical information collected as a part of routine clinical care may be inadequate to determine the true nature of a patient’s clinical condition. Third, because of the inconsistency of clinical record entries created by different clinicians across time, we could not correlate the volume changes we identified with consistently measured clinical outcomes. Fourth, decline in brain volume with time could confound our results, particularly in older individuals. However, reference to the published changes in volume due to aging during 12 months makes it highly implausible that we would detect a change due to aging in our follow-up interval (mTBI median, 267 days; control median, 295 days). Moreover, to detect a decline due to normal aging during only 1 of the 2 sequential follow-up intervals we examined would be very unlikely. Last, our CT protocols could not be optimized prospectively, creating the potential for bias due to image quality. We were, however, able to confirm that there were no changes in imaging equipment during the timeframe of each subject’s 3 CT examinations, that voxel size was identical across all CT examinations, and that imaging parameters for the studies we included differed minimally, if at all. Moreover, minimal changes in imaging parameters limitation is ubiquitous in long-term studies, especially retrospective analyses.

CONCLUSIONS

Semiautomated volumetric analysis of the NCCTHs is reliable and detected a significant decrease in normalized brain volumes following mTBI in the setting of preinjury volumetric stability, suggesting neurodegeneration as the primary mechanism for volume loss. Further study is warranted to determine how these changes correlate with clinical outcomes.

REFERENCES


3D Black-Blood Luminal Angiography Derived from High-Resolution MR Vessel Wall Imaging in Detecting MCA Stenosis: A Preliminary Study

X. Bai, P. Lv, K. Liu, Q. Li, J. Ding, J. Qu, and J. Lin

ABSTRACT

BACKGROUND AND PURPOSE: 3D high-resolution vessel wall imaging is increasingly used for intracranial arterial diseases. This study compared the diagnostic performance of black-blood luminal angiography derived from 3D vessel wall imaging with source images of vessel wall imaging and TOF-MRA in detecting middle cerebral artery stenosis.

MATERIALS AND METHODS: Sixty-two patients with suspected MCA atherosclerosis underwent TOF-MRA, vessel wall imaging, and CTA. Intracranial black-blood luminal angiography was created from source images of vessel wall imaging using minimum intensity projection. The degree and length of MCA stenosis were measured on source images of vessel wall imaging, TOF-MRA, and black-blood luminal angiography and compared using CTA as a reference standard.

RESULTS: The image quality of black-blood luminal angiography was diagnostic in most patients. The intra- and interobserver agreement for both stenosis degree and length measurements was excellent for black-blood luminal angiography. It was comparable with that of source images of vessel wall imaging in grading stenosis. Compared with TOF-MRA, black-blood luminal angiography showed significantly higher sensitivity for the detection of severe stenosis (89.3% versus 64.3%, \( P < .0039 \)) and higher specificity for the detection of occlusion (95.4% versus 84.6%, \( P = .039 \)). Lesion length estimated on source images of vessel wall imaging was significantly greater than that measured by CTA and black-blood luminal angiography (\( P < .0001 \) and \( P = .010 \)).

CONCLUSIONS: Black-blood luminal angiography is better than TOF-MRA in detecting severe stenosis and occlusion of the MCA. Compared with source images of vessel wall imaging, it is more accurate in evaluating stenosis length. Black-blood luminal angiography can be produced as a derivative from vessel wall imaging and implemented as an adjunct to vessel wall imaging and TOF-MRA without extra acquisition time.

ABBREVIATIONS: BBLA = black-blood luminal angiography; VWI = vessel wall imaging

Intracranial atherosclerotic stenosis is an important cause of ischemic stroke and most often involves the proximal segment of the middle cerebral artery (MCA).\(^1\) Imaging evaluation of the stenosis is critical to treatment planning. Traditionally, the degree of stenosis is evaluated by luminal angiography techniques, including DSA, CTA, and MRA. DSA, although long considered the criterion standard in assessing intracranial atherosclerotic stenosis, is limited by invasiveness, ionizing radiation, and high cost. CTA has been widely used for evaluating intracranial artery stenosis with high accuracy,\(^2\) but it also exposes patients to ionizing radiation. Both DSA and CTA use iodinated contrast material and are restricted in patients who have previous adverse reactions to iodine and in patients with impaired renal function. Time-of-flight imaging is the most commonly used unenhanced MRA technique for intracranial arterial imaging; however, its major weakness is local signal loss resulting from slow and turbulent flow.\(^3\)

Recently, intracranial high-resolution vessel wall imaging (VWI) has been implemented for direct depiction of the intracranial arterial wall. Beyond the features of the arterial wall, dimensions of the vessel lumen could be reliably measured on the cross-sectional images of VWI.\(^4-6\) With large spatial coverage, high signal-to-noise ratio, and flexible orientation of the reformatted images, 3D VWI is often chosen to demonstrate tortuous cerebral arteries. Clear demonstration of the vessel wall requires suppression of the MR signal of the luminal spins to yield minimum or no
signal from the flowing blood. Based on these characteristics of 3D VWI, we attempted to convert the already acquired vessel wall images, referred to as source images of VWI, into 3D black-blood luminal angiography (BBLA) using minimum intensity projection. In this study, we used CTA as the reference standard and prospectively compared BBLA with source images of VWI and with TOF-MRA for measuring both the stenosis degree and stenosis length of the MCA. Our hypothesis was that this product of 3D VWI would be superior to source images of VWI and TOF-MRA in the evaluation of MCA stenosis.

MATERIALS AND METHODS

Patients
From February 2017 to January 2018, sixty-two consecutive patients suspected of having intracranial atherosclerotic stenosis involving the MCA were prospectively included in the study. Patients were excluded if they had contraindications to MR imaging or nonatherosclerotic intracranial diseases, such as vasculitis, Moyamoya disease, and arterial dissection, which were clinically diagnosed according to their diagnostic criteria.7–9 Patients with extracranial carotid artery stenosis of >50% were also excluded. All patients underwent intracranial TOF-MRA and 3D VWI, followed by CTA within 1 week. Institutional review board approval for the study design was obtained, and informed consent was obtained from all patients.

MR Imaging
All examinations were performed on a 3T MR imaging system (Discovery MR750; GE Healthcare, Milwaukee, Wisconsin) with an 8-channel head coil. The MR imaging protocol included TOF-MRA and high-resolution VWI. TOF-MRA was acquired first through the circle of Willis with the following parameters: TR, 23 ms; TE, 3.4 ms; flip angle, 20°; FOV, 22 × 22.7 cm; matrix size, 320 × 224; slice thickness, 1.4 mm; mean acquisition time, 5 minutes 8 seconds. Using TOF-MRA as a localizer, we performed pulse sequences with different image contrast weightings, including 2D FSE T2WI, 3D-Cube T1WI (GE Healthcare), and 3D-Cube proton-density-weighted imaging, for assessing the intracranial vessel walls. The parameters for targeted 2D FSE T2WI were the following: TR, 2150 ms; TE, 50 ms; flip angle, 111°; FOV, 10 × 10 cm; matrix, 384 × 384; slice thickness, 2 mm; slice gap, 0.5 mm; NEX, 12; mean acquisition time, 4 minutes 23 seconds. The parameters for 3D-Cube T1WI were as follows: TR, 600 ms; TE, 13 ms; FOV, 20 × 20 cm; matrix size, 288 × 288; slice thickness, 1 mm; NEX, 0.5; mean acquisition time, 4 minutes 16 seconds. 3D-Cube proton-density-weighted imaging using variable flip angle refocusing pulses was performed with the following parameters: TR, 2500 ms; TE, 25 ms; FOV, 20 × 20 cm; matrix size, 288 × 288, slice thickness, 1 mm; NEX, 0.5; and mean acquisition time, 5 minutes 13 seconds. 3D-Cube proton-density-weighted imaging was performed in the sagittal plane orthogonal to the proximal MCA with intravoxel dephasing for suppression of the signal from the blood. The imaging volume was centered at the circle of Willis with voxel dimensions of 0.69 × 0.69 × 1 mm and displayed with a reconstructed resolution of 0.39 × 0.39 × 0.5 mm.

Source images from both TOF-MRA and 3D-Cube proton-density-weighted imaging were transferred to an ADW4.6 workstation (GE Healthcare). MIP and MPR images were created from TOF images. Minimum intensity projection was used to generate BBLA from 3D-Cube proton-density-weighted imaging.

CTA
All craniocervical CTAAs, from the aortic arch to the distal intracranial arteries, were performed on a 320-detector row CT system (Aquilion ONE CT scanner; Toshiba Medical Systems, Tokyo, Japan) with the following scanning parameters: tube voltage, 120 kV(peak); tube current, 146–210 mA; slice thickness, 0.5 mm; FOV, 22 × 22 cm; and matrix size, 512 × 512. With a power injector, 50–70 mL of nonionic iodinated contrast media (Ultra- viest 370, iopromide; Bayer HealthCare, Berlin, Germany) was injected intravenously at a rate of 5 mL/s followed by a 30-mL saline bolus. CT scanning was initiated using a bolus-tracking technique at the level of the aortic arch with a trigger threshold of 150 HU.

CTA raw images were reformatted on a Vitrea fX workstation (Version 3.10; Toshiba Medical Systems) using MIP and MPR reconstructions for stenosis evaluation.

Image Analysis
First, an experienced neuroradiologist assessed the image quality of BBLA in demonstrating the MCA. It was graded on a 3-point scale as follows15: 1, poor visualization or nondiagnostic; 2, moderate visualization, adequate for diagnostic purposes; and 3, good, high quality for diagnostic purposes. Patients in whom image quality was grade 1 were excluded from the comparison. Second, 2 other experienced neuroradiologists independently quantified both the degree and length of the stenosis of the MCA from 3D-Cube proton-density-weighted images, BBLA, TOF-MRA, and CTA. All cases were reviewed in random order. The measurements were performed using electronic calipers on the workstation. For all imaging modalities, the Warfarin Aspirin Symptomatic Intracranial Disease criteria were used to measure the MCA stenosis11,12: luminal stenosis = [1 - (Dstenosis / Dnormal)] × 100%, where Dstenosis indicates the diameter of the residual lumen at the site of the most severe degree of stenosis, and Dnormal, the diameter of the proximal normal artery. If the proximal artery was diseased, the diameter of the distal portion of the artery at its widest point could be used instead. In case of multiple stenosis in the MCA, the most severe one was chosen for evaluation. The degree of stenosis was graded as normal (<30%), mild (30%–49%), moderate (50%–69%), severe stenosis (70%–99%), and occlusion (100%).4,13 Arteries with stenosis of <30% on all these modalities were excluded from analysis.14 Stenosis length was defined as the distance from the proximal-to-distal end of the stenotic artery.15 On source images of VWI, however, the lesion length could not be measured directly but was estimated by multiplying the slice thickness and the number of the slices showing the luminal stenosis in the MCA. The lesion length was not measured in cases with total occlusion. One of the 2 neuroradiologists re-evaluated the BBLA 4 weeks later to assess intraobserver agreement. For the degree and length of stenosis, the average of the 2 radiologists’ measurements was used for the final analysis.
Comparison of the degree of stenosis with source images of VWI, BBLA, and TOF-MRA with CTA

<table>
<thead>
<tr>
<th>Stenosis Degree on VWI/BBLA/TOF-MRA</th>
<th>Stenosis Degree on CTA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30%–49%</td>
</tr>
<tr>
<td>30%–49%</td>
<td>15/15/14</td>
</tr>
<tr>
<td>50%–69%</td>
<td>1/1/2</td>
</tr>
<tr>
<td>70%–99%</td>
<td>0/0/0</td>
</tr>
<tr>
<td>100%</td>
<td>0/0/0</td>
</tr>
</tbody>
</table>

**RESULTS**

The image quality of BBLA in demonstrating the MCA was graded as good in 33 patients and moderate in 27 patients. Two of 62 (3.2%) patients were excluded because of poor image quality from motion artifacts. A total of 80 diseased middle cerebral arteries from 60 patients (20 patients had bilateral MCA lesions) were further examined in this study. Among these patients, there were 33 men (55%) and 27 women (45%) with a mean age of 61.3 years (range, 45–80 years). Thirty-nine patients (65%) had ischemic stroke or transient ischemic attack with hemiparesis, language disorder, or perceptual deficits. The remaining 21 patients (35%) presented with dizziness or headache.

The intra- and interobserver agreement for both stenosis degree (intraclass correlation coefficient = 0.982–0.929, respectively) and length measurements (intraclass correlation coefficient = 0.953–0.939, respectively) were excellent for BBLA.

The Table summarizes the results of MCA stenosis degree measured with source images of VWI, BBLA, and TOF-MRA in comparison with CTA. For stenosis degree, there was excellent agreement between source images of VWI and CTA (κ = 0.956; 95% CI, 0.913–0.998), excellent agreement between BBLA and CTA (κ = 0.934; 95% CI, 0.882–0.986), and good agreement between TOF-MRA and CTA (κ = 0.800; 95% CI, 0.717–0.883) (Fig 1). Sensitivity and specificity in detecting severe stenosis were 92.9% and 98.1% with source images of VWI, 89.3% and 96.2% with BBLA, and 64.3% and 88.5% with TOF-MRA. Sensitivity and specificity in detecting occlusion were 100% and 96.9% with source images of VWI, 100% and 95.4% with BBLA, and 100% and 84.6% with TOF-MRA. The sensitivity and specificity of source images of VWI and BBLA for the detection of severe stenosis or occlusion are comparable (sensitivity, P = .500 and P = 1.000; specificity, P = 1.000 and P = .500). Compared with TOF-MRA, BBLA showed significantly higher sensitivity for the detection of severe stenosis (89.3% versus 64.3%, P = .039) and higher specificity for the detection of occlusion (95.4% versus 84.6%, P = .039). BBLA demonstrated arteries distal to severe stenosis and occlusion in 11 cases in which TOF-MRA did not (Fig 2). The average length of the stenosis was 4.25 ± 1.58 mm estimated with source images of VWI, 4.10 ± 1.48 mm measured on BBLA, 4.11 ± 1.65 mm on TOF-MRA, and 3.97 ± 1.63 mm on CTA. The measurement of stenosis length between BBLA and CTA (P = .060) and between TOF-MRA and CTA (P = .054) was not significantly different.

**FIG 1.** 3D proton-density-weighted vessel wall source image (A), black-blood luminal angiography (B), TOF-MRA (C), and CTA (D) all depict moderate stenosis of the left M1 MCA (arrow). The 3D vessel wall image (A) shows eccentric plaque (arrow) on the MCA wall. The inset is a magnified vessel wall image. These images are used as source images to generate luminal angiography with a minimum intensity projection.

**Statistical Analysis**

Commercially available software (MedCalc for Windows, Version 18; MedCalc Software, Mariakerke, Belgium) was used for the statistical analysis. Intra- and interobserver agreement in the measurement of stenosis degree and length by BBLA was assessed using the intraclass correlation coefficient. Agreement between the source images of VWI and CTA, between BBLA and CTA, as well as between TOF-MRA and CTA for the evaluation of MCA stenosis degree was calculated with the Cohen κ test. Sensitivity and specificity in the detection of severe stenosis and occlusion of the MCA were calculated for source images of VWI, BBLA, and TOF-MRA using CTA as the reference standard. A Wilcoxon matched pairs test was performed to test whether stenosis lengths obtained from different imaging modalities were statistically different. The strength of agreement of the κ and intraclass correlation coefficient was categorized as follows: poor, <0.20; fair, 0.21–0.40; moderate, 0.41–0.60; good, 0.61–0.80; and excellent, 0.81–1.00. A 2-tailed P value <.05 was considered indicative of a significant difference.
However, compared with CTA and BBLA, the estimated stenosis length was significantly larger with source images of VWI ($P < .001$ and $P = .010$, respectively).

Twelve of 60 patients underwent DSA before possible interventional angioplasty. Fourteen diseased MCAs were identified on DSA. They were 9 occlusions, 3 severe stenoses, 1 moderate stenosis, and 1 mild stenosis. CTA, BBLA, and source images of VWI correctly diagnosed all of the 14 diseased MCAs (Fig 3) compared with DSA.

**DISCUSSION**

The current common practice for delineating the intracranial vessel wall is VWI; the volumetric proton intensity–weighted images obtained from VWI were used as source images in this study. With the application of minimum intensity projection to these source images, we further obtained BBLA, which provided additional advantages compared with VWI alone. Our study demonstrates that the presentation of stenosis degree and length of the MCA on BBLA was in excellent agreement with the observations on CTA, which supported the high diagnostic quality of BBLA. Compared with source images of VWI, with comparable performance in the grading of stenosis, BBLA was noticeably more accurate in the measurement of stenosis length. We also found that BBLA was superior to TOF-MRA in terms of its higher sensitivity for the detection of severe stenosis and higher specificity for occlusion of the MCA. Furthermore, since BBLA did not take any additional acquisition time, we recommend its use as an adjunct to VWI and TOF-MRA in the evaluation of MCA stenosis during a single session of intracranial VWI.

In addition to its accuracy as source images of VWI in the grading of stenosis, BBLA also showed the capacity to present combined information from both VWI and MRA. By selectively suppressing the signal from blood flow, 3D black-blood MRA has been used for the assessment of vessel lumens with medium-to-large diameters, such as the carotid artery.\textsuperscript{15,16} With technical development, the application of 3D black-blood MRA has been extended to arteries of smaller diameters such as small intracranial arteries.\textsuperscript{17,18} Despite the improved resolution in luminal imaging, 3D black-blood MRA cannot provide sufficient information on the vessel wall. However, 2D/3D high-resolution
VWI is a sequence specifically tailored to the intracranial vessel walls by sampling the cross-sections of the target vessels. Another application of VWI is that it can evaluate the luminal stenosis of the MCA with high accuracy, as already reported by some earlier studies.

Until now, no studies have yet reported the use of VWI for evaluating MCA stenosis length, which is also a critical parameter for potential interventional treatment. It is well-known from previous studies of CTA and MRA that multiple postprocessing techniques can provide an overview of the entire target vessel in a much more straightforward manner than the cross-sectional source images. Similarly, source VWI provides only the 2D cross-sectional profile of a vessel, making it challenging to form a full picture of the diseased vessel based on individual source images. Without a direct measurement of the stenosis length, however, an intuitive estimation from VWI source images is inaccurate. As shown in this study, this estimation was significantly greater than the results obtained from CTA or BBLA. In contrast, BBLA shows the entire MCA lumen from different perspectives, whereby both the degree and the length of the stenosis can be evaluated easily and accurately. More important, BBLA can be presented like conventional TOF-MRA, CTA, and DSA, with which our clinical colleagues are already familiar.

3D TOF-MRA is conventionally used for detecting luminal narrowing and serves as a localizer for subsequent VWI. However, its imaging quality depends on the prominent inflow effect of blood spins. It has intravascular signal loss at the stenosis, where blood flow is slow and turbulent, especially in small intracranial arteries. Therefore, to accurately define the stenosis, especially more severe ones, additional gadolinium-enhanced MRA or CTA is sometimes required. In this study, TOF-MRA was found to be less accurate than BBLA in detecting and differentiating severe stenosis and occlusion. We recorded 18 cases of overestimation of the stenosis when TOF-MRA was used, compared with 6 when BBLA was used. Furthermore, BBLA was superior to TOF-MRA in the demonstration of arteries or collaterals distal to the severe stenosis or occlusion, which is crucial for clinical intervention. Nevertheless, BBLA still overestimated stenoses in 6 cases, while source images of VWI overestimated stenoses in 4 cases in this study. The possible contributing factors were the suboptimal suppression of the blood signal and the anisotropic resolution in our study. Nevertheless, BBLA still overestimated stenoses in 6 cases, while source images of VWI overestimated stenoses in 4 cases in this study. The possible contributing factors were the suboptimal suppression of the blood signal and the anisotropic resolution in our reconstructed luminal images.

This study has several limitations. First, a small number of patients were enrolled, which may limit the statistical significance. Second, CTA rather than DSA was performed for comparison. However, CTA has been well-established as a highly accurate approach in identifying intracranial atherosclerotic stenosis. Using 16–detector row CT, a prior study reported that CTA performed very well compared with DSA for the detection of >50% intracranial stenosis with 97.1% sensitivity and 99.5% specificity. In another earlier study with a 4–detector row technique, Bash et al found that CTA had a higher sensitivity and positive predictive value than TOF-MRA for the detection of intracranial stenosis and occlusion. Compared with DSA, which was available in 12 patients in our study, CTA provided identical information on MCA stenosis. Accordingly, invasive DSA is now largely reserved for use before interventional treatment rather than for diagnosis alone. Third, our study was based on a single vascular pathology (ie, MCA stenosis). Although the MCA is the most frequently involved intracranial artery in stroke, the value of BBLA in other diseased vessel territories warrants further investigation. Fourth, owing to technical restrictions, 3D-Cube proton-density VWI was acquired with an anisotropic resolution, which may impair the demonstration of stenosis on reformatted BBLA.

CONCLUSIONS
BBLA is comparable with CTA in the evaluation of stenosis degree and the length of the MCA. It is better than TOF-MRA in detecting severe stenosis and occlusion and more accurate than VWI in measuring the stenosis length. BBLA can be produced as a derivative from VWI without extra acquisition time; therefore, it could be implemented as an adjunct to VWI and TOF-MRA.

REFERENCES
Clinical Evaluation of Highly Accelerated Compressed Sensing Time-of-Flight MR Angiography for Intracranial Arterial Stenosis

S.s. Lu, M. Qi, X. Zhang, X.h. Mu, M. Schmidt, Y. Sun, C. Forman, P. Speier, and X.n. Hong

ABSTRACT

BACKGROUND AND PURPOSE: Time-of-flight MR angiography is the preferred imaging technique to assess intracranial arterial stenosis but is limited by a relatively long acquisition time. Compressed sensing provides an innovative approach in undersampling k-space to minimize the data-acquisition time. We aimed to evaluate the diagnostic accuracy of compressed sensing TOF for detecting intracranial arterial stenosis by comparison with conventional parallel imaging TOF-MRA.

MATERIALS AND METHODS: Compressed sensing TOF and parallel imaging TOF were performed in 22 patients with intracranial arterial stenosis. The MRA scan times were 2 minutes and 31 seconds and 4 minutes and 48 seconds for compressed sensing TOF and parallel imaging TOF, respectively. The reconstructed resolutions were 0.4×0.4×0.4 and 0.4×0.4×0.6 mm³ for compressed sensing TOF and parallel imaging TOF, respectively. The diagnostic quality of the images and visibility of the stenoses were independently ranked by 2 neuroradiologists blinded to the type of method and were compared using the Wilcoxon signed rank test. Concordance was calculated with the Cohen κ. Edge sharpness of the arteries and the luminal stenosis ratio were analyzed and compared using a paired-sample t test.

RESULTS: The interrater agreement was good to excellent. Compressed sensing TOF resulted in image quality comparable with that of parallel imaging TOF but boosted confidence in diagnosing arterial stenoses (P = .025). The edge sharpness of the intracranial arteries for compressed sensing TOF was significantly higher than that for parallel imaging TOF (P < .001). The luminal stenosis ratio on compressed sensing TOF showed no significant difference compared with that on parallel imaging TOF.

CONCLUSIONS: Compressed sensing TOF both remarkably reduced the scan time and provided adequate image quality for the diagnosis of intracranial arterial stenosis.

ABBREVIATIONS: CS = compressed sensing; GRAPPA = generalized autocalibrating partially parallel acquisition; PI = parallel imaging

Cerebrovascular disease is a major cause of morbidity and mortality worldwide and can arise from several intracranial vessel wall pathologies, such as atherosclerosis, dissection, and vasculitis.1,2 Studies have found that intracranial arterial stenosis is highly prevalent in fatal stroke.3 The imaging of intracranial vessels is an important tool for the clinical evaluation of cerebrovascular disease.

The preferred imaging techniques used to assess intracranial arterial stenosis in clinical practice include CTA and MRA. The limitations of CTA include exposure to radiation, the use of iodinated contrast agents, and impaired accuracy in the presence of vascular calcifications.4,5 Time-of-flight MRA is widely used and is a noninvasive technique for intracranial vascular evaluation that requires no exogenous contrast agent. Because conventional TOF-MRA is a rather slow imaging technique, pursuing high spatial resolution is challenging. The spatial coverage is often compromised to achieve a half-millimeter resolution and good signal-to-noise ratio, while keeping the scan time clinically acceptable.6 Furthermore, relatively long acquisition times may lead to motion artifacts, which disrupt the detection of vascular lesions. Further limitations of conventional TOF-MRA include a loss of signal intensity related to turbulent and slow flow.7

Parallel imaging (PI) is routinely used as a method for k-space undersampling during the acquisition of TOF-MRA. However, the acceleration factor of PI is often only 2- or 3-fold because of the rapid increase in noise or aliasing at higher acceleration factors.8-10 Compressed sensing (CS) provides an innovative ap-
Table 1: MR imaging parameters for CS-TOF and PI-TOF

<table>
<thead>
<tr>
<th>Parameters</th>
<th>CS-TOF</th>
<th>PI-TOF</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOV (mm²)</td>
<td>220 × 200</td>
<td>220 × 200</td>
</tr>
<tr>
<td>TR/TE (ms)</td>
<td>21/3.49</td>
<td>21/3.49</td>
</tr>
<tr>
<td>Flip angle</td>
<td>18°</td>
<td>18°</td>
</tr>
<tr>
<td>Matrix</td>
<td>368 × 334</td>
<td>368 × 334</td>
</tr>
<tr>
<td>Slice thickness (mm)</td>
<td>0.4</td>
<td>0.6</td>
</tr>
<tr>
<td>No. of slabs</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Slices per slab</td>
<td>60</td>
<td>40</td>
</tr>
<tr>
<td>Slice oversampling</td>
<td>20%</td>
<td>20%</td>
</tr>
<tr>
<td>Phase partial Fourier factor</td>
<td>None</td>
<td>6/8</td>
</tr>
<tr>
<td>Slice partial Fourier factor</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Acceleration factor</td>
<td>10.3</td>
<td>GRAPPA 2</td>
</tr>
<tr>
<td>Reconstructed voxel size (mm³)</td>
<td>0.4 × 0.4 × 0.4</td>
<td>0.4 × 0.4 × 0.6</td>
</tr>
</tbody>
</table>

Data were reconstructed using 10 iterations of the Modified Fast Iterative Shrinkage-Thresholding Algorithm according to a previous report. Maximum-intensity-projection images of the axial, coronal, and sagittal views were reconstructed.

**Imaging Evaluation**

Two senior neuroradiologists (M.Q. and S.S.L., with 10 and 6 years’ experience, respectively), who were blinded to both the patients’ information and the type of reconstruction undertaken, independently assessed the diagnostic quality of the images. All MIP images of PI-TOF and CS-TOF were presented to the 2 neuroradiologists in random order. The diagnostic quality of the images was graded on an ordinal scale from 0 to 3, with 0 indicating completely blurred arteries and severe artifacts; 1 indicating partially obscured arteries and moderate artifacts; 2 indicating good and clear arteries and slight artifacts; and 3 indicating excellent arteries and no artifacts. For visualization of the arterial stenosis, the readers first identified a stenosis on the MIP images, and then declared their level of confidence using a 3-point scoring scheme as follows: grade 2, definite stenosis and sufficiently recognized, high confidence; grade 1, probable stenosis, moderately confident; grade 0, low confidence. The location of the arterial stenosis was also recorded. For any discrepancy between the 2 readers, another senior neuroradiologist (X.N.H. with 20 years’ experience) re-evaluated the images and assisted in reaching a consensus agreement. The consensus scores were used for the subsequent analyses.

The degree of luminal stenosis was calculated using the following formula: stenosis ratio = (1 – Narrow Lumen Diameter/Reference Lumen Diameter) × 100%, according to the Warfarin-Aspirin Symptomatic Intracranial Disease criterion. The reference lumen was defined as the neighboring segment of normal appearance proximal to the stenotic site. The measurement results of the luminal stenosis ratio from the 2 neuroradiologists were averaged for subsequent analysis.

After that, the MIP images were presented to the 2 neuroradiologists in randomized order for a side-by-side comparison the next day. The readers were blinded to the type of reconstruction. They viewed the MIP images of CS-TOF and PI-TOF simultaneously and ranked the images in order of diagnostic quality preference (CS-TOF better than PI-TOF, or equivalent, or PI-TOF better than CS-TOF) on the basis of the criteria that included delineation of the cerebral arteries and recognition of arterial stenosis.

**Arterial Sharpness Evaluation**

The edge sharpness of the intracranial arteries was calculated by a perceptual image sharpness metric called the Perceptual Sharpness Index based on MIP images, according to a previous report, using Matlab (2013b; MathWorks, Natick, Massachusetts). The Perceptual Sharpness Index represents the perceived sharpness in an image. Briefly, the method estimates the sharpness on the basis of a statistical analysis of local edge gradients. It is a no-reference metric and takes properties of the human visual system into account. Based on perceptual properties, a relationship between the extracted statistical features and the metric score is established to form a Perceptual Sharpness Index.
**Statistical Analysis**

The interrater reliability was performed using the Cohen $\kappa$ analysis for grading the diagnostic quality and visualization of arterial stenosis. The intraclass correlation coefficients were calculated for measurement of the luminal stenosis ratio. Reliabilities $<$0.4 were characterized as poor; 0.4–0.6, fair; 0.6–0.8, good; and those $>$0.8 were deemed excellent. Continuous data were summarized as mean ± SD, and they were assessed for normality by the Kolmogorov-Smirnov test before further comparison. Categorical data were recorded as counts and percentages. The diagnostic quality and visualization of arterial stenoses on PI-TOF and CS-TOF were compared using the Wilcoxon rank test. The luminal stenosis ratio and the edge sharpness of the intracranial arteries were compared using a paired-sample $t$ test if the data were normally distributed or the Wilcoxon signed rank test as appropriate. All the statistical analyses were performed using SPSS (Version 16.0; IBM, Armonk, New York). The $P$ value was 2-sided, and $P < .05$ was considered statistically significant.

**RESULTS**

**Patient Characteristics**

Among the 22 patients (11 men, 11 women; 28–86 years of age), 12 had multiple segments of intracranial arterial stenosis and 3 were eventually diagnosed with Moyamoya disease. No patient with a single stenosis seen only on 1 of the TOF sequences was found. In total, 48 arterial segments with stenosis were diagnosed. The demographic details of the patients in this study are listed in Table 2.

**Imaging Evaluation**

The interrater agreement for diagnostic-quality grading was 0.776 and 0.753 for CS-TOF and PI-TOF, respectively. The senior neuroradiologist had to decide on a consensus for 1 of 22 (4.5%) cases of CS-TOF and 2 of 22 (9.1%) cases of PI-TOF. The diagnostic quality of most CS-TOF (90.9%) and PI-TOF (77.3%) images was graded as excellent. The CS-TOF images provided comparable diagnostic quality with the PI-TOF images ($P = .046$). The mean edge sharpness of the intracranial arteries for CS-TOF was $0.358 ± 0.038$, significantly higher than that for PI-TOF ($0.267 ± 0.042$) ($P < .001$).

The interrater agreement for visualization of arterial stenosis was 1.000 and 0.778 for CS-TOF and PI-TOF, respectively. The senior neuroradiologist had to decide on a consensus for 2 of 48 (4.2%) segments on PI-TOF. Forty-eight (100%) and 43 (89.6%) segments of arterial stenosis were sufficiently recognized on CS-TOF and PI-TOF, respectively. Seven stenosed segments were recorded as probable stenosis on PI-TOF (grade 1), of which 5 were in the intracranial internal carotid artery, and 2 in the proximal M2 segment of the middle cerebral artery. Two segments of the intracranial ICA with suspicious stenosis on PI-TOF (Fig 1B) were eventually considered normal after comparison with CS-TOF in the same patients (Fig 1D) by the senior neuroradiologist. These segments were excluded from further analysis. The other 5 segments were recorded as grade 2 (definite stenosis) on CS-TOF. CS-TOF resulted in more confidence in diagnosing intracranial arterial stenosis than PI-TOF ($P = .025$).

The intraclass correlation coefficient for luminal stenosis measurement was 0.980 and 0.979 for CS-TOF and PI-TOF, respectively. The mean luminal stenosis ratio measured on CS-TOF ($0.579 ± 0.305$) showed no significant difference from that of PI-TOF ($0.649 ± 0.305$) ($P = .923$).

**Table 2: Patient demographics ($N = 22$)**

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Mean ± SD or Number (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>11 (50.0%)</td>
</tr>
<tr>
<td>Age (yr)</td>
<td>61.8 ± 16.8</td>
</tr>
<tr>
<td>Stenosis location (R/L)</td>
<td>48</td>
</tr>
<tr>
<td>Intracranial internal carotid artery</td>
<td>6/5</td>
</tr>
<tr>
<td>Middle cerebral artery</td>
<td>6/13</td>
</tr>
<tr>
<td>Anterior cerebral artery</td>
<td>5/2</td>
</tr>
<tr>
<td>Posterior cerebral artery</td>
<td>3/4</td>
</tr>
<tr>
<td>Basilar artery</td>
<td>1</td>
</tr>
<tr>
<td>Intracranial vertebral artery</td>
<td>2/1</td>
</tr>
</tbody>
</table>

Note:—R indicates right; L, left; SD, standard deviation.

**FIG 1.** Source images and coronal view of MIP images in a 42-year-old patient. The speckled noise in the center can be seen on the source image of conventional PI-TOF (A), whereas some artifacts with a curved stripe pattern can be seen on the source image of CS-TOF (C). These artifacts are eliminated on the MIP images and have little effect on the visualization of the stenosis. An obvious stenosis located in the M1 segment of the left middle cerebral artery is sufficiently visualized on both PI-TOF (B) and CS-TOF (D) (arrowheads). The edge sharpness of vessels on CS-TOF (D) is higher than that on PI-TOF (B) (short arrows). The image quality of the right intracranial internal carotid artery (long arrow, B) is improved on CS-TOF (D).
difference from that measured on PI-TOF (58.9% ± 31.0%) (P = .241). Table 3 shows detailed comparison results between CS-TOF and PI-TOF for the evaluation of intracranial arterial stenosis. Representative cases are shown in Figs 1 and 2. The degree of each luminal stenosis measured on CS-TOF and PI-TOF is shown in Fig 3.

**DISCUSSION**

In the present study, we evaluated the diagnostic accuracy of CS-TOF for intracranial arterial stenosis by comparison with conventional PI-TOF. Our results showed that CS-TOF resulted in more clear visualization of arteries and more confidence in diagnosing intracranial arterial stenosis than PI-TOF, even in a reduced acquisition time of 2 minutes and 31 seconds.

3D TOF MRA is an effective and widely used tool to noninvasively evaluate and follow-up patients with cerebrovascular disease. However, to maintain a reasonable spatial resolution and provide detailed depictions of the vessels, the acquisition time is relatively long for conventional PI-TOF. Reduction of the scan time is clinically significant because motion artifacts would be decreased in a shortened scan time, which would particularly benefit patients with acute ischemic stroke.

CS is a novel technique that uses random undersampling. TOF-MRA is well-suited to CS because high-signal vessels are sparse in space. Previously, sparse TOF has been investigated in patients with cerebral aneurysms by Fushimi et al. They reported that cerebral aneurysms were visible with equivalent clarity in sparse TOF and PI-TOF. The measured neck height and width of aneurysms were not significantly different when either method was used. Most recently, the same group studied the reliability of CS-TOF in the evaluation of Moyamoya disease. The group found that CS-TOF could improve the visualization of small collaterals in the same amount of time or produce the same results in a shorter acquisition time compared with PI-TOF. In our study, variable degrees of intracranial arterial stenosis (from mild stenosis to occlusion) were included. An acceleration factor of 10.3 for CS-TOF significantly shortened the acquisition time, while providing comparable image quality and more clear visualization of arteries than PI-TOF. The possibly stenosed segments recorded on PI-TOF included the intracranial
ICA and proximal M2 segment of the MCA, of which 2 segments were considered artifactual narrowing caused by signal loss due to turbulent and slow flow according to previous studies.\textsuperscript{27} The diagnostically confidence of stenosis in these segments was improved on CS-TOF. The luminal stenosis ratio based on PI-TOF and CS-TOF was not significantly different; this finding suggests the absence of either over- or underestimation of the luminal stenosis grade due to irregular undersampling with CS-TOF.

We found that CS-TOF provided higher edge sharpness of the intracranial arteries than PI-TOF and enhanced the contrast of the vessels. In previous studies, the apparent contrast-to-background deviation of the sparse TOF images was reported to be stable at various acceleration factors and was significantly higher than with PI-TOF.\textsuperscript{11,26} Our results were consistent with these findings. The high apparent contrast-to-background deviations in CS-TOF may contribute to the better visualization of stenosis in segments of the intracranial ICA and proximal M2 segment of the MCA, as well as small branches, which are more prone to turbulent or slow flow.

Curved stripe pattern artifacts associated with sparse undersampling could be observed on all the source images of CS-TOF. Such artifacts are considered ghost artifacts originating from the skull boundaries.\textsuperscript{13} However, these artifacts are hardly noticeable on the MIP images of CS-TOF because the MIP connects the high-intensity dots of the blood vessels in 3D. Each point in the MIP represents the highest intensity experienced in that location on any partition within the imaging volume. Besides, the artifacts introduce an additional modulation of the already inhomogeneous background signal. The additional modulation is spatially slowly varying and of small amplitude compared with the normal signal variations in the background. Therefore, the performance of vessel segmentation algorithms performing well on the artifact-free image should not be degraded significantly by the artifacts. Moreover, the speckled noise on conventional PI-TOF in the central parts of the source images was reduced in CS-TOF. The total reconstruction time after scanning was approximately 1 minute 40 seconds for CS-TOF because we used a graphic processing unit in the current study, shortened to clinically acceptable times, compared with 10 minutes in a previous report.\textsuperscript{13}

There are several limitations to our study. First, DSA is considered the criterion standard. However, only 4 patients had DSA within 4 months before or after the MRA examination. DSA is invasive, and when MRA can depict reliable arterial stenosis, DSA is not considered essential in clinical practice. Previous studies have verified a good correlation between TOF-MRA and DSA for detecting arterial stenosis, though the stenosis ratio may be overestimated on TOF-MRA.\textsuperscript{28} Patients with mild (<50%) or moderate intracranial arterial stenosis (50%–69%) account for a large proportion of all those with cerebrovascular diseases. However, such patients usually do not undergo DSA examinations. Because TOF-MRA is a good noninvasive screening tool, we considered that including patients with different stenosis ratios varying from mild stenosis to occlusion in our study would be more clinically significant. The evaluation of all the images was performed by 3 experienced neuroradiologists to avoid any false-positive or -negative judgment. In the case of stenosis seen only on 1 of the TOF sequences, the 3 neuroradiologists would confer to obtain a consensus. However, such a dilemma was very rare in our study (only 2 segments). Second, the in-plane resolution between PI-TOF and CS-TOF was identical (0.4 × 0.4 mm), while the slice thickness was not. Keeping the same voxel size (0.4 × 0.4 × 0.4 mm) as CS-TOF would lead to a scan time of around 11 minutes for PI-TOF. We therefore increased the slice thickness while keeping the slab thickness constant in the PI-TOF protocols to achieve an acceptable acquisition time of around 5 minutes. Third, the number of iterations was fixed at 10 in the current study, consistent with a previous report by Yamamoto et al.\textsuperscript{24} Fushimi et al\textsuperscript{11} reported that CS-TOF with 10 iterations could provide adequate image quality for the clinical diagnosis of cerebral aneurysms. Vessel sharpness and the visibility of small vessels can be improved with an increasing number of iterations,\textsuperscript{14} resulting in a largely extended reconstruction time. The reconstruction time would be 4 minutes 50 seconds if the iterations were increased to 20 in our study, which may affect the clinical examination workflow.

CONCLUSIONS

The image quality of highly accelerated CS-TOF is comparable with that of PI-TOF, while CS-TOF has the obvious benefit of reducing imaging time and boosts confidence in diagnosing intracranial arterial stenosis.

ACKNOWLEDGMENTS

We sincerely thank Long-quan Dai from Nanjing University of Sciences and Technology for his kind and warm help during the analysis of edge sharpness of the intracranial arteries. We also thank Min-lin Zhou from the National Clinical Research Center of Kidney Diseases, Jinling Hospital, for reviewing all the statistics.


REFERENCES


Cerebrovascular Reactivity during Prolonged Breath-Hold in Experienced Freedivers

ABSTRACT

BACKGROUND AND PURPOSE: Experienced freedivers can endure prolonged breath-holds despite severe hypoxemia and are therefore ideal subjects to study apnea-induced cerebrovascular reactivity. This multiparametric study investigated CBF, the spatial coefficient of variation as a correlate of arterial transit time and brain metabolism, dynamics during prolonged apnea.

MATERIALS AND METHODS: Fifteen male freedivers (age range, 20–64 years; cumulative previous prolonged breath-holds >2 minutes and 30 seconds: 4–79,200) underwent repetitive 3T pseudocontinuous arterial spin-labeling and 31P-/1H-MR spectroscopy before, during, and after a 5-minute breath-hold (split into early and late phases) and gave temporally matching venous blood gas samples. Correlation of temporal and regional cerebrovascular reactivity to blood gases and cumulative previous breath-holds of >2 minutes and 30 seconds in a lifetime was assessed.

RESULTS: The spatial coefficient of variation of CBF (by arterial spin-labeling) decreased during the early breath-hold phase (−30.0%, \( P = .002 \)), whereas CBF remained almost stable during this phase and increased in the late phase ( + 51.8%, \( P = .001 \)). CBF differed between the anterior and the posterior circulation during all phases (eg, during late breath-hold: MCA, 57.3 ± 14.2 versus posterior cerebral artery, 42.7 ± 10.8 mL/100 g/min; \( P = .001 \)). There was an association between breath-hold experience and lower CBF (1000 previous breath-holds reduced WM CBF by 0.6 mL/100 g/min; 95% CI, 0.15–1.1 mL/100 g/min; \( P = .01 \)). While breath-hold caused peripheral lactate rise ( + 18.5%) and hypoxemia (oxygenduration, − 24.0%), cerebral lactate and adenosine diphosphate remained within physiologic ranges despite early signs of oxidative stress (− 6.4% phosphocreatine / (adenosine triphosphate + adenosine diphosphate); \( P = .02 \)).

CONCLUSIONS: This study revealed that the cerebral energy metabolism of trained freedivers withstands severe hypoxic hypercarbia in prolonged breath-hold due to a complex cerebrovascular hemodynamic response.

ABBREVIATIONS: ATP = adenosine triphosphate; ASL = arterial spin-labeling; ASL-sCoV = spatial coefficient of variation of CBF [by ASL]; ATT = arterial transit time; CVR = cerebrovascular reactivity; HR = heart rate; P = inorganic phosphates; PCr = phosphocreatine; pO2 = partial pressure of oxygen; SpO2 = oxygen saturation

Freedivers acquire the ability to voluntarily breath-hold for several minutes: One breath-hold can be extended up to the world record of 11 minutes and 30 seconds.\(^1\) However, animal studies have suggested that threshold times for hypoxia-induced neuronal cell death can be as low as 2 minutes and 30 seconds.\(^2\) The regulation and sufficiency of cerebrovascular reactivity (CVR) to prevent ischemic brain damage during prolonged breath-hold are unclear. Such knowledge may hold clues to explain patterns of brain damage in adverse diving outcomes and may provide suggestions for targeted therapies. It can also be useful to deduce CVR in medical conditions of repetitive hypoxemia such as obstructive sleep apnea syndrome.\(^3\)

Previous MR imaging breath-hold experiments involved hy-
The sufficiency of CVR can be studied indirectly by 31P-MR spectrometry—that is, to prevent the depletion of adenosine triphosphate (ATP) and the maintenance of an aerobic brain metabolism under hypoxemia. Arterial spin-labeling (ASL) perfusion MR imaging allows an in vivo assessment of absolute CBF both regionally and on a vessel-selective level. ASL studies already successfully identified chronic CVR alterations due to obstructive sleep apnea as well as acute alterations during very short breath-holds. A corollary ASL measure is the spatial coefficient of variation (ASL-sCoV) of the CBF image itself, which was recently identified as a correlate on CVR. The putative effect of experience from previous prolonged breath-holds on CBF must be mentioned in this context and has not been explored.

Arterial spin-labeling (ASL) perfusion MR imaging allows an in vivo assessment of absolute CBF both regionally and on a vessel-selective level. ASL studies already successfully identified chronic CVR alterations due to obstructive sleep apnea as well as acute alterations during very short breath-holds. A corollary ASL measure is the spatial coefficient of variation (ASL-sCoV) of the CBF image itself, which was recently identified as a correlate on CVR. The putative effect of experience from previous prolonged breath-holds on CBF must be mentioned in this context and has not been explored.

The goal of the present study in experienced freedivers was to explore the following: 1) acute CVR during a prolonged breath-hold challenge with ASL, and 2) the sufficiency of CVR to maintain aerobic cerebral energy metabolism as measured by 31P- and 1H-MR spectroscopy. In addition, this study investigated whether pre-existing experience with prolonged breath-holds has an influence on the CVR during the breath-hold experiment.

**Materials and Methods**

**Participants**

Fifteen experienced male freedivers (median age, 36.0 years; 95% CI, 32.0–50.0 years) participated (Table). Inclusion criteria were adult age and an ability to breath-hold for >4 minutes without stress symptoms (ie, tachycardia, oxygen saturation [SpO2] below 60 %, delirium). Exclusion criteria were pre-existing cardiac or neurologic disorders and current smoking. The participant’s age and cumulative number of breath-holds longer than 2 minutes and 30 seconds during the volunteer’s lifetime (estimated by interview report) were registered as potential influential factors.

**Study Design**

Preparatory evaluations involved questionnaires regarding claustrophobia, noise tolerance, or issues with restraints. All participants were instructed to refrain from meals and caffeine for at least 2 hours before MR imaging to reduce perfusion confounders. Preparation (eg, meditation, test placement on the MR imaging table) was allowed. MR imaging–electrocardiography and finger oximetry guaranteed continuous monitoring of SpO2 and heart rate (HR). Arterial CO2 measurements were not permitted by the hospital ethics committee, but venous blood gas was repetitively analyzed. Continuity of breath-hold and consciousness were visually monitored (L.E.) to guarantee safety and correct measurements.

### Table: Anthropometric data of freediver volunteers

<table>
<thead>
<tr>
<th>Participant</th>
<th>Longest Breath-Hold until MRI (min)</th>
<th>Start of Sport (yr before MRI)</th>
<th>Estimated Breath-Holds &gt;2 min 30 s per Session</th>
<th>Sessions per Month</th>
<th>Estimated Cumulative Lifetime Breath-Holds &gt;2 min 30 s</th>
<th>Age (yr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10.6</td>
<td>11.3</td>
<td>30</td>
<td>0.4</td>
<td>1632</td>
<td>31</td>
</tr>
<tr>
<td>2</td>
<td>5.5</td>
<td>1.5</td>
<td>7</td>
<td>8</td>
<td>1008</td>
<td>36</td>
</tr>
<tr>
<td>3</td>
<td>5.4</td>
<td>2.0</td>
<td>1</td>
<td>1</td>
<td>24</td>
<td>20</td>
</tr>
<tr>
<td>4</td>
<td>5.4</td>
<td>11.0</td>
<td>~0</td>
<td>8</td>
<td>4</td>
<td>33</td>
</tr>
<tr>
<td>5</td>
<td>5.0</td>
<td>2.0</td>
<td>0.125</td>
<td>8</td>
<td>24</td>
<td>40</td>
</tr>
<tr>
<td>6</td>
<td>5.5</td>
<td>10.0</td>
<td>~0</td>
<td>4</td>
<td>4</td>
<td>38</td>
</tr>
<tr>
<td>7</td>
<td>5.0</td>
<td>8.0</td>
<td>5</td>
<td>8</td>
<td>3840</td>
<td>61</td>
</tr>
<tr>
<td>8</td>
<td>5.5</td>
<td>1.5</td>
<td>0.5</td>
<td>4</td>
<td>36</td>
<td>34</td>
</tr>
<tr>
<td>9</td>
<td>5.0</td>
<td>7.0</td>
<td>6</td>
<td>16</td>
<td>8064</td>
<td>47</td>
</tr>
<tr>
<td>10</td>
<td>6.2</td>
<td>9.0</td>
<td>6</td>
<td>16</td>
<td>4536</td>
<td>26</td>
</tr>
<tr>
<td>11</td>
<td>5.3</td>
<td>27.5</td>
<td>15</td>
<td>16</td>
<td>79,200</td>
<td>64</td>
</tr>
<tr>
<td>12</td>
<td>6.5</td>
<td>10.2</td>
<td>12</td>
<td>28</td>
<td>11,712</td>
<td>51</td>
</tr>
<tr>
<td>13</td>
<td>5.5</td>
<td>11.3</td>
<td>10</td>
<td>16</td>
<td>21,760</td>
<td>35</td>
</tr>
<tr>
<td>14</td>
<td>5.5</td>
<td>1.6</td>
<td>18</td>
<td>10</td>
<td>3420</td>
<td>32</td>
</tr>
<tr>
<td>15</td>
<td>5.0</td>
<td>5.0</td>
<td>0.083</td>
<td>4</td>
<td>20</td>
<td>50</td>
</tr>
<tr>
<td>Median</td>
<td>5.5</td>
<td>8.0</td>
<td>6</td>
<td>8</td>
<td>2526</td>
<td>36.0</td>
</tr>
<tr>
<td>95% CI</td>
<td>5.3–5.5</td>
<td>2.0–11.0</td>
<td>0.14–12.0</td>
<td>4.0–9.96</td>
<td>30–6300</td>
<td>32–50</td>
</tr>
<tr>
<td>Minimum</td>
<td>5.0</td>
<td>1.5</td>
<td>0</td>
<td>0.41</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Maximum</td>
<td>10.6</td>
<td>27.5</td>
<td>30</td>
<td>28</td>
<td>79,200</td>
<td>64</td>
</tr>
</tbody>
</table>

*Participants were interviewed concerning the duration (start of sport) and intensity (freediving sessions per month) and the estimated frequency of breath-holding for >2 minutes and 30 seconds per session. The latter could be near zero in freedivers specializing in short high-frequency breath-holds. Some participants breath-held for 2 minutes and 30 seconds only as a qualification for this study; hence, their cumulative lifetime breath-holds of >2 minutes and 30 seconds before MRI were 4, while others very frequently underwent prolonged breath-hold.*
The MR imaging protocol consisted of a 3D T1-weighted acquisition followed by 5 ASL scans. Participants were instructed to breathe with normal frequency to obtain the baseline CBF (baseline phase). Hyperventilation to reduce blood CO₂ or increase O₂ concentrations was a forbidden confounder. The participants gave an acoustic signal when starting the 5-minute breath-hold, when 2 consecutive ASL scans were acquired (early and late breath-hold phases). After the second breath-hold scan, the participant was instructed to breathe again at normal frequency (approximately 16/min.). Without delay, 2 normal-breathing ASL scans (early and late recovery phases) were obtained. A subset of participants was available for ³¹P- and ¹H-MR spectroscopy (n = 11 and 8 participants, respectively) during separate sessions to determine relative brain metabolites and, again, venous blood gas. The study was approved by the University Hospital Bonn, Germany ethics committee. All participants provided written informed consent.

**MR Imaging Sequences**

All imaging was performed on a 3T Ingenia MR imaging scanner (Philips Healthcare, Best, the Netherlands). The T1-weighted MPRAGE sequence was 1 × 1 × 1 mm³. The 5 identical pseudo-continuous ASL sequences were acquired with a 3D gradient- and spin-echo readout (5 segments; acquisition voxel size, 3.75 × 3.75 × 6 mm³; FOV, 240 × 240 × 96 mm³; TE/TR, 8/4.28 s; labeling duration, 1.8 s; radiofrequency labeling pulse duration/interval, 0.7/1.41 ms; postlabeling delay, 2 s; 4 background suppression pulses; scan time, 2 minutes and 30 seconds per scan phase). Each sequence consisted of 2 M0 images and 2 control-label pairs. A labeling distance of 13 cm (middle slice of the ASL stack to labeling plane) was chosen.

**Image Processing**

Image processing was performed with ExploreASL (www.ExploreASL.com), a toolbox based on SPM12 routines (http://www.fil.ion.ucl.ac.uk/spm/software/spm12), which was initiated through the European Union–funded European Cooperation in Science and Technology (COST) Action ASL In Dementia (http://s434060124.online.de/aslindementiacms/), aiming at harmonizing ASL imaging for single- and multicenter ASL studies. Image-processing steps were the following: automated segmentation of 3D T1-weighted images using the Computational Anatomy Toolbox (CAT12 toolbox) rigid-body registration of CBF to the gray matter partial volume map and spatial normalization into a common space using the Diffeomorphic Anatomical Registration Through Exponentiated Lie Algebra Toolbox (DARTEL, part of SPM). M0 images were masked, iteratively smoothed, and extrapolated outside the mask, and CBF was quantified using a single-compartment quantification model. The ROIs analyzed were the total (cortical) GM and the total white matter. The GM was subsegmented into vascular territories of the anterior, middle, and posterior cerebral arteries. The vascular territories were delineated in the common space on the Montreal Neurological Institute atlas according to Tatu et al.

The ASL-sCoV was defined as the SD of CBF divided by the mean CBF, within the total cortical GM (gray matter probability > 0.7): 

\[
\text{ASL-sCoV}_{ROI} = \frac{\sigma(\text{CBF}_{GM})}{\mu(\text{CBF}_{GM})} \times 100\%.
\]

Partial volume effects within the GM and WM ROIs were corrected using the method of Asllani et al.

**MR Spectroscopy**

Proton-decoupled ³¹P spectra were acquired with a dual-tunable ³¹P/¹H birdcage transmit/receive head coil (1024 data points; 3-kHz sampling; TR, 4 seconds; 4 signal averages) as a continuous time-series with 16-second duration for each spectrum, comprising a total of 40 scans before, during, and directly after breath-hold in a 25-mm-thick axial slice on the level of the basal ganglia. ³¹P signals were processed by the AMARES algorithm of the JAVA-MRUI software (http://www.jmrui.eu/features/quantitation/) quantifying 15 peaks in each spectrum arising from phosphocreatine (PCr), inorganic phosphates (Pᵢ), ATP, and phosphomono- and diesters. The cerebral pH was determined from the frequency separation between Pᵢ and PCr. Five consecutive spectra sets within the ³¹P time-series were averaged for 2 minutes and 30 seconds, which represent the metabolic baseline, first and second halves of breath-hold, as well as early and late metabolic recovery phases. This step was also to improve the signal-to-noise ratio. Thus, the time course of pH and of the ratios Pᵢ/PCr, Pᵢ/β-ATP, PCr/β-ATP, as well as the ratios of all ³¹P metabolites relative to (ATP + adenosine diphosphate), calculated from the mean of the γ- and α-ATP peaks, were obtained in the same 2 minutes and 30 seconds intervals as the ASL sequences. γ- and α-ATP peaks include the signals from nucleotide diphosphates and thus are expected as near-constant.

In a consecutive single-volume ¹H-MR spectroscopy breath-hold experiment, ratios of the ¹H-MR signals of N-acetylaspartate, total creatine, choline, and lactate were determined from an 8-ml volume in the left basal ganglia (point-resolved spectroscopy sequence localized spectra: TR 2 s; TE 0.14 s; 128 signal averages). Recording periods were split to match the ASL-MR imaging and ³¹P-MR spectroscopy phases of the breath-hold experiment.

**Blood Gas Analysis**

We drew 10 mL of venous blood before breath-hold, after 2 minutes and 30 seconds of breath-hold, at first breath after breath-hold, and after 2 minutes and 30 seconds of recovery. Immediate analysis (RAPIDLab 1265; Siemens, Erlangen, Germany) involved partial pressure of CO₂ and oxygen (pO₂), glucose, and lactate levels. With the exception of partial pressure of CO₂, these parameters were shown to correlate well with arterial values.

**Statistics**

Overall CBF variability was analyzed by a mixed linear model with the participant as a random factor, while differences between single time points were analyzed by a paired t test, which considers CBF differences compared with baseline (Δ CBF/CBF₀) as the expression of CVR without division by a fixed rate of stimulus such as per unit CO₂. The same tests were applied for the partial volume–corrected ASL-sCoV. MR spectroscopy parameters were analyzed in a mixed linear model and with the Pearson R correlation. The GM CBF in the vascular territories was analyzed separa-
phase CBF was indeed lower than baseline CBF (mean, −22.4%; P = .01). Baseline-to-late recovery phase CBF differences were not significant (P = .55 for the total GM ROI). Return to baseline CBF or below occurred during the early recovery phase with a mild secondary CBF increase during the second 2 minutes and 30 seconds of recovery (P = .03 in total GM ROI).

The absolute CBF and CVR of the anterior circulation (anterior cerebral artery, MCA) were, at all times, higher than in the posterior circulation (P = .001; On-line Table). The mean difference between CBF of the MCA and the posterior cerebral artery increased steadily during breath-hold from 8.8 ± 6.6 mL/100 g/min at baseline to 14.6 ± 6.1 mL/100 g/min at late breath-hold (P = .001).

During all scan phases, between-participant CBF variability showed a narrower range in the GM than in the WM as well as a lower overall CVR (CVR variability in GM: 83.8%–152.8% from baseline; CVR in WM: 74.2%–231.5% increase from baseline CBF; P = .001; On-line Table).

Spatial Coefficient of Variation

The ASL-sCoV varied over the course of the experiment (P = .001). In most cases (n = 12/15), there was an ASL-sCoV decrease between baseline and the early breath-hold phase (mean decrease: −30.0% ± 21.6%; P = .002; Fig 2C). The ASL-sCoV remained reduced during the 5-minute breath-hold (P = .81 for the difference between early and late breath-hold) and rose again during recovery (P = .01). There was no difference between the baseline and recovery phases (P = .29).

The range of GM ASL-sCoV among participants was smaller during breath-hold than during normal breathing: mean, 2.8% ± 1.0%; range, 1.9%–5.3% versus mean, 1.8% ± 0.4%; range, 1.5%–3.0%, for baseline versus early breath-hold phases, respectively. This range equates to a 35.7% decrease of ASL-sCoV variability among participants (P = .02).

Physiologic Correlations

The dynamics of SpO2 and HR are presented in the On-line Table. SpO2 correlated with CVR, with an estimated increase of CBF of 0.82 mL/100 g/min with each 1% SpO2 drop (95% CI, 0.6–1.1 mL/100 g/min; P = .001; Fig 3A, -B). HR was not correlated with CVR (P = .36; Fig 3C, -D). Similarly, age was not identified as an influential cofactor on CVR (P = .32).

If one took the entire group of 15 freedivers into account, previous prolonged breath-hold events were not correlated with absolute CBF values (P = .56). However, there were 2 outlier participants (divers 11 and 13) with exceptionally extensive experience and comparatively high CBF. After their exclusion, a relationship between previous breath-hold experience and absolute CBF was found with 1000 previous prolonged breath-holds, reducing CBF in GM by 2.2 mL/100 g/min (95% CI, 0.7–3.7 mL/100 g/min; P = .01; Fig 4) for the remaining 13 cases. This finding was similar for CBF in WM: One thousand previous breath-holds reduced WM CBF by 0.6 mL/100 g/min; 95% CI, 0.15–1.1 mL/100 g/min (P = .01). Mean CVR, however, as defined by the Δ CBF/ CBF0, was not correlated with previous experience with prolonged breath-holds (P = .23). Similarly, ASL-sCoV was not cor-

RESULTS

Cerebrovascular Reactivity

All participants showed a significant CVR by an increase in CBF until 5 minutes of breath-hold with a subsequent decline at recovery (P = .001 for all vessel territories, GM, and WM; On-line Table and Figs 1 and 2A, -B). While the CBF increase from baseline was substantial after 5 minutes (late breath-hold scan phase for the total GM ROI: mean Δ CBF, 18.3 ± 14.4 mL/100 g/min [±51.8%]; P = .001), it was only subtle within the first 2.5 minutes of breath-hold (total GM ROI: mean Δ CBF, 6.3 ± 11.1 mL/100 g/min [±17.8%]; P = .04). In 4 volunteers, early breath-hold-
related with SpO₂, HR, age, or diver experience with prolonged breath-holds \( (P = .45, .53, .90, \text{ and .69, respectively).} \)

**MR Spectroscopy and Blood Analyses**

\(^{31}\)P-MR spectroscopy (Fig 5A) revealed minor fluctuations in pH (Fig 5B) and ATP metabolites within physiologic ranges during the course of breath-hold \( (P = .07 \text{ for pH; } P > .05 \text{ for dynamics in relative PCr; } P; \text{ ATP-α, -β, -γ; and phospho-} \text{monoo- and diester levels). There was a small-but-significant decrease of the PCr / (ATP + adenosine diphosphate) ratio between baseline and late-phase breath-hold in the matched-pairs analysis of the individuals \((−6.4\%; P = .02; \text{Fig 5C}).\) PCr and β-ATP differences relative to their baseline values \((ΔPCr \text{ and } Δβ-ATP)\) were correlated with the differences in pH from baseline \((R = 0.53; P < .001 \text{ for } ΔPCr \text{ and } R = 0.45; P = .003 \text{ for } Δβ-ATP, \text{ respectively; Fig 5D).}\)

\(^{1}\)H-MR spectroscopy never showed the CH₃ doublet of lactate at 1.34 ppm in any of the participants. No significant changes in the levels of N-acetylaspartate, total creatine, or choline occurred \((P > .05 \text{ for all).}\)

Venous blood samples showed a development of hypoxemia and hypercapnia during breath-hold (On-line Table, Fig 5E). A significant partial pressure of CO₂ increase was noted only at late breath-hold \((P = .02), \) while pO₂ had already dropped significantly after the early breath-hold phase \((P = .002).\) Both parameters returned to baseline after breath-hold, while venous lactate and glucose levels increased until the end of the experiment during the recovery phase \((P = .001 \text{ and } .01, \text{ respectively; Fig 5E).}\)

**DISCUSSION**

This study provides 3 key findings. First, despite individual variability, freedivers show a relatively consistent and vessel-territory-specific CVR during a breath-hold challenge, which is measurable with ASL. Second, CBF and the ATT correlate ASL-sCoV deliver independent aspects of the cerebrovascular response to breath-hold. Finally, this study identified indicators for an influence of earlier experience with prolonged breath-hold on absolute CBF during the breath-hold challenge but not on CVR itself. Physiologic responses can apparently withstand the extreme biochemical challenge induced by a prolonged breath-hold of 5 minutes, which can be detected by ASL and MR spectroscopy.

Multiple methods exist to assess CBF, including ASL, phase-contrast MR imaging, PET, and Doppler sonography. Maximum breath-hold studies are rare, and CBF evaluations were, until now, exclusively performed with Doppler sonography, which revealed a continuous elevation of flow velocity in the MCA of around 100%. \(^{19,20}\) The mean maximum CBF increase after 5 minutes of breath-hold observed in this ASL study was 51.8%, which is very close to values observed in ASL studies using hypercarbic gas inhalation or short breath-hold, but indeed in some cases lower than values measured with ASL in a recent maximum breath-hold study \((+107\%).^{21,22}\) Previous studies comparing CBF assessments with different methods similarly revealed substantial intermethod differences in absolute CBF but otherwise confirmed a high correlation between methods. PET-estimated absolute CBF was, for example, consistently lower than phase-contrast MR imaging, while Doppler and ASL differed substantially in relative CBF change in a drug-stimulation trial. \(^{23,24}\) These differences are not surprising and can be explained by different influential factors acting on the respective flow parameters. While ASL can measure absolute CBF, Doppler provides flow velocity in a local vessel segment as a surrogate parameter for CBF. CO₂ is a strong vasodilating agent in cerebral tissue causing an increase in CBF due to increased blood volume (CBV) based on the equation CBF = CBV / MTT. The blood flow velocity rises despite vessel dilation also due to reduced ATT. However, earlier Doppler studies revealed a strong neuromuscular response in proximal vessel segments of the anterior circulation, while more distal segments and the posterior circulation seemed less responsive. \(^{25,26}\)
may explain why Doppler-assessed breath-hold experiments identify a higher absolute and relative CVR, mostly in the M1 segments of the MCA, than most of those applying ASL, which captures perfusion in most distal and nonmuscular vascular segments.

A heterogeneous CVR between vessels was also observed in the present ASL study with lower CBF and CVR in the posterior cerebral artery territories. This finding can be explained by differences in labeling efficiency between vascular territories as well as the longer ATT in the posterior vascular territory. Microanatomic differences leading to a diverse autoregulation capacity between the anterior and posterior circulation as an underlying reason for a diverse CVR are, on the other hand, more controversially discussed even beyond the field of perfusion studies under extreme conditions. A recent Doppler-monitored breath-hold study supports our findings of a lower CVR in the posterior cerebral artery and suggests a different sympathetic activation between the 2 territories as an additional explanation beyond the labeling aspect, which needs to be considered in ASL. An awareness of a vessel-selective CVR is, however, crucial when interpreting ASL measurements in focal ischemic lesions after prolonged clinical conditions of apnea.

Another finding of this study is that CBF decreased during the first 2 minutes and 30 seconds of breath-hold in 4 of 15 participants. The counter-suggestive relatively higher CBF at baseline compared with later time points might be the effect of anticipation anxiety toward the upcoming breath-hold challenge. This well-known mental phenomenon among freedivers is currently not sufficiently addressed in sports physiologic research. Indeed, our own results may only suggest that some freedivers experienced an early sympathetic activation before the breath-hold challenge. Due to the limited temporal resolution of the ASL sequence, which delivers a mean CBF over each of the 2 minutes and 30 seconds phases, we cannot readily assess how long the reduction of CBF persists and when exactly an elevation of the CBF above took place during the early breath-hold phase in these 4 volunteers.

We further confirmed a difference between the cortical GM and the deep WM CBF, with the total GM CBF known to be at least 2 times higher than the WM CBF. Low WM signal is an obstacle for WM CBF assessment in ASL, despite the availability of background suppression. However, CVR differences between GM and WM tissue can be interesting because they may help to better understand morphologic findings in brain diseases such as obstructive sleep apnea syndrome. We noted a larger CVR variability and relative increase in WM than in GM, which could be explained by the later arrival of blood in the relatively more distal WM vessels. WM CBF is difficult to measure at baseline due to longer ATT. Direct ATT assessment was technically not possible as part of our experimental setting due to temporal restrictions and a resulting monopost-labeling delay ASL sequence. However, our ATT approximation based on ASL-sCoV confirmed a breath-hold-induced decrease in ATT, which might increase the SNR of WM CBF, which inflates the measured ΔCBF to a certain extent.

For the interpretation of ASL-sCoV, it is viable to consider the methodologic peculiarities of ASL. An ATT increase will cause the ASL signal to appear in larger vessels, resulting in vascular artifacts. The ASL signal can, at the same time, decrease in areas with higher baseline ATT as in, for example, the perfusion watershed. These 2 effects both increase ASL-sCoV. However, ASL-sCoV...
FIG 5. Energy metabolites and blood gases in brain and venous blood. A, Sample $^3$P-MR spectrum of 1 participant averaged over 2 minutes and 30 seconds of the late breath-hold phase (above), displayed together with AMARES-fitted spectral components (below). B, Brain pH as assessed by $^3$P-MR spectroscopy is near-constant during the entire breath-hold experiment. C, The PCR/(ATP + adenosine diphosphate) ratio measured by $^3$P-MR spectroscopy slightly decreases during breath-hold. D, PCR and β-ATP differences (Δ) from individual baseline values correlate with the pH differences from baseline, indicating that small tendencies toward acidosis and ATP depletion occur during breath-hold. E, Venous blood analyses reveal significant hypoxemia and hypercapnia development (left axis in millimeters of mercury) during breath-hold with fast recovery. Blood glucose (in milligrams/deciliter, left axis) and lactate (in mmol/L, right axis) rose during breath-hold and did not return to baseline. All values are expressed as differences from baseline.

has a theoretic lower limit attained when all labeled blood has arrived in the tissue and all vascular artifacts have already disappeared. Further ATT decrease beyond this limit will have only minimal effect on the ASL-sCoV. Reaching this lower limit of ASL-sCoV during the early breath-hold could explain the low ASL-sCoV variability in some participants. The ASL-sCoV decrease, however, appears to occur earlier than the CBF increase and levels off during the second breath-hold phase, while CBF further increases. An earlier response to breath-hold leading to decreasing ATT before the CBF increase in the late breath-hold phase can be the explanation, denoting that the effect is mainly vascular and perfusion changes as detected by ASL follow later. ASL-sCoV may be an earlier CVR marker of hypoxemia than absolute CBF.

CVR is predominantly triggered by changes in blood CO$_2$, while, for example, hypoxemia detected by peripheral chemoreceptors is considered to play an independent-but-inferior role in cerebral vasodilation. While Willie et al.$^{16}$ already reported that O$_2$ metabolism has a crucial influence on the breath-hold capability of freedivers, stressing the role of O$_2$ for breath-hold tolerance and CVR, Cross et al.$^{18}$ could not confirm an influence of O$_2$ on cerebral autoregulation in their prolonged breath-hold study. We identified falling O$_2$ as an influential factor of CVR in this study, which can usually not be observed in CBF studies applying hypercarbic-normoxic gas despite otherwise comparable CVR between breath-hold and hypercarbic-normoxic gas studies.$^{5,26}$ Due to the design of this study, which could not rely on arterial CO$_2$ measurements, it remains, however, impossible to discern the relative contribution of hypercarbia and hypoxia to CBF increase. CVR increased faster in the second half of the breath-hold experiment in correlation with the secondarily faster SpO$_2$ decline, which can be explained by pulmonary O$_2$ stores that allow normal hemodynamic conditions during the first minutes of breath-hold. Cerebral near-infrared measurements in elite freedivers showed that cerebral O$_2$ desaturation tends to occur within a mean of 175 ± 50 seconds, but not before, which supports our finding.$^{14}$

In this study, CBF itself was not strongly correlated to HR. This finding does not allow concluding that CBF in breath-hold is not modulated by cardiac causes. It is, however, beyond the scope of this study to evaluate cardiac cofactors to CVR such as heart stroke volume or the diving reflex in detail. The factor age (an indirect measure of the cardiac and vascular influence on CVR) was assessed but was not associated with any of the flow parameters, which is not surprising considering that most volunteers in this study were younger than 40 years of age. Also, a selection bias of outstandingly healthy and well-trained freediver volunteers must be considered a further contributing factor.

A physiologic adaptation to breath-hold was another hypothesis to be tested in this study. A diminished CVR during hypercarbia/hypoxemia was reported for patients with chronic obstructive pulmonary disease as well as sleep apnea.$^{15}$ In an attempt to investigate a similar association between experience with hypercarbic/hypoxic states and CVR in freedivers, we estimated the total amount of previous prolonged breath-holds of the participants before participation in the current breath-hold experiment (defined as cumulative breath-holds of ≥2 minutes and 30 seconds in a lifetime). Our findings do not unconditionally corroborate that breath-hold experience accounts for an adaptation effect on CBF. First, CVR did not differ between more breath-hold-experienced freedivers and their less experienced counterparts. Second, by far, the 2 most prolonged breath-hold-experienced freedivers, who additionally stated a high frequency of longer breath-holds per training session, showed relatively high CBF values. However, for the remaining cohort, an association between experience and lower absolute CBF in all phases of the experiment could be observed, which may indicate that repetitive previous hypercarbia and hypoxemia have an acute cerebrovas-
cicular effect during a breath-hold challenge. Due to the limited cohort size, which also included freedivers normally specialized in shorter breath-holds, this interesting and also clinically relevant observation will need to undergo further critical evaluation in the future, favorably in a more homogeneous group regarding age and freediving specialization.

In prolonged breath-hold, the efficacy of the cerebrovascular but also the cardiac response to maintain a stable O$_2$ supply to the brain despite decreasing availability is a crucial health aspect and can be assessed spectroscopically regarding energy metabolism. Direct noninvasive in vivo measurements of brain energy metabolism during prolonged breath-hold are extremely rare.$^{22,36}$ Cerebral lactate accumulation or acidosis was observed in none of our participants, suggesting a sufficient compensation of limited O$_2$ supply by recruitment of ATP stores and increasing CBF. However, we identified a substantial decrease in the PCr / (ATP + adenosine diphosphate) ratio during breath-hold. This can be interpreted as a compensatory PCr decrease to provide ATP by PCr hydrolysis as a consequence of declining O$_2$ availability and reduced aerobic ATP production capacity in prolonged breath-hold.

Rising peripheral venous lactate levels during the breath-hold challenge in contrast to stable cerebral lactate stresses the shift toward a preferred cerebral O$_2$ supply in breath-hold, including a CBF increase and, simultaneously, peripheral vasoconstriction. The rising glucose levels are likely a consequence of adrenaline-induced glucose mobilization and underline the exceptional metabolic and mental challenge of prolonged breath-hold.$^{27}$ The correlative venous pO$_2$ analyses documented pathologically low O$_2$ levels after 5 minutes of breath-hold (down to 60% SpO$_2$) and corroborated similarities of this breath-hold experiment with clinical settings of hypoxemia. On the other hand, untrained persons may encounter life-threatening consequences under these circumstances, while freedivers face hypoxemia under voluntary and trained conditions.

Maximum breath-hold without contact with water is a particular challenge for freedivers because regular training sessions mostly include water immersion, which causes an augmented diving reflex. Only very few participants fulfilled the inclusion criteria and were able to perform sufficient breath-hold in the noisy MR imaging environment. Multiple candidates declined due to the lack of silence, or they would not volunteer for additional MR spectroscopy breath-hold experiments, thus the limited cohort size. Furthermore, participants did not tolerate a CO$_2$ mask, which interfered with their meditative state. We therefore decided to analyze venous CO$_2$. However, while venous CO$_2$ can be used to confirm hypercapnia, it is, however, too variable to study the CO$_2$ influence on CVR due to known substantial deviations from arterial CO$_2$ in the brain.$^{38}$ For this reason, we can also only assume that all normoventilating volunteers started at normal CO$_2$ blood levels in the experiment. The unavailability of these otherwise valuable data limits the evaluation of the physiologic processes behind our observations.

CONCLUSIONS
This study revealed that experienced freedivers develop a CVR, which is sufficient to maintain a physiologic cerebral energy metabolism even during a prolonged breath-hold period of 5 minutes and severely diminishing blood O$_2$. Furthermore, ASL parameters, which are determined by blood flow and vessel diameter alterations alike, serve as excellent candidate MR imaging parameters to reveal this response, while $^{31}$P-MR spectroscopy revealed its utility to dynamically study acute changes in cerebral energy metabolism. ASL may also provide evidence for long-term adaptation of cerebral vasculature following repetitive hypoxia-hypercapnia. Imaging and metabolic findings of the present freediver study can be used to better understand CVR during hypoxia-hypercapnia in critical care and sleep apnea conditions.

ACKNOWLEDGMENTS
The authors acknowledge Dipl.-Phys. Jürgen Gieseke for his assistance with the ASL sequence implementation, and we thank all freedivers for their outstanding effort to make this study possible.


REFERENCES
24. Vestergaard MB, Larsson HB. Regulating cerebral blood flow and vasoreactivity to CO2 measured by arterial spin labeling after 6 days at 4350m. Neuroimage 2013;72:272–79 CrossRef Medline
Reasons for Reperfusion Failures in Stent-Retriever-Based Thrombectomy: Registry Analysis and Proposal of a Classification System


ABSTRACT

BACKGROUND AND PURPOSE: In 5%–10% of patients with acute ischemic stroke with an intention to treat with mechanical thrombectomy, no reperfusion can be achieved (Thrombolysis in Cerebral Infarction score = 0/1). Purpose of this analysis was a systematic assessment of underlying reasons for reperfusion failures.

MATERIALS AND METHODS: An intention-to-treat single-center cohort (n = 592) was re-evaluated for all patients in whom no reperfusion could be achieved (n = 63). Baseline characteristics of patients were compared between patients with and without reperfusion failures. After qualitative review of all cases with reperfusion failures, a classification system was proposed and relative frequencies were reported. In a second step, occurrence of delayed recanalization at 24 hours after reperfusion failure and dependency on IV-tPA were evaluated.

RESULTS: In 63/592 patients with an intention to perform stent-retriever thrombectomy, no reperfusion was achieved (TICI 0/1, 10.6%, 95% CI, 8.2%–13.1%). Older patients (adjusted OR per yr = 1.03; 95% CI, 1.01–1.05) and patients with M2 occlusion (adjusted OR = 3.36; 95% CI, 1.82–6.21) were at higher risk for reperfusion failure. In most cases, no reperfusion was a consequence of technical difficulties (56/63, 88.9%). In one-third of these cases, reperfusion failures were due to the inability to reach the target occlusion (20/63, 31.7%), while “stent-retriever failure” occurred in 39.7% (25/63) of patients. Delayed recanalization was very rare (18.2%), without dependence on IV-tPA pretreatment status.

CONCLUSIONS: Reasons for reperfusion failure in stent-retriever thrombectomy are heterogeneous. The failure to establish intracranial or cervical access is almost as common as stent-retriever failure after establishing intracranial access. Systematic reporting standards of reasons may help to further estimate relative frequencies and thereby guide priorities for technical development and scientific effort.

The technical success of endovascular stroke therapy is one of the most important modifiable predictors of therapy benefit in patients presenting with large-vessel-occlusion acute ischemic stroke.1,2 The American Stroke Association/American Heart Association guidelines continue to support stent retrievers as the dominant technical platform for thrombectomy; however, this may change with the final publication of A Comparison of Direct Aspiration Versus Stent Retriever as a First Approach (COMPASS) trial.3-5 Recently published large-cohort registries of patients treated with stent retrievers have shown that contemporary endovascular interventions in acute ischemic stroke are angiographically successful in up to 80%–90% of cases (Thrombolysis in Cerebral Infarction 2b/3).6,7 While unsuccessful reperfusion (≤TICI 2a) often results from incomplete retrieval due to distal embolization and/or clot fragmentation (TICI 2a), in some patients, no reperfusion (TICI 0/1) can be achieved (“reperfusion failure”). Reasons for such reperfusion failures may range from difficulty establishing cervical or intracranial access to the inability to dislocate and retrieve the clot despite having reached the target location and having established intracranial access.8,9 Further conceivable explanations for failures to reestablish flow are underlying nonembolic vessel diseases (eg, vasculitis, intracranial atherosclerosis)10,11 or thrombi of nonthrombotic origin and extraordinary composition, such as calcified or neoplastic thrombi.12,13
The aim of this analysis was to provide estimates of the relative frequencies of underlying causes of reperfusion failure in patients with acute ischemic stroke who underwent angiography with an intention to perform stent-retriever-based thrombectomy.

MATERIALS AND METHODS
The prospective Bernese Stroke Registry was accessed to find all directly admitted patients with acute ischemic stroke with an intention to perform stent-retriever thrombectomy from January 2012 to July 2017 (n = 592). For the classification of reasons for reperfusion failure, all patients with postinterventional final Thrombolysis in Cerebral Infarction grades 0 and 1 were included.14 Most important, patients in whom no stent retriever was deployed (eg, because of access difficulties or vessel elongation) were also analyzed. The Bernese Stroke Registry was approved by the local ethics committee (Kantonal Ethics Committee Bern, amendment access number: 231/2014).

Classification of Reasons for Reperfusion Failures
After qualitative review of all cases with reperfusion failure, the underlying reasons for TICI 0/1 were classified into the following categories by a consensus of a neuroradiologist in training (J.K., 3 years of experience) and an interventional neuroradiologist (P. Mordasini, 15 years of experience).

Technical Reasons: Target Occlusion Not Reached (Category I)
- IA: Intracranial target occlusion was not reached due to marked cervical vessel tortuosity including twisted, looped, or kinked vessels.15 The proximal cervical vessels were successfully catheterized.
- IB: Target occlusion was not reached owing to failed catheterization of proximal supra-aortic vessels. The proximal cervical vessels were not successfully catheterized owing to difficult aortic arch anatomy.16
- IC: Target occlusion was not reached owing to the inability to pass a cervical ICA occlusion (eg, unpassable tandem lesion).

Technical Reasons: Target Occlusion Reached (Category II)
- IIA: Target occlusion was reached, but the operator was unable to pass the thrombus with the microwire/microcatheter. In these cases, no stent retriever is deployed.
- IIB: Target occlusion was reached, the stent retriever was deployed, but no reperfusion occurred after multiple retrievals (no clot retrieval or dislocation), thus, stent-retriever failure.
- IIC: Initial reperfusion was achieved, followed by spontaneous or iatrogenic reclosure (eg, intracranial stenosis, intracranial dissection, or perforation with subsequent vessel sacrifice).

Nontechnical Reasons: Other (Category III)
- Presumed futility.
- Adverse non-neurologic event with the need to stop mechanical thrombectomy.
- Signs of contrast extravasation without perforation (early hemorrhagic transformation).

For all cases, categorization of reasons and the evaluation regarding the primary intention of treating with a stent-retriever-based thrombectomy were based on the following information:
- Review of the initial radiologic report with respect to a written decision and interdisciplinary consensus for stent-retriever-based thrombectomy.
- Review of materials used along with an angiographic report concerning catheter changes and a description of interventional difficulties
- Review of all preinterventional and angiographic images.

Patient Characteristics
Baseline parameters, clinical outcomes, and information on recanalization status at 24 ± 6-hour follow-up are provided. Follow-up recanalization status was graded using the 4-step arterial occlusive lesion score14 on postinterventional intracranial vessel imaging, if available (44/63, 13 missing because only noncontrast CT was performed, 6 missing due to early death).

Endovascular Therapy
Endovascular therapy was performed immediately after CT or MR imaging under the following conditions: 1) The diagnosis of ischemic stroke was established; 2) the NIHSS score on admission assessed by a neurologist was ≥4 points, isolated aphasia or hemianopia or severe paresis of 1 hand was present, or neurologic deficits recurred; 3) CT or MR angiography showed occlusion of a large intracranial artery; 4) hemorrhage was excluded; 5) neurologic deficits correlated with the vessel occlusion; and 6) no individual clinical or premorbid conditions or laboratory findings were contraindications. When the criteria for endovascular therapy were fulfilled, digital subtraction angiography was performed via a transfemoral approach using a biplane, high-resolution angiography system (Axiom Artis zee; Siemens, Erlangen, Germany) using iopamidol (Iopamiro 300; Bracco, Milan, Italy) as a contrast agent. The choice of access and retrieval technique was left to the discretion of the neurointerventionalist, taking access anatomy and occlusion pattern into account. For anterior circulation strokes, a first-line recanalization technique consists of placing an 8F or 9F balloon-guiding catheter over a long exchange wire as high into the cervical ICA as possible followed by stent-retriever thrombectomy under proximal balloon occlusion and manual aspiration. In cases in which placement of a balloon-guiding catheter was deemed suboptimal (eg, low position in the cervical ICA or distal common carotid artery or highly tortuous cervical vessels) or stent-retriever thrombectomy through the balloon-guiding catheter alone was unsuccessful, a 5F or 6F intermediate catheter for concomitant distal aspiration during stent-retriever thrombectomy was used as a second-line technique. Stand-alone aspiration was only used as a third-line technique after failure of stent-retriever thrombectomy. During the study period, several different stent retriever models were used, the mainstay consisting of the Solitaire device (Covidien, Irvine, California). The decision of when to abandon the procedure and whether to administer intra-arterial thrombolysis with urokinase was left to the discretion of the neurointerventionalist.
Statistical Analysis
Categoric group comparisons were performed applying the Fisher exact test. Normally distributed data were presented as mean ± SD, while non-normally distributed data are shown as median (interquartile range). Comparison of continuous or ordinally scaled variables was performed using the Mann-Whitney U test or the Welch t test for independent samples (for non-normally and normally distributed variables, respectively). Estimated 95% confidence intervals of prevalences were calculated using the normal approximation interval (Wald interval). Variables with univariate significant differences between groups were entered into a multivariate logistic regression model. Results from multivariate logistic regression analysis were presented as adjusted ORs and respective 95% confidence intervals. Predicted probabilities were analyzed with receiver operator characteristic analysis with calculation of the area under the curve to evaluate the discriminative power of the model.

RESULTS
Study Cohort
Five hundred ninety-two patients were included (mean age, 72.2 ± 14.3 years; 47.3% female). Of these patients, 93.1% (n = 551) were treated for anterior circulation occlusions. The median symptom-onset to diagnosis interval (witnessed or last seen well) was 126 minutes (interquartile range, 90–222 minutes), and patients presented with severe neurologic deficits (median admission NIHSS, 15; interquartile range, 9–20). Other baseline characteristics are shown in the On-line Table. In 63 of 592 patients with an intention to treat with stent-retriever thrombectomy, no reperfusion was achieved (TICI 0/1, 10.6%; 95% CI, 8.2%–13.1%). Patients in whom no reperfusion was achieved were older (mean age, 76.9 versus 71.6 years; P = .004), were treated later (median symptom-onset to diagnosis 155 versus 121 minutes, P = .007), and had more M2 and posterior circulation occlusions (P for overall difference < .001). Other baseline characteristics did not reveal significant differences (On-line Table). When we entered significant variables derived from univariate comparison into a multivariable logistic regression model, only M2 occlusions and increased age were associated with reperfusion failure (adjusted OR = 3.36; 95% CI, 1.82–6.21; and 1.03; 95% CI, 1.01–1.05, respectively, with the Nagelkerke R² = 0.074). The discriminative power of the multivariate logistic regression model derived from the included variables age, time to diagnosis, and M2 occlusions was weak (area under the curve = 0.665; 95% CI, 0.585–0.746).

Distribution of Reasons for TICI 0/1
In most cases, no reperfusion was a consequence of technical difficulties (categories I and II: 56/63, 88.9%; 95% CI, 81.1%–96.6%, Figure). One-third of reperfusion failures were due to the inability to reach the target occlusion (category I: 20/63, 31.7%; 95% CI, 20.3%–43.2%). Reasons for not reaching the target occlusion were subcategorized into cervical vessel tortuosity (the proximal cervical vessel was catheterized, 10/20), failed catheterization of proximal supra-aortic vessels because of difficult acrotic arch anatomy (8/20), or, rarely, the inability to pass a proximal ICA occlusion/stenosis in patients presenting with extracranial-intracranial tandem lesions (2/20). If the target occlusion was reached (category II: 36/63, 57.1%; 95% CI, 44.9%–69.4%), reperfusion failure was due to the inability to pass the intracranial occlusion with the microwire/microcatheter in 7/36 cases. When stent retrievers were deployed (29/36), the most common reason for failure was the inability to retrieve/dislocate the thrombus after multiple attempts (stent-retriever failure, 25/36; median attempts, 3), while in the remaining cases, iatrogenic perforation and subsequent coil herniation were causative (4/36). Nontechnical reasons for TICI 0/1 (category III: 7/63, 11.1%; 95% CI, 3.4%–18.9%) were an active consensus decision to stop treatment after diagnostic angiography or an initial thrombectomy attempt (presumed futile, 5/7), a non-neurologic event with the need to stop endovascular therapy (1/7), and signs of contrast extravasation without perforation, interpreted as early hemorrhagic transformation (1/7).

Rescue and Patency at 24-Hour Follow-Up
Follow-up imaging at 24 hours was available for 44/63 (69.8%) patients without initial reperfusion. Of those patients, 8 (18.2%) revealed substantial vessel recanalization at follow-up (arterial occlusive lesion score, 2/3), while all other patients (36/44, 81.8%) showed evidence of persistent occlusion without or with minimal
recanalization (arterial occlusive lesion score, 0/1). Recanalization at follow-up tended to be observed less often if the reason for no reperfusion was a category II failure (11.5% versus 33.3%, \( P = .176 \)). If intra-arterial thrombolysis was applied as a rescue, a non-significant trend toward higher rates of recanalization at follow-up was observed (33.3% versus 11.5%, \( P = .125 \)). No difference in 24-hour recanalization in patients with reperfusion failure was observed when comparing those treated with IV-tPA with those treated without IV-tPA (21.4% versus 16.7%, \( P = .501 \)).

**DISCUSSION**

The presented study suggests that reasons for reperfusion failure in patients with an intention to perform stent-retriever-based thrombectomy are not homogeneous and underline the notion that reperfusion failure in contemporary stroke thrombectomy is a multifactorial problem. The failure to establish intracranial or cervical access is approximately as common as the inability to retrieve the clot once adequate intracranial positioning and access to the target occlusion are established (true stent-retriever failure). The presented data highlight the need to distinguish among potential factors associated with reperfusion failure and advocate for technical development and scientific effort equally focusing on stent-retriever efficacy as well as tools and alternative access routes to improve cervical and intracranial access.

**Access Failures and Alternative Access Routes**

The two main domains of difficulty are the inability to reach the target occlusion and the failure to dislocate or retrieve the clot. A recent study has suggested that carotid tortuosity relates to technical failure; however, another study did not find a significant association between carotid elongation and technical success or procedure length. One reason for the discrepant results of these studies may be because vessel elongation predominantly causes category I reperfusion failures, while intracranial geometry and thrombus properties may primarily relate to stent-retriever failures (category IIB). Distribution of these reasons in such cohorts may thus severely affect the sensitivity and respective power of the aforementioned analyses. Although no femoral access failure has been observed in the presented consecutive cohort, brachial or carotid access may serve as a reliable rescue approach in cases of femoral access failure or other category I failures (eg, in the presence of vessel tortuosity or the inability to adequately catheterize the proximal supra-aortic vessels). Until the date of this analysis, we had not used alternative access routes in a standardized fashion if the femoral artery was successfully cannulated but a category I failure had occurred. One consequence of this analysis is the institutional introduction of a more widely applied and standardized use of alternative access routes. We now consider these alternative routes not only when femoral access cannot be achieved (failed femoral artery cannulation) but also in all cases of category I failures.

**Retrieval Failures**

Several factors influencing stent-retriever failure have been proposed. So far, there is some preliminary evidence that MCA geometric anatomy, along with thrombus composition and shape, may predict the inability to dislocate or retrieve the clot. Other factors may include underlying vessel disease such as atherosclerosis, dissection, or vasculitis. According to current observational studies, thrombus length has not been shown to effect retrieval efficacy. Besides clot and occlusion site characteristics, stent-retriever dwelling time and respective device-clot interaction may serve as additional contributing factors. The present analysis included only patients with an intention to treat with stent-retriever-based thrombectomy. Recent randomized-controlled trial results have suggested that similar technical and clinical success rates could be achieved with aspiration techniques using large-bore catheters (COMPASS, International Stroke Conference 2018). A different mechanical retrieval approach is likely to increase the probability of achieving reperfusion, because some occlusions, which are refractory to one technique may respond favorably to an other approach. Currently, there is emerging evidence that the relative effectiveness of one approach over the other may depend on the localization of the occlusion (eg, anterior-versus-posterior circulation) or the shape of the proximal occlusion site.

**Delayed Recanalization and Rescue Options**

While reperfusion rates in intracranial vessel occlusions after IV-tPA may be as high as 60% after 7 hours, only a few patients had reperfusion at 24 hours after endovascular reperfusion failure, particularly after stent-retriever failure (~10%). These results provide preliminary evidence that clots not responding to mechanical treatment also show low response rates to intravenous thrombolysis and spontaneous clot lysis, corroborating findings that they may represent clots of distinct composition (eg, prominent calcification). In the presented cohort, rescue administration of intra-arterial thrombolytics showed a trend toward higher rates of recanalization at follow-up, without reaching statistical significance. Recently, low-dose tirofiban has been associated with increased reperfusion rates in patients treated with mechanical thrombectomy. However, the results are mainly derived from Asian cohorts, which have higher rates of underlying intracranial atherosclerosis and higher rates of early re-occlusions, thus severely limiting its transferability to other cohorts. Besides medical rescue approaches, stent placement and Y-stent-retriever maneuvers may serve as mechanical rescue approaches if an intracranial access was successfully established. A recent cohort study has suggested that the higher recanalization rates achievable with rescue permanent intracranial stent placement translated into clinical benefit without increasing the risk of symptomatic intracerebral hemorrhage. Although further studies in other populations are needed to confirm these findings, the results promote permanent stent placement as a feasible rescue technique in cases of stent-retriever failures.

**Identifying Patients with a High Risk of Reperfusion Failure**

Preinterventional stratification of patients with a high chance of mechanical reperfusion failure is desirable. This is important not only because in these cases, inclusion into currently enrolling randomized-controlled trials (Multicenter Randomized Clinical Trial of Endovascular Treatment for Acute Ischemic Stroke in the Netherlands—NO IV, [ISRCTN80619088]; and Solitaire With the Intention For Thrombectomy Plus In-
travenous t-PA versus DIRECT Solitaire Stent-Retriever Thrombectomy in Acute Anterior Circulation Stroke [SWIFT DIRECT, NCT03192332, clinicaltrials.gov]) may raise potential ethical considerations but also because primary alternative access routes might be considered beforehand. However, no reliable classification algorithm with high sensitivity and specificity is currently available. In our cohort, patients with reperfusion failure were older, treated later, and frequently had M2 occlusions. Combining those parameters, however, yielded low accuracy in correctly identifying patients with reperfusion failures. The issue of more reperfusion failures in M2 occlusions observed in the presented cohort deserves attention because a recent meta-analysis has suggested that endovascular therapy is successful in up to 85% of M2 occlusion. In this meta-analysis, recanalization rates were comparable between patients with M1 and M2 occlusions. However, the author acknowledged that the results can only be "interpreted in the context of patients with M2 occlusions that can be safely accessed by mechanical thrombectomy." In the present analysis, the initial intention for stent-retriever thrombectomy was crucial for inclusion, and patients were included even if no stent retriever was deployed at any time point during angiography. This criterion may differ from that in most of the retrospective studies included into the aforementioned quantitative synopsis because the major inclusion criterion was treatment with a stent retriever, rather than intention to treat with stent retrievers. Furthermore, the overall frequency of M2 occlusions treated with stent retrievers was small in our cohort, giving it scope for confounding due to the operator’s experience and associated learning curves.

**Strengths and Limitations**

While the present data allow a real-world intention-to-treat analysis of the relative frequencies of reasons for reperfusion failures, there are several limitations to this study: First, category IIB is defined as reperfusion failure after multiple attempts. Thus, there may be different thresholds for different interventionalists and centers to stop at different time points, and aggressiveness of rescue approaches vary. Second, futility criteria based on a consensus between the treating neurologists and neuroradiologists may change with time (eg, advanced time windows of >10 hours may no longer serve as a futility reason in the era of published Endovascular Therapy Following Imaging Evaluation for Acute Ischemic Stroke 3 (DEFUSE 3) and Clinical Mismatch in the Triage of Wake Up and Late Presenting Strokes Undergoing Neurointervention with Trevo (DAWN)) trial results. Third, intracranial access stability and torqueing forces on the stent retriever may impact the vessel and clot geometry and the respective clot interaction of the device. Thus, a classification into the true stent-retriever failure category may neglect the partial contribution of unstable intracranial access, potentiating difficulties to retrieve the clot or impeding the device efficacy. Fourth, standardized use of more aggressive rescue approaches for access failures may have averted some reperfusion failures in other centers. Last, distal aspiration techniques with large-bore catheters have gained popularity and have shown increasing effectiveness during the past years. Recent randomized-controlled trial results endorsed them as technically and clinically equally effective approaches (COMPASS/Combined Use of Contact Aspiration and the Stent Retriever Technique Versus Stent Retriever Alone for Recanalization in Acute Cerebral Infarction [ASTER]). Although aspiration attempts were implemented in the standard rescue approach at our center, overall distribution of reperfusion failure reasons may differ in centers that more commonly use aspiration techniques (ie, as first attempt) and other categories must be added.

**CONCLUSIONS**

Reasons for reperfusion failure in stent-retriever thrombectomy are heterogeneous. The failure to establish intracranial or cervical access is nearly as common as the inability to retrieve the clot despite the clot having been passed and adequate intracranial positioning having been established (true stent-retriever failure). Systematic reporting standards of those reasons may help to elucidate relative frequencies and thereby guide priorities for technical development and scientific effort (eg, adequate subgroup analyses regarding predictive factors). The low rates of delayed recanalization after mechanical reperfusion failures underline the benefit of alternative access routes and the need for a systematic evaluation of other medical (intra-arterial thrombolytics, antiplatelets, and so forth), or, if feasible, mechanical (eg, stent placement) rescue approaches.

**REFERENCES**


6. Froehle MT, Saver JL, Zaidat OO, et al; on behalf of the STRATIS Investigators. Interhospital transfer prior to thrombectomy is associated with delayed treatment and worse outcome in the STRATIS Registry. Circulation 2017;136:2311–21

7. Zaidat OO, Castonguay AC, Nogueira RG, et al. TREVO stent-retriever mechanical thrombectomy for acute ischemic stroke sec-
BACKGROUND AND PURPOSE: Thrombus permeability assessed on conventional CTA is associated with neurologic outcome in patients with acute ischemic stroke. We aimed to investigate whether dynamic CTA can improve the accuracy of thrombus permeability assessment and its predictive value for outcome.

MATERIALS AND METHODS: We reviewed consecutive patients with acute ischemic stroke who had occlusion of the M1 segment of the middle artery cerebral artery and underwent pretreatment perfusion CT. Thrombus permeability, determined by thrombus attenuation increase (TAI), was assessed on 26-phase dynamic CTA derived from perfusion CT. TAI_max was defined as the maximum TAI among phases; TAI_peak as TAI of peak arterial phase; TAI_con as TAI on phase 13. Good outcome was defined as a 3-month mRS score of ≤2.

RESULTS: One hundred four patients were enrolled in the final analysis. The median TAI_max, TAI_peak, and TAI_con were 30.1 HU (interquartile range, 13.0–50.2 HU), 9.5 HU (interquartile range, −1.6–28.7 HU), and 6.6 HU (interquartile range, −5.1–24.4 HU), respectively. Multivariable regression analyses showed that TAI_max (OR = 1.027; 95% CI, 1.007–1.048; P = .008), TAI_peak (OR = 1.029; 95% CI, 1.005–1.054; P = .020), and TAI_con (OR = 1.026; 95% CI, 1.002–1.051; P = .037) were independently associated with good outcome. The areas under the ROC curve of TAI_max, TAI_peak, and TAI_con, in predicting good outcome were 0.734, 0.701, and 0.658, respectively.

CONCLUSIONS: Thrombus permeability assessed on dynamic CTA could be a better predictor of outcome after reperfusion therapy than that assessed on conventional single-phase CTA.

Cerebral large-artery occlusion accounts for about one-third of acute ischemic strokes (AISs), which may cause severe disability and high mortality rates. Effort has been made on neuroimaging to predict the outcome of acute ischemic stroke because improved neurologic outcome is the goal for treatment. Thrombus characteristics on admission imaging, such as clot length, density, and location, may have the potential to predict the outcome of patients with AIS under different treatments. It has been suggested that thrombus permeability might be related to the physical porosity of thrombus and might reflect the ability of soluble molecules to move within the gaps among adjacent platelets, fibrin filaments, and red blood cells. Preclinical studies have demonstrated that high permeability of the thrombus within the occluded artery allowed residual blood to flow through the thrombus, which may have a positive effect on neurologic outcome after acute ischemia due to the compensating oxygenation of brain tissue distal to the occluded artery. Recently, thrombus attenuation increase (TAI) was used to assess the thrombus permeability on conventional single-phase CT angiography and noncontrast CT. A pervious thrombus with a high TAI

http://dx.doi.org/10.3174/ajnr.A5785
was found to be associated with better neurologic outcome after reperfusion therapy in patients with AIS.8,12

The enhancement of arteries on conventional single-phase CTA is affected by the acquisition time because the contrast filling of the occlusion site is delayed compared with that in the normal condition.14 Therefore, the permeability of the thrombi may be underestimated if single-phase CTA is acquired before contrast completely penetrates the thrombus. Compared with single-phase CTA, dynamically acquired CTA provides a broad temporal coverage that spans from unenhanced through to the arterial and, subsequently, venous phases. In addition, previous studies have demonstrated that dynamic CTA can better characterize the intracranial thrombus burden than single-phase CTA.14

We thus hypothesized that dynamic CTA could overcome the timing limitation and improve the assessment accuracy of thrombus permeability. We then assessed thrombus permeability on 26-phase dynamic CTA derived from CT perfusion and aimed to investigate its predictive value for neurologic outcome in patients with AIS after reperfusion therapy.

MATERIALS AND METHODS

Patient Inclusion

We retrospectively reviewed our prospectively collected data base for consecutive patients with AIS who received intravenous thrombolysis with or without endovascular thrombectomy from May 2009 to February 2017. Then, we enrolled patients who fulfilled the following criteria: 1) They had a diagnosis of AIS confirmed by diffusion-weighted imaging or CT at 24 hours after symptom onset, 2) underwent CTP within 8 hours after stroke onset, 3) had occlusion of the M1 segment of the middle cerebral artery without involvement of the internal carotid artery, 4) underwent CTA or time-of-flight MR angiography at 24 hours after treatment, and 5) underwent follow-up NCCT or susceptibility-weighted imaging at 24 hours after treatment. We excluded patients who had poor image quality due to motion artifacts or incomplete consecutive acquisitions.

Ethics Statement

Ethical approval was obtained from the human ethics committee of our center. The clinical investigation was conducted according to the principles expressed in the Declaration of Helsinki. Written informed consent was obtained from all patients.

Imaging Parameters

CTP was performed on a 64-slice CT scanner (Somatom Definition Flash; Siemens, Erlangen, Germany), including an NCCT scan (120 kV, 320 mA, contiguous 5-mm axial slices, 7-second acquisition time) and volume CTP (100 mm in the z-axis, 32 × 1.2 mm collimation). Volume CTP consisted of 26 consecutive spiral acquisitions of the brain. All 26 scans were divided into 4 parts: 1) two scans with 3-second cycle time, 2) fifteen scans with 1.5-second cycle time, 3) four scans with 3-second cycle time, and 4) five scans with 6-second cycle time. Axial slice coverage was 150 mm. A 60-ml bolus of contrast medium (iopamidol, Imeron; Bracco Sine, Shanghai, China) with a single injection was used at a flow rate of 6 mL/s, followed by a 20-ml saline chaser at 6 mL/s.

Imaging Analysis

The Arterial Occlusive Lesion Scale was used to assess recanalization on 24-hour MRA or CTA (grade 0: complete occlusion of the target artery; grade 1: incomplete occlusion or partial local recanalization at the target artery with no distal flow; grade 2: incomplete occlusion or partial local recanalization at the target artery with any distal flow; grade 3: complete recanalization and restoration of the target artery with any distal flow).15 Recanalization and no recanalization were defined as arterial occlusive lesion grades 2–3 and 0–1, respectively. Hemorrhagic transformation was assessed on 24-hour SWI or NCCT according to the second European-Australasian Acute Stroke Study (ECASS II), including hemorrhagic infarction and parenchymal hemorrhage (PH).16 Previously validated thresholds were applied to measure the baseline hypoperfusion volume (time-to-maximum > 6 seconds)17 and infarct core volume (relative cerebral blood flow of <30%) on CTP.18

Thrombus Permeability Assessment

Thrombus permeability assessments were conducted on commercial software (MIStar; Apollo Medical Imaging Technology, Melbourne, Australia). The 26-phase dynamic CTA with 4.5-mm-thick maximum intensity projection was reconstructed from CTP source images. The proximal artery of the contralateral hemisphere was selected to generate the arterial input function curve. The measurements of TAI were performed as described previously.8,12,19 Three ROIs with a radius of 1 mm were placed on the thrombus by 2 experienced neuroradiologists blinded to the patients’ information, with rater discrepancies settled by consensus discussion; the mean attenuation on each phase was calculated. The ROIs might partially overlap each other in case of a small thrombus. The mean attenuation of the thrombus on phase 1 was set as the reference value. The TAI of phase 2 to phase 26 was defined as the increase of the mean attenuation of the thrombus from phase 1 to each phase, respectively. Three parameters of TAI, including $\text{TAI}_{\text{max}}$, $\text{TAI}_\text{peak}$, and $\text{TAI}_{\text{cont}}$, were used for the final analyses (Fig 1). $\text{TAI}_{\text{max}}$ was defined as the maximal TAI among 25 phases (phases 2–26), $\text{TAI}_\text{peak}$ was defined as the TAI of the arterial peak phase according to the arterial input function curve. The mean time between contrast injection and peak concentration on the normal side of the middle cerebral artery was 25 seconds (phase 13) in current study; thus, $\text{TAI}_{\text{cont}}$ was simulated as TAI on phase 13 of dynamic CTA.

Clinical Data

We reviewed demographic, clinical, and radiologic data, including age; sex; prior antiplatelet use; risk factors such as smoking, hypertension, diabetes mellitus, hyperlipidemia, smoking, and atrial fibrillation; time interval from stroke onset to intravenous thrombolysis (ONT); National Institutes of Health Stroke Scale score on admission; baseline hypoperfusion volume and infarct core volume on admission; hemorrhagic transformation and recanalization after intravenous thrombolysis; and modified Rankin Scale score after 3 months. Patients were dichotomized...
into good (mRS score of ≤2) and poor outcome (mRS score of ≥3) at 90 days.

Statistical Analysis
Mean with SD, medians with interquartile range (IQR), and percentages were used to describe the distribution of continuous and categoric variables. The Fisher exact test was used to compare categoric variables among groups, whereas the independent-samples 2-tailed t test or the Mann-Whitney U test was used for the continuous variables, as appropriate. The Spearman rank correlation test was used to test the association of TAI with clinical and imaging variables. The strength of these associations was compared using the area under the receiver operating characteristic (ROC) curve. The ROC-derived optimal cutoff was determined at the maximal Youden Index. Variables with a P value < .1 in univariate analyses were enrolled in the multivariable regression model. A P value < .05 was considered statistically significant. All statistical analyses were performed with SPSS, Version 22.0 (IBM, Armonk, New York).

RESULTS

Overall Characteristics
One hundred four patients with AIS with MCA-M1 were enrolled in the final analysis. The mean age was 68 ± 14 years, the mean NIHSS score was 14 ± 6 on admission, and 61 (58.7%) patients were male. The median ONT was 215 (IQR, 140–295) minutes. The median TAImax, TAIpeak, and TAIcon were 30.1 HU (IQR, 13.0–50.2 HU), 9.5 HU (IQR, –1.6–28.7 HU), and 6.6 HU (IQR, –5.1–24.4 HU), respectively.

Association of TAI with Radiologic and Neurologic Outcome
As shown in Table 1, 44 (42.3%) patients achieved good outcome. Patients with good outcome had higher TAImax, TAIpeak, and TAIcon than those with poor outcome (P < .001, P < .001, and P = .006, respectively). Patients with good outcome were younger (P = .005) and had a lower baseline NIHSS score (P < .001), lower baseline infarct core volume (P = .001), lower baseline hypoperfusion volume (P = .004), and higher recanalization rates (84% versus 64%, P = .023) compared with those with poor outcome.

After we adjusted for age, baseline NIHSS score, atrial fibrillation, ONT, baseline infarct core volume, TAImax (OR = 1.027; 95% CI, 1.007–1.048; P = .008), TAIpeak (OR = 1.029; 95% CI, 1.005–1.054; P = .020), and TAIcon (OR = 1.026; 95% CI, 1.002–1.051; P = .037) were independently associated with good outcome, respectively. Patients with PH had lower TAImax and TAIpeak than those without PH, while TAIcon did not show a significant difference between patients with and without PH (P = .308).

The ROC curves of TAImax, TAIpeak, and TAIcon in predicting good outcome are shown in Fig 2, and the areas under the curve (AUCs) were 0.734, 0.701, and 0.658, respectively. The AUCs of each single-phase CTA derived from CTP are shown in On-line Table. The optimal cutoffs were 30.9, 9.0, and 14.6 HU for TAImax, TAIpeak, and TAIcon, respectively. Patients were then dichotomized into a pervious thrombus group and an impervious thrombus group according to the optimal cutoff values of TAImax, TAIpeak, and TAIcon, respectively. After we adjusted for age, baseline NIHSS score, atrial fibrillation, ONT, and baseline infarct core volume, patients with a pervious thrombus had a higher rate of good outcome than those with an impervious thrombus group according to the optimal cutoff values of TAImax, TAIpeak, and TAIcon, respectively. After we adjusted for age, baseline NIHSS score, atrial fibrillation, ONT, and baseline infarct core volume, patients with a pervious thrombus had a higher rate of good outcome than those with an impervious thrombus when dichotomized by TAImax (OR = 3.957; 95% CI, 1.463–10.705; P = .007) and TAIpeak (OR = 2.887; 95% CI, 1.075–7.754; P = .035) (Table 2). However, when dichotomized by TAIcon, patients with a pervious thrombus were not independently associated with...
good outcome (OR = 2.323; 95% CI, 0.815–6.621; P = .115) (Table 2).

**DISCUSSION**

In the present study, we found that high thrombus permeability on dynamic CTA is an independent predictor of good outcome in patients with AIS. Most interesting, TAImax and TAIpeak derived from CTP data are better imaging markers than TAIcon for predicting neurologic outcome.

Our finding that the permeable thrombus was strongly associated with good outcome is consistent with previous studies using conventional single-phase CTA.8,12 These studies also reported that the association between permeable thrombi and good outcome was independent of recanalization status.8 High thrombus permeability promotes anterograde filling of blood, and, therefore, less severely ischemic brain tissue.

Most important, our study revealed that the AUCs of TAImax and TAIpeak for predicting good outcome (0.734 and 0.701) were higher than those of TAIcon (AUC = 0.658) and previously reported values (AUC = 0.67) using conventional single-phase CTA; these findings indicate that thrombus permeability assessed on dynamic CTA might be a better biomarker for predicting outcome. Usually, single-phase CTA is used to assess thrombus permeability by calculating the attenuation increase before and after contrast penetrating the thrombus. The thrombus permeability on single-phase CTA might be underestimated due to timing limitations, hemodynamic restriction, or even pseudo-occlusion.14,20 It was reported that delayed phases after the arterial peak phase on dynamic CTA could provide better thrombus depiction and prognostic information and thus might affect treatment decisions in the acute setting.14,21 A previous study using 3-phase CTA found that arterial phase CTA was superior to venous phase CTA (8 seconds after arterial phase CTA) or delayed-phase CTA.

| Table 1: Comparison of baseline characteristics according to clinical outcome |
|---------------------------------|-------------------|-----------------|-----|
|                                | Poor Outcome      | Good Outcome    | P   |
|                                | (mRS > 2) n = 60  | (mRS ≤ 2) n = 44|     |
| Age (yr)                        | 74 (65–81)        | 66 (58–76)      | .005|
| Male (No.) (%)                  | 33 (55.0%)        | 28 (63.6%)      | .424|
| Baseline NIHSS (mean)           | 16 ± 5            | 11 ± 6          | <.001|
| Onset-to-needle time (mean) (min)| 230 (168–297)     | 207 (119–259)   | .059|
| Endovascular thrombectomy (No.) (%)| 22 (56.7%)       | 16 (36.4%)      | 1.000|
| Prior antiplatelet usage (No.) (%) | 12 (20.0%)       | 9 (20.5%)       | 1.000|
| Risk factors                    |                   |                 |     |
| Smoking (No.) (%)               | 18 (30.0%)        | 13 (29.5%)      | 1.000|
| Hypertension (No.) (%)          | 35 (58.3%)        | 22 (50.0%)      | .431|
| Diabetes mellitus (No.) (%)     | 16 (26.7%)        | 6 (3.6%)        | .146|
| Hyperlipidemia (No.) (%)        | 22 (36.7%)        | 21 (47.7%)      | .315|
| History of stroke/TIA (No.) (%) | 11 (18.3%)        | 5 (11.4%)       | .415|
| Atrial fibrillation (No.) (%)   | 38 (63.3%)        | 20 (45.5%)      | .076|
| Radiologic data                 |                   |                 |     |
| Baseline infarct core volume (mean) (mL) | 69.27 ± 49.51 | 40.72 ± 30.91 | .001|
| Baseline hypoperfusion volume (mean) (mL) | 134.28 ± 61.07 | 98.22 ± 60.43 | .004|
| TAImax (median) (IQR) (HU)     | 4.6 (4.2–17.5)    | 18.7 (5.9–43.0) | <.001|
| TAIpeak (median) (IQR) (HU)    | 2.2 (–5.5–12.4)  | 10.4 (–2.0–36.7) | .006|
| Thrombus perviousness by dichotomized TAImax (No.) (%) | 18 (30.0%) | 31 (70.5%) | <.001|
| Thrombus perviousness by dichotomized TAIpeak (No.) (%) | 22 (36.7%) | 31 (70.5%) | .001|
| Thrombus perviousness by dichotomized TAIcon (No.) (%) | 13 (21.7%) | 21 (45.5%) | .005|

| Table 2: Multivariable regression for good outcome |
|---------------------------------|-------------------|-----|
|                                | OR 95% CI | P   |
| Model 1                         |           |     |
| Age                             | 0.966     | 0.926–1.008 | .114|
| Baseline NIHSS                  | 0.911     | 0.823–1.008 | .071|
| Atrial fibrillation             | 0.872     | 0.298–2.555 | .803|
| ONT                             | 0.992     | 0.987–0.997 | .004|
| Baseline infarct core volume    | 0.986     | 0.972–1.001 | .061|
| Thrombus perviousness by dichotomized TAImax | 3.957 | 1.463–10.705 | .007|
| Model 2                         |           |     |
| Age                             | 0.963     | 0.924–1.004 | .074|
| Baseline NIHSS                  | 0.912     | 0.826–1.006 | .066|
| Atrial fibrillation             | 0.736     | 0.260–2.082 | .564|
| ONT                             | 0.992     | 0.987–0.997 | .003|
| Baseline infarct core volume    | 0.986     | 0.974–1.002 | .093|
| Thrombus perviousness by dichotomized TAIpeak | 2.887 | 1.075–7.754 | .035|
| Model 3                         |           |     |
| Age                             | 0.965     | 0.926–1.004 | .080|
| Baseline NIHSS                  | 0.917     | 0.827–1.016 | .098|
| Atrial fibrillation             | 0.751     | 0.269–2.102 | .586|
| ONT                             | 0.992     | 0.987–0.997 | .003|
| Baseline infarct core volume    | 0.986     | 0.972–1.000 | .060|
| Thrombus perviousness by dichotomized TAIcon | 2.323 | 0.815–6.621 | .115|

FIG 2. ROC curve of thrombus attenuation increase for TAImax, TAIpeak, and TAIcon, respectively.
(16 seconds after arterial phase CTA) to assess the TAI.\textsuperscript{22} However, 3-phase CTA did not cover enough time points, and the optimal phase could be located between the arterial phase and venous phase. Thus, we believe dynamic CTA derived from CTP was better for evaluating clot perviousness. Moreover, the characteristics of the arterial input function curve were affected by cardiac function, vascular curvature, and vascular stenosis. Thus, parameters based on individual hemodynamic characteristics derived from CTP data can better represent thrombus permeability than conventional single-phase CTA. Furthermore, the evaluation of dynamic CTA has no additional x-ray exposure and contrast medium usage because it is based on CTP, which can also provide information about infarct core and penumbra.

To the best of our knowledge, we are the first to have found that low thrombus permeability was associated with a high rate of PH after reperfusion therapy. It has been shown that compensating oxygenation to ischemic brain tissue could reduce hemorraghic transformation after thrombolysis.\textsuperscript{23–25} Therefore, we postulate that permeable thrombi may enhance tissue oxygenation distal to the occlusion due to increased passage of blood, which reduces the rate of PH. Additionally, the negative correlation between TAI and baseline NIHSS score, infarct core volume, and hypoperfusion volume might imply less severity of ischemia in patients with pervious thrombi, which could be related to lower rates of PH.

We did not find a significant association between thrombus permeability and recanalization, inconsistent with previous studies,\textsuperscript{8,12} which may be due to the inclusion of patients who received endovascular thrombectomy because endovascular thrombectomy usually has higher recanalization with mechanical retrieval devices. Another reason may be due to the assessment of recanalization at 24 hours after stroke in our study, while previous studies underwent follow-up imaging at 3 days or even later.\textsuperscript{8,12} Besides, the sample size of the current study was relatively small, and all patients were from a single center. Thus, further investigations in larger and multicenter cohorts are needed.

This study has several limitations. First, this was a retrospective design and might have a potential risk of selection bias, though the data were prospectively collected using the same stroke registry and CTP protocol. Second, although the ROI placement was performed by 2 experts blinded to the patients’ information with rater discrepancies settled by consensus discussion, variation still exists within a thrombus. Therefore, the mean attenuation of the 3 ROIs selected may not represent the entire thrombus. Third, the overlapping vessels, calcified clots, and vascular calcification might influence the attenuation measurements, though the ROIs were placed very carefully to avoid them. Fourth, further research about blood flow status on DSA is needed to explain the association between thrombus perviousness and good outcome. Finally, the sample size was modest, and all the samples were from a single center. Confirmation and extension of these findings in larger and multicenter cohorts are needed.

CONCLUSIONS

Our study suggests that thrombus permeability assessed on dynamic CTA could be a better predictor of outcome after reperfusion therapy than that assessed on conventional single-phase CTA.

Disclosures: Zhicai Chen—RELATED: Grant: National Natural Science Foundation of China (81600071) and Zhejiang Provincial Natural Science Foundation of China (LJQ16H040003)* Min Lou—RELATED: Grant: National Natural Science Foundation of China (81620207 and 81471707), Comments: National Natural Science Foundation of China (81620207 and 81471707), the National Key Research and Development Program of China (2016YFC0101500), and Fundamental Research Funds for the Central Universities (2017XZX002-09).* *Money paid to the institution.

REFERENCES


ABSTRACT

BACKGROUND AND PURPOSE: Hostile hemodynamic conditions and geometries are thought to predispose aneurysms for instability and rupture. This study compares stable, unstable, and ruptured aneurysms while controlling for location and patient characteristics.

MATERIALS AND METHODS: The hemodynamics and geometries of 165 stable, 65 unstable, and 554 ruptured aneurysms were compared. Hemodynamics was modeled using image-based computational fluid dynamics. Case-control pairs were selected matching aneurysm location, patient age, and sex. Paired Wilcoxon tests were used to compare hemodynamic and geometric variables among different aneurysm groups. The pairing was repeated 100 times, and the combined P values were calculated and adjusted for multiple testing.

RESULTS: Ruptured aneurysms had lower minimum wall shear stress (P = .03), higher maximum wall shear stress (P = .03), more concentrated (P = .03) and mean oscillatory shear stress (P = .03), higher maximum velocity (P = .03), and more complex flows (vortex core-line length, P = .03) than stable aneurysms. Similarly, unstable aneurysms had more concentrated shear stress (P = .04) and more complex flows (vortex core-line length, P = .04) than stable aneurysms. Compared with stable aneurysms, ruptured aneurysms were larger (size ratio, P = .03), more elongated (aspect ratio, P = .03), and irregular (nonsphericity index, P = .03). Similarly, unstable aneurysms were larger (size ratio, P = .04), more elongated (aspect ratio, P = .04), and irregular (bulge location, P = .04; area-weighted Gaussian curvature; P = .04) than stable aneurysms. No significant differences were found between unstable and ruptured aneurysms.

CONCLUSIONS: Unstable and ruptured aneurysms have more complex flows with concentrated wall shear stress and are larger, more elongated, and irregular than stable aneurysms, independent of aneurysm location and patient sex and age.

ABBREVIATIONS: AR = aspect ratio; Asize = aneurysm maximum size; BL = bulge location; CORELEN = vortex core-line length; flow complexity; CP = conicity parameter; GAA = area-weighted Gaussian curvature; LSA = percentage area under low WSS; max = maximum; min = minimum; OSI = nonsphericity index; Osc = oscillatory shear stress; SCI = concentrated shear stress; SizeR = size ratio (aneurysm size/vessel size); Vmax = maximum velocity; VOR = volume-to-ostium ratio; WSS = wall shear stress

Identification of Hostile Hemodynamics and Geometries of Cerebral Aneurysms: A Case-Control Study

B.J. Chung, F. Mut, C.M. Putman, F. Hamzei-Sichani, W. Brinjikji, D. Kallmes, C.M. Jimenez, and J.R. Cebral

ABSTRACT

BACKGROUND AND PURPOSE: Hostile hemodynamic conditions and geometries are thought to predispose aneurysms for instability and rupture. This study compares stable, unstable, and ruptured aneurysms while controlling for location and patient characteristics.

MATERIALS AND METHODS: The hemodynamics and geometries of 165 stable, 65 unstable, and 554 ruptured aneurysms were compared. Hemodynamics was modeled using image-based computational fluid dynamics. Case-control pairs were selected matching aneurysm location, patient age, and sex. Paired Wilcoxon tests were used to compare hemodynamic and geometric variables among different aneurysm groups. The pairing was repeated 100 times, and the combined P values were calculated and adjusted for multiple testing.

RESULTS: Ruptured aneurysms had lower minimum wall shear stress (P = .03), higher maximum wall shear stress (P = .03), more concentrated (P = .03) and mean oscillatory shear stress (P = .03), higher maximum velocity (P = .03), and more complex flows (vortex core-line length, P = .03) than stable aneurysms. Similarly, unstable aneurysms had more concentrated shear stress (P = .04) and more complex flows (vortex core-line length, P = .04) than stable aneurysms. Compared with stable aneurysms, ruptured aneurysms were larger (size ratio, P = .03), more elongated (aspect ratio, P = .03), and irregular (nonsphericity index, P = .03). Similarly, unstable aneurysms were larger (size ratio, P = .04), more elongated (aspect ratio, P = .04), and irregular (bulge location, P = .04; area-weighted Gaussian curvature; P = .04) than stable aneurysms. No significant differences were found between unstable and ruptured aneurysms.

CONCLUSIONS: Unstable and ruptured aneurysms have more complex flows with concentrated wall shear stress and are larger, more elongated, and irregular than stable aneurysms, independent of aneurysm location and patient sex and age.

ABBREVIATIONS: AR = aspect ratio; Asize = aneurysm maximum size; BL = bulge location; CORELEN = vortex core-line length; flow complexity; CP = conicity parameter; GAA = area-weighted Gaussian curvature; LSA = percentage area under low WSS; max = maximum; min = minimum; OSI = nonsphericity index; Osc = oscillatory shear stress; SCI = concentrated shear stress; SizeR = size ratio (aneurysm size/vessel size); Vmax = maximum velocity; VOR = volume-to-ostium ratio; WSS = wall shear stress

Many factors appear to be related to aneurysm rupture risk: location, size, morphology, hemodynamics, perianeurysmal environment, and patient history, among others. It is well-understood that aneurysm rupture-risk estimation is based on the combination of such factors and is a complex process. The American Heart Association and American Stroke Association guidelines for intracranial aneurysm management recommend that in addition to the size and location of the aneurysm and patient’s age and health status, it may be reasonable to consider the morphologic and hemodynamic characteristics of the aneurysm when discussing its rupture risk. Thus, identification of “hostile” hemodynamic conditions and geometries that predispose aneurysms to instability and rupture is important to discriminate high- and low-risk aneurysms and guide the management strategy for each individual patient. Additionally, understanding adverse conditions and their connections to the underlying mechanisms of aneurysm progression, instability, and rupture is valuable for designing new therapeutic strategies targeting those mechanisms.

Several previous studies have analyzed possible associations between hemodynamics and rupture and between aneurysm

http://dx.doi.org/10.3174/ajnr.A5764
morphology and rupture. One of the most important limitations of most cross-sectional studies comparing ruptured and unruptured aneurysms is the inability to discriminate stable and unstable unruptured aneurysms. On the other hand, the most important limitation of most longitudinal studies is the small sample size drawn from a highly selected population. Additionally, only a few studies control for aneurysm location and/or patient characteristics. These limitations may cause confusion and apparent conflicts among the results of different studies.

The goal of this study was to compare the hemodynamics and geometries of stable, unstable, and ruptured aneurysms while controlling for location and patient characteristics.

**MATERIALS AND METHODS**

**Data**

Our data base contains >2000 cerebral aneurysms imaged with 3D rotational angiography and basic information including aneurysm status, location, size, patient age, and sex. The main contributors have been Inova Fairfax Hospital (Virginia), Mt. Sinai Medical Center (New York), and the Mayo Clinic (Minnesota).

Age was divided into 3 groups: young (younger than 40 years), middle age (40–60 years), and old (older than 60 years). The images and data have been anonymized, and the study has been approved by the institutional review board of George Mason University. In the present study, a subset of 784 aneurysms was included, subdivided into the following subgroups (the distribution of aneurysms by location is presented in On-line Table 1):

1) **Stable Aneurysms.** Untreated unruptured aneurysms followed longitudinally without noticeable enlargement, shape change, new symptoms, or rupture for at least 1 year were considered stable. A total of 165 stable aneurysms in 85 patients were considered.

2) **Unstable Aneurysms.** Unruptured aneurysms were considered unstable if they either enlarged during follow-up (as determined by Sforza et al) or presented with symptoms such as cranial nerve palsy indicative of aneurysm instability. A total of 65 aneurysms in 55 patients were classified as unstable, with 53 aneurysms classified as growing and 12 with instabilities related to symptoms.

3) **Ruptured Aneurysms.** The ruptured aneurysms group included all aneurysms confirmed as the source of subarachnoid hemorrhage. A total of 554 ruptured aneurysms in 301 patients were considered.

**Models**

Computational fluid dynamics models of all 2000 aneurysms in our data base have been created from the corresponding 3D rotational angiography images (at baseline for longitudinally followed aneurysms) using previously described methods. These models include the patient-specific vascular geometry, but because patient-specific flow conditions were not available, typical pulsatile flow conditions were derived from phase-contrast MR imaging measurements in healthy subjects and scaled with the area of the inflow vessel. Outflow boundary conditions consistent with flow splits given by the principle of minimal work (Murray law) were prescribed at the outlets. Vessel walls were approximated as rigid, and blood viscosity was approximated as Newtonian. The unsteady incompressible Navier-Stokes equations were numerically solved with a finite-element code developed in-house. All simulations were run in parallel shared-memory computers for 2 cardiac cycles. Data from the second cycle were used to quantitatively characterize the aneurysm hemodynamic environment by computing several flow variables over the aneurysm cavity, the aneurysm orifice, and the aneurysm surface. Additionally, several geometric variables were also computed to characterize the aneurysm morphology. A list of the variables considered is provided in On-line Table 2.

**Analysis**

Case-control studies are a common and efficient means of studying rare diseases with long latency periods, such as ruptured/unruptured aneurysms. Matching of cases and controls is frequently used to control the effects of known potential confounding variables. The analysis of matched data requires specific statistical methods. The nonparametric Wilcoxon signed rank test was used for measured outcomes with 1:1 matching was used in the following 4 case-control studies: 1) stable-versus-ruptured aneurysms, 2) stable-versus-unstable aneurysms, 3) unstable-versus-ruptured aneurysms, and 4) unruptured-versus-ruptured aneurysms. In each study, case-control pairs were created and the mean values of hemodynamic and geometric variables of the 2 groups were statistically compared using a paired Wilcoxon test. The matching of cases and controls in each study was performed as detailed below:

1) **Stable-versus-Ruptured Aneurysms.** In this case, the controls (stable aneurysms) were fewer than the cases (ruptured aneurysms); therefore, the matching was performed as follows: For each stable aneurysm, a list of ruptured aneurysms with matching location, sex, and age groups was created, and one of these matching aneurysms was randomly selected. A total of 134 aneurysm pairs were thus created.

2) **Stable-versus-Unstable Aneurysms.** In this study, the cases (unstable aneurysms) were fewer than the controls (stable aneurysms); therefore, for each unstable aneurysm, a stable one was randomly selected from the list of stable aneurysms at the same location with matching sex and age groups. Sixty-five aneurysm pairs were created.

3) **Unstable-versus-Ruptured Aneurysms.** In this study, for each unstable aneurysm, a ruptured one was randomly selected from the list of ruptured aneurysms matching location, sex, and age. Sixty pairs were created.

4) **Unruptured-versus-Ruptured Aneurysms.** In this study, the controls (unruptured aneurysms) were fewer than the cases (ruptured aneurysms). Thus, for each unruptured aneurysm, a ruptured one was randomly selected from the list of unruptured aneurysms at the same location with matching sex and age groups. A total of 180 pairs were created.

Once the case-control pairs were created, the mean values of the hemodynamic and geometric variables of the 2 groups were statistically compared using a paired Wilcoxon test. Because the pairing processes described above involve random selection of a
case-control pair from the list of available matches, a bootstrapping or random sampling with a replacement approach was used to improve the robustness of the analysis. Specifically, the pairing and tests were repeated 100 times, and the P values were combined with the Wilkinson method. A difference was then considered statistically significant if the combined P value was <.05. The combined P values were then adjusted for multiple testing using the false discovery rate method. It was verified that 100 pairings were enough by performing 200 pairings and verifying that the significance of the comparisons did not change. All statistical analyses were performed by using R scripts (R statistical computing software; http://www.r-project.org).

RESULTS
The general hemodynamic and geometric characteristics of the aneurysms groups considered in this study are summarized in the Table. This Table lists the mean values and SDs of each variable over each group.  

Stable-versus-Ruptured Aneurysms
The statistically significant differences (P < .05 before correcting for multiple testing) between stable and ruptured aneurysms are summarized in Fig 1. The bars represent the ratio of the mean values of the ruptured aneurysm group over the mean values of the stable aneurysm group, and the error bars represent the variability of these ratios over the 100 random pairings. The mean and SDs of all variables over the 2 groups, along with the corresponding P values, are listed in On-line Table 3.

These results indicate that compared with stable aneurysms, after adjustment for multiple testing, ruptured aneurysms tend to have hemodynamic environments characterized by lower minimum wall shear stress (WSSmin, P = .03), higher maximum WSS (WSSmax, P = .03), more concentrated shear stress distributions (SCI, P = .03), more mean oscillatory shear stress (OSImax, P = .03), higher maximum velocity (Vmax, P = .03), and more complex flow patterns (vortex core-line length [CORELEN], P = .03). Additionally, ruptured aneurysms had geometries characterized by larger size (size ratio, aneurysm size/vessel size [SizeR], P = .03), more elongated shapes (aspect ratio [AR], P = .03), and more irregular shapes (nonsphericity index [NSI], P = .03).

Stable-versus-Unstable Unruptured Aneurysms
Comparisons of hemodynamic and geometric variables between stable and unstable unruptured aneurysms are presented in Online Table 4 and are summarized in Fig 2. These results indicate that unstable aneurysms tend to have more SCI (P = .04) and more complex flow patterns (CORELEN, P = .04) than stable aneurysms. Unstable aneurysms also tend to have larger areas

<table>
<thead>
<tr>
<th>Variable</th>
<th>Unruptured</th>
<th>Unstable</th>
<th>Growing</th>
<th>Ruptured</th>
</tr>
</thead>
<tbody>
<tr>
<td>WSSmin</td>
<td>0.8 ± 1.9</td>
<td>0.3 ± 0.5</td>
<td>1.0 ± 4.5</td>
<td>0.4 ± 1.3</td>
</tr>
<tr>
<td>WSSmax</td>
<td>223.7 ± 187.5</td>
<td>235.0 ± 101.8</td>
<td>253.7 ± 170.2</td>
<td>377.5 ± 948.6</td>
</tr>
<tr>
<td>WSSmean</td>
<td>21.4 ± 18.5</td>
<td>16.6 ± 13.8</td>
<td>22.0 ± 22.1</td>
<td>23.2 ± 28.9</td>
</tr>
<tr>
<td>WSSnorm</td>
<td>0.6 ± 0.3</td>
<td>0.4 ± 0.3</td>
<td>0.5 ± 0.3</td>
<td>0.4 ± 0.3</td>
</tr>
<tr>
<td>SCI</td>
<td>3.78 ± 4.76</td>
<td>8.52 ± 9.44</td>
<td>5.52 ± 4.53</td>
<td>6.23 ± 6.88</td>
</tr>
<tr>
<td>LSA</td>
<td>45.8 ± 33.6</td>
<td>62.9 ± 29.7</td>
<td>58.8 ± 30.9</td>
<td>51.6 ± 33.0</td>
</tr>
<tr>
<td>OSI</td>
<td>0.259 ± 0.137</td>
<td>0.358 ± 0.089</td>
<td>0.303 ± 0.125</td>
<td>0.317 ± 0.134</td>
</tr>
<tr>
<td>OSImean</td>
<td>0.012 ± 0.011</td>
<td>0.014 ± 0.008</td>
<td>0.017 ± 0.016</td>
<td>0.017 ± 0.017</td>
</tr>
<tr>
<td>Q</td>
<td>0.746 ± 0.805</td>
<td>1.360 ± 1.368</td>
<td>1.081 ± 0.957</td>
<td>0.730 ± 0.807</td>
</tr>
<tr>
<td>CI</td>
<td>0.716 ± 0.763</td>
<td>1.193 ± 1.309</td>
<td>1.088 ± 0.941</td>
<td>0.791 ± 0.764</td>
</tr>
<tr>
<td>Vmax</td>
<td>62.9 ± 31.8</td>
<td>70.8 ± 19.6</td>
<td>80.4 ± 50.0</td>
<td>88.8 ± 62.6</td>
</tr>
<tr>
<td>VE</td>
<td>9.7 ± 6.4</td>
<td>8.7 ± 5.7</td>
<td>10.1 ± 6.1</td>
<td>9.8 ± 8.1</td>
</tr>
<tr>
<td>SR</td>
<td>226.2 ± 176.1</td>
<td>155.0 ± 131.4</td>
<td>202.9 ± 221.2</td>
<td>231.5 ± 260.7</td>
</tr>
<tr>
<td>VO</td>
<td>307.7 ± 241.1</td>
<td>218.0 ± 183.7</td>
<td>278.9 ± 279.0</td>
<td>317.4 ± 344.4</td>
</tr>
<tr>
<td>CORELEN</td>
<td>1.753 ± 3.313</td>
<td>2.726 ± 1.827</td>
<td>2.504 ± 3.043</td>
<td>2.378 ± 2.593</td>
</tr>
<tr>
<td>PODENT</td>
<td>0.176 ± 0.125</td>
<td>0.227 ± 0.115</td>
<td>0.228 ± 0.155</td>
<td>0.207 ± 0.137</td>
</tr>
<tr>
<td>Asize</td>
<td>0.687 ± 0.524</td>
<td>1.109 ± 0.603</td>
<td>0.933 ± 0.550</td>
<td>0.770 ± 0.398</td>
</tr>
<tr>
<td>Nsize</td>
<td>0.482 ± 0.309</td>
<td>0.690 ± 0.461</td>
<td>0.577 ± 0.311</td>
<td>0.429 ± 0.172</td>
</tr>
<tr>
<td>SizeR</td>
<td>1.834 ± 1.395</td>
<td>2.440 ± 1.127</td>
<td>2.533 ± 1.487</td>
<td>2.492 ± 1.373</td>
</tr>
<tr>
<td>AR</td>
<td>0.879 ± 0.607</td>
<td>1.308 ± 0.828</td>
<td>1.200 ± 0.831</td>
<td>1.304 ± 0.691</td>
</tr>
<tr>
<td>VOR</td>
<td>0.894 ± 3.355</td>
<td>1.220 ± 1.310</td>
<td>1.460 ± 2.906</td>
<td>0.999 ± 1.559</td>
</tr>
<tr>
<td>EI</td>
<td>0.273 ± 0.036</td>
<td>0.268 ± 0.021</td>
<td>0.263 ± 0.026</td>
<td>0.273 ± 0.025</td>
</tr>
<tr>
<td>NSI</td>
<td>0.202 ± 0.051</td>
<td>0.229 ± 0.080</td>
<td>0.211 ± 0.041</td>
<td>0.246 ± 0.054</td>
</tr>
<tr>
<td>BF</td>
<td>1.167 ± 0.465</td>
<td>1.414 ± 0.478</td>
<td>1.377 ± 0.574</td>
<td>1.322 ± 0.441</td>
</tr>
<tr>
<td>BL</td>
<td>0.290 ± 0.161</td>
<td>0.431 ± 0.170</td>
<td>0.361 ± 0.144</td>
<td>0.355 ± 0.163</td>
</tr>
<tr>
<td>CR</td>
<td>0.762 ± 0.144</td>
<td>0.793 ± 0.083</td>
<td>0.816 ± 0.118</td>
<td>0.785 ± 0.100</td>
</tr>
<tr>
<td>CP</td>
<td>0.210 ± 0.161</td>
<td>0.069 ± 0.170</td>
<td>0.139 ± 0.144</td>
<td>0.145 ± 0.163</td>
</tr>
<tr>
<td>UI</td>
<td>0.238 ± 0.144</td>
<td>0.207 ± 0.083</td>
<td>0.184 ± 0.118</td>
<td>0.215 ± 0.100</td>
</tr>
</tbody>
</table>

Note: P indicates aneurysm flow rate; CI, inflow concentration index; VE, mean aneurysm velocity; SR, mean aneurysm shear rate; VO, mean aneurysm velocity; PODENT, POD entropy; flow instability; POD, proper orthogonal decomposition; Nsize, neck maximum size; EI, ellipticity index; BF, bottleneck factor; CR, convexity ratio; UI, undulation index; WSSnorm, normalized wall shear stress.

6 Values are given as mean ± SD.
under low WSS (percentage area under low WSS [LSA], $P = .10$ before adjustment) and more concentrated inflow jets (inflow concentration index, $P = .07$ before adjustment), but these associations did not reach statistical significance. Geometrically, unstable aneurysms are larger (aneurysm maximum size [Asize], $P = .04$; SizeR, $P = .04$), more elongated (AR, $P = .04$; volume-

**FIG 1.** Ratios of the mean values of the ruptured aneurysm group over mean values of the stable aneurysm group. Only hemodynamic (dark gray) and geometric (light gray) variables that were significantly different between these groups ($P < .05$ before adjusting for multiple testing) are included.

**FIG 2.** Ratios of mean values of the unstable aneurysm group over mean values of the stable aneurysm group. Only hemodynamic (dark gray) and geometric (light gray) variables that were significantly different ($P < .05$ before adjusting for multiple testing) or marginally significant ($P < .1$ before adjusting for multiple testing; marked with a plus sign) are included. ICI indicates inflow concentration index; BF, bottleneck factor; CR, convexity ratio.

to-ostium ratio [VOR], $P = .04$), and more irregular in shape (bulge location [BL], $P = .04$; conicity parameter [CP], $P = .04$; area-weighted Gaussian curvature; [GAA], $P = .04$) than stable aneurysms.

**Unstable-versus-Ruptured Aneurysms**

No significant differences between unstable unruptured aneurysms and matched ruptured aneurysms were found, including both hemodynamic and geometric variables (see the results in On-line Table 5).

**Unruptured-versus-Ruptured Aneurysms**

Results of the comparison of all unruptured aneurysms included in this study (stable and unstable combined) against matching ruptured aneurysms are presented in On-line Table 6 and summarized in Fig 3. Again, ruptured aneurysms had more SCI ($P = .04$) and more complex flow patterns (CORELEN, $P = .04$) with higher Vmax ($P = .04$) compared with unruptured aneurysms. They also tended to have lower WSSmin ($P = .02$ before adjustment) and larger WSSmax ($P = .02$ before adjustment), as well as more maximum oscillatory shear stress (OSImax, $P = .02$; OSImean, $P = .02$ before adjustment), but these associations became only marginally significant ($P = .06$) after adjusting for multiple testing. Geometrically, ruptured aneurysms were more elongated (AR, $P = .04$) and more irregular (NSI, $P = .04$) and tended to be larger (SizeR, $P = .02$ before adjustment), but the latter association became marginally significant ($P = .06$) after adjusting for multiple testing.

**DISCUSSION**

The current study attempts to characterize hostile aneurysm hemodynamics and geometry that could predispose aneurysms to instability and rupture while controlling for location, sex, and age. This knowledge is important for improved understanding of the mechanisms of wall degeneration and aneurysm progression.

The differences observed between stable and ruptured aneurysm pairs allow us to identify hemodynamic and geometric characteristics that are more prevalent in ruptured aneurysms. Our results indicate that these hostile hemodynamic characteristics include extreme values of wall shear stress (low and high), concentrated and oscillatory shear stress distributions, and complex intrasaccular flow patterns. Adverse aneurysm geometries include large, elongated, and irregular shapes. These characteristics are illustrated with 3 examples presented in Fig 4. Additionally, a simple correlation analysis was performed to identify variables correlated to one another. Groups of correlated variables (whose pair-wise regression coefficient was greater than 0.80) are presented in On-line Tables 7–9. Most interesting, no correlations were found between hemodynamic and geometric variables in this sample.

Differences between stable and unstable aneurysms were less pronounced but in the same general direction as those between stable and ruptured aneurysms. Specifically, unstable aneurysms had more complex flow patterns with concentrated wall shear stress than stable aneurysms, and they were larger and more elongated and irregular. These features suggest that unstable and ruptured aneurysms differ from stable aneurysms in similar ways.
Furthermore, no significant differences were found between unstable and ruptured aneurysms. These observations are consistent with a study\textsuperscript{25} that found differences between symptomatic and asymptomatic posterior communicating artery aneurysms that were similar to the differences between symptomatic and ruptured aneurysms at this location. This finding suggests that hemodynamic environments and shapes of unstable aneurysms may resemble those of ruptured aneurysms, but further studies with larger samples are needed to confirm this conjecture. Finally, when comparing all unruptured aneurysms (stable and unstable combined) against ruptured aneurysms, differences similar to those observed between stable and ruptured aneurysms were observed. This is likely due to the larger proportion of stable aneurysms in the unruptured group.

Several studies have compared stable and unstable unruptured aneurysms, but without controlling for location or patient characteristics. One such study did not find significant geometric or hemodynamic differences (149 stable, 20 unstable),\textsuperscript{26} while 2 other studies found differences in geometric (70 stable, 23 unstable)\textsuperscript{27} and hemodynamic (17 stable, 16 unstable) characteristics.\textsuperscript{11} Our findings agree with these latter studies. Two other studies found larger LSAs in unstable aneurysms compared with stable aneurysms after matching location and size (12 stable, 12 unstable),\textsuperscript{14} as well as sex and age (12 stable, 12 unstable).\textsuperscript{28} In general, our results show similar trends, but LSA did not reach statistical significance.

In summary, the results of our study suggest that hostile hemodynamic environments (ie, more prevalent in ruptured and unstable aneurysms than in stable aneurysms) are characterized by complex flows and concentrated wall shear stress distributions, while adverse geometries include larger and more elongated and irregular shapes. This finding leads us to propose that through geometric and flow analysis, there are ways to a priori identify unruptured aneurysms that will evolve toward instability and will therefore be at a higher risk of rupture. Additionally, our study suggests that lumping stable and unstable aneurysms into a single group of unruptured aneurysms (as in most cross-sectional studies) still allows characterization of adverse conditions likely because of the higher prevalence of stable aneurysms in the unruptured population.

Our study has several limitations. Computational fluid dynamics models are based on several assumptions and approximations.\textsuperscript{19} Our sample size and available data were not large enough to allow us to match and control for more patient characteristics or to obtain more statistical significance. The unstable group contained growing and cerebral palsy cases, which should be studied separately with larger samples. This study identified several trends that suggest that hostile hemodynamics and geometries could

**FIG 4.** Examples of stable (left), unstable (middle), and ruptured (right) aneurysms at a single location (posterior communicating artery). Visualizations show, from upper to lower: inflow jet, flow pattern, vortex core lines at 4 instants during the cardiac cycle, wall shear stress, and oscillatory shear index.
help identify aneurysms at higher risk of instability and rupture and should be considered in studies aiming at further understanding the underlying mechanisms that contribute to the evolution of a cerebral aneurysm from stable to unstable and ultimately ruptured status.

CONCLUSIONS

Unstable and ruptured aneurysms have more complex flows with concentrated wall shear stress and are larger, more elongated, and irregular than stable aneurysms, independent of aneurysm location, sex, and age. These adverse conditions could be used to identify unruptured aneurysms at higher risk of rupture and should be taken into account in studies of the mechanisms responsible for aneurysm wall degradation and progression to instability and rupture.


REFERENCES

Integrating 3D Rotational Angiography into Gamma Knife Planning


ABSTRACT

SUMMARY: 3D rotational angiography provides remarkable spatial resolution for cerebrovascular disorders; however, it cannot be integrated directly into gamma knife planning due to the discrepancy of DICOM “tag” information, and most physicians still cannot benefit from 3D rotational angiography. Here, we describe a simple and easy technique to enable the integration of 3D rotational angiography.

ABBREVIATIONS: GKRS = gamma knife stereotactic radiosurgery; 3DRA = 3D rotational angiography

Gamma knife stereotactic radiosurgery (GKRS) is an image-guided radiation therapy characterized by its high geometric accuracy; thus, no treatment margin is usually required when circumscribing the target. This feature, in conjunction with its sharp dose fall-off, enables high-dose irradiation in a single session; however, successful radiosurgery is highly dependent on the quality of radiographic images used.

GKRS has been accepted as one of the standard therapeutic modalities for small-to-medium arteriovenous malformations.1-9 Currently, biplanar DSA and CT angiography or MR imaging or both are commonly used in most institutions.10-15 Recently, advances in modern endovascular suites and newer generation flat panel detectors with C-arm systems have enabled acquisition of 3D rotational angiography (3DRA), providing remarkable spatial resolution for cerebrovascular disorders. Indeed, 3DRA is beginning to be used in the treatment planning of other modalities of stereotactic radiation therapy, including CyberKnife (Accuray, Sunnyvale, California), XKnife (Integra LifeSciences, Plainsboro, New Jersey), and Trilogy (Varian Medical Systems, Palo Alto, California), contributing to improved accuracy of the treatment planning.16-18 One remaining issue is that 3DRA cannot be integrated directly into GKRS planning because the planning software (Leksell GammaPlan; Elekta Instruments, Stockholm, Sweden) does not accept 3DRA. Moreover, very few studies have described the effectiveness of 3DRA on GKRS planning, though they do not state a detailed integration method.19-21 Thus, most physicians still cannot benefit from 3DRA. Here, we show a very simple method to integrate 3DRA into GKRS planning.

Techniques

Before the day of treatment, we usually perform MR imaging (mainly time-of-flight MR angiography, supplemented by T2 and gadolinium-enhanced T1 images) for preplanning. As in the usual preparations for GKRS, the Leksell frame (Elekta Instruments) is set on the patient’s head with the patient under sedation with local anesthesia. Then, the patient is transferred to the angiographic suite (Allura Xper FD20/10; Philips Healthcare, Best, Netherlands). Along with conventional DSA, 3DRA is acquired using the programmed acquisition protocol (3DRA mode). The amount of contrast medium used and preinjection delay are individually determined by neuroendovascular surgeons; briefly, 1–3 mL/s of contrast medium is continuously injected with a 1.5- to 2.0-second preinjection delay during the rotation of the C-arm. Because 3DRA could be easily coregistered to stereotactic CT if 3DRA contained enough bony tissue, a large-sized detector is generally preferred for precise image coregistration (Fig 1A). On the contrary, the smallest 8-inch detector could be used when the nidus is located near the skull base (basal ganglia, posterior fossa, and so forth) because acquired images spontaneously contain a large portion of bony tissues (Fig 1B). The obtained volume dataset is automatically transferred to the preinstalled workstation (XtraVision; Philips Healthcare), with which further reconstruction is performed with a 256³- or 512³-resolution voxel matrix in a
planned cube-shaped FOV with preset side lengths of 34.96, 52.18, 69.92, or 104.36 mm. We mostly use a 69.92-mm side-length cube with a 256³-resolution voxel matrix to maintain the balance between spatial resolution and contrast resolution; thus, spatial resolution is roughly calculated as 0.27 mm. Then, the DICOM “tag” technique is corrected from XA into CT in our software so that the GammaPlan can recognize the 3DRA properly as a CT-like image. Once the 3DRA is installed in the GammaPlan, image coregistration to stereotactic CT will be automated using the preinstalled coregistration function. A step-by-step instruction manual to integrate 3DRA is shown in On-line Figure. Detailed case illustrations are shown in Fig 2.

**DISCUSSION**

With the above-described simple contrivance, 3DRA can readily be used for GKRS planning. The strength of this technique is its accessibility. GammaPlan cannot directly accept 3DRA because of the discrepancy of DICOM tag information; accordingly, we must address this issue. No changes are required in the geometric or patient information; thus, the image quality itself is intact. Because 3DRA is a CT-like image, being characteristic of the finest spatial resolution for high-contrast objects with poor contrast resolution and thus having many features common to CT,22,23 stereotactic CT would be the best reference image for the coregistration.19 By means of high-definition images, physicians can not only reduce unwanted waste radiation to the surrounding brain tissues but also enable a safe prescription of high radiosurgical doses, enough to obliterate the nidus, which might theoretically lead to improvement in the obliteration rate as well as a decrease in radiation-induced adverse events. However, this article is only a technical report, and the actual clinical outcomes should be further examined.

Although 3DRA mode is preferred in our institution, contrast-enhanced conebeam CT with a small targeted FOV (high-resolution XperCT mode; Philips Healthcare) is also available in our angiographic suite. This acquisition mode provides superior spatial resolution with a long acquisition time (20 seconds), using a slow C-arm rotation and an 8-inch detector, which might be better for very small arteriovenous malformations. However, use of the high-resolution XperCT mode might raise a concern about coregistration. Notably, images containing a large portion of the cranium as well as skull base bone are important for precise coregistration; the use of a smaller detector contributes to further increased spatial resolution but also loses images of the surrounding cranium and skull base, leading to difficulty in coregistration.

We usually perform stereotactic CT after finishing angiography so that the contrast medium remaining in the blood vessels can provide additional information of vascular anatomies, which may enhance the quality of image coregistration. Further research is desirable to examine the coregistration accuracy to ensure the quality of the prescription of the therapeutic radiation dose.

Moreover, when a nidus receives blood flow from >2 vessels, the precise nidus contour is shown as the summation of parts of the nidus obtained by cannulation in each vessel. Thus, it is quite important to perform 3- or 4-vessel angiography and judge the involvement to avoid underestimation of the whole nidus angioarchitecture.

Although 3DRA provides superior resolution for angioarchitectures of vascular lesions, we recommend creating radiosurgical plans by meticulously comparing all the available imaging modalities. Particularly, MR imaging exhibits excellent contrast resolution and provides a better understanding of the surrounding functional brain anatomies, and DSA enables surgeons to instinctively recognize a spatial expanse of the nidus. Notably, surgeons should manipulate the DICOM header at their own risk because carelessly manipulating DICOM could spoil the quality of the images. Meticulous care must be taken not to change important information in the DICOM header other than technique because it could alter spatial relationships. Given the above reasons, 3DRA should not be used as a main technique but as reference information during radiosurgical planning.

**CONCLUSIONS**

We describe a simple, easy-to-access technique to enable integration of 3DRA into GKRS planning, which could provide the highest resolution for angioarchitectures of vascular lesions. The present method remains preliminary; thus, the created treatment plans should be validated in comparison with the conventional planning method. Further research is desirable to assess the effect of this technique on the actual radiosurgical outcomes.

Disclosures: Hirotaka Hasegawa—RELATED: Grant: JSPS KAKENHI, Comments: grant No. JP17K06628.
REFERENCES


FIG 2. A 34-year-old woman with an unruptured, small, left medial frontal arteriovenous malformation, once treated with gamma knife (A). The small remnant nidus persisted at 3.5 years from the initial radiosurgery (B, a red arrowhead shows the remnant nidus). The actual treatment planning of the secondary treatment (C) shows that 3D rotational angiography (left column) successfully depicts the faint remnant, while both time-of-flight (middle column) and gadolinium-enhanced T1 images (right column) fail to depict it. The nidus was finally obliterated (D) at 1.5 years from the secondary treatment. Yellow lines show the prescription isodose lines.


Quantification of Blood Velocity with 4D Digital Subtraction Angiography Using the Shifted Least-Squares Method


ABSTRACT

BACKGROUND AND PURPOSE: 4D-DSA provides time-resolved 3D-DSA volumes with high temporal and spatial resolutions. The purpose of this study is to investigate a shifted least squares method to estimate the blood velocity from the 4D DSA images. Quantitative validation was performed using a flow phantom with an ultrasonic flow probe as ground truth. Quantification of blood velocity in human internal carotid arteries was compared with measurements generated from 3D phase-contrast MR imaging.

MATERIALS AND METHODS: The centerlines of selected vascular segments and the time concentration curves of each voxel along the centerlines were determined from the 4D-DSA dataset. The temporal shift required to achieve a minimum difference between any point and other points along the centerline of a segment was calculated. The temporal shift as a function of centerline point position was fit to a straight line to generate the velocity. The proposed shifted least-squares method was first validated using a flow phantom study. Blood velocities were also estimated in the 14 ICAs of human subjects who had both 4D-DSA and phase-contrast MR imaging studies. Linear regression and correlation analysis were performed on both the phantom study and clinical study, respectively.

RESULTS: Mean velocities of the flow phantom calculated from 4D-DSA matched very well with ultrasonic flow probe measurements with 11% relative root mean square error. Mean blood velocities of ICAs calculated from 4D-DSA correlated well with phase-contrast MR imaging measurements with Pearson correlation coefficient $r = 0.835$.

CONCLUSIONS: The availability of 4D-DSA provides the opportunity to use the shifted least-squares method to estimate velocity in vessels within a 3D volume.

ABBREVIATIONS: PC = phase-contrast; SBR = sideband ratio; TCC = time concentration curve; VIPR = vastly undersampled isotropic projection reconstruction

2D digital subtraction angiography and a 3D rotational angiography acquisition followed by 3D-DSA reconstruction are the standards for vascular morphology assessment. There is an increasing demand for hemodynamic information, including blood flow rate and velocity for diagnosis, treatment planning, and evaluation. Many algorithms have been proposed to estimate blood flow in arteries from 2D-DSA; these are divided into 2 major classes: bolus-tracking and computational methods. A thorough review of these techniques can be found in Shpilfogel et al. However, 2D-DSA based methods are challenged by vessel overlap and vessels being aligned with the primary x-ray beam direction.

Recently, a new reconstruction method has been applied to a single 3D rotational angiography acquisition to generate time-resolved 3D-DSA at frame rates up to 30 frames/second (4D DSA). The availability of the geometric and temporal data in a 4D-DSA reconstruction provides the opportunity to estimate velocity and flow more accurately than has previously been possible with 2D DSA. In this article, a shifted least-squares based technique was used with the 4D-DSA data to estimate blood velocity. The proposed algorithm was first validated using flow phantom studies, in which the velocity could be documented using an ultrasonic flow probe. Blood velocity was also determined in 4D-DSA datasets from 14 ICAs of volunteers who also underwent 3D fast phase-contrast with vastly undersampled isotropic projection reconstruction (PC VIPR) studies.
 MATERIALS AND METHODS

4D DSA Acquisition and Data Preparation

Noncontrast (mask) and contrast-enhanced (fill) rotational acquisitions were performed using a conventional flat panel detector angiographic system (Artis zee; Siemens, Erlangen, Germany) with the following settings: 70 kV (peak), 0.36 μGy/frame, 304 images, 260° arc. Contrast injection (iohexol, Omnipaque 300; GE Healthcare, Piscataway, New Jersey) into the subject was started at the beginning of the fill run (3 mL/second, 7-second injection duration). The rotational acquisition protocol yields x-ray projection images, which are used to automatically reconstruct both 3D-DSA and 4D-DSA with an isotropic spatial resolution of 0.46 × 0.46 × 0.46 mm³.

The centerline of each vascular segment was determined from the static 3D volume using an efficient 3D parallel-thinning algorithm.20 The contrast waveform map, which is the time concentration curves (TCCs) along the centerline path (curvilinear length z), was extracted from the 4D-DSA and stored for each vascular segment for velocity calculation.

Shifted Least-Squares Algorithm

In an arterial contrast injection, there is a temporal oscillation in iodine concentration that arises from the mixing of contrast medium, which is injected at a fixed rate, and nonopacified blood, which flows at a variable rate driven by the cardiac cycle. This temporal variation in contrast, referred to as pulsatility, appears at points downstream of the injection with a time delay. The time delay is related to the distance downstream and the blood velocity. Therefore, velocity can be estimated by measuring the distance along the vessel centerline and the time delay.

For any 2 points, i and j, along the centerline, the shifted least-squares difference between their TCCs, \( c_i(t) \) and \( c_j(t) \), can be calculated as

\[
e(\tau_{ij}) = \frac{1}{T} \sum_{t=1}^{T} [c_i(t - \tau_{ij}) - c_j(t)]^2.
\]

The value \( \tau_{ij} \in [0, T] \) that minimizes the shifted least-squares difference, \( e(\tau_{ij}) \), is regarded as the time of bolus transport between these 2 points. The spatial distance \( z \) between the 2 points is calculated from the 3D path length along the vessel centerline. The time-shift \( \tau^0 \) as a function of the spatial distance \( z \) can then be fit to a linear relation:

\[
\tau^0 = \alpha \times z + b,
\]

where the slope \( \alpha \) is the inverse of the velocity \( v \).

Optimizing Waveform Selection

It has been found that the shifted least-squares algorithm provides the most reliable velocity calculation in vessel segments where pulsatility is strong and consistent. To automatically select the waveform regions with strong and consistent pulsatility, we generated a sideband ratio (SBR) map.21

For each point in the waveform map, a short-time Fourier transform was applied to generate the local power spectrum \( PS_{\text{local}} \), where \( f_0 \) is the 10th characteristic frequency. The SBR at this point can be calculated as

\[
SBR = \frac{|PS_{ij}|^2}{N \sum |PS_{ij}|^2}
\]

where the denominator summation includes the neighboring frequency points and excludes the fundamental frequency \( f_0 \). From this SBR map, a low-pass filter was applied and a threshold of one-fourth of the median SBR was enforced to generate a mask. This mask excludes low pulsatile signals from consideration in the velocity calculation.

Phantom Study

To validate the proposed method for velocity estimation, we used a flow phantom containing a loop of plastic tubing to perform 4D-DSA studies. Pulsatile water flow with different mean flow rates was generated using a perfusion pump with a controllable flow setting (L; Storkert, Freiburg, Germany). A 5F catheter was introduced into the tubing for injection of iodinated contrast medium (Omnipaque 300 diluted with 25% water) with a power injector. 4D-DSA-derived flow was compared with flow measurements made with an ultrasonic flow probe (400-Series Multi-Channel Flowmeter; Transonic, Ithaca, New York) attached to the tubing 5 inches away from the catheter tip.

Rotational projection images were acquired using the mask and iodine-filled scans with 12-second acquisition over a 260° arc. The frame rate was 30 frames per second. The 4D-DSA volume data were reconstructed using the vendor’s prototype software (Siemens Healthineers, Forchheim, Germany).

A time-resolved flow profile was recorded during each fill scan using the ultrasonic probe. The reference flow measurement was taken as the average of the flow profile over the same time window that was used for 4D-DSA-derived flow.

Five different flow settings were selected to provide average flow rates ranging from 285 to 1140 mL/min (corresponding to the mean velocity of 15–60 cm/s for the tubing with a 0.64-cm diameter). A 3D-DSA scan and reference flow measurements were made for each flow setting.

Clinical Study

Nine subjects who had both DSA and PC VIPR examinations were selected from a database generated under University of Wisconsin review board approval. Because 5 of 9 subjects had bilateral studies, a total of 14 ICAs were available for evaluation. Scanning parameters for postcontrast PC VIPR were the following: FOV = 22 × 22 × 22 cm³, TR/TE = 12.5/4.8 ms, velocity-encoded gradient-echo imaging = 80–100 cm/s, bandwidth = 83.3 kHz, readout matrix = 320 points per projections, spatial resolution for the complex difference image = 0.68 × 0.68 × 0.68 mm³. The complex difference images from the PC VIPR scan were segmented for vessel content using Materialise Mimics (Materialise NV, Leuven, Belgium). The segmented MR imaging data were then registered to the 3D-DSA using a correlated point registration with manual point placement, followed by an affine rigid registration. Centerline coordinates of the ICAs generated from the 3D-DSA volumes were used for velocity estimation. An automated workflow for the PC VIPR analysis was developed in Matlab (MathWorks, Natick, Massachusetts). This software performed cross-sectional plane placement for every voxel along the
FIG 1. A representative phantom study demonstrates the flow calculation from 4D-DSA. A. MIP image of the reconstructed phantom with centerline (red) overlaid. B. The contrast waveform map of the voxels along the centerline. The horizontal direction is the timeframe in the 4D-DSA scan, and the vertical direction is the position along the vessel centerline. Each point in the map $M(t,z)$ represents the signal intensity of the voxel at distance $z$ along the centerline and at time frame $t$. The highlighted area is the optimized waveform region where the pulsatility is strong and consistent. TCCs of 2 selected voxels along the centerline (marked as blue and red stars on A and blue and red curves on B) are shown in C. D. The least-squares differences of these 2 signals as a function of the time-shift, in which the minimal appears at $r^2 = 5$ shown as red arrow on D is considered as the time of bolus transport from the blue voxel to the red voxel. The time-shift as a function of the centerline position was fit to a linear relation ($f$), where the slope is the inverse of the velocity. E. The flow profile recorded from the flow probe (downsampled to 1/30 second to correspond to the 4D-DSA timeframes). Flow measurement is the average taken between the red lines, which correspond to the optimized window shown in B. AUI indicates arbitrary unit of intensity.
3D centerline path, segmentation of the vessel boundaries in each plane, and averaging of the normal velocities for all points inside the segmented boundaries. The reported velocity value for a vessel was then the average result over all the centerline points that were present in both the PC VIPR and 3D-DSA volumes.

**Statistical Analysis**

The validation process was performed using linear regression fit between flow probe ($V_{fp}$) and $V_{dsa}$ (calculated using 4D-DSA) for the phantom study. Correlation analysis and Bland-Altman plots were used to compare the velocity measurements using PC VIPR versus 4D-DSA for the clinical study. The Pearson $r$ was used for correlation analysis, and a $P$ value < .05 for the correlation coefficient was considered statistically significant.

**RESULTS**

**Phantom Study**

Figure 1 shows sample data from the phantom study. First, the centerline was generated from the static 3D volume (Fig 1A). Figure 1B is the contrast waveform map (TCCs along the centerline path) of the selected centerline. Each point in the map $M(t,z)$ represents the signal intensity of the voxel at distance $z$ along the centerline and at timeframe $t$ ($C_z(t)$).

The intensity variations along the vertical axis reflect the concentration of iodine contrast medium versus the distance from the injection site for a given time point. Variations along the horizontal axis reflect variations in concentration versus time for a given point along the centerline. The slope of the intensity ridges reflects the blood velocity. In the displayed format, higher blood velocities produce ridges that are closer to vertical, whereas lower blood velocities produce ridges closer to horizontal. The highlighted region is the optimized waveform region for the flow calculation determined using the SBR method. TCCs of 2 selected voxels along the centerline (marked as blue and red stars on Fig 1A and blue and red dashed lines on

**FIG 2.** Linear regression between the flow probe measurements and the 4D-DSA calculations from 15 phantom studies. Evaluations have been obtained with significance levels $P = 4.521E-11$ for the slope and $P = 0.18$ for the intercept.

**FIG 3.** A representative ICA study. A, The centerline positions of the ICA. B, The contrast waveform map of the voxels along the centerline. The highlighted area is the optimized waveform region for the flow calculation. C, The time-shift as a function of the centerline position was fit to a linear relation in which the slope is the inverse of the velocity.
Fig 1B) are shown in Fig 1C. Figure 1D shows the least-squares differences of these 2 signals as a function of the time-shift, where the minimal appears at $t^0 = 5$ is considered as the time of bolus transport from the blue voxel to the red voxel. Figure 1F demonstrates the linear fitting between the time-shift of each voxel along the centerline and its corresponding centerline position. Figure 1E is the flow profile recorded from the flow probe (downsampled to 1/30 second to correspond to the 4D-DSA timeframes). Flow measurement was taken as an average over the same temporal window used in the 4D-DSA flow calculation.

The linear regression fit between the flow probe measurement and the 4D-DSA calculation is shown in Fig 2. The linear regression equation is as follows: $V_{d\text{sa}} = 1.0069V_{fp} + 2.8482$, where $V_{d\text{sa}}$ is the velocity calculated from 4D-DSA and $V_{fp}$ is the measurement using the flow probe. The determination coefficient $R^2$ of the fit was 0.9677. The 95% confidence interval slope was 0.8966 – 1.1172 and the 95% CI intercept was $-1.4917 – 7.1881$. These evaluations have been obtained with significance levels of $P = 4.521E-11$ and $P = 0.18$, respectively. The relative root mean square error (normalized by mean velocity) was 11.3%.

**Clinical Study**

The average ICA flow rates were calculated from 13 of the 14 ICAs that had 4D-DSA examinations. One subject was excluded due to the poor pulsatility in the TCCs. Figure 3 shows an example of the ICA studies. Average ICA flow rates were also measured with the PC VIPR studies. Correlation between 2 measurements using 4D-DSA and PC VIPR is shown in Fig 4. Average ICA flow rates calculated using 4D-DSA and measured using PC VIPR are well-correlated, with the Pearson $r = 0.835$ and a significance level $P = .0002$. Figure 5 is the Bland–Altman analysis of 4D-DSA compared with PC VIPR, which shows all points within 2 SDs, with a bias of 6.5 cm/s and 2-SD limits of agreement of $-3.56$ cm/s and 16.5 cm/s.

**DISCUSSION**

The shifted least-squares algorithm is a simple-but-robust method to calculate flow rates by finding the bolus transport time between any 2 points along the centerline and performing linear regression with the corresponding positions. The combination of the spatial and temporal data available in the 4D-DSA reconstruction provides the opportunity to apply this technique to measurement of blood flow in routine 4D-DSA acquisitions. The flow phantom study demonstrated that by using this method, we were able to calculate the flow rates with 11% relative root mean square error compared with ultrasonic flow probe measurements. Results from the 13 human ICAs showed that the calculated velocities from 4D-DSA studies correlated well with the measurements.
using the 3D phase-contrast MR imaging method. We believe that the reason for the higher velocities in the 4D-DSA studies compared with the PC VIPR measurements (6.5 cm/s bias) is the increment in flow caused by the intra-arterial power injection of contrast medium.

The proposed method calculates the mean flow rate within a selected window defined by the SBR map. This truncated average of pulsatile flow varies depending on the window size and location. This variation can be demonstrated in Fig 6A and B, where the average of the flow profile was performed with a gradually expanding average window. The truncated average oscillates depending on where the truncation window ends and eventually converges to the mean signal as the window expands. Similar oscillation has been observed in the flow calculation using 4D-DSA. Figure 6D shows the estimated average velocity changes as the right side of the window gradually expands as shown in Fig 6C. The current algorithm automatically selects a window purely on the basis of the SBR map so that the selected window range may fall anywhere between 4 and 6 pulsatile cycles. Further criteria could be included to trim the window to an integer number of pulsatile cycles to minimize this error.

The proposed algorithm worked best when there was good and consistent pulsatility in the 4D-DSA TCCs. With each cardiac cycle, as an intra-arterial contrast bolus is injected, there is an oscillation in the ratio of contrast medium to nonopacified blood. This pulsatility can be identified and extracted from the contrast waveform map by thresholding the SBR map. For vessels more distal to the contrast injection site where the pulsatility can be low, injection protocols should be modified to enhance the pulsatility strength or contrast kinetics. The present work focused on the evaluation of blood velocity within the ICAs. Investigation of more downstream segments, including the MCAs and basilar arteries, should be performed in the future.

CONCLUSIONS

4D DSA is a novel technique that provides both the geometric and temporal information required to measure blood flow velocity and flow. In this study, calculated velocity values from 4D-DSA TCCs correlated well with human ICA measurements derived from phase-contrast MR imaging and flow phantom measurements using an ultrasonic flow probe.

**FIG 6.** Fluctuation of the velocity estimation with a different average range. A. The flow profile recorded from the flow probe. Red lines define the starting average window. This window was gradually expanded to the purple line. The corresponding mean velocity was shown as the blue curve. A zoom-in of the green window is shown in B. Similarly, velocity calculations have been performed by gradually expanding the window coverage from yellow to purple as shown in the waveform map (C). Estimated velocity varies (D) as the window edge slides from position A toward B. AU indicates arbitrary units.
REFERENCES


Correlation between Human Papillomavirus Status and Quantitative MR Imaging Parameters including Diffusion-Weighted Imaging and Texture Features in Oropharyngeal Carcinoma


ABSTRACT

BACKGROUND AND PURPOSE: The incidence of Oropharyngeal Squamous Cell Carcinoma (OPSCC) cases is increasing especially in the Western countries due to the spreading of human papilloma virus (HPV) infection. Radiological investigations, MRI in particular, are used in the daily clinical practice to stage OPSCC. The aim of this study was to investigate the association of quantitative MR imaging features including diffusion-weighted imaging and human papillomavirus status in oropharyngeal squamous cell carcinoma.

MATERIALS AND METHODS: We retrospectively analyzed 59 patients with untreated histologically proved T2–T4 oropharyngeal squamous cell carcinoma. Human papillomavirus status was determined by viral DNA detection on tissue samples. MR imaging protocol included T2-weighted, contrast-enhanced T1-weighted (volumetric interpolated brain examination), and DWI sequences. Parametric maps of apparent diffusion coefficient were obtained from DWI sequences. Texture analysis was performed on T2 and volumetric-interpolated brain examination sequences and on ADC maps. Differences in quantitative MR imaging features between tumors positive and negative for human papillomavirus and among subgroups of patients stratified by smoking status were tested using the nonparametric Mann-Whitney U test; the false discovery rate was controlled using the Benjamini-Hochberg correction; and a predictive model for human papillomavirus status was built using multivariable logistic regression.

RESULTS: Twenty-eight patients had human papillomavirus-positive oropharyngeal squamous cell carcinoma, while 31 patients had human papillomavirus-negative oropharyngeal squamous cell carcinoma. Tumors positive for human papillomavirus had a significantly lower mean ADC compared with those negative for it (median, 850.87 versus median, 1033.68; P < .001). Texture features had a lower discriminatory power for human papillomavirus status. Skewness on volumetric interpolated brain examination sequences was significantly higher in the subgroup of patients positive for human papillomavirus and smokers (P = .003). A predictive model based on smoking status and mean ADC yielded a sensitivity of 83.3% and specificity 92.6% in classifying human papillomavirus status.

CONCLUSIONS: ADC is significantly lower in oropharyngeal squamous cell carcinoma positive for human papillomavirus compared with oropharyngeal squamous cell carcinoma negative for it. ADC and smoking status allowed noninvasive prediction of human papillomavirus status with a good accuracy. These results should be validated and further investigated on larger prospective studies.

ABBREVIATIONS: HPV = human papillomavirus; OPSCC = oropharyngeal squamous cell carcinoma; SSF = spatial scaling factor; VIBE = volumetric interpolated brain examination

Oropharyngeal squamous cell carcinoma (OPSCC) is probably the most studied head and neck neoplasm in the past decade, mainly due to the discovery of the role of human papillomavirus infection in carcinogenesis. This led to identification of a subset of patients with human papillomavirus (HPV)-positive OPSCC with specific demographic characteristics (younger and with fewer comorbidities) and better prognosis compared with those with HPV-negative OPSCC (usually older with more comorbidities, smokers, and alcohol consumers).1–4 Other characteristics of HPV-positive OPSCC are the high ratio between nodal and primary tumor burden5 and the greater presence of intranodal cystic degeneration.6–8 MR imaging is routinely used for staging OPSCC in many centers. In addition to standard morphologic sequences, diffusion-weighted imaging offers an insight in tumor ultrastructure. Some recent studies have suggested that HPV-positive tumors are associated with lower apparent diffusion coefficient compared with HPV-negative ones.9–11 However, these results are equivocal.12

Texture analysis allows quantitative parameters to be ex-
tracted from CT, MR imaging, or PET images by applying various mathematic computations and algorithms. The technique has been recently introduced in medical imaging research and has provided promising results for cancer prognostication and a noninvasive signature of relevant genotypic and phenotypic tumor patterns. Bogowicz et al. recently demonstrated that a radiomic CT signature can predict HPV status in head and neck cancer, even if its accuracy was not high enough to represent a valid alternative to p16 immunohistochemistry (a surrogate biomarker) and direct viral DNA or messenger RNA studies, which remain the criterion standard. To the best of our knowledge, no study has used MR imaging texture analysis to noninvasively determine HPV status in oropharyngeal cancer. The first objective of this study was to test the correlation between ADC and HPV status on a larger and more homogeneous OPSCC sample compared with previous studies. The second objective was to investigate correlations between HPV status and a set of texture features extracted from both morphologic and DWI sequences.

**MATERIALS AND METHODS**

**Patients**

This retrospective study included patients with untreated histologically demonstrated T2–T4 OPSCC assessed by the Multidisciplinary Head and Neck Group of our institution between March 2010 and April 2017. HPV status was determined by direct viral DNA study on tissue samples. Pretreatment MR imaging was available for all patients. Patients with primary tumors too small to be analyzed or with a low quality of MR images due to artifacts were excluded from image analysis. Tumor, Node, Metastasis classification was performed according to the eighth edition of the TNM classification of head and neck cancer.

**HPV Status Assessment**

HPV status was evaluated with the _digene_ HC2 High-Risk HPV DNA Test (QIAGEN; https://www.qiagen.com/us/shop/detection-solutions/human-pathogens/digene-hc2-high-risk-hpv-dna-test/#orderinginformation), an in vitro nonradioactive nucleic acid hybridization assay with signal amplification using a chemiluminescent microtiter plate. This test is able to detect 18 HPV types, including high-risk (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68) and low-risk types (6, 11, 42, 43, 44). The Hybrid Capture 2 test shows a sensitivity and specificity equivalent to that of the polymerase chain reaction.

**MR Imaging Protocol**

MR imaging was performed using a 1.5T scanner (Magnetom Aera; Siemens, Erlangen, Germany). The MR imaging protocol included: axial TSE T2 (TR, 5570 ms; TE, 102 ms; slice thickness, 3 mm; matrix, 448 × 224), axial echo-planar DWI (TR, 3900 ms; TE, 59 ms; slice thickness, 3 mm; matrix, 132 × 132) with b-values of 0 and 1000 s/mm², and contrast-enhanced 3D fat-saturated gradient-echo T1 (volumetric interpolated brain examination [VIBE]) with an isotropic resolution of 0.7 mm. Apparent diffusion coefficient maps were automatically generated using an optimized noise filter.

**Texture Analysis**

Texture analysis was performed on primary tumors. MR images were transferred to an off-line PC and analyzed using proprietary texture analysis software TexRAD (TexRAD; Cambridge, UK). Three head and neck radiologists in consensus (M.R., M.L., E.T.), blinded to HPV status, drew an ROI on a single section of axial T2, ADC map, and contrast-enhanced VIBE images (Fig 1), encompassing the primary tumor on its largest axial area. TexRAD software uses the filtration-histogram method as described by Miles et al. The filtration-histogram method comprises an initial filtration step that highlights image features of a specified size, followed by histogram analysis of the filtered image. The size of highlighted image features is denoted by the spatial scaling factor (SSF), expressed in millimeters. Histograms generated from unfiltered and filtered images were quantified by the following parameters: mean (average value of the pixels within the ROI), SD (a measure of how much variation or dispersion exists from the average), mean of positive pixels (average value of the pixels greater than zero within the ROI), entropy (a metric that reflects texture irregularity and positively correlates with tumor heterogeneity), skewness (a measure of histogram asymmetry; positive values indicate histograms skewed to the right, while negative values indicate histograms skewed to the left; skewness = 0 indicates a perfectly symmetric histogram), and kurtosis (a measure of the peakedness of the histogram; positive kurtosis indicates a histogram that is more peaked than a Gaussian distribution, while negative kurtosis in-
RESULTS
Sixty-eight patients with OPSCC were enrolled; however, 4 were excluded because of the small size of the primary tumor and 5 for the low quality of MR images due to motion artifacts. Thus, 59 patients were analyzed (43 men and 16 women); 28 were HPV-positive (47%), and 31 patients were HPV-negative (53%). Table 1 summarizes baseline patient characteristics. Twenty-six HPV-negative lesions were found in smokers (class 1); 20 HPV-positive lesions, in nonsmokers (class 2); eight HPV-positive lesions, in smokers (class 3); and 5 HPV-negative lesions, in nonsmokers (class 4). Use of tobacco was observed in a significantly higher proportion of low image quality. In 1 patient, neither T2 nor DWI was analyzed because of artifacts (class 1); in 2, both were not available (class 2); in 4, DWI was not analyzed in 1 patient (class 3); and in 1 patient, neither T2 nor DWI was analyzed because of low image quality.

MR imaging texture analysis was performed on all 3 sequences (T2, DWI, and VIBE) in 50 patients. DWI was not analyzed in 5 patients (because of artifacts in 3 and because the sequence was not available in 2). The VIBE sequence was not analyzed in 3 patients (because of artifacts in 3 and because the sequence was not available). In 1 patient, neither T2 nor DWI was analyzed because of low image quality.

Table 2 summarizes the texture features that were significantly associated with HPV status. Mean ADC was the parameter with the highest discriminatory power, while other parameters had significant but higher P values. After Benjamini-Hochberg correction for the false discovery rate, only mean ADC maintained statistical significance. If one combined mean ADC and smoking habit in multivariable logistic regression, the resulting model allowed correctly classifying 88.2% of cases, corresponding to an area under the receiver operating characteristic curve of 0.944 (sensitivity, 83.3%; specificity, 92.6%) (Fig 2). When subclasses...
The relationship between texture parameters and HPV status was determined by a combination of HPV status and history of smoking being considered, some texture parameters were significantly different between class 3 (HPV-positive, smoking) and classes 1 and 2; class 4 (HPV-negative, nonsmokers) was excluded because of the very low patient numbers. In particular, patients in class 3 (HPV-positive, smokers) had a significantly higher skewness: $SSF = 2$ mm on VIBE sequences compared with the other 2 classes (Table 3). This association maintained its statistical significance after Benjamini-Hochberg correction.

**DISCUSSION**

Our study confirms the preliminary results obtained by previous publications regarding the correlation between ADC derived from DWI and HPV status in OPSCC; finding that HPV-positive tumors had a significantly lower ADC than HPV-negative tumors. Even if the reason for this finding remains unknown, several possible explanations may be conceived. On the basis of their previous study, Driessen et al.\(^27\) hypothesized that the low stromal volume observed in HPV-positive head and neck cancer could explain the cancer lower ADC. Furthermore, cancer cell nests in HPV-driven cancer are often surrounded by zones of lymphoid cells,\(^28\) which could increase tissue cell density, thus leading to lower ADC. This hypothesis is supported by a very recent pilot study published by Swartz et al.,\(^29\) who found a strongly significant negative correlation between ADC and CD3-positive cell count in 20 patients with OPSCC.

In addition, HPV-positive OPSCC seems to be more frequently associated with higher Ki-67 levels,\(^30\) which have been found to be negatively correlated to ADC in several cancer types.\(^31-35\) The same aforementioned study by Swartz et al.\(^29\) also demonstrated a negative correlation between ADC and Ki-67 expression. The increasing evidence for the relationship between the ADC value and HPV status is relevant given that the results from several preliminary studies\(^36-43\) suggest an association between higher tumor pretreatment ADC and a poor response to chemoradiation and prognosis in head and neck cancer, though none considered HPV status in multivariable analysis. Therefore, our results emphasize the need for future studies on DWI to include HPV status as a possible important confounder for ADC in OPSCC. The present study was performed on a larger patient cohort than previous studies.\(^10-12\) The number of HPV-positive and HPV-negative tumors is better balanced (28 and 31, respectively) compared with the group analyzed by Driessen et al.,\(^9\) which included only 6 HPV-positive tumors. Furthermore, the latter study included a miscellanea of head and neck cancers, while our study was selectively directed to analyze oropharyngeal cancer. Different from articles published by Chan et al.,\(^10\) Nakahira et al.,\(^11\) and Schouten et al.,\(^12\) HPV status in our study was defined by direct viral DNA and messenger RNA studies rather than by p16 immunohistochemistry. This is an important strength because p16 immunostaining is a surrogate marker of HPV status and, despite its high sensitivity, has only moderate specificity because p16 may be constitutively activated in HPV-negative OPSCC.\(^44\) Remarkably, the absolute ADC values that we found are compellingly lower than those obtained in the previously cited studies in both HPV-positive and HPV-negative OPSCC. Even if the reason for these discrepancies cannot be fully explained, they may be conceivably imputed to differences in sequences and segmentation strategy. Kolff-Gart et al.\(^45\) demonstrated that ADC values measured on different head and neck tissues differ significantly among different MR imaging systems and sequences. Juan et al.\(^46\) demonstrated that non-echo-planar sequences (like those used by Schouten et al.)\(^12\) produce significantly higher ADC values for the major salivary glands compared with echo-planar sequences. Choices of different b-values and intrinsic signal-to-noise ratio might also influence ADC. Finally, ROI segmentation on B0 images (as performed by Driessen et al.\(^9\)and Chan et al.\(^10\)) might include peritumoral edema, leading to increased ADC mean values.

This is the first study investigating the correlation between MR imaging texture features and HPV status in head and neck cancer. The relationship between texture parameters and HPV status was weaker than that observed between mean ADC and HPV status. Even if some parameters showed raw P values <.05 (1 below .01), they lost significance after correction for the false discovery rate, which is strongly recommendable when a large number of

![FIG 2. Receiver operating characteristic (ROC) curve of the predictive model based on mean ADC and smoking status. The classification variable was HPV status, while the variable was the predicted probability produced by multivariable logistic regression. AUC indicates area under the curve.](image)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Class 1 or 2</th>
<th>Class 3</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>Mean</td>
<td>No.</td>
</tr>
<tr>
<td>Skewness VIBE SSF 2</td>
<td>42</td>
<td>0.165</td>
<td>8</td>
</tr>
<tr>
<td>Kurtosis VIBE SSF 2</td>
<td>42</td>
<td>0.4544</td>
<td>8</td>
</tr>
<tr>
<td>Skewness VIBE SSF 3</td>
<td>42</td>
<td>0.1713</td>
<td>8</td>
</tr>
<tr>
<td>Skewness VIBE SSF 1</td>
<td>42</td>
<td>0.1119</td>
<td>8</td>
</tr>
</tbody>
</table>

**Table 3:** Quantitative MR imaging features significantly different between patients in class 3 (HPV+ and smokers) and those in classes 1 and 2 (smokers only and HPV+ only, respectively)
variables without a priori validation are tested. Conversely, a strongly significant association was found between skewness, which is in line with previous observations that the combination of HPV infection and tobacco smoking leads to a higher risk of treatment failure and an accumulation of genetic mutations.

This study has some limitations. It is retrospective and based on ROIs traced on a single slice rather than on volumetric tumor analysis. Due to the relatively low number of patients, analysis of texture data did not include complex classification algorithms, which would have required both training and validation subgroups, each with sufficient numbers of patients. Thus, further information regarding relevant texture patterns could still be studied within our data. A more comprehensive approach will be adopted when our sample size reaches an adequate size. Even if the identification of a noninvasive detection method for HPV status was not the objective of the present study, the good performance of the simple regression model based on mean ADC and smoking status (which led to correct classification in 88% of cases) is an interesting result that deserves to be validated on external datasets. Furthermore, the accuracy of the model could be further improved by the addition of texture information.

CONCLUSIONS

Our study confirms the correlation between ADC derived from DWI and HPV status in OPSCC with a significantly lower ADC in HPV-positive tumors compared to HPV-negative.

Bases on these findings, we developed a simple predictive model based on ADC and smoking status that can be used as a non-invasive and cost-effective detection method for HPV status. This model deserves to be validated on external datasets and to be perfected with other parameters to increase its sensitivity.

A further noteworthy observation was that patients who are both HPV-positive and smokers have significantly higher MRI skewness, which is in line with published literature. This result does not yet have an obvious explanation and therefore would require confirmation by adding more patients and to be validated on external datasets.

REFERENCES

Negative Predictive Value of NI-RADS Category 2 in the First Posttreatment FDG-PET/CT in Head and Neck Squamous Cell Carcinoma

P. Wangaryattawanich, B.F. Branstetter IV, M. Hughes, D.A. Clump II, D.E. Heron, and T.J. Rath

ABSTRACT

BACKGROUND AND PURPOSE: FDG PET/CT has a high negative predictive value in patients with head and neck squamous cell carcinoma who respond completely to non-operative therapy. However, the treatment failure rate in patients with a partial but incomplete response is unclear. Our aim was to investigate the negative predictive value of the first posttreatment FDG-PET/CT in patients with head and neck squamous cell carcinoma with incomplete response interpreted as Neck Imaging Reporting and Data System (NI-RADS) category 2.

MATERIALS AND METHODS: We retrospectively identified patients with head and neck squamous cell carcinoma treated with chemoradiation or radiation therapy with curative intent in our institution between 2008 and 2016. We included patients whose first posttreatment FDG-PET/CT was interpreted as showing marked improvement of disease but who had a mild residual mass or FDG avidity in either the primary tumor bed or lymph nodes (NI-RADS 2). The negative predictive value of FDG-PET/CT was calculated, including the 95% CI, using the Newcombe method. Two-year disease-free survival was the reference standard.

RESULTS: Seventeen of 110 patients (15%) experienced locoregional treatment failure within 2 years of completing treatment, yielding a negative predictive value of 85% (95% CI, 77%–90%). The most common location of tumor recurrence was the cervical lymph nodes (59%). The median time interval between completion of therapy and treatment failure was 10 months (range, 5–24 months).

CONCLUSIONS: In patients with an incomplete response after treatment of head and neck squamous cell carcinoma, the negative predictive value of the first posttreatment FDG-PET/CT was 85%, which is lower than the 91% negative predictive value of FDG-PET/CT in patients with an initial complete response. Patients with an incomplete response (NI-RADS 2) should undergo more frequent clinical and imaging surveillance than patients with an initial complete response (NI-RADS 1).

ABBREVIATIONS: AJCC = American Joint Committee on Cancer; HNSCC = head and neck squamous cell carcinoma; IR = incomplete response; NI-RADS = Neck Imaging Reporting and Data System; NPV = negative predictive value; TF = treatment failure

PET/CT is critical to the management of patients with head and neck squamous cell carcinoma (HNSCC), given its high accuracy in pretreatment staging and the detection of persistent and recurrent disease after therapy compared with CT or MR imaging, allowing salvage treatment to be initiated in a timely manner.1–12 The high negative predictive value (NPV) of posttreatment PET/CT in patients with a complete response has been established.13 It was demonstrated in a prospective, randomized controlled trial that concluded that planned neck dissection can be deferred in patients with HNSCC with a complete response on initial posttreatment PET/CT following definitive chemoradiation therapy.11,13 However, the treatment failure (TF) rate in patients who have a partial, incomplete response (IR) on the initial posttreatment PET/CT remains controversial, and management of an initial IR is inconsistent.14,15

The American College of Radiology has recently published the Neck Imaging Reporting and Data System (NI-RADS), a standardized radiologic reporting system for head and neck cancer imaging to facilitate communication between radiologists and referring physicians and to determine the appropriate next management steps for an individual patient.16 This reporting system also assists in sharing data among institutions, which may facilitate the advancement of head and neck cancer research. In NI-RADS, the results of posttreat-
ment imaging surveillance are classified into 4 numeric categories based on imaging suspicion for residual or recurrent tumors (category 1, no evidence of recurrence; category 2, low suspicion; category 3, high suspicion; category 4, definitive disease recurrence).15-18

The purpose of this study was to investigate the NPV of the first posttreatment FDG-PET/CT in patients with HNSCC with an IR interpreted as NI-RADS category 2.

MATERIALS AND METHODS
Study Design and Patient Selection
We conducted a retrospective study that was approved by our institutional review board (PRO08120419) and was in compliance with the Health Insurance Portability and Accountability Act. Medical records from our institutional Head and Neck Oncologic Data Repository were reviewed to include patients with histologically confirmed HNSCC treated with primary definitive chemoradiation therapy or radiation therapy at our institution between 2008 and 2016 with available contrast-enhanced staging imaging and at least 24 months of clinical and radiographic follow-up at our institution after the conclusion of treatment. Patients who had non-squamous cell malignancies or a history of previously treated head and neck cancer were excluded. Pretreatment tumor staging of all patients was performed using the American Joint Committee on Cancer (AJCC) Cancer Staging Manual, 7th edition.19 We selected patients whose first posttreatment PET/CT was performed 2–3 months after the completion of treatment and was interpreted as showing marked improvement in size and/or FDG avidity of a locoregional tumor, but with persistent mild residual soft-tissue abnormality and/or FDG uptake at either the primary tumor site or regional nodes. These PET/CT findings were categorized as NI-RADS 2 (low suspicion for residual viable tumor). These patients can be divided into 3 major groups: 1) patients who had a partial IR at the primary tumor beds, with complete response at regional nodal sites (P2N1); 2) patients who had a complete response at primary tumor beds, but partial IR at regional nodal sites (P1N2); and 3) patients who had a partial IR at both primary tumor beds and regional nodal sites (P2N2).

These patients were followed clinically and radiologically for 2 years after the conclusion of treatment to determine TF rates; TF was defined as the persistence of viable residual tumor or locoregional tumor recurrence after a disease-free interval. TF was confirmed by means of histopathology or unequivocal evidence of disease progression on subsequent imaging and clinical evaluation.

PET/CT Parameters
PET/CTs were performed on several PET/CT scanners (Discovery, GE Healthcare, Milwaukee, Wisconsin; and Biograph mCT, Siemens, Erlangen, Germany), ranging from 2 to 64 input channels. Except for water, patients fasted for at least 4–6 hours before the PET/CT scan and were instructed to avoid strenuous exercise before the test. Serum glucose levels were obtained, and imaging was deferred if glucose levels were >200 mg/dL. Each patient received 10–20 mCi of [18F] FDG dosed by body weight, after which the patient remained seated or recumbent in a relaxed environment during the 50-minute radiotracer uptake phase. Axial PET and diagnostic contrast-enhanced CT images were obtained from the calvarial vertex through the upper thighs after urinary voiding. Emission images were obtained at 50–60 minutes after radiopharmaceutical injection. Diagnostic CT images were obtained 45 seconds after administration of 125 mL of intravenous contrast (iopamidol 370 mg/mL, Isovue-370; Bracco, Princeton, New Jersey). CT parameters included the following: 120–30 kV-peak; variable milliampere (AutomA; GE Healthcare); pitch, 1.5–2; and collimation, 3.75-mm. The images of the head and neck were reconstructed in 2.5-mm slice thicknesses with a small FOV, whereas the images of the thorax, abdomen, and pelvis were reconstructed in 3.75-mm slice thicknesses with a full-body FOV.

Clinical Assessment, Treatment, and Surveillance Protocol
All patients had staging contrast-enhanced imaging using PET/CT, neck CT, or MR imaging. Patient treatment protocols, including radiation dose and chemotherapy regimen, were determined by the standard practice guidelines of the multidisciplinary head and neck oncology team at our institution. After completion of treatment, patients had clinical surveillance by means of physical examination and endoscopy every 6–8 weeks for the first year per our institutional protocol. The first PET/CT scan was performed approximately 8 weeks after the last course of chemoradiation therapy or radiation therapy, and subsequent PET/CT scans were then performed at 5, 8, and 14 months after therapy.20 If patients had clinical signs or symptoms suspicious for TF, PET/CT and histologic confirmation were pursued outside the usual imaging surveillance regimen.

Image Interpretation
All PET/CT scans were interpreted in routine clinical workflow by board-certified neuroradiologists with dedicated experience in head and neck radiology. Postprocessing fusion software (Mirada; Mirada Medical, Denver, Colorado) was used to assist in interpretation. The images were interpreted qualitatively, without specific standard uptake value thresholds. Patients were considered appropriate for inclusion in this study if their first surveillance scan showed a marked decrease in the size and FDG avidity of the documented primary tumor and metastatic lymph nodes with only mild residual soft-tissue abnormality and/or FDG uptake at either the primary tumor or regional nodes.

Statistical Methods
The primary outcome measure was 2-year disease-free survival. A 95% confidence interval for the negative predictive value for the first surveillance PET/CT was calculated using the Newcombe method.21 Disease-free survival was visualized with Kaplan-Meier survival curves. Statistical analyses were performed with SAS Version 9.4 (SAS Institute, Cary, North Carolina).

RESULTS
We performed a first posttreatment PET/CT on 2077 patients in our institution between 2008 and 2016. Of these, 464 patients (22%) were classified as having a NI-RADS 2 response. One hundred ten patients met the inclusion criteria, with a mean age of 59
Patient characteristics \((N = 110)\)

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Male</th>
<th>Female</th>
<th>Age (range) (mean) (median) (yr)</th>
<th>Oropharynx</th>
<th>Oral cavity</th>
<th>Larynx</th>
<th>Hypopharynx</th>
<th>Nasopharynx</th>
<th>Unknown primary</th>
<th>Paranasal sinuses/nasal cavity</th>
<th>HPV status (oropharynx) ((n = 51))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>85 (77%)</td>
<td>25 (23%)</td>
<td>30–87, 59, 58</td>
<td>19 (17%)</td>
<td>19 (17%)</td>
<td>6 (6%)</td>
<td>6 (6%)</td>
<td>5 (5%)</td>
<td>4 (3%)</td>
<td></td>
<td>Positive 35 (69%), Negative 8 (15.5%), Unknown 8 (15.5%)</td>
</tr>
<tr>
<td>Primary tumor location</td>
<td>Oropharynx 51 (46%), Oral cavity 19 (17%), Larynx 19 (17%), Hypopharynx 6 (6%), Nasopharynx 6 (6%), Unknown primary 5 (5%), Paranasal sinuses/nasal cavity 4 (3%)</td>
<td></td>
<td>T-stage</td>
<td>T0 5 (5%), T1 22 (20%), T2 30 (27%), T3 25 (23%), T4 28 (25%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N-stage</td>
<td>N0 28 (25%), N1 17 (15%), N2 61 (57%), N3 4 (3%)</td>
<td></td>
<td>M-stage</td>
<td>M0 107 (97%), M1 3 (3%), MX 0 (0%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNM stage</td>
<td>I 1 (1%), II 9 (8%), III 22 (20%), IV 78 (71%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interval between the first posttreatment PET/CT and completion of chemoradiation (range) (mean) (median) (wk)</td>
<td>3–46, 11, 9</td>
<td></td>
<td>Duration of follow-up (range) (mean) (median) (yr)</td>
<td>13–97, 4.9, 4.4</td>
<td></td>
<td>Total patients with residual or recurrent disease within 2 yr after completion of primary chemoradiation therapy*</td>
<td>17/110 (15%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: HPV indicates human papillomavirus; TNM, Tumor, Node, Metastasis.

*NPV = 85% [95% CI, 77%–90%].

DISCUSSION

The NI-RADS classification system for surveillance imaging in HNSCC is used to convey the degree of radiologic certainty regarding the presence of recurrent or residual disease in treated patients. Previous work has demonstrated that patients classified in the NI-RADS category 1 on their first surveillance PET/CT have an NPV of 91%. The results from the current study indicate that the NPV in patients with NI-RADS category 2 is lower at 85%. Due to the higher incidence of TF, patients with an IR should undergo more frequent clinical and imaging surveillance than patients with a complete response. For example, at our institution, patients who are categorized as NI-RADS 1 undergo their next imaging surveillance after 6 months, whereas patients categorized as NI-RADS 2 undergo imaging at 3 months.

The literature regarding the TF rates in patients with HNSCC with an initial IR is limited. There are no established treatment or surveillance protocols for these patients. The most recent National Comprehensive Cancer Network guidelines of 2018 recommend either observation, fine needle aspiration, or planned neck dissection if the initial 12-week posttreatment PET/CT shows an equivocal response of nodal disease (ie, size of lymph node <1 cm with abnormal FDG uptake which is suspicious for disease or size of the lymph node of >1 cm with no FDG uptake).
Current evidence supports using PET/CT surveillance in patients with HNSCC with an initial complete response following definitive treatment.9,11,14,23,24 Mehanna et al11 recently demonstrated, in a multicenter prospective randomized controlled trial, that there was no statistical difference in 2-year overall survival between patients who underwent planned neck dissection versus those who underwent PET/CT-guided surveillance following definitive nonoperative therapy. However, imaging surveillance resulted in fewer operations and substantial cost savings for the nonoperative group.

Porceddu et al14 prospectively investigated the performance of PET-directed management of the neck in 112 patients with node-positive HNSCC whose primary tumors showed complete response to therapy. The results of this study suggest that residual PET uptake is more relevant than a residual nodal mass in assessing therapy response. However, there was no difference in the TF rates between the patients with and without residual FDG avidity in our study (16% [13/82] versus 14% [4/28]). In our series, 9 of 51 patients (18%) with oropharyngeal cancer who had an IR developed locoregional failure within 2 years. More than 90% of patients in our series had TF within 5–16 months after the conclusion of treatment. Our results suggest that closer imaging and clinical surveillance in patients with an IR on initial PET/CT is warranted and may need to extend to 16 months to detect TF early with the goal of optimizing patient outcomes.

The TF rate of patients with NI-RADS 2 in our study is like that in the initial published performance of NI-RADS.17 Krieger et al17 analyzed a local recurrence rate of 58 of 618 targets that were scored NI-RADS 2. The overall rate of recurrence was 17.2%, with similar rates for the primary tumor bed and nodes. However, in our study, TF most frequently occurred in regional lymph nodes (65% [11/17]) versus the primary tumor site (41% [7/17]), with 1 patient having both local and regional recurrence. In addition, the results of our study reflect the capability of NI-RADS in predicting TF rates in patients with NI-RADS category 2. We support the current effort to standardize radiology reporting of PET/CT in patients with HNSCC with the goals of improving communication among physicians and guiding the next step in management.16-18

This study has several limitations. It is inherently limited by the retrospective study design resulting in variation in the timing of the first posttreatment PET/CT, which can impact the false-positive rate, mainly due to radiation-induced inflammation causing FDG avidity.1,8 In our series, most patients had the first posttreatment PET/CT at 8–9 weeks, which has been shown to have an accuracy similar to that of PET/CT performed 11–14 weeks after treatment.20 Regarding tumor staging, we used the AJCC Cancer Staging Manual, 7th edition, because of the heterogeneity of the clinical data and because it was used in the management of the patients. Approximately 15% of the patients with oropharyngeal cancer were also not tested for human papillomavirus status, limiting the application of the AJCC Cancer Staging Manual, 8th edition,25 in some patients. Moreover, a standard uptake value analysis was not used in PET/CT interpretation;

FIG 1. Patient 8. A 69-year-old woman with advanced oropharyngeal squamous cell carcinoma. A, Pretreatment PET/CT shows a large FDG-avid right faucial tonsil and lateral oropharyngeal wall tumor with an FDG-avid right level II nodal metastasis. B, At 8 months after completion of therapy, there is complete response of the primary tumor but mild residual FDG uptake in the treated right level II node. C, Surveillance PET/CT scan obtained at 16 months after treatment shows increased FDG avidity and size of the right level II node, consistent with regional treatment failure, confirmed by salvage neck dissection. There is no disease recurrence at the primary tumor site.

FIG 2. Two-year disease-free survival curve of patients with HNSCC with an incomplete response on the first posttreatment PET/CT scan. Time zero is defined as completion of treatment.
however, the repeatability of quantitative standard uptake value measurements, particularly in lesions with low FDG uptake, has been proved to be poor, and there is no established standard uptake value cutoff reliably distinguishing benign from malignant tissue. Furthermore, the true benefits of this PET/CT surveillance guideline for early detection of recurrent or residual disease and the potential impact on patients’ overall survival require further investigation, preferably with an additional long-term prospective study. Last, our patients were derived from a single institution with extensive experience in PET/CT imaging of head and neck cancer; therefore, the results may not be universally applicable.

CONCLUSIONS

In patients with incomplete response (NI-RADS 2) after treatment of HNSCC, the NPV of the first posttreatment FDG-PET/CT was 85%, which was lower than the 91% NPV of FDG-PET/CT in patients with a complete response (NI-RADS 1). Patients with an incomplete response should undergo more frequent clinical and imaging surveillance than patients with a complete response.

REFERENCES

The Diagnostic Value of Diffusion-Weighted Imaging in Differentiating Metastatic Lymph Nodes of Head and Neck Squamous Cell Carcinoma: A Systematic Review and Meta-Analysis

C.H. Suh, Y.J. Choi, J.H. Baek, and J.H. Lee

ABSTRACT

BACKGROUND: Accurate lymph node staging is crucial for proper treatment planning for metastasis in patients with head and neck squamous cell carcinoma.

PURPOSE: Our aim was to evaluate the diagnostic performance of DWI for differentiating metastatic cervical lymph nodes from benign cervical lymph nodes in patients with head and neck squamous cell carcinoma and to identify optimal cutoff values for ADC.

DATA SOURCES: A computerized literature search was performed to identify relevant original articles in Ovid MEDLINE and EMBASE.

STUDY SELECTION: Studies evaluating the diagnostic performance of DWI for differentiating metastatic cervical lymph nodes from benign cervical lymph nodes were selected.

DATA ANALYSIS: Diagnostic meta-analysis was conducted with a bivariate random-effects model, and a hierarchical summary receiver operating characteristic curve was obtained. Meta-regression was also performed.

DATA SYNTHESIS: Nine studies with 337 patients were included. In all studies, ADC values derived from metastatic lymph nodes were significantly lower than ADC values derived from benign lymph nodes. The median ADC cutoff value was 0.965 x 10^-3 mm^2/s. The pooled sensitivity and specificity for the diagnostic performance of DWI in differentiating metastatic lymph nodes from benign lymph nodes were 90% (95% CI, 84%–94%) and 88% (95% CI, 80%–93%), respectively. In the meta-regression, sensitivity was significantly higher in the studies using a 3-mm slice thickness (93% [95% CI, 88%–98%]) than in studies using a slice thickness of >3 mm (86% [95% CI, 77%–95%], P < .01).

LIMITATIONS: A small number of studies were included in our meta-analysis.

CONCLUSIONS: DWI demonstrated high diagnostic performance for differentiating metastatic lymph nodes from benign lymph nodes in patients with head and neck squamous cell carcinoma, and the median ADC cutoff value was 0.965 x 10^-3 mm^2/s. A 3-mm DWI slice thickness can provide a slight improvement in sensitivity.

ABBREVIATIONS: HNSCC = head and neck squamous cell carcinoma; LN = lymph node; NCCN = National Comprehensive Cancer Network; QUADAS-2 = Quality Assessment of Diagnostic Accuracy Studies-2

Lymph node (LN) metastasis is an adverse prognostic factor in patients with head and neck squamous cell carcinoma (HNSCC), and accurate LN staging is crucial for proper treatment planning. The National Comprehensive Cancer Network (NCCN) guidelines recommend CT and/or MR imaging with contrast for the initial work-up in patients with HNSCC. CT and MR imaging are useful for determining morphologic criteria, including shape, size, internal architecture, extracapsular extension, and vascular features associated with LN metastasis; however, diagnostic performance is limited, especially in normal-sized non-necrotic LNs.

During the past decade, diffusion-weighted imaging has been used for differentiating metastatic cervical LNs from benign cervical LNs in patients with HNSCC. Some studies have reported a high diagnostic performance for DWI, whereas other studies have demonstrated disappointing results. In addition, various cutoff values have been proposed for the apparent diffusion coefficient.

We considered it timely to review the DWI protocols, parameters, and reported diagnostic performances because there are no
Materials and Methods
This study adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines.15

Search Strategy
A preliminary literature search demonstrated that various terminology was used to indicate cervical LN metastasis, including “cervical nodal metastasis,” “malignant cervical lymph nodes,” “metastatic cervical lymph node,” and “cervical lymphadenopathy.” These synonyms for cervical LN metastasis were used in the search terms for Ovid MEDLINE and EMBASE: ((cervical lymph node metastasis) OR (cervical nodal metastasis) OR (malignant cervical lymph nodes) OR (metastatic cervical lymph node) OR (cervical lymphadenopathy)) AND ((diffusion lesion) OR (cervical lymph node metastasis)) OR (cervical lymphadenopathy) OR (cervical lymphadenopathy)). The literature search was not limited to a publication date or study setting but was limited to English-language publications. Any additional relevant studies identified were investigated, and the literature search was updated until January 3, 2018.

Study Selection
We used the following eligibility criteria: 1) patients with biopsy-proved HNSCC who underwent preoperative MR imaging including DWI, 2) histopathology as a reference standard, 3) provision of the diagnostic performance and corresponding ADC cutoff value for differentiating metastatic cervical LNs from benign cervical LNs, and 4) published original articles. Case reports/series (including <10 patients), reviews, conference abstracts, and studies including other types of tumor (including nasopharyngeal carcinoma or lymphoma), or a study population overlapping other studies were excluded. Authors of the studies were contacted for further information when 2 × 2 tables could not be acquired.

Data Extraction and Quality Assessment
The following information was extracted from the selected studies using a standardized form: 1) study characteristics: authors, publication years, institution, study period, study design, and data analysis (per LN versus per patient); 2) demographic characteristics: sample size, mean age, age range, sex, and proportion of metastatic LNs; 3) MR imaging characteristics: magnetic field strength, MR imaging vendor, MR imaging scanner, coil, DWI sequence, b-values (seconds/square millimeter), TR, TE, slice thickness, interslice gap, matrix, FOV, number of signal acquisitions, scan time, number of readers, experience of readers; and 4) outcomes: a 2 × 2 contingency table (number of true-positive, false-positive, false-negative, and true-negative results) demonstrating the presence of metastatic LNs according to the ADC values and optimal ADC cutoff values for differentiating metastatic from benign LNs.

The risk of bias was assessed for each selected study, according to the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) criteria.16 Two reviewers (C.H.S. and Y.J.C.) independently performed study selection, data extraction, and quality assessment.

Statistical Analysis
A diagnostic meta-analysis of the DWI was conducted with a bi-variate random-effects model.17-19 Individual study sensitivity/specificity and pooled sensitivity/specificity were plotted using a coupled forest plot. The pooled positive likelihood ratio, negative likelihood ratio, and diagnostic odds ratio were calculated. “Positive likelihood ratio” was defined as the likelihood that a DWI result positive for differentiating metastatic LNs from benign LNs would occur in patients with metastatic LNs. “Negative likelihood ratio” was defined as the likelihood that a DWI result negative for differentiating metastatic LNs from benign LNs would occur in patients without metastatic LNs. The “diagnostic odds ratio” was defined as the odds of having a positive DWI result in patients with metastatic LNs compared with the odds of having a positive DWI result in patients without metastatic LNs. A hierarchical summary receiver operating characteristic curve with 95% confidence and prediction regions was obtained, and the area under the hierarchical summary receiver operating characteristic curve was calculated.

Heterogeneity across the studies was explored using the inconsistency index (I2) and Cochran Q-statistics.20 I2 values of >50% indicated the presence of heterogeneity across the studies.21 Visual assessment of a coupled forest plot (inverse correlation indicating the presence of a threshold effect) and a Spearman correlation (a coefficient of >0.6 indicating the presence of a threshold effect) were performed to evaluate any threshold effect (positive correlation between sensitivity and the false-positive rate).22 Visual assessment of the difference between the 95% confidence and prediction regions in the hierarchical summary receiver operating characteristic curve (a large difference indicating heterogeneity) was also performed. The presence of publication bias was assessed by a Deeks funnel plot asymmetry test,23 and a slope coefficient with P < .1 was considered significant small-study bias.

Meta-regression was performed using the following covariates: 1) analysis method (per LN versus per patient); 2) the percentage of metastatic LNs (<36.7% [median value of the included studies] versus ≥36.7%); 3) underlying disease (HNSCC versus oral squamous cell carcinoma); 4) study design (prospective versus retrospective); 5) consecutive enrollment; 6) number of readers (2 versus 1); 7) magnetic field strength (3T versus 1.5T); 8) maximum b-value (<1000 versus 1000 s/mm²); 9) slice thickness (3 versus >3 mm); and 10) ADC measurement (whole node versus single section).

All statistical analyses were conducted by one of the reviewers (C.H.S., with 5 years of experience in conducting systematic reviews and meta-analyses) using commercially available software (STATA 15.0, StataCorp, College Station, Texas; and R statistical and computing software, Version 3.4.1; http://www.r-project.org/). P < .05 indicated statistical significance.
RESULTS

Study Search

Figure 1 provides an overview of the search strategy and study-selection procedure. After 15 non-English studies were excluded, our search yielded 214 records, of which 35 articles remained after screening of the titles and abstracts. The full text of these studies was reviewed, and 26 studies were excluded as follows (On-line Appendix): studies evaluating patients with enlarged cervical LNs (not all patients had HNSCC [n = 9]), studies including patients with non-HNSCC malignancy (nasopharyngeal carcinoma or lymphoma [n = 5]), a study population partially overlapping other studies (n = 4), studies that did not allow a 2 × 2 contingency table to be obtained (n = 4), and studies not in the field of interest (n = 4). There were no studies reporting the diagnostic performance of DWI without mentioning the corresponding ADC cutoff value. Ultimately, 9 studies with 337 patients were included in this meta-analysis and were considered for further analyses.4-12

Study Characteristics and Quality Assessment

The relevant study characteristics are summarized in On-line Table 1. Eight of 9 studies analyzed the diagnostic performance of DWI per LN,4,5,7-12 whereas 1 study performed analysis on a per-patient basis.9 The number of included patients ranged from 16 to 80, and the number of LNs ranged from 34 to 651. There were 7 prospective studies4-8,10,11 and 1 retrospective one,9 with the study design not being explicit in a further study.12 Informed consent was obtained in 8 studies,4,6-12 as was approval by an ethics committee or institutional review board.4-8,10-12

The results of the methodologic quality assessment according to QUADAS-2 are presented in Fig 2. Most studies were considered to have a low risk of bias and minimal concerns regarding applicability. Common weaknesses involved uncertainties in blinded to the reference standard when analyzing the MR imaging results and a poorly documented time interval between MR imaging and the reference standard. In the patient-selection domain, 2 studies had a high risk of bias due to a case-control design9 or inappropriate exclusion criteria.9 In the index test domain, 3 studies had an unclear risk of bias because no information was provided on blinded to the reference standard.6,7,12 In the reference standard and flow/timing domain, 1 study had a high risk of bias and a high concern regarding applicability because both histopathology and follow-up imaging results were used as a reference standard.9 No studies were excluded from the meta-analysis on the basis of the quality assessment.

FIG 1. Flowchart depicting the literature search and study selection.

FIG 2. Grouped bar charts indicating methodologic quality according to the QUADAS-2 criteria and expressed as the percentage of studies meeting each criterion. For each quality domain, the proportions of studies suggesting a low, high, or unclear risk of bias and/or concerns regarding applicability are illustrated in green, red, and blue, respectively.

**MR Imaging Characteristics**

The detailed MR imaging parameters are summarized in On-line Table 2. Four studies used 3T, and 5 studies used 1.5T MR imaging. Six studies used echo-planar imaging as a DWI sequence, with the other studies not being explicit. Two b-values were used for DWI in 4 studies; 3 b-values, in 4 studies; and 6 b-values, in 1 study. Four studies used a 3-mm slice thickness, and 4 studies used a 4- or 5-mm slice thickness.

ROIs were drawn manually around each LN in all studies, with the ROIs being placed on solid portions and avoiding cystic or necrotic portions. An ADC value covering the whole node was obtained in 5 studies, while an ADC value for a single section was obtained in 4 studies. The mean ADC was calculated in 8 studies, and the minimum ADC (the lowest value) was calculated in 1 study. The smallest LN sizes for ADC calculation were set at a minimal axial diameter of 2 mm, 3 mm, 4 mm, or 10 mm.

**Data Analysis**

In all studies, ADC values derived from metastatic LNs were significantly lower than ADC values derived from benign LNs. The optimal ADC cutoff values varied slightly among individual studies, ranging from $0.851 \times 10^{-3}$ mm²/s to $1.038 \times 10^{-3}$ mm²/s. The median ADC cutoff value was $0.965 \times 10^{-3}$ mm²/s. The individual sensitivities ranged from 80% to 97%, and the individual specificities ranged from 65% to 96%.

The pooled sensitivity and specificity for the diagnostic performance of DWI in differentiating metastatic LNs from benign cervical LNs were 90% (95% CI, 84%–94%) and 88% (95% CI, 80%–93%), respectively (Fig 3). The pooled positive likelihood ratio, negative likelihood ratio, and diagnostic odds ratio were 7.3 (95% CI, 4.4–12.1), 0.11 (95% CI, 0.07–0.19), and 64 (95% CI, 27–156), respectively. The area under the hierarchical summary receiver operating characteristic curve was 0.95 (95% CI, 0.93–0.97), which suggests high diagnostic performance (Fig 4).

Heterogeneity was present, with $I^2$ values exceeding 50% for both sensitivity and specificity. Visual assessment of the coupled forest plots revealed no threshold effect, and the Spearman correlation coefficient was $-0.471$ (95% CI, $-0.865–0.281$), also indicating no threshold effect. The slope coefficient for the Deeks funnel plot for differentiating metastatic LNs from benign LNs is presented in Fig 5 and suggests slight asymmetry in the data ($P = 0.03$) and possible publication bias.

The Table shows the results of the meta-regression to explore the influence of 10 covariates on pooled sensitivity and specificity. Slice thickness was revealed to be a significant factor affecting study heterogeneity. Sensitivity was significantly higher in studies...
using a 3-mm slice thickness (93% [95% CI, 88%–98%]) than in studies using a slice thickness of 3 mm (86% [95% CI, 77%–95%], \( P < .01 \)). Otherwise, the analysis method, percentage of metastatic LNs, underlying disease, study design, consecutive enrollment, number of readers, magnetic field strength, maximum b-value, and ROI used for ADC measurement were not significant factors affecting heterogeneity. MR imaging using a 3T scanner, a maximum b-value of 1000 s/mm\(^2\), and an ADC measurement of the whole node all showed slightly higher sensitivity; however, the differences did not reach statistical significance.

**DISCUSSION**

The present systematic review and meta-analysis demonstrated that in patients with HNSCC, the ADC derived from metastatic LNs was significantly lower than the ADC derived from benign LNs. The median ADC cutoff value was \( 0.965 \times 10^{-3} \text{ mm}^2/\text{s} \). In addition, our study demonstrated a high pooled sensitivity and specificity for the diagnostic performance of DWI for differentiating metastatic from benign LNs in patients with HNSCC. The meta-regression revealed that studies using a 3-mm slice thickness had higher sensitivity than studies using a slice thickness of >3 mm. Therefore, DWI using a 3-mm slice thickness should be optimally considered for differentiating metastatic from benign cervical LNs.

A dedicated sequence optimization is essential to obtain optimized DWI. In the meta-regression, sensitivity was significantly higher in studies using a 3-mm slice thickness (93% [95% CI, 88%–98%]) than in studies using a slice thickness of 3 mm (86% [95% CI, 77%–95%], \( P < .01 \)). A 3-mm slice thickness may help to detect smaller sized LNs. In addition, 3T MR imaging (92% [95% CI, 86%–98%]), the use of a maximum b-value of 1000 s/mm\(^2\) (91% [95% CI, 86%–97%]), and ADC measurement of the whole node (93% [95% CI, 89–97]) showed slightly higher sensitivity, though the differences did not reach statistical significance. A previous report also mentioned that a high gradient strength substantially increases the signal-to-noise ratio and that applying a larger number of b-values not only reduces the influence of noise propagation in ADC calculations but also decreases the risk of motion-related artifacts.\(^{10}\) When one uses DWI to differentiate metastatic from benign LNs in patients with HNSCC, use of a 3-mm slice thickness, a 3T scanner, a maximum b-value of 1000 s/mm\(^2\), and ADC measurement of the whole node should all be considered to obtain a high diagnostic performance. Considerable effort is required to achieve standardization, and further studies are needed.

Among the included studies, the ADCs derived from metastatic LNs were consistently lower than the ADCs derived from benign LNs. The lower ADC values are probably due to the tumor microstructure in metastatic LNs, which typically show a larger number of cells, cellular polymorphism, and increased mitosis in comparison with benign LNs; these characteristics may reduce the extracellular extravascular space and decrease the ADC value.\(^{24}\) In our study, the optimal ADC cutoff values ranged from \( 0.851 \times 10^{-3} \text{ mm}^2/\text{s} \) to \( 1.038 \times 10^{-3} \text{ mm}^2/\text{s} \), with 7 of 9 studies reporting optimal cutoff values between \( 0.94 \times 10^{-3} \text{ mm}^2/\text{s} \) and \( 1.038 \times 10^{-3} \text{ mm}^2/\text{s} \), which is a relatively small variation. In addition, the median ADC cutoff value was \( 0.965 \times 10^{-3} \text{ mm}^2/\text{s} \).

We recognize that our study has several limitations. First, a small number of studies were included in our meta-analysis; therefore, we cannot evaluate all potential causes of heterogeneity. Although we found that slice thickness was a significant factor affecting study heterogeneity, other technical aspects, including different TRs/TEs and different sets of b-values, may

**FIG 4.** Hierarchical summary receiver operating characteristic curve with 95% confidence and prediction regions of DWI for differentiating metastatic LNs from benign LNs in patients with HNSCC. Each circle indicates one included study.

**FIG 5.** The Deeks funnel plot showing the presence of publication bias. Numbers in circles refer to study number. ESS indicates effective sample size.
account for some portion of the heterogeneity. In addition, the low number of included studies may limit the power to achieve statistical significance. Second, publication bias was reported. One possible reason is that 2 studies showing negative results were excluded because of the nonavailability of 2 × 2 contingency tables.13,14 Therefore, our results should be interpreted cautiously, and the high diagnostic performance of DWI may have been overestimated. To overcome these limitations, we included a relatively homogeneous study population (ie, biopsy-proved HNSCC) and performed an extensive meta-regression using 10 covariates. Moreover, we applied recent robust methodology (hierarchical logistic regression modeling17–19) and reported our results according to prestigious guidelines (the Preferred Reporting Items for Systematic Review and Meta-Analysis15 and the Handbook for Diagnostic Test Accuracy Reviews published by the Cochrane Collaboration25). Nevertheless, caution should be used when applying our results to daily clinical practice.

CONCLUSIONS

DWI demonstrated a high diagnostic performance for differentiating metastatic from benign cervical LN in patients with HNSCC, and the median ADC cutoff value was 0.965 × 10⁻³ mm²/s. A 3-mm slice thickness for DWI can slightly improve sensitivity. Further large prospective multicenter studies are required to confirm these findings.


REFERENCES


diffusion-weighted magnetic resonance imaging and computed to-
mography perfusion. 


neck squamous carcinoma: efficacy of intravoxel incoherent mo-
tion magnetic resonance imaging for the differential diagnosis. 
Eur J Radiol 2017;90:159–65 CrossRef Medline

differentiation of small non-necrotic cervical lymph nodes in pa-
tients with head and neck malignancies? Eur J Radiol 2017;15: 
810–16 CrossRef Medline

reporting systematic reviews and meta-analyses of studies that
evaluate health care interventions: explanation and elaboration. 

16. Whiting PF, Rutjes AW, Westwood ME, et al; QUADAS-2 Group, 
QUADAS-2: a revised tool for the quality assessment of diagnos-

17. Suh CH, Park SH. Successful publication of systematic review and
meta-analysis of studies evaluating diagnostic test accuracy. 
Korean J Radiol 2016;17:5–6 CrossRef Medline

of studies evaluating diagnostic test accuracy: a practical review for 
clinical researchers, Part I: general guidance and tips. Korean J Ra-
diol 2015;16:1175–87 CrossRef Medline

of studies evaluating diagnostic test accuracy: a practical review for 
clinical researchers, Part II: statistical methods of meta-analysis. 
Korean J Radiol 2015;16:1188–96 CrossRef Medline

20. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-

in meta-analyses. BMJ 2003;327:557–60 CrossRef Medline

reviews of diagnostic studies: didactic guidelines. 

23. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publica-
    tion bias and other sample size effects in systematic reviews of di-
    agnostic test accuracy was assessed. J Clin Epidemiol 2005;58:882–93
CrossRef Medline

24. Koh DM, Collins DJ. Diffusion-weighted MRI in the body: applica-
    tions and challenges in oncology. AJR Am J Roentgenol 2007;188: 
1622–35 CrossRef Medline

Reviews of Diagnostic Test Accuracy, Version 1.0.0. The Cochrane Col-
January 8, 2017
Radiologic-Pathologic Correlation of Tumor Thickness and Its Prognostic Importance in Squamous Cell Carcinoma of the Oral Cavity: Implications for the Eighth Edition Tumor, Node, Metastasis Classification


ABSTRACT

BACKGROUND AND PURPOSE: Addressing the performance of an imaging-based parameter compared to a “gold standard” pathologic measurement is essential to achieve accurate clinical T-classification. Our aim was to determine the radiologic-pathologic tumor thickness correlation and its prognostic value in oral squamous cell carcinoma.

MATERIALS AND METHODS: All pathologic T1–T3 (seventh edition of the Cancer Staging Manual of the American Joint Committee on Cancer) oral squamous cell carcinomas diagnosed between 2010 and 2015 were reviewed. Radiologic tumor thickness was measured on preoperative CT or MR imaging blinded to pathology. The radiologic-pathologic tumor thickness correlation was calculated. The impact of the imaging-to-surgery time interval and imaging technique on the correlation was explored. Intra-/interrater reliability on radiologic tumor thickness was calculated. The correlation of radiologic-versus-pathologic tumor thickness and its performance as the seventh edition T-category modifier was evaluated. Multivariable analysis assessed the prognostic value of the radiologic tumor thickness for overall survival adjusted for age, seventh edition T-category, and performance status.

RESULTS: For 354 consecutive patients, the radiologic-pathologic tumor thickness correlation was similar for the image-to-surgery interval of ≤4.0 weeks (ρ = 0.76) versus 4–8 weeks (ρ = 0.80) but lower in those with more than an 8-week interval (ρ = 0.62). CT and MR imaging had similar correlations (0.76 and 0.80). Intraobserver and interobserver reliability was excellent (0.88 and 0.84). Excluding 19 cases with an imaging-to-surgery interval of >8 weeks, 335 patients were eligible for further analysis. The radiologic-pathologic tumor thickness correlation was 0.78. The accuracy for upstaging the T-classification based on radiologic tumor thickness was 83% for pathologic T1 and 74% for pathologic T2 tumors. Multivariable analysis confirmed the prognostic value of radiologic tumor thickness (hazard ratio = 1.5, P = .02) for overall survival.

CONCLUSIONS: This study demonstrates a good radiologic-pathologic tumor thickness correlation. Intraobserver and interobserver reliability for radiologic tumor thickness was excellent. Radiologically thicker tumor was predictive of inferior survival.

ABBREVIATIONS: DOI = depth of invasion; HR = hazard ratio; OS = overall survival; OSCC = oral cavity squamous cell carcinoma; pDOI = pathologic depth of invasion; pTT = pathologic tumor thickness; rDOI = radiologic depth of invasion; rTT = radiologic tumor thickness; TNM = tumor, node, metastasis; TT = tumor thickness

Tumor thickness (TT) and depth of invasion (DOI) are independently prognostic in oral cavity squamous cell carcinoma (OSCC). Although often used interchangeably, the precise definitions of TT and DOI differ. DOI assesses primary tumor invasiveness (measured from the adjacent normal mucosal basement membrane to the deepest point of tumor invasion), while TT represents the distance from the tumor surface to the deepest point of invasion. Recognizing the prognostic significance and clinical relevance, the eighth edition of the Cancer Staging Manual of the American Joint Committee on Cancer tumor, node, metastasis classification was released in 2017. This study demonstrates a good radiologic-pathologic tumor thickness correlation. Intraobserver and interobserver reliability for radiologic tumor thickness was excellent. Radiologically thicker tumor was predictive of inferior survival.
tasis (TNM) classification has differentiated both terms and introduced DOI into the OSCC T-classification. Inclusion of DOI is applicable to both clinical and pathologic T-classifications, though its prognostic value is primarily derived from surgical specimens. Although mainly managed via an operation, some patients with OSCC do not undergo an operation due to the high risk, functional considerations, and personal choice. In these cases, radiologic measurement combined with clinical assessment is the only way to assess TT and DOI to stage a tumor when an operation is not undertaken. Therefore, addressing the performance of an imaging-based parameter compared with a criterion standard pathologic measurement is essential to achieve an accurate T-classification. Robust data confirming the reliability of measuring radiologic depth of invasion (rDOI) versus pathologic depth of invasion (pDOI) do not exist, likely due to the unavailability of the latter because institutions traditionally only reported pathologic TT (pTT) not pDOI. Several studies have reported that TT measured on MR imaging or CT correlates well with pTT. However, the interrater and intrarater reliability and prognostic value of radiologic tumor thickness (rTT) remain elusive.

Confirming the reliability of radiologic-versus-pathologic measurement of a parameter and its prognostic value is paramount to ensuring feasible implementation of the eighth edition clinical T-classification for OSCC. Ideally, this requires a comparison of rDOI versus pDOI. However, because only pTT was available in our institution during the study period, we confined the radiologic-pathologic correlation to TT, though both rTT and rDOI were recorded. We hypothesized that the rTT-pTT correlation could be indicative of the rDOI-pDOI correlation. Because pTT has a similar implication for the T-classification compared with pDOI, we further evaluated the prognostic value of both rTT and rDOI for overall survival (OS).

**MATERIALS AND METHODS**

**Study Population**

Following ethics board approval, we reviewed all newly diagnosed pathologic T1-T3 (seventh edition) OSCCs treated with definitive surgery from 2010 to 2015. We included all OSCC subites except the lip (typically different etiology [ie, sun exposure] rather than smoking/drinking). Exclusion criteria included unavailable imaging or pTT, a >12-week imaging-to-surgery time interval, or unassessable rTT due to imaging artifacts. Clinical and pathologic information was obtained from our institutional data base, in which pTT and outcomes were prospectively recorded.

**Image Analysis**

Occasionally (<1% of cases), preoperative CT was provided from referring institutions using 5-mm collimation. Standard MR imaging protocol in our center includes nonenhanced T1, T2, and T2 fat-saturated sequences in axial, coronal, and sagittal planes with 3-mm thicknesses. We used outside MR imaging studies that included contrast-enhanced T1 sequences. However, rTT was usually assessed via nonenhanced T1 and T2 sequences in the most appropriate plane perpendicular to the mucosal surface. T2 fat-saturated images helped to identify tumor, especially if the lesion was small because such lesions are accentuated by their brighter T2 signal against the saturated background. Nonenhanced T1 images best delineated tumor margins with lower intermediate signal intensity contrasted against the brighter (fatty) signal of the adjacent tissues. To appreciate the difference between rTT and rDOI, we also recorded rDOI, measured from an “interpreted mucosal plane” across the closest intact surface of the normal mucosa (Fig 1). If both CT and MR imaging were available, rTT and rDOI were measured on both imaging modalities. If the same imaging technique was available at multiple time points, rTT and rDOI were measured on the examination most closely approximating the date of the operation.

MR imaging or CT or both were reviewed, and rTT and rDOI were measured by the first author (E.A.M.W.) blinded to the histopathologic findings. To ensure consistency of radiology-pathology rTT measurements, we consulted our pathologists (B.P.-O. and I.W.), who confirmed that the maximum pTT value recorded on synoptic pathology reports represented the value measured on the slice with the thickest tumor chosen after evaluating the entire gross tumor and all slices of a specimen. In the case of rTT measurements, we followed the same process (ie, going through the entire series of scans to find the best orientation, axial, coronal, and sagittal, and the image slice that visually represented the “thickest” portion of tumor to measure).

**Statistical Analysis**

To appreciate the practicality, we evaluated the difference in rTT and rDOI in “exophytic,” “ulcerated,” and “flat” tumor. To avoid potential confounding from tumor growth during the “wait time period” to an operation, we calculated the rTT-pTT Spearman correlation coefficient ($\rho$) among ≤4.0-, 4.0- to 8.0-, and >8.0-week subgroups to determine the acceptable time interval. To justify whether CT and MR imaging rTT measurements can be combined to increase study power, we compared the performance of CT-versus-MR imaging on the rTT assessment. Finally, to determine the reliability and reproducibility of rTT measurements, interrater and intrarater reliability was assessed using the Cohen $\kappa$ coefficient in a subset of patients. Blinded rTT re-assessment by the initial interpreter (E.A.M.W.) and a second experienced neuroradiologist (E.Y.) was undertaken after a 3-month interval.

After excluding cases with unacceptably protracted imaging-to-surgery time intervals, a valid study cohort for rTT-pTT correlation analysis and prognostic assessment was assembled. We calculated the Spearman correlation coefficient of rTT versus pTT and the shrinkage factor (dividing the mean of pTT by the mean of rTT) for the entire cohort, oral tongue subgroup, and other OSCC subgroup. To evaluate the performance of rTT as a potential T-classification modifier to upstage the seventh edition T-category to the eighth edition, we calculated the diagnostic accuracy of rTT (with or without adjusting for the shrinkage factor) versus pTT within the seventh edition T1 (TT $\leq$ 5 versus >5 mm) and T2 (TT $\leq$ 10 versus >10 mm) tumors. Shrinkage factor was used to account for potential tumor shrinkage during specimen processing and fixation. Finally, to assess the prognostic value of rTT and rDOI and its implication for staging refinement, we calculate OS using Kaplan-Meier methods and compared within the seventh edition T1 (cutoff: ≤5 versus >5 mm) and T2 tumors (cutoff: ≤10 versus >10 mm) using the log-rank test. Multivariable analysis calculated the hazard ratio (HR) of the risk of death for rTT
and rDOI separately, adjusting for age, seventh edition Tumor/Node category, and Eastern Cooperative Oncology Group performance status. We also calculated the HR of the eighth edition T-category using rTT and rDOI as the seventh edition T-category modifier, separately adjusted for the aforementioned covariates.

All statistical analyses were 2-sided, and a $P$ value of $<0.05$ was considered statistically significant.

RESULTS

Of 463 consecutive OSCCs during the study period, 109 were excluded (lip tumors: $n = 6$; imaging-to-surgery time interval $>12$ weeks: $n = 12$; unavailable pathologic reports: $n = 26$; nonassessable tumor due to imaging artifacts: $n = 65$). The remaining 354 were eligible for exploratory analyses (On-line Figure).

Primary Tumor Type and Difference in rTT-versus-rDOI Measurements

Both rTT and rDOI were measured on all scans. Most (311/354, 87.9%) were flat tumors, where rTT and rDOI yielded the same measurement. Only 36 (10%) were exophytic (rTT > rDOI) and 7 (2%) were ulcerated tumors (rTT < rDOI). The median differences between rTT and rDOI were 4.4 mm (range, 0.2–17.6 mm) and 1.7 mm (range, 0.1–14.4 mm) for exophytic and ulcerated tumors, respectively.

Influence of the Imaging-to-Surgery Time Interval on the rTT-pTT Correlation

To explore the potential impact of the imaging-to-surgery time interval on the rTT-pTT correlation, we stratified the 354 cases into 3 subgroups: 0–4 weeks ($n = 205$, 58%), 4.1–8.0 weeks ($n = 130$, 37%), and >8 weeks ($n = 19$, 5%). The rTT-pTT correlation was similar between <4.0 weeks ($p = 0.76$) and 4.1–8.0 weeks ($p = 0.80$) ($P = .83$), but it was nonsignificantly lower for >8 weeks ($p = 0.62$, $P = .69$).

Comparison of the rTT-pTT Correlation on CT versus MR Imaging

A total of 206 patients had preoperative CT, and 187 had MR imaging (49 had both CT and MR). While MR imaging showed slightly better correlation, the difference was minor ($p = 0.80$ versus 0.78 for all cases; 0.75 versus 0.67 for MR imaging/CT both available cases) after adjusting for the imaging-to-surgery interval ($P = .83$). Thus, we combined rTT on CT and MR imaging as a composite rTT for subsequent analyses.

Interrater and Intrarater Reliability for rTT Measurements

On the basis of the power calculation, 85 cases (provided at least 85% power to detect significant difference) were randomly selected for the intrarater and interrater reliability assessment. The intrarater and interrater concordance of rTT was 0.88 (95% CI, 0.83–0.92) and 0.84 (95% CI, 0.77–0.90), respectively.

rTT-pTT Correlation and rTT as the T-Category Modifier

After we excluded 19 cases with more than an 8-week imaging-to-surgery interval due to their suboptimal rTT-pTT concordance, the remaining 335 cases (189 oral tongue and 146 other oral cavity subsites) were eligible for further analyses. The clinical characteristic of these 335 cases are listed in Table 1. The distribution of rTT versus pTT showed a clear linear correlation for the entire cohort as well as tongue and other subsites (Fig 2). The rTT-pTT correlation adjusted for the imaging-to-
surgery time interval was 0.78 for the entire cohort and 0.74 for both oral tongue and other subsite subgroups with shrinkage factors of 0.81, 0.90, and 0.71, respectively (Table 2). Because 0.80 was the most commonly used shrinkage factor\(^1\) and almost identical to 0.81 derived herein, we used 0.80 as a shrinkage factor to account for potential shrinkage of tumor during specimen processing and fixation.

Sensitivity, specificity, positive predictive value, negative predictive value, and accuracy for the ability of rTT to upstage the original size-based T-category from T1 to T2 (TT\(_{\text{rTT}}\)/H\(\geq 0.5\) cm) and from T2 to T3 (TT\(_{\text{rTT}}\)/H\(\geq 1.0\) cm), with and without correction for the shrinkage factor of 0.80, are summarized in Table 3. The overall accuracy in the entire cohort was high for both T1 and T2 tumors (83% and 82% and 74% and 70% with and without correction for the shrinkage factor, respectively).

**The Prognostic Value of rTT and rDOI for Overall Survival**

The median follow-up was 3.6 years. A trend toward lower OS was observed in thicker tumors within each seventh edition T-category: Three-year OS for T1\(_{\text{rTT}}\)/H\(\geq 5\) mm (\(n = 55\)) versus T1\(_{\text{rTT}}\)/H\(\geq 10\) mm (\(n = 64\)) was 78% versus 92% (\(P = .13\)); T2\(_{\text{rTT}}\)/H\(\geq 10\) mm (\(n = 95\)) versus T2\(_{\text{rTT}}\)/H\(\geq 1.0\) cm (\(n = 66\)) was 67% versus 82% (\(P = .19\)). Only 2 T3 tumors were ≤10 mm, and no deaths occurred. Three-year OS for T3\(_{\text{rTT}}\)/H\(\geq 1.0\) cm (\(n = 53\)) was 49% (\(P = .23\)). If one replaced rTT with rDOI, the results were almost identical: T1\(_{\text{rDOI}}\)/H\(\geq 5\) mm (\(n = 49\)) versus T1\(_{\text{rDOI}}\)/H\(\geq 10\) mm (\(n = 70\)) was 78% versus 91% (\(P = .16\)); T2\(_{\text{rDOI}}\)/H\(\geq 10\) mm (\(n = 92\)) versus T2\(_{\text{rDOI}}\)/H\(\geq 1.0\) cm (\(n = 69\)) was 66% versus 83% (\(P = .14\)). Only 3 T3 tumors had rDOI ≤10 mm, and no deaths occurred. Three-year OS for T3\(_{\text{rDOI}}\)/H\(\geq 10\) mm (\(n = 52\)) was 50% (\(P = .23\)). Multivariable analysis confirmed, similar to pTT (HR 1.38, \(P = .01\)), that both rTT and rDOI were prognostic for OS with HRs of 1.50 (1.06–

---

### Table 1: Clinical characteristics of 335 patients

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Entire Cohort</th>
<th>Oral Tongue</th>
<th>Other Subsite(^a)</th>
<th>(P) Value(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total case No.</td>
<td>335</td>
<td>189 (56%)</td>
<td>146 (44%)</td>
<td></td>
</tr>
<tr>
<td>Age (median) (range) (yr)</td>
<td>62 (22–96)</td>
<td>60 (22–96)</td>
<td>64 (28–96.4)</td>
<td>.004(^c)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td>.260</td>
</tr>
<tr>
<td>Female</td>
<td>129 (39%)</td>
<td>78 (41%)</td>
<td>51 (35%)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>206 (61%)</td>
<td>111 (59%)</td>
<td>95 (65%)</td>
<td></td>
</tr>
<tr>
<td>Smoking PY (median) (range)</td>
<td>10 (0–100)</td>
<td>5 (0–100)</td>
<td>20 (0–86)</td>
<td>&lt;.001(^c)</td>
</tr>
<tr>
<td>pT Category (7th edition)</td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001(^c)</td>
</tr>
<tr>
<td>T1</td>
<td>119 (36%)</td>
<td>47 (25%)</td>
<td>72 (49%)</td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>161 (48%)</td>
<td>101 (53%)</td>
<td>60 (41%)</td>
<td></td>
</tr>
<tr>
<td>T3</td>
<td>55 (16%)</td>
<td>41 (22%)</td>
<td>14 (10%)</td>
<td></td>
</tr>
<tr>
<td>pN Category (7th edition)</td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001(^c)</td>
</tr>
<tr>
<td>N0</td>
<td>181 (54%)</td>
<td>83 (44%)</td>
<td>98 (67%)</td>
<td></td>
</tr>
<tr>
<td>pN+</td>
<td>154 (46%)</td>
<td>106 (56%)</td>
<td>48 (33%)</td>
<td></td>
</tr>
<tr>
<td>pTT (median) (range) (cm)</td>
<td>0.9 (0.1–4.0)</td>
<td>1.1 (0.1–4.0)</td>
<td>0.6 (0.1–3.9)</td>
<td>&lt;.001(^c)</td>
</tr>
<tr>
<td>rTT (median) (range) (cm)</td>
<td>1.0 (1.3–3.4)</td>
<td>1.2 (1.3–3.0)</td>
<td>0.6 (1.3–3.4)</td>
<td>&lt;.001(^c)</td>
</tr>
<tr>
<td>rTT on CT (median) (range) (cm)</td>
<td>1.0 (1.3–3.4)</td>
<td>1.2 (1.3–3.0)</td>
<td>0.3 (1.3–3.4)</td>
<td>&lt;.001(^c)</td>
</tr>
<tr>
<td>rTT on MR (median) (range) (cm)</td>
<td>1.2 (1.3–3.0)</td>
<td>1.3 (1.3–3.0)</td>
<td>0.6 (1.3–3.0)</td>
<td>&lt;.001(^c)</td>
</tr>
<tr>
<td>rDOI on CT (median) (range) (cm)</td>
<td>0.9 (1.0–3.3)</td>
<td>1.3 (1.0–3.0)</td>
<td>0.3 (1.3–3.0)</td>
<td>&lt;.001(^c)</td>
</tr>
<tr>
<td>rDOI on MR (median) (range) (cm)</td>
<td>1.1 (1.3–3.0)</td>
<td>1.3 (1.3–3.0)</td>
<td>0.3 (1.3–3.0)</td>
<td>&lt;.001(^c)</td>
</tr>
</tbody>
</table>

Note: PY indicates pack-year; pN\(^+\), pathologic-positive lymph nodes.

\(^a\) Other oral cavity subsite included the following: \(n = 75\), floor of mouth; \(n = 37\), buccal mucosa; \(n = 16\), lower alveolar and gingiva; \(n = 8\), retromolar trigone; \(n = 6\), upper alveolar and gingiva; \(n = 4\), hard palate.

\(^b\) \(P\) value was for comparison between oral tongue and other oral cavity subsites.

\(^c\) Significant.
Table 3: Diagnostic accuracy of rTT as the seventh edition T-category modifier

<table>
<thead>
<tr>
<th>Variable</th>
<th>No Shrinkage Factor</th>
<th>With Shrinkage Factor 0.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>rTT identifying 7th edition T1 tumor with pTT &gt; 0.5 cm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total case No.</td>
<td>119</td>
<td>119</td>
</tr>
<tr>
<td>True positive</td>
<td>36</td>
<td>34</td>
</tr>
<tr>
<td>True negative</td>
<td>59</td>
<td>65</td>
</tr>
<tr>
<td>False positive</td>
<td>17</td>
<td>11</td>
</tr>
<tr>
<td>False negative</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>88% (75–96)</td>
<td>79% (64–90)</td>
</tr>
<tr>
<td>Specificity</td>
<td>78% (67–86)</td>
<td>86% (76–93)</td>
</tr>
<tr>
<td>Positive predictive value</td>
<td>69% (55–81)</td>
<td>76% (60–87)</td>
</tr>
<tr>
<td>Negative predictive value</td>
<td>92% (83–97)</td>
<td>88% (78–94)</td>
</tr>
<tr>
<td>Accuracy</td>
<td>82%</td>
<td>83%</td>
</tr>
</tbody>
</table>

rTT identifying 7th edition T2 tumor with pTT > 1.0 cm

Total case No. | 161 | 161 |
True positive | 64 | 54 |
True negative | 49 | 65 |
False positive | 31 | 15 |
False negative | 17 | 27 |
Sensitivity | 79% (69–87) | 67% (55–77) |
Specificity | 61% (50–72) | 81% (71–89) |
Positive predictive value | 67% (57–77) | 78% (67–87) |
Negative predictive value | 74% (62–84) | 71% (60–80) |
Accuracy | 70% | 74% |

2.12) (P = .021) and 1.77 (1.22–2.56) (P = .003), respectively (Fig 3). Finally, when the T-category was reclassified to the eighth edition T-category using either rTT or rDOI as T-modifiers (adjusted for age, N-category, and ECOG performance status), an incremental HR with higher eighth edition T-category was apparent: HRs for eighth edition T2, T3, and T4 compared with T1 = 1.88 (0.76–4.68), 2.86 (1.20–6.84), and 4.65 (1.86–11.6) based on rTT (P < .001) and 1.67 (0.72–3.88), 2.88 (1.26–6.58), and 4.2 (1.77–9.95) based on rDOI, respectively (P < .001).

DISCUSSION

This large cohort study shows a high rTT-pTT correlation. By means of pTT as a reference, the diagnostic accuracy of rTT for upstaging seventh edition T1 and T2 tumors is good (>70%). An excellent intrarater and interrater reliability of measuring rTT confirms the reliability of recording this parameter in clinical practice. In addition, most OSCCs are flat tumors, in which the rTT and rDOI are identical. Both rTT and rDOI confer independent prognostic significance in addition to a size-based T-category, supporting inclusion of either parameter in the eighth edition TNM classification. Suboptimal rTT-pTT correlation when imaging is performed ≥8 weeks before an operation suggests that repeat staging imaging might be required to accurately depict tumor extent when a protracted interval to treatment occurs. While MR imaging–based rTT seems to have a slightly higher correlation with pTT compared with CT-based rTT, the difference was nonsignificant, permitting the combining of CT and MR imaging measurements to ensure a sufficiently large sample size.

The high rTT-pTT correlation (0.78) in this study is consistent with others.11-15,19-21 The correlation was similar for both oral tongue and other OSCC subsites. We found that pTT is generally thinner than rTT, potentially attributable to tumor shrinkage after formaldehyde fixation, like findings in other studies. Most interesting, the shrinkage factor was smaller for oral tongue compared with other subsite tumors (0.91 versus 0.70). This was also observed by Lwin et al,13 who reported shrinkage factors of 0.87, 0.65, and 0.59 for oral tongue, floor of mouth, and others, respectively. We hypothesized that the tongue, an organ with more free margins, has less propensity to shrink than tumors that are more deeply embedded in surrounding tissues.

Because the eighth edition TNM includes DOI for the clinical T-classification, confirming its reliability and prognostication clinically and radiologically is important because not all patients undergo an operation. Clinicians need to use both clinical assessment and imaging to best determine the clinical T-classification for this population. A practical challenge in assessing rDOI is the starting point of the “plumb line.” Pathologic assessment used the adjacent mucosal basement membrane, which is invisible on imaging because the thickness of the oral mucosal epithelium is <0.5 mm,11 representing a negligible difference between the potential originating points of measurement (mucosal surface versus basement membrane). Correspondingly, for practical reasons, we proposed that imaging could use an interpreted mucosal plane across the “surface” of the adjacent normal mucosa for rDOI measurement.

Our study confirmed that both rDOI and rTT are independently associated with inferior OS in addition to seventh edition T-category. When one applies rTT and rDOI to modify seventh edition T1 and T2, the separation in OS is evident (though nonsignificant due to an insufficient sample size). Nonetheless, the trend supports consideration of either for modification of a previously size-based T-classification. Furthermore, the multivariable analysis confirmed that both the rTT- and rDOI-based eighth edition T-categories demonstrate a clear distinction in HRs between each T-category, an essential requirement for staging. Thus, rTT can be a surrogate if rDOI is unavailable. The similar prognostic performance of rTT and rDOI echoes pathology-based findings. Dirven et al16 compared pTT and pDOI in 927 patients with OSCC and found that 79% of cases had a <1-mm difference between both parameters and prognostic performance, like those of T-category modifiers, and suggested that TT can be used as a surrogate in retrospective studies for eighth edition TNM classification.

Study limitations include its retrospective nature and unavailability of pDOI. pTT was obtained prospectively from synoptic reports and by convention, measured from the tumor surface to the deepest point of invasion. pDOI was unavailable as the reference for rDOI. However, it was difficult to recognize whether a tumor had an exophytic or ulcerative component on imaging, and in most, rTT and rDOI were similar. Because spatial resolution remains a disadvantage on imaging compared with pathology, very thin tumors that were not reliably measurable on imaging were coded as rTT <1 mm for this study. Because the smallest cutoff for rDOI in the eighth edition T-category is 5 mm, this arbitrary coding is not expected to affect the reclassification of the T-category.

Although most CT scans were obtained at our institution using 2-mm slice thickness, rarely (<1% cases) did preoperative CT from referring institutions use a 5-mm slice thickness. On careful
review of the data, our statisticians (W.X., L.L.) determined that this difference would not significantly influence our results.

CONCLUSIONS

rTT measurement assessed by either CT or MR imaging is an acceptable representation for pTT in OSCC. rTT can upstage the seventh edition size-based T-category to the eighth-edition T-category with good accuracy. Both rTT and rDOI are independent survival predictors and can stratify risk of death in addition to traditional tumor size. Similar to pTT for pDOI, rTT can be a surrogate for rDOI. Finally, we propose using the interpreted mucosal plane (ie, a plane crossing an adjacent normal mucosal surface) to measure rDOI.

ACKNOWLEDGMENTS

The author (B.O.) would like to acknowledge the O. Harold Warwick Prize of the Canadian Cancer Society for supporting his academic activities.

REFERENCES


FIG 3. Prognostic value of rTT and rDOI for overall survival. AHR indicates adjusted hazard ratio, adjusted for age, seventh edition T-category, N-category, and Eastern Cooperative Oncology Group performance status.


MR Imaging of the Facial Nerve through the Temporal Bone at 3T with a Noncontrast Ultrashort Echo Time Sequence

J.P. Guenette, R.T. Seethamraju, J. Jayender, C.E. Corrales, and T.C. Lee

ABSTRACT

SUMMARY: The pointwise encoding time reduction with radial acquisition (PETRA) ultrashort echo time MR imaging sequence at 3T enables visualization of the facial nerve from the brain stem, through the temporal bone, to the stylomastoid foramen without intravenous contrast. Use of the PETRA sequence, or other ultrashort echo time sequences, should be considered in the MR imaging evaluation of certain skull base tumors and perhaps other facial nerve and temporal bone pathologies.

ABBREVIATIONS: CNR = contrast-to-noise ratio; IR-FSPGR = inversion recovery-prepared fast spoiled gradient recalled-echo; PETRA = pointwise encoding time reduction with radial acquisition; UTE = ultrashort echo time

The cisternal, canalicular, labyrinthine, geniculate, tympanic, and mastoid segments of the facial nerve are either not detectable or only faintly visible on noncontrast T1-weighted MR images. Moreover, although contrast enhancement of the surrounding venous plexus can aid in facial nerve visualization, the extent of enhancement is variable and unreliable. We hypothesized that visibility of the facial nerve could be improved with ultrashort echo time (UTE) imaging. UTE sequences capture signal from rapidly decaying short-T2 tissue, such as cortical bone and middle ear ossicles and, therefore, should capture signal from the petrous and mastoid portions of the temporal bone. Moreover, peripheral nerves have detectable signal in the ultrashort T2 spectrum, so UTE imaging should provide visualization of the facial nerve. Finally, UTE imaging minimizes air-related susceptibility artifacts and would therefore minimize artifacts from the middle ear cavity and mastoid air cells that may contribute to nonvisualization of the facial nerve on spin-echo and gradient-echo sequences. The pointwise encoding time reduction with radial acquisition (PETRA) UTE sequence provides more consistent image quality over a wider range of conditions than similar sequences. This report describes the evaluation of the facial nerve segments through the temporal bone using both the default PETRA protocol and a shortened 4-minute PETRA protocol.

MATERIALS AND METHODS
As part of a prospective study designed to evaluate the extracranial facial nerve, an MR imaging examination of the brain and face that included a PETRA sequence was performed in 8 healthy subjects (6 men, 2 women; 31 ± 8 years of age) between January and March 2018. All subjects provided informed consent. The study was approved by our institutional review board and performed in compliance with the Health Insurance Portability and Accountability Act.

Imaging Protocols
All imaging was performed on a Magnetom Prisma 3T MR imaging system with a 64-channel head/neck coil (Siemens, Erlangen, Germany). The default PETRA protocol as installed by Siemens (Table 1) was performed on 7 subjects and 14 facial nerves. Due to the 6-minute duration of the protocol, parameter modifications were tested to reduce imaging time while retaining the signal-to-noise ratio and contrast-to-noise ratio (CNR) of the facial nerve and surrounding structures. A modified 4-minute protocol (Table 1) was performed on 5 subjects and 10 facial nerves. Imaging with both PETRA protocols was performed on 4 subjects.

Imaging Analysis
Two independent reviewers scored the visibility of the cisternal, canalicular, labyrinthine, geniculate, tympanic, and mastoid seg-
ments of the facial nerve and the greater superficial petrosal nerve on each side of each subject on all series of default PETRA and 4-minute PETRA images. One reviewer was a neuroradiologist with 10 years’ attending-level experience, and 1 reviewer was a postgraduate year-5 radiology resident completing a year-long neuroradiology mini-fellowship. To allow indirect comparison with a prior study that evaluated these facial nerve segments using a 3D inversion recovery fast spoiled gradient recalled (IR-FSPGR) sequence, we scored the signal intensity as follows: 0, no signal; 1, faint visualization; 2, signal equivalent to that of the normal cerebellum. The scored data from both reviewers were summed and averaged for each facial nerve segment for both the default PETRA and 4-minute PETRA images. Using sample size, means, and SDs, we calculated independent-sample t tests to compare the signal intensity of the facial nerve on the 4-minute PETRA protocol images, as scored for this study, with the signal intensity of the facial nerve on the IR-FSPGR images, as scored in the study by Dehkharshani et al.1

To determine the SNR and CNR, in all 4 subjects in whom both protocols were performed, we drew ROIs on a single slice over the following structures of the default and 4-minute PETRA images: facial nerve tympanic segment, ossicle, middle ear cavity, mastoid, brain stem, and extracorporeal background region lateral to the temporal bone. The SNR and CNR were calculated as follows: \( \text{SNR} = \frac{\text{Mean Signal Intensity of the Region}}{\text{SD of Background Noise}}; \text{CNR} = \frac{\text{SNR Region 1}}{\text{SNR Region 2}}. \)

RESULTS

All evaluated segments of the facial nerve were visible by both reviewers (Figure) with a score of either 1 or 2 with 90.5% overall reviewer agreement. The signal intensity of all facial nerve segments was significantly greater for the 4-minute PETRA protocol images compared the IR-FSPGR images (Table 2).

SNR and CNR values were consistently similar or higher with the 4-minute PETRA protocol compared with the default PETRA protocol (Tables 3 and 4).

DISCUSSION

This study demonstrates that noncontrast UTE imaging with the PETRA sequence can provide visualization of the facial nerve from the brain stem to the stylomastoid foramen. Such visualization is not possible with more routine sequences currently used in skull base or internal auditory canal imaging.1,2 The PETRA sequence is a proved alternative to an IR-FSPGR sequence in pediatric brain imaging8 and could perhaps similarly substitute for IR-FSPGR sequences in skull base imaging.

It is known that ultrashort T2 (<1 ms) MR signal can be detected in peripheral nerves and imaged. This signal originates from protons in myelin phospholipids5 and likely accounts, at least in part, for the consistent visibility of the facial nerve with the PETRA sequence. In addition, it is likely that artifacts related to the air/tissue interfaces in the temporal bone interfere with detection of signal from the facial nerve on more routine sequences. Minimization of air-related susceptibility artifacts with the

<table>
<thead>
<tr>
<th>Table 1: Parameters for default and 4-minute PETRA sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Default</strong></td>
</tr>
<tr>
<td>TR 1</td>
</tr>
<tr>
<td>TR 2</td>
</tr>
<tr>
<td>TE</td>
</tr>
<tr>
<td>TI 1</td>
</tr>
<tr>
<td>TI 2</td>
</tr>
<tr>
<td>Averages</td>
</tr>
<tr>
<td>Slice thickness</td>
</tr>
<tr>
<td>FOV</td>
</tr>
<tr>
<td>Matrix</td>
</tr>
<tr>
<td>Voxel size</td>
</tr>
<tr>
<td>Bandwidth</td>
</tr>
<tr>
<td>Flip angle</td>
</tr>
<tr>
<td>Radial views</td>
</tr>
<tr>
<td>Acquisition time</td>
</tr>
</tbody>
</table>

FIGURE. Normal right facial nerve in a 27-year-old healthy female volunteer. Axial oblique (A and B) and sagittal oblique (C) noncontrast MR images obtained with the 4-minute PETRA protocol show the cisternal, canalicular, labyrinthine, geniculate, tympanic, and mastoid segments of the right facial nerve (white arrows) and a portion of an ossicle (black arrow). D. Sagittal oblique image obtained with the default PETRA protocol shows a slightly sharper but similar appearance to C.
UTE sequences\(^6\) may reduce such interference and may also contribute to the consistent visibility of the facial nerve with the PETRA sequence. Comparison of temporal bone imaging with multiple different UTE sequences could help elucidate the biophysical properties that allow imaging of the facial nerve through the temporal bone and help optimize the technique.

The default PETRA protocol runs nearly 6 minutes. We shortened the protocol to 4 minutes by reducing the radial acquisitions from 60,000 to 40,000. To recover SNR and CNR, we increased the first TI, which is used in the pointwise acquisition of central k-space data.\(^6,8\) Our increase of the TI from 1300 to 2000 ms would be expected to increase signal in the brain and nerve tissues at some expense to gray and white matter contrast. These parameter changes resulted in mild signal artifacts in the middle ear cavity, demonstrated by higher middle ear cavity SNR. The major limitation of this study is that imaging was performed with a single MR imaging system on a small sample of healthy subjects and does not prove reproducibility on other MR imaging systems or in a clinical setting. In addition, the comparison of signal intensities between PETRA and IR-FSPGR sequences was made across studies performed by different research groups with different subjects and equipment.

**CONCLUSIONS**

A 3T MR imaging 4-minute noncontrast PETRA protocol enables visualization of the facial nerve from the brain stem, through the temporal bone, to the stylomastoid foramen. Use of the PETRA sequence, or other UTE sequences, should be considered in the MR imaging evaluation of certain skull base tumors and perhaps other facial nerve and temporal bone pathologies.

Disclosures: Jeffrey P. Guenette—RELATED: Grant: American Society of Head and Neck Radiology.\(^*\) Ravi Teja Seethamraju—UNRELATED: Employment: Siemens Medical Solutions USA, Inc.; Stock/Stock Options: Siemens Medical Solutions USA, Inc. Jayender Jagadeesan—RELATED: Grant: National Institutes of Health, Comments: through the National Institute of Biomedical Imaging and Bioengineering grant number P41EB015898*; UNRELATED: Board Membership—Nuclear Medicine; Consultancy: Navigation Sciences; Grants/Grants Pending: Siemens Research Grant*; Patents (Planned, Pending or Issued): system for localizing deformable tumors*; Stock/ Stock Options: Navigation Sciences. Thomas C. Lee—RELATED: Grant: American Society of Head and Neck Radiology. Comments: This study was supported by the American Society of Head and Neck Radiology through the 2017 William N. Hanafee Research Grant*. *Money paid to the institution.

**REFERENCES**

5. Hortch RA1, Gore JC, Does MD. Origins of the ultrashort-T2 1H NMR

---

**Table 2: Signal intensity comparison of PETRA and IR-FSPGR sequences as previously reported\(^a\)**

<table>
<thead>
<tr>
<th></th>
<th>Default PETRA</th>
<th>4-Minute PETRA</th>
<th>IR-FSPGR (from Dehkharghani et al)</th>
<th>P Value (4-Minute PETRA vs IR-FSPGR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cisternal</td>
<td>1.89 (0.31)</td>
<td>2.00 (0.00)</td>
<td>0.86 (0.42)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Canalicul</td>
<td>1.86 (0.36)</td>
<td>2.00 (0.00)</td>
<td>0.83 (0.44)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Labyrinthine</td>
<td>1.79 (0.42)</td>
<td>1.85 (0.37)</td>
<td>0.88 (0.39)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Geniculate</td>
<td>2.00 (0.00)</td>
<td>2.00 (0.00)</td>
<td>1.03 (0.22)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Tympnic</td>
<td>1.96 (0.19)</td>
<td>2.00 (0.00)</td>
<td>0.95 (0.32)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Superficial Petrosal</td>
<td>2.00 (0.00)</td>
<td>2.00 (0.00)</td>
<td>0.96 (0.30)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Mastoid</td>
<td>2.00 (0.00)</td>
<td>2.00 (0.00)</td>
<td>1.01 (0.30)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

\(^a\) To allow direct comparison with the data published by Dehkharghani et al,\(^1\) except for the P values, we report data as mean (SD) of assigned signal intensity, in which each facial nerve segment was assigned a value of 0–2 (0, no detectable signal; 1, faint visualization; 2, signal equivalent to normal brain).

**Table 3: Signal-to-noise ratio for default and 4-minute PETRA sequences**

<table>
<thead>
<tr>
<th></th>
<th>Subject C</th>
<th>Subject D</th>
<th>Subject F</th>
<th>Subject G</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default PETRA</td>
<td>4-Minute PETRA</td>
<td>Default PETRA</td>
<td>4-Minute PETRA</td>
<td>Default PETRA</td>
</tr>
<tr>
<td>Facial nerve</td>
<td>125</td>
<td>165</td>
<td>113</td>
<td>159</td>
</tr>
<tr>
<td>Ossicle</td>
<td>58</td>
<td>116</td>
<td>65</td>
<td>114</td>
</tr>
<tr>
<td>Middle ear</td>
<td>5</td>
<td>51</td>
<td>31</td>
<td>75</td>
</tr>
<tr>
<td>Mastoid</td>
<td>36</td>
<td>55</td>
<td>62</td>
<td>106</td>
</tr>
<tr>
<td>Brain stem</td>
<td>168</td>
<td>281</td>
<td>155</td>
<td>290</td>
</tr>
</tbody>
</table>

**Table 4: Contrast-to-noise ratio for default and 4-minute PETRA sequences**

<table>
<thead>
<tr>
<th></th>
<th>Subject C</th>
<th>Subject D</th>
<th>Subject F</th>
<th>Subject G</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default PETRA</td>
<td>4-Minute PETRA</td>
<td>Default PETRA</td>
<td>4-Minute PETRA</td>
<td>Default PETRA</td>
</tr>
<tr>
<td>Facial nerve/mastoid</td>
<td>89</td>
<td>111</td>
<td>51</td>
<td>43</td>
</tr>
<tr>
<td>Facial nerve/middle ear</td>
<td>120</td>
<td>114</td>
<td>82</td>
<td>67</td>
</tr>
<tr>
<td>Ossicle/mastoid</td>
<td>22</td>
<td>62</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Ossicle/middle ear</td>
<td>53</td>
<td>65</td>
<td>34</td>
<td>31</td>
</tr>
</tbody>
</table>
ABSTRACT

BACKGROUND AND PURPOSE: Leksell stereotactic radiosurgery is an effective option for patients with vestibular schwannomas. Some centers use a combination of stereotactic CT fused with stereotactic MR imaging to achieve an optimal target definition as well as minimize the radiation dose delivered to adjacent structures that correlate with hearing outcomes. The present prospective study was designed to determine whether there is cochlear dose variability between MR imaging and CT.

MATERIALS AND METHODS: Fifty consecutive patients underwent stereotactic radiosurgery for vestibular schwannomas. Dose-planning was performed using high-definition fused stereotactic MR imaging and stereotactic CT images. The 3D cochlear volume was determined by delineating the cochlea on both CT and T2-weighted MR imaging. The mean radiation dose, maximum dose, and 3- and 4.20-Gy cochlear volumes were identified using standard Leksell Gamma Knife software.

RESULTS: The median mean radiation dose delivered to the cochlea was 3.50 Gy (range, 1.20–6.80 Gy) on CT and 3.40 Gy (range, 1–6.70 Gy) on MR imaging (concordance correlation coefficient $r = 0.86$, $r^2 = 0.9, P \leq .001$). The median maximum dose delivered to the cochlea was 6.7 Gy on CT and 6.6 Gy on MR imaging (concordance correlation coefficient $r = 0.89$, $r^2 = 0.90, P \leq .001$). Dose-volume histograms generated from CT and MR imaging demonstrated a strong level of correlation in estimating the 3- and 4.20-Gy volumes (concordance correlation coefficient $r = 0.81$, $r^2 = 0.82, P \leq .001$ and concordance correlation coefficient $r = 0.87$, $r^2 = 0.89, P \leq .001$).

CONCLUSIONS: Both MR imaging and CT provide similar cochlear dose parameters. Despite the reported superiority of CT in identifying bony structures, high-definition MR imaging alone is sufficient to identify the radiation doses delivered to the cochlea.

ABBREVIATIONS: CCC = concordance correlation coefficient; SRS = stereotactic radiosurgery; VS = vestibular schwannomas

Vestibular schwannomas (VS), also known as acoustic neuromas, are benign tumors that most commonly arise from the vestibular portion of cranial nerve VIII, the vestibulocochlear nerve. The most common presenting symptoms of VS are hearing loss, tinnitus, and imbalance. Depending on the presentation, the options currently available for the management of this tumor include observation with serial imaging, surgical resection, and radiosurgery. Leksell gamma knife stereotactic radiosurgery (SRS) (Elekta Instruments, Stockholm, Sweden) is a widely accepted treatment technique for VS. SRS involves delivering highly focused radiation to the 3D tumor volume in a single session, with rapid radiation fall-off in the structures surrounding the tumor target. The aim of SRS is tumor control with minimal collateral damage to adjacent cranial nerve and brain stem structures. Multiple reports demonstrate successful long-term SRS outcomes.

Factors such as patient age at the time of SRS, hearing status before the procedure, tumor size, the interval between diagnosis and treatment, and cochlear radiation dose have been found to influence hearing preservation rates. Cochlear dose remains the only variable that can be modified during treatment planning to improve hearing-preservation rates.

The use of MR imaging for dose-planning in SRS for VS has been shown to be safe and efficient. However, the concern that MR imaging accuracy may be affected by magnetic susceptibility issues has led many centers to use both stereotactic CT and stereotactic MR imaging to achieve a more accurate target as well as cochlear definition. Whether bone window CT provides superior resolution of the cochlea compared with high-definition T2 MR imaging is not known.

Received December 7, 2017; accepted after revision July 1, 2018.
From the Department of Neurological Surgery and Radiation Oncology, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania.
Please address correspondence to Hideyuki Kano, MD, PhD, Department of Neurological Surgery, University of Pittsburgh, Suite B-400, UPMC Presbyterian, 200 Lothrop St, Pittsburgh, PA 15213; e-mail: Kanoh@upmc.edu
http://dx.doi.org/10.3174/ajnr.A5808
In this study, we compared the definition of the cochlear volume determined by MR imaging with the cochlear volume defined by CT to detect whether any variance in the radiation dose delivered to the cochlea was detectable.

**MATERIALS AND METHODS**

**Patient Population**

Between May 2016 and October 2017, fifty consecutive patients with VS underwent MR imaging and CT-guided SRS at the University of Pittsburgh Medical Center. There were 26 men and 24 women with a median age of 60 years (range, 28–77 years) at the time of SRS. The most common presentation was unilateral hearing loss (94%). The median duration of symptoms before SRS was 18 months. The median speech discrimination score and pure-tone average at the time of SRS were 64% and 37 dB, respectively. Hearing level was classified as Gardner Robertson grades I and II in 32 patients (64%). SRS was the primary management in 46 patients (92%). Three patients had SRS for residual tumor after surgery, and 1 patient, for recurrent tumor after prior SRS.

**Radiosurgery Technique**

Radiosurgery was performed using the Perfexion or ICON surgery, and 1 patient, for recurrent tumor after prior SRS. The median prescription isodose was 50% (range, 15.5–25 Gy). The median prescription isodose was 25 Gy (range, 30–70 mm³). The median margin dose was 12 Gy (range, 11–12.50 Gy), and the median maximum dose was 24 Gy (range, 15.5–25 Gy). The median prescription isodose was 50% (range, 50%–70%). After the procedure, all patients were given 20–40-mg of IV methylprednisolone. All patients were discharged on the same day. Only 4-mm isocenters were used in the internal auditory canal.

**Statistical Analysis**

Continuous features were summarized using median and range. The Pearson \( r^2 \) and the Lin concordance correlation coefficient (CCC) were used to assess the correlation of volumes and doses between the 2 imaging modalities. The CCC ranged between −1 and 1, with values indicating perfect negative or positive concordance, respectively. Statistical analysis was performed with SPSS Statistics 24 (IBM, Armonk, New York). A \( P \) value < .05 was set for statistical significance.

**RESULTS**

**Cochlear Dose Parameters**

The median cochlear volume identified by CT was 36.8 mm³ (range, 4.90–77 mm³). The median cochlear volume as identified on T2-weighted MR imaging was 41 mm³ (range, 4.90–70 mm³). On the basis of CCC analysis, we found a poor correlation in cochlear volume between CT and MR imaging (CCC = 0.16, \( r^2 = 0.07 \), \( P = .245 \)) (Fig 1).

The median of the average radiation dose delivered to the cochlea was 3.50 Gy (range, 1.20–6.80 Gy) on CT and 3.40 Gy (range, 1–6.70 Gy) on MR imaging (CCC = 0.86, \( r^2 = 0.90, P \leq .001 \)) (Fig 2). The median maximum dose delivered to the cochlea was 6.70 Gy on CT and 6.60 Gy on MR imaging (CCC = 0.89, \( r^2 = 0.90, P \leq .001 \)) (Fig 3). Based on CCC analysis, an almost perfect correlation was observed between CT and MR imaging in estimating the mean and maximum doses of radiation delivered to the cochlea.

The median cochlear volume receiving a 3-Gy radiation dose as identified on CT was 25 mm³ (range, 0–47 mm³), and on MR imaging, it was 25 mm³ (range, 0–55.40 mm³). Based on CCC analysis, an almost perfect level of correlation in estimating the cochlear volume receiving a 3-Gy radiation dose (CCC = 0.81, \( r^2 = 0.82, P \leq .001 \)).

The median cochlear volume receiving a 4.2-Gy radiation dose as identified on CT was 12.7 mm³ (range, 0–63 mm³), and on MR imaging, it was 7.60 mm³ (range, 0–46 mm³). Based on CCC analysis, an almost perfect correlation was observed between CT and MR imaging in estimating the cochlear volume receiving a 4.2-Gy radiation dose (CCC = 0.87, \( r^2 = 0.89, P \leq .001 \)).
Prevention of further tumor growth and the preservation of cranial nerve function are the main goals of SRS management of VS. During the past several years, multiple factors have been found to influence hearing-preservation rates. These include the age of the patient, hearing status using scales such as the Gardner Robertson classification, length of the time between diagnosis and treatment, difference in the pure-tone average between the tumor and nontumor ear, and estimated length of the vestibulocochlear nerve. The correlation between cochlear dose and hearing preservation has been studied extensively. Multiple reports have demonstrated that a higher radiation dose to the cochlea is associated with a higher chance of hearing decline during long-term follow-up. In a review of the outcomes of 53 patients who underwent SRS for the management of vestibular schwannomas, Brown et al found that patient age and the percentage of the cochlea that receives >5.30 Gy are the main predictors of hearing preservation. Kano et al found that patients who received an average cochlear dose of ≤4.20 Gy to the center of the cochlea had better hearing preservation rates.

Hearing preservation is an important consideration in patients undergoing management. Yang et al performed a literature review looking at the hearing-preservation rates in patients who underwent SRS for the management of VS. In their review, among 4234 patients with VS treated with SRS, there was a 51% chance of hearing preservation at pre–gamma knife levels at a mean follow-up of 3 years. Yomo et al retrospectively reviewed the outcomes of 154 patients who underwent SRS for the management of vestibular schwannomas. After a mean audiologic follow-up of 52 months post-SRS, a maximum cochlear dose of <4 Gy was found to be the sole prognostic factor for hearing preservation. Hasegawa et al reported hearing outcomes in 92 patients who underwent SRS for the management of vestibular schwannomas. After a mean audiologic follow-up of 52 months post-SRS, a maximum cochlear dose of <4 Gy was found to be the sole prognostic factor for hearing preservation. Hasegawa et al reported hearing outcomes in 92 patients who underwent SRS for the management of vestibular schwannomas. After a mean audiologic follow-up of 52 months post-SRS, a maximum cochlear dose of <4 Gy was found to be the sole prognostic factor for hearing preservation.

Both MR imaging and CT are commonly used in treatment planning. However, it is widely accepted that MR imaging is superior when it comes to identifying soft-tissue structures, whereas CT provides superior resolution of bony structures. Using CCC analysis, we found a poor correlation in cochlear volume identification between CT and MR imaging. This variance may also relate to the windowing level used to define the cochlea using both CT and MR imaging. Kulkarni et al performed a comparison of gross target volumes as delineated independently on contrast-enhanced CT and T1- and T2-weighted MR imaging in vestibular schwannomas. In their analysis, they found that cochlear volume as identified on T2 images was significantly larger (23.9 mm³) than the cochlear volume identified on CT (15 mm³). Jacob et al compared CT- and MR imaging–based modiolus point dose measurements and found a moderate level of correlation between CT and MR imaging in identifying the dose point of the cochlear modiolus. Treatment planning is performed after the fusion of CT and MR images. This may explain how, despite the significant difference in cochlear volumes between the 2 modalities, the dose delivered to the cochlea was found to be the same.

The importance of accurately identifying the mean and maximum doses received by the cochlea is related to the fact that multiple reports observed that the radiation dose of 3–5.3 Gy was...
and MR imaging in the same patient are not different indicates that we can eliminate the need for additional radiation delivered by spiral CT in patients with VS undergoing Leksell gamma knife SRS.

**Weaknesses of the Present Study**

The overall findings of this report may be limited by the small number of patients included in the analysis. In this study, both the responsible surgeon and the radiation oncologist independently found that the cochlear volume drawn by each had no impact on the dose delivered to the cochlea. Future studies may allow a more robust analysis of variance. Windowing of the cochlear volume by CT may influence the tracing of the cochlear volume in each axial slice. Variation in the slice thickness of CT (1.25 mm) and T2 volume MR imaging (1 mm) may also affect these results.

**CONCLUSIONS**

In this study, both CT and MR imaging provided similar cochlear dose parameters during Leksell radiosurgery for VS. Despite the differences in cochlear volume identified by either CT or MR imaging, the dose delivered to the cochlea in this study was not different. In patients undergoing Leksell gamma knife SRS for VS, high-definition MR imaging alone provides superior 3D tumor volume definition and a satisfactory depiction of cochlear volume.

Disclosures: Hideyuki Kano—UNRELATED: Comments: supported by Elekta AB. L. Dade Lunsford—UNRELATED: Consultancy: Insightec, Data and Safety Monitoring Board; Stock/Stock Options: Elekta AB.

**REFERENCES**

4. Bailo M, Boari N, Franzin A, et al. Gamma knife radiosurgery as primary treatment for large vestibular schwannomas: clinical re-
Multidelay Arterial Spin-Labeling MRI in Neonates and Infants: Cerebral Perfusion Changes during Brain Maturation

H.G. Kim, J.H. Lee, J.W. Choi, M. Han, S.-M. Gho, and Y. Moon

ABSTRACT

BACKGROUND AND PURPOSE: Arterial spin-labeling with multiple postlabeling delays can correct transit times. We tried to evaluate CBF in neonates and infants using multidelay arterial spin-labeling.

MATERIALS AND METHODS: Multidelay arterial spin-labeling was applied to 13 preterm neonates (mean postmenstrual age, 34.9 weeks), 13 term-equivalent-age neonates (mean postmenstrual age, 39.2 weeks), and 6 infants (mean postmenstrual age, 57.8 weeks). Transit time–corrected CBF in the caudate, thalamus, frontal GM, occipital GM, frontal WM, and occipital WM was measured, and relative CBF compared with the whole-brain CBF was calculated. Inter- and intragroup comparisons were performed among the 3 age groups. A correlation and nonlinear regression analysis were performed between postmenstrual age and CBF.

RESULTS: Intergroup comparisons showed significantly higher whole-brain CBF in infants (38.3 mL/100 g/min) compared with preterm (15.5 mL/100 g/min) and term-equivalent-age (18.3 mL/100 g/min) neonates (P < .001). In the intragroup comparison, all 3 groups showed significantly higher relative CBF values in the occipital WM (63.6%–90.3%) compared with the frontal WM (46.3%–73.9%). In term-equivalent-age neonates, the occipital GM (120.8%) had significantly higher relative CBF values than the frontal GM (103.5%). There was a significant negative correlation between postmenstrual age and the relative CBF of the thalamus (r = −0.449, P = .010). There were significant positive relationships between postmenstrual age and the relative CBF of the frontal WM (R² = 0.298, P = .001) and occipital WM (R² = 0.452, P < .001).

CONCLUSIONS: Multidelay arterial spin-labeling with transit time–corrected CBF showed developmental changes and regional differences of CBF in neonates and infants.

ABSTRACT

BACKGROUND AND PURPOSE: Arterial spin-labeling with multiple postlabeling delays can correct transit times. We tried to evaluate CBF in neonates and infants using multidelay arterial spin-labeling.

MATERIALS AND METHODS: Multidelay arterial spin-labeling was applied to 13 preterm neonates (mean postmenstrual age, 34.9 weeks), 13 term-equivalent-age neonates (mean postmenstrual age, 39.2 weeks), and 6 infants (mean postmenstrual age, 57.8 weeks). Transit time–corrected CBF in the caudate, thalamus, frontal GM, occipital GM, frontal WM, and occipital WM was measured, and relative CBF compared with the whole-brain CBF was calculated. Inter- and intragroup comparisons were performed among the 3 age groups. A correlation and nonlinear regression analysis were performed between postmenstrual age and CBF.

RESULTS: Intergroup comparisons showed significantly higher whole-brain CBF in infants (38.3 mL/100 g/min) compared with preterm (15.5 mL/100 g/min) and term-equivalent-age (18.3 mL/100 g/min) neonates (P < .001). In the intragroup comparison, all 3 groups showed significantly higher relative CBF values in the occipital WM (63.6%–90.3%) compared with the frontal WM (46.3%–73.9%). In term-equivalent-age neonates, the occipital GM (120.8%) had significantly higher relative CBF values than the frontal GM (103.5%). There was a significant negative correlation between postmenstrual age and the relative CBF of the thalamus (r = −0.449, P = .010). There were significant positive relationships between postmenstrual age and the relative CBF of the frontal WM (R² = 0.298, P = .001) and occipital WM (R² = 0.452, P < .001).

CONCLUSIONS: Multidelay arterial spin-labeling with transit time–corrected CBF showed developmental changes and regional differences of CBF in neonates and infants.

ABBREVIATIONS: ASL = arterial spin-labeling; F-GM = frontal GM; F-WM = frontal WM; O-GM = occipital GM; O-WM = occipital WM; PLD = postlabeling delay; PMA = postmenstrual age; rCBF = relative CBF; T1b = longitudinal relaxation time of blood; TEA = term-equivalent-age; wbCBF = whole-brain CBF.

Brain maturation is one of the most vital processes occurring during neonatal life, and imaging studies can potentially provide insight into normal brain development. MR imaging can provide high-resolution structural and functional images. As an increasing number of neonates undergo routine brain MR imaging, advanced tools have been applied to observe maturation processes in the neonatal brain. These include DWI,1,2 DTI,3 magnetization transfer imaging,4 and arterial spin-labeling (ASL).5 Among these methods, ASL is a noninvasive method that measures CBF. ASL uses endogenous blood water as a diffusible tracer5 and is increasingly applied in neurologic studies. By using a radiofrequency inversion pulse, arterial blood protons are magnetically labeled in the carotid artery level.5 ASL provides a quantitative measure of regional brain function and can show changes in baseline function associated with aging.6 In neonates, ASL has been applied to see changes in perfusion during brain maturation7 and to see the predictive value of these changes in patients who undergo hypothermia treatment after hypoxic-ischemic injury.8

However, there are some limitations to applying ASL in neonates. First, neonates have different blood and brain tissue T1...
values than older children and adults. Adjusting these values to neonates is essential because inadequate application of ASL will lead to inaccurate calculation of CBF. Second, neonates have decreased cerebral blood perfusion and low flow velocity. In ASL, labeled blood signals are acquired at the cerebral level after enough time has passed to allow the labeled spins to reach the imaging section, and this time lapse is called the postlabeling delay (PLD). If the PLD is not long enough to reflect the brain perfusion of neonates with low flow velocity, the acquired perfusion values will be inaccurately low. Inaccurately low perfusion values will lead to inadequate interpretation of CBF in neonates combined with the inherent problem of a low signal-to-noise ratio in ASL.

ASL with multiple PLD acquisitions (multidelay ASL) has been applied in adults. CBF values calculated from multidelay ASL showed good correlation with the CBF results of PET, dynamic susceptibility contrast perfusion MR imaging, and CT perfusion. Compared with single-PLD ASL imaging, multidelay ASL showed improved CBF quantification in patients with Moyamoya disease, who have a delay between labeling in the feeding arteries and the arrival of the labeled blood in tissue.

The major trade-off of multidelay ASL used to be longer scanning time. However, recent application of Hadamard encoding has made it possible to acquire multiple PLDs in a much shorter time, and a more upgraded version of the multidelay ASL technique (enhanced ASL technique; GE Healthcare, Milwaukee, Wisconsin) has been introduced. To the best of our knowledge, there has been no study that applies multidelay ASL to neonates. Therefore, the purpose of this study was to show cerebral perfusion in neonates and infants using multidelay ASL with optimized T1 values.

MATERIALS AND METHODS

Patients

This retrospective study was approved by the institutional review board of Ajou University Hospital. Brain MRIs obtained between July 2016 and March 2017 were reviewed. In our institution, brain MR imaging is performed in preterm neonates born before the 32nd week of gestation or in neonates with a birth weight of <1500 g. These neonates stay in the neonatal intensive care unit during hospitalization and undergo brain MR imaging before discharge. In addition to the above indications, infants undergo brain MR imaging for clinical suspicion of brain injury. We excluded infants who had abnormalities on brain MR imaging, including hydrocephalus, hemorrhage, or hypoxic-ischemic injury.

Thirty-two patients were included, and we divided the subjects into 3 groups according to age at the time of MR imaging: preterm neonates, term-equivalent-age (TEA) neonates, and infants. All those in the infant group were full-term with a postnatal age of 1 month or older at the time of MR imaging. Therefore, there were 13 preterm neonates (34.9 ± 1.1 weeks’ postmenstrual age [PMA]), 13 TEA neonates (39.2 ± 2.1 weeks’ PMA), and 6 infants (57.8 ± 5.4 weeks’ PMA; 111.0 ± 62.8 days’ postnatal age).

Several patients underwent brain MR imaging for indications that were not routine. One preterm neonate underwent MR imaging to rule out cranial complications of asphyxia. Among TEA neonates, 3 underwent MR imaging for apnea; 2, for macrocephaly; 2, for postresuscitation evaluation; and 1, for seizure. Among infants, 2 underwent MR imaging for suspicion of delayed development; 2, for follow-up of meningitis; 1, for early closure of the anterior fontanelle; and 1, for transient apnea. None of the included subjects showed abnormalities on conventional MR imaging. All patients were sedated for the MR imaging examination with oral chloral hydrate (0.5 mL/kg). An additional single dose of IV midazolam (0.1 mg/kg) was administered to patients who awoke during scanning.

Image Acquisition

Multidelay ASL was included in our routine imaging protocol for infants who underwent brain MR imaging with a 3T scanner (750w; GE Healthcare). The routine brain MR imaging protocol for neonates and infants included 3D T1-weighted spoiled gradient-echo imaging, axial T2WI, DWI, gradient-echo imaging, and multidelay ASL.

We used a multidelay ASL sequence based on Hadamard encoding to obtain 7 perfusion-weighted images with different PLDs and effective labeling durations. For neonates and infants, PLDs were the following: 2.00, 2.22, 2.48, 2.8, 3.15, 3.63, 4.32 seconds and 1.50, 1.72, 1.98, 2.28, 2.65, 3.13, 3.82 seconds, respectively. Effective labeling durations were the following: 0.22, 0.26, 0.30, 0.37, 0.48, 0.68, 1.18 seconds, regardless of age. By means of Hadamard encoding, CBF images from 7 different PLDs could be generated from 8 acquisitions, and each CBF image was an average of 4 images.

The arterial transit time was estimated using the weighted-delay method of Dai et al. On the arterial transit time map, the median arterial transit times of preterm neonates, TEA neonates, and infants were 2.27 seconds (interquartile range, 0.11 seconds), 2.26 seconds (interquartile range, 0.12 seconds), and 1.24 seconds (interquartile range, 0.53 seconds), respectively. Transit time-corrected CBF maps were computed from the cumulative perfusion-weighted image. Other parameters for multidelay ASL imaging were as follows: TR, 6791 ms; TE, 11 ms; FOV, 24 cm; 640 sampling points on 5 spirals (matrix size, 640 × 5); section thickness, 5 mm; number of sections, 20.

Image Analysis

Acquired raw images of multidelay ASL were processed using dedicated software to adjust for the longitudinal relaxation time of blood (T1b) and tissue. This procedure was used because the quantified perfusion of the acquired multidelay ASL was assumed to have the T1b and longitudinal relaxation time of tissue values of adults, which are different from those of neonates. The longitudinal relaxation time of tissue was assumed to be 1.5 seconds. Because neonates show a large individual variation for T1b, we individualized the T1b values for each neonate. The T1b was calculated using the following equation:

\[ \frac{1}{T1b} = a \times \text{Hematocrit} + b, \]

where \( a = 0.50 \text{ seconds}^{-1} \), the specific relaxivity, and \( b = 0.37 \text{ seconds}^{-1} \), the native relaxivity. Hematocrit values were acquired from blood tests performed within a week of MR imaging. The mean values of hematocrit in preterm neonates, TEA neonates, and infants were 31.1% ± 8.0%, 39.0% ± 8.3%, and 33.1% ± 5.1%, respectively. The mean values of T1b in preterm neonates,
TEA neonates, and infants were 1.91 ± 0.14, 1.78 ± 0.13, and 1.87 ± 0.09, respectively.

ROIs were manually drawn by 1 radiologist with 8 years of experience in brain imaging on the PACS at the following brain regions for CBF in mL/100 g/min: the caudate, thalamus, frontal GM (F-GM), occipital GM (O-GM), frontal WM (F-WM), and occipital WM (O-WM) (Fig 1). The size range of the ROIs was 16.6–24.9 mm². The whole-brain CBF (wbCBF) was obtained, and we calculated relative CBF (rCBF) for each brain region by comparing the specific CBF of the region with the wbCBF.

**Results**

Regional CBF and rCBF for each age group are summarized in Table 1. WM (7.3–35.6 mL/100 g/min) showed lower CBF than cortical GM (16.8–49.8 mL/100 g/min) in all age groups. In preterm and TEA neonates, deep GM structures of the caudate and thalamus showed the highest CBF (21.2–32.2 mL/100 g/min) with a rCBF of 136.4%–177.2% for all brain regions. In infants, the highest CBF was observed in the O-GM (49.8 mL/100 g/min) followed by the caudate (42.9 mL/100 g/min) and F-GM (41.8 mL/100 g/min).

**Table 1. Regional CBF in the age groups of preterm neonates, term-equivalent-age neonates, and infants**

<table>
<thead>
<tr>
<th>Brain Region</th>
<th>Preterm (n = 13)</th>
<th>TEA (n = 13)</th>
<th>Infants (n = 6)</th>
<th>P Value</th>
<th>Preterm vs TEA</th>
<th>Preterm vs Infants</th>
<th>Preterm vs Infants</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBF (mL/100 g/min)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>constante</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caudate</td>
<td>212.6 (6.9)</td>
<td>255.6 (6.6)</td>
<td>42.6 (7.1)</td>
<td>&lt;.001</td>
<td>.260</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Thalamus</td>
<td>277.7 (11.4)</td>
<td>322.7 (7.1)</td>
<td>36.2 (7.5)</td>
<td>.168</td>
<td>.437</td>
<td>.166</td>
<td>.658</td>
</tr>
<tr>
<td>F-GM</td>
<td>16.8 (4.1)</td>
<td>18.8 (3.8)</td>
<td>41.8 (12.0)</td>
<td>&lt;.001</td>
<td>.688</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>O-GM</td>
<td>19.1 (6.2)</td>
<td>21.8 (3.6)</td>
<td>49.8 (19.8)</td>
<td>&lt;.001</td>
<td>.755</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>F-WM</td>
<td>7.3 (3.1)</td>
<td>8.4 (1.9)</td>
<td>28.3 (7.5)</td>
<td>&lt;.001</td>
<td>.744</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>O-WM</td>
<td>9.9 (3.5)</td>
<td>11.9 (3.3)</td>
<td>35.6 (4.1)</td>
<td>&lt;.001</td>
<td>.718</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Whole-brain</td>
<td>15.5 (4.2)</td>
<td>18.3 (3.3)</td>
<td>38.3 (9.5)</td>
<td>&lt;.001</td>
<td>.388</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>rCBF (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caudate</td>
<td>136.4 (25.6)</td>
<td>140.8 (28.9)</td>
<td>116.0 (25.3)</td>
<td>.184</td>
<td>.917</td>
<td>.291</td>
<td>.368</td>
</tr>
<tr>
<td>Thalamus</td>
<td>174.1 (33.1)</td>
<td>177.2 (28.1)</td>
<td>96.5 (17.6)</td>
<td>&lt;.001</td>
<td>.960</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>F-GM</td>
<td>110.0 (24.9)</td>
<td>103.5 (15.2)</td>
<td>108.5 (11.6)</td>
<td>.614</td>
<td>.592</td>
<td>.965</td>
<td>.858</td>
</tr>
<tr>
<td>O-GM</td>
<td>122.8 (19.8)</td>
<td>120.8 (17.5)</td>
<td>127.6 (26.1)</td>
<td>.791</td>
<td>.956</td>
<td>.879</td>
<td>.773</td>
</tr>
<tr>
<td>F-WM</td>
<td>46.7 (6.5)</td>
<td>46.3 (7.9)</td>
<td>73.9 (7.1)</td>
<td>&lt;.001</td>
<td>.995</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>O-WM</td>
<td>63.6 (15.4)</td>
<td>65.6 (10.4)</td>
<td>90.3 (13.7)</td>
<td>.001</td>
<td>.980</td>
<td>.001</td>
<td>.001</td>
</tr>
</tbody>
</table>

a Data are mean (standard deviation).

Multiple linear regression models with a significance level of .05 were used to determine the effects of PMA on CBF. The best-fit nonlinear regression model (ie, the quadratic, logarithmic, and linear regression models) was selected among quadratic, logarithmic, and linear regression models. A P < .05 indicated statistical significance. SPSS, Version 25.0 (IBM, Armonk, New York) and the R software package (Version 3.4.3; www.r-project.org) were used for analysis.

**Statistical Analysis**

Continuous variables were first tested for normality using the Kolmogorov-Smirnov test. For the intergroup comparison of CBF for the 3 age groups, normally distributed independent variables were compared using one-way analysis of variance followed by the Tukey post hoc test. For the intragroup comparison of rCBF in the frontal and occipital lobes, the paired t test was used. PMA had a non-normal distribution. Correlations between PMA and CBF were analyzed using the Spearman correlation. For the CBF of WM, the best-fit nonlinear regression model (ie, the lowest Bayesian information criterion) was selected among quadratic, logarithmic, and linear regression models. A P < .05 indicated statistical significance.
Intragroup Comparison

Results of the intragroup comparison are summarized in Table 2 and Fig 2. When the rCBF of the F-WM and O-WM was compared, all 3 age groups showed significantly higher rCBF values in the O-WM compared with the F-WM. In infants, the O-WM showed a rCBF value 1.2 times higher than the rCBF value of the F-WM. In preterm and TEA neonates, the O-WM showed a rCBF value 1.4 times higher than the F-WM. Regarding GM, the O-GM showed a significantly higher rCBF value compared with the F-GM in TEA neonates (120.8% versus 103.5%, \(P = 0.026\)). In preterm neonates and infants, the O-GM (122.8%–127.6%) showed a higher rCBF value than the F-GM (108.5%–111.0%), but there was no statistical significance.

Relationship between Postmenstrual Age and CBF

There was a significant positive correlation between PMA and wbCBF (\(r = 0.732, P < 0.001\)) (Fig 3). For PMA and the rCBF of the thalamus, there was a significant negative correlation between values (\(r = -0.449, P = 0.010\)). There was a significant positive correlation between PMA and the rCBF of the F-WM (\(r = 0.402, P = 0.023\)) and O-WM (\(r = 0.410, P = 0.020\)). Correlation between PMA and the rCBF of the caudate (\(r = -0.152, P = 0.406\)), F-GM (\(r = -0.045, P = 0.806\)), and O-GM (\(r = -0.157, P = 0.392\)) showed no statistical significance. Among nonlinear regression models, the logarithmic regression model was the best fit for the rCBF of the F-WM and PMA with an \(R^2\) value of 0.298 (\(P = 0.001\)).

DISCUSSION

This study showed that multidelay ASL could be applied to neonates and infants by optimizing PLD and T1 values. By multidelay ASL, absolute cerebral perfusion values in neonates and infants were estimated and a significantly higher wbCBF was observed in the older age groups. The rCBF of brain regions significantly differed according to age group. Intragroup analysis of rCBF showed higher perfusion in the occipital lobes compared with the frontal lobes. There was a positive correlation between the 2 variables (\([r_{CBF} \text{of O-WM}] = 24.15 + 1.096 \times [PMA]\)) (Fig 3).

There was a positive relationship between the rCBF of the F-WM and PMA (\([r_{CBF} \text{of F-WM}] = \text{log}_{10}(PMA) - 107.48 - 43.13 \times [PMA]\)). For the rCBF of the O-WM and PMA, the linear regression model was the best fit with an \(R^2\) value of 0.452 (\(P < 0.001\), and there was a positive relationship between the 2 variables (\([r_{CBF} \text{of O-WM}] = \text{log}_{10}(PMA) - 107.48 - 43.13 \times [PMA]\)) (Fig 3).

Table 2: Intragroup comparison between rCBF in the frontal and occipital lobesa

<table>
<thead>
<tr>
<th>Group</th>
<th>F-GM (mean [standard deviation])</th>
<th>O-GM (mean [standard deviation])</th>
<th>P Value</th>
<th>F-WM (mean [standard deviation])</th>
<th>O-WM (mean [standard deviation])</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preterm</td>
<td>111.0 (24.9)</td>
<td>122.8 (19.8)</td>
<td>.730</td>
<td>46.7 (16.5)</td>
<td>63.6 (15.4)</td>
<td>.004</td>
</tr>
<tr>
<td>TEA</td>
<td>103.5 (15.2)</td>
<td>120.8 (17.5)</td>
<td>.026</td>
<td>46.3 (7.9)</td>
<td>65.6 (10.4)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Infants</td>
<td>108.5 (11.6)</td>
<td>127.6 (26.1)</td>
<td>.190</td>
<td>73.9 (7.1)</td>
<td>90.3 (13.7)</td>
<td>.046</td>
</tr>
</tbody>
</table>

aData are mean (standard deviation) [%].

FIG 2. Intragroup comparison of rCBF in the frontal and occipital lobes. A single asterisk indicates \(P < .05\); double asterisks, \(P < .001\).

FIG 3. Relationship between postmenstrual age and CBF. A single asterisk indicates \(P < .05\); double asterisks, \(P < .001\).
Because functional activity and brain maturation can be indirectly measured by the cerebral metabolic rate, numerous attempts have been made to measure CBF in neonates. In the past, PET\textsuperscript{25-28} and xenon-enhanced CT\textsuperscript{29} were used to estimate CBF in neonates. With these methods, a relatively low wbCBF was identified in neonates compared with adults. However, these past studies had inherent limitations due to radiation use and the invasive properties of their methods. The subsequent development and application of ASL in neonates showed results that were in agreement with former PET results.\textsuperscript{30} Still, an accurate estimation of CBF is challenging with ASL in neonates due to their low blood content and restricted water extraction in brain tissue along with physiologic low CBF, which all contribute to a low signal-to-noise ratio.\textsuperscript{31}

The absolute wbCBF found in our study (15.5–18.3 mL/100 g/min) was higher than that in a previous ASL study of neonates (7–12 mL/100 g/min).\textsuperscript{7} We speculate that this difference is due to arterial transit-time correction with multidelay ASL. ASL incorporates a similar fundamental mechanism with PET, but it uses protons in the blood as tracers and is easily affected by arterial transit time. For example, the CBF of WM calculated from ASL was underestimated due to a longer arterial transit time compared with GM.\textsuperscript{31} In a similar manner, in occlusive diseases such as Moyamoya disease, arterial transit time is elongated in the affected hemispheres and results in underestimated CBF.\textsuperscript{32} With multiple PLDs, we could generate an arterial transit time map that was corrected for CBF values. It was possible to overcome delayed transit time in the adult population using this transit time–corrected CBF map made with multidelay ASL.\textsuperscript{33} While we did acquire arterial transit-time maps and showed median (interquartile range) values according to the age groups, analyzing regional and age-dependent arterial transit times was beyond the scope of this study. Still, investigating arterial transit times obtained by multidelay ASL and comparing single-delay and multidelay ASL in the neonatal population would be an interesting topic for future studies.

A prior neonatal ASL study measured a wbCBF of 16–21 mL/100 g/min,\textsuperscript{34} which is relatively higher than the wbCBF of our study. However, the study did not optimize the T1 value for neonates, which could result in underestimated CBF values. Hematocrit levels vary in neonates according to gestational age and continue to change,\textsuperscript{34} with the T1b being influenced by the hematocrit levels.\textsuperscript{35} Without augmentation of the T1 value for neonates, ASL could result in overestimated CBF.\textsuperscript{10,35}

The higher wbCBF values observed in the older age groups and the positive correlation between PMA and wbCBF found in our study were consistent with those in previous studies.\textsuperscript{7,27,36} Unlike in the older age groups, which show a reversed pattern of decreasing CBF according to age,\textsuperscript{4} infants show a rapid increase in CBF according to age. By means of PET, infants with a PMA of 32–60 weeks were reported to have CBF from 5.5 to 18.7 μmol/100 g/min.\textsuperscript{27} This prior study showed a significant positive correlation between wbCBF and PMA as was seen in our results.\textsuperscript{27} In another study using phase-contrast MR imaging, a CBF of 18–30 mL/100 mL/min at birth rapidly increased with age and reached 60 mL/100 mL/min in 1 year.\textsuperscript{36} Compared with the number of studies in older populations, there have been very few studies on the physiologic evolution of CBF within 1 year of birth. Deeper insight into this field could be possible with future studies using noninvasive ASL.

Deep GM showed the highest rCBF values among the brain regions, and there was a negative trend of values with age, which is in line with findings in prior studies.\textsuperscript{7,27,29,33} Prior studies on brain metabolism in neonates showed that resting metabolism is not identical across the brain.\textsuperscript{7} Metabolism is generally low in the cortex at term but increases in the parietal, temporal, and occipital lobes by 3 months. The last region to increase its metabolism is the frontal cortex.\textsuperscript{37} A previous neonatal ASL study showed higher occipital cortical rCBF compared with frontal cortical rCBF in both preterm (68% versus 56%) and TEA neonates (92% versus 74%), though the study did not evaluate statistical significance.\textsuperscript{7} Still, 1 study using PET showed comparable occipital and frontal cortical CBF values in preterm (5.8 versus 6.5 μmol/100 g/min) and term neonates (8.7 versus 7.6 μmol/100 g/min).\textsuperscript{27} Except for in TEA neonates, the rCBF of the cortical GM in our study did not show significant difference between the frontal and occipital lobes. These discrepancies could be attributed to the small sample size of our study.

WM showed lower CBF values compared with GM, which is consistent with previous studies on neonates.\textsuperscript{33} In all age groups, the O-WM showed higher rCBF values compared with the F-WM. One possible explanation for this finding is the higher development of the O-WM compared with the F-WM. A number of studies have performed regional WM evaluation using MR imaging, and the most widely accepted tools are DWI\textsuperscript{1,2} and DTI.\textsuperscript{3} One DWI study evaluating the development of WM in different brain regions showed lower ADC values in the occipital peritrigonal WM compared with the F-WM.\textsuperscript{3} In addition, a DTI study showed lower isotropic diffusion values in the O-WM (1.46 × 10^{-3} mm^2/s) compared with the F-WM (1.56 × 10^{-3} mm^2/s).\textsuperscript{3} Because lower ADC and lower isotropic diffusion values suggest higher WM development, the higher rCBF in the O-WM in our study is in accordance with these studies. Still, we think that this result needs to be interpreted with caution. A meta-analysis of DWI studies performed on neonate brains showed similar ADC values in the O-WM (146.4–164.2 × 10^{-3} mm^2/s) and F-WM (147.9–161.9 × 10^{-3} mm^2/s), both of which were measured in the cortical WM.\textsuperscript{2} In that study, a relatively lower ADC value was observed in the subcortical WM (105.4–149.5 × 10^{-3} mm^2/s). Because we manually drew ROIs and the spatial resolution of ASL is generally low, including the subcortical WM of the occipital lobe might have resulted in significantly higher CBF compared with F-WM.

There are several limitations to this study. First, there were a limited number of subjects in each age group. The insignificant results found for some of the parameters might have been due to the small number of included subjects. In addition, the significant correlation between PMA and rCBF values could have been highly influenced by the infant group. Future studies with a larger number of subjects would offer additional information on the age-related changes of CBF in this young population. Second, the study population cannot fully represent healthy neonates and infants because there were subjects with clinical suspicion of brain injury. Still, we tried to minimize the effect of abnormal CBF by
excluding subjects with visible abnormalities. In addition, TEA neonates cannot represent full-term neonates. We know that PMA rather than postnatal age is the chief factor in glucose use. However, preterm birth might disrupt the maturational program in a regionally specific manner, so TEA neonates cannot completely represent healthy full-term neonates. Third, the ASL signal could have been affected by head positioning in the coil. The infant brain has a larger variability of labeling efficiency due to variability in positioning within the standard head coil. We tried to position heads as centrally as possible in the coil during scanning. Still, the coil itself was not dedicated to neonates, and there was inevitable variability in head positioning. Fourth, sedation during scanning may have affected the CBF. Although chloral hydrate does not affect brain cortex activity, no thorough investigation has been performed on human CBF. On the other hand, midazolam causes dose-related changes in CBF. Natural sleep during scanning using the “feed and wrap” technique on neonates would be a better way of assessing normal development using CBF. Last, we used a single longitudinal relaxation value for brain tissue. The intrinsic longitudinal relaxation value, which is used for CBF estimation, differs by region, age, and subject. In neonates, the longitudinal relaxation value for WM is higher than for GM, and both WM and GM values decrease with PMA. Region-, age-, and subject-specific CBF estimations would be ideal; however, these are not yet applicable in this study as well as in most other clinical studies.

CONCLUSIONS
Multidelay ASL results in transit time-corrected CBF maps, which can show the developmental changes and differences occurring in neonates and infants. PMA was significantly correlated with regional changes in CBF. Because multidelay ASL could overcome transit delay in this young population, it could be a promising tool for imaging brain maturation processes.

Disclosures: Hyun Gi Kim—RELATED: Grant: National Research Foundation of Korea grant was received. Detailed information is in the submitted work.

REFERENCES
27. Kinnala A, Suonen-Holvi H, Aarimaa T, et al. Cerebral metabolic...


The Impact of Persistent Leukoencephalopathy on Brain White Matter Microstructure in Long-Term Survivors of Acute Lymphoblastic Leukemia Treated with Chemotherapy Only


ABSTRACT

BACKGROUND AND PURPOSE: Survivors of acute lymphoblastic leukemia are at risk for neurocognitive deficits and leukoencephalopathy. We performed a longitudinal assessment of leukoencephalopathy and its associations with long-term brain microstructural white matter integrity and neurocognitive outcomes in survivors of childhood acute lymphoblastic leukemia treated on a modern chemotherapy-only protocol.

MATERIALS AND METHODS: One hundred seventy-three survivors of acute lymphoblastic leukemia (49% female), treated on a chemotherapy-only protocol, underwent brain MR imaging during active therapy and repeat imaging and neurocognitive testing at follow-up (median, 13.5 years of age; interquartile range, 10.7–17.6 years; median time since diagnosis, 7.5 years; interquartile range, 6.3–9.1 years). Persistence of leukoencephalopathy was examined in relation to demographic and treatment data and to brain DTI in major fiber tracts and neurocognitive testing at follow-up.

RESULTS: Leukoencephalopathy was found in 52 of 173 long-term survivors (30.0%) and persisted in 41 of 52 (78.8%) who developed it during therapy. DTI parameters were associated with leukoencephalopathy in multiple brain regions, including the corona radiata (fractional anisotropy, P = .001; mean diffusivity, P < .001), superior longitudinal fasciculi (fractional anisotropy, P = .02; mean diffusivity, P < .001), and superior fronto-occipital fasciculi (fractional anisotropy, P = .006; mean diffusivity, P < .001). Mean diffusivity was associated with neurocognitive impairment including in the genu of the corpus callosum (P = .04), corona radiata (P = .02), and superior fronto-occipital fasciculi (P = .02).

CONCLUSIONS: Leukoencephalopathy during active therapy and neurocognitive impairment at long-term follow-up are associated with microstructural white matter integrity. DTI may be more sensitive than standard MR imaging for detection of clinically consequential white matter abnormalities in childhood acute lymphoblastic leukemia survivors treated with chemotherapy and in children undergoing treatment.

ABBREVIATION: ALL = acute lymphoblastic leukemia

Original Research

Original Research

The Impact of Persistent Leukoencephalopathy on Brain White Matter Microstructure in Long-Term Survivors of Acute Lymphoblastic Leukemia Treated with Chemotherapy Only


ABSTRACT

BACKGROUND AND PURPOSE: Survivors of acute lymphoblastic leukemia are at risk for neurocognitive deficits and leukoencephalopathy. We performed a longitudinal assessment of leukoencephalopathy and its associations with long-term brain microstructural white matter integrity and neurocognitive outcomes in survivors of childhood acute lymphoblastic leukemia treated on a modern chemotherapy-only protocol.

MATERIALS AND METHODS: One hundred seventy-three survivors of acute lymphoblastic leukemia (49% female), treated on a chemotherapy-only protocol, underwent brain MR imaging during active therapy and repeat imaging and neurocognitive testing at follow-up (median, 13.5 years of age; interquartile range, 10.7–17.6 years; median time since diagnosis, 7.5 years; interquartile range, 6.3–9.1 years). Persistence of leukoencephalopathy was examined in relation to demographic and treatment data and to brain DTI in major fiber tracts and neurocognitive testing at follow-up.

RESULTS: Leukoencephalopathy was found in 52 of 173 long-term survivors (30.0%) and persisted in 41 of 52 (78.8%) who developed it during therapy. DTI parameters were associated with leukoencephalopathy in multiple brain regions, including the corona radiata (fractional anisotropy, P = .001; mean diffusivity, P < .001), superior longitudinal fasciculi (fractional anisotropy, P = .02; mean diffusivity, P < .001), and superior fronto-occipital fasciculi (fractional anisotropy, P = .006; mean diffusivity, P < .001). Mean diffusivity was associated with neurocognitive impairment including in the genu of the corpus callosum (P = .04), corona radiata (P = .02), and superior fronto-occipital fasciculi (P = .02).

CONCLUSIONS: Leukoencephalopathy during active therapy and neurocognitive impairment at long-term follow-up are associated with microstructural white matter integrity. DTI may be more sensitive than standard MR imaging for detection of clinically consequential white matter abnormalities in childhood acute lymphoblastic leukemia survivors treated with chemotherapy and in children undergoing treatment.

ABBREVIATION: ALL = acute lymphoblastic leukemia

Contemporary treatment for pediatric acute lymphoblastic leukemia (ALL) has largely eliminated the use of prophylactic cranial radiation therapy1,2 and has substantially reduced many adverse effects such as neurocognitive deficits, metabolic and endocrine disorders, and subsequent neoplasms.1,3,4 Parallel to this advance has been the use of risk-directed systemic and intrathecal chemotherapy, which has increased survival rates to >90% in some studies.1 Although the neurocognitive function of survivors treated with chemotherapy only is better preserved than in those who underwent cranial radiation therapy, these survivors do demonstrate deficiencies.2,5

High doses of methotrexate, one of the primary chemotherapeutic agents used for consolidation treatment of ALL, are associated with leukoencephalopathy (ie, white matter hyperintensities on brain MR imaging).6,7 These changes may be transient or
persists over multiple MR imaging examinations conducted during and after chemotherapy. Leukoencephalopathy is sometimes associated with clinical findings of neurotoxicity such as stroke-like symptoms, seizures, or aphasia. 6,8

DTI is an MR imaging technique to evaluate microstructural white matter integrity based on the diffusion of water in the brain. Water diffusion in intact white matter is anisotropic (ie, diffusion is limited in direction, presumably due to cell membranes and myelin sheaths in parallel axonal fibers associated with white matter tracts). 9 Fractional anisotropy is a measure of unidirectionality of water diffusion on a scale of 0–1, with 0 representing no limitation in direction and 1 indicating diffusion completely confined to 1 direction. Mean diffusivity represents the average amount of diffusion regardless of direction. 10–12 Injury to a white matter tract is associated with lower fractional anisotropy and higher mean diffusivity compared with healthy white matter. 13,14

In a previous study of long-term survivors of childhood ALL treated with chemotherapy only, we reported associations between neurobehavioral outcomes and on-treatment leukoencephalopathy and DTI parameters in the frontostriatal tract, a tract we a priori predicted to be associated with executive function problems. 15 We also examined pharmacologic determinants of neurocognitive performance, cortical thickness, and functional MR imaging in this cohort. 16 In the current study, we more broadly examined the association between white matter integrity and neurocognitive outcomes in survivors by evaluating persistent leukoencephalopathy and DTI parameters in all major white matter tracts in the brain. We hypothesized that survivors who demonstrated persistent leukoencephalopathy would demonstrate poor microstructural white matter integrity as measured by DTI parameters and that poor white matter integrity would be associated with neurocognitive impairment.

MATERIALS AND METHODS

Study Participants

Four hundred eight children with ALL were treated at St. Jude Children’s Research Hospital on a single protocol (June 2000 to October 2010; Total Therapy XV, Total Therapy Study XV for Newly Diagnosed Patients with Acute Lymphoblastic Leukemia; clinicaltrials.gov No. NCT00137111) without prophylactic cranial radiation therapy. Of these, 369 underwent MR imaging examinations conducted during therapy to assess leukoencephalopathy of the brain at 4 time points during therapy to assess leukoencephalopathy. Of these, 369 underwent MR imaging examinations conducted during therapy to assess leukoencephalopathy. 16 Survivors were recruited during a long-term follow-up clinical evaluation when they were at least 8 years of age and ≥5 years from diagnosis. Exclusion criteria included the following: relapse, diagnosis of a subsequent neoplasm, lack of proficiency in English, or the presence of an unrelated neurologic disorder associated with cognitive impairment. Among the 295 survivors who fulfilled the eligibility criteria, 189 (64%) participated in long-term outcome studies between June 2009 and October 2014. Follow-up MR imaging of the brain was successfully performed for 173 of the survivors (On-line Figure). This study was approved by the institutional review board, and written informed consent was obtained from the patients, their parents, or guardians, as appropriate.

Treatment

The details of protocol therapy have been previously reported. 1,17 Briefly, central nervous system–directed therapy comprised 13–18 triple intrathecal treatments with methotrexate, hydrocortisone, and cytarabine and 4 doses of high-dose intravenous methotrexate at an average of 2.5 g/m2 per dose for patients with low-risk ALL. Patients with standard- or high-risk ALL received 16–25 triple intrathecal treatments and 4 doses of high-dose methotrexate at an average of 5.0 g/m2 per dose. No patient received prophylactic central nervous system radiation therapy, even in the presence of central nervous system leukemia at diagnosis.

MR Imaging

MR imaging examinations for the follow-up study included non-contrast 3D sagittal T1WI, axial T2WI, and axial proton-density and axial T2 FLAIR images obtained on a 1.5T platform. DTI of the brain was successfully obtained on a 1.5T platform using a double spin-echo EPI pulse sequence with 12 noncollinear, non-coplanar diffusion-gradient directions. Imaging sets were acquired with a spatial resolution of 1.7 × 1.7 × 3.0 mm and 4 acquisitions to ensure the highest signal-to-noise ratio possible within a limited amount of time. Voxelwise tensor calculations were performed with the Diffusion II toolkit under SPM8 (https://www.fil.ion.ucl.ac.uk/spm/). DTI measures of fractional anisotropy and mean diffusivity were extracted from voxels within fiber tracts throughout the brain.

All active therapy and long-term follow-up MR imaging examinations were reviewed for leukoencephalopathy by a board-certified neuroradiologist with a Certificate of Added Qualification in neuroradiology and graded according to the radiographic criteria of the Common Terminology Criteria for Adverse Events (Version 4.0; https://www.eortc.be/services/doc/ctc/CTCAE_4.03_2010-06-14_QuickReference_5x7.pdf). The neuroradiologist was blinded to patient risk stratification, neurocognitive test performance, and DTI results. Hyperintensity on the T2-weighted and T2 FLAIR images in the supratentorial white matter was considered leukoencephalopathy if located in the supraventricular white matter and/or periventricular white matter and more prominent than expected for terminal zones of myelination, normal developmental areas of T2 hyperintensity, for each subject’s age. 18 A second board-certified neuroradiologist with a Certificate of Added Qualification in neuroradiology separately graded 30 of the MR imaging examinations to estimate interrater reliability, yielding a k statistic of 0.75 (95% CI, 0.64–0.85), indicating substantial agreement.

Neurocognitive Testing

Neurocognitive testing was performed by certified examiners under the supervision of a board-certified clinical neuropsychologist. Testing procedures followed standard clinical guidelines, with fixed test order and a schedule to reduce the effects of interference and fatigue. Testing evaluated executive function (cognitive flexibility, cognitive fluency, working memory, organization, and problem-solving abilities), 19 intelligence (intelligence quotient), 20 processing speed, 21,22 attention, 23 memory, 24 and fine-motor dexterity. 25 Raw scores for these neurocognitive tests were...
Table 1: Demographics and treatment characteristics (N = 173)

<table>
<thead>
<tr>
<th>Demographics</th>
<th>No. (%)</th>
<th>Mean (SD)</th>
<th>Median (IQR)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>89 (51)</td>
<td>14.4 (4.6)</td>
<td>13.5 (10.7–17.6)</td>
</tr>
<tr>
<td>Female</td>
<td>84 (49)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Race/ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>124 (72)</td>
<td>6.7 (4.3)</td>
<td>5.3 (3.5–8.6)</td>
</tr>
<tr>
<td>Asian</td>
<td>3 (2)</td>
<td>7.7 (3.9)</td>
<td>7.0 (4.0–11.0)</td>
</tr>
<tr>
<td>Black</td>
<td>21 (12)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>18 (10)</td>
<td>13.6 (3.1)</td>
<td>13.0 (12.0–16.0)</td>
</tr>
<tr>
<td>Others</td>
<td>7 (4)</td>
<td>13.6 (3.1)</td>
<td>12.0 (12.0–16.0)</td>
</tr>
<tr>
<td><strong>Current age (yr)</strong></td>
<td></td>
<td>14.4 (4.6)</td>
<td>13.5 (10.7–17.6)</td>
</tr>
<tr>
<td><strong>Patient’s highest education (yr)</strong></td>
<td></td>
<td>7.7 (3.9)</td>
<td>7.0 (4.0–11.0)</td>
</tr>
<tr>
<td><strong>Maternal education (yr)</strong></td>
<td></td>
<td>13.6 (2.5)</td>
<td>13.0 (12.0–16.0)</td>
</tr>
<tr>
<td><strong>Paternal education (yr)</strong></td>
<td></td>
<td>13.6 (3.1)</td>
<td>12.0 (12.0–16.0)</td>
</tr>
<tr>
<td><strong>Time since diagnosis (yr)</strong></td>
<td></td>
<td>7.7 (1.8)</td>
<td>7.5 (6.3–9.1)</td>
</tr>
<tr>
<td><strong>Treatment risk stratum</strong></td>
<td></td>
<td>102 (59)</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard</td>
<td>71 (41)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Chemotherapy doses</strong> (^a)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oral dexamethasone (mg/m²)</td>
<td>1096.4 (303.2)</td>
<td>1099.9 (985.3–1246.1)</td>
<td></td>
</tr>
<tr>
<td>IV high-dose cytarabine (g/m²)</td>
<td>8.5 (3.5)</td>
<td>8.0 (8.0–8.0)</td>
<td></td>
</tr>
<tr>
<td>IV leucovorin (mg/m²)</td>
<td>343.5 (207.1)</td>
<td>300.0 (220.0–390.0)</td>
<td></td>
</tr>
<tr>
<td>IV high-dose methotrexate (^b) (g/m²)</td>
<td>15.4 (6.7)</td>
<td>14.2 (11.4–19.0)</td>
<td></td>
</tr>
<tr>
<td>IT MHA (No. of counts)</td>
<td>14.4 (4.0)</td>
<td>13.0 (12.0–16.0)</td>
<td></td>
</tr>
</tbody>
</table>

Note: \(\text{IQR}\) indicates interquartile range; IT MHA, intrathecal injection of methotrexate plus hydrocortisone plus cytarabine.

\(^a\) Except for IT MHA, all drug doses are presented as cumulative doses (g/m² or mg/m²).

\(^b\) High-dose IV methotrexate was defined as a daily dose of \(>1\) g/m² of IV methotrexate.

FIG 1. Examples of leukoencephalopathy grading. Left, mild hyperintensity in the bilateral periatrial white matter in this survivor of childhood ALL is compatible with grade 1 leukoencephalopathy according to the Common Terminology Criteria for Adverse Events (Version 4.0). Right, more extensive and confluent hyperintensity in the periventricular white matter that extends into the bilateral supraventricular frontoparietal white matter is considered grade 2 leukoencephalopathy.

RESULTS

Demographic and treatment-related data are provided in Table 1, and their associations with leukoencephalopathy are shown in Online Table 1. Race/ethnicity for each subject was obtained from the medical record as self-reported by the patients’ parents at the time of treatment. Of the 41 survivors with persistent leukoencephalopathy (leukoencephalopathy that was present on both active therapy and long-term follow-up MR imaging examinations), 37 had grade 1 and 4 had grade 2 according to the Common Terminology Criteria for Adverse Events (Version 4.0) (Fig 1). Because of the small number of participants with grade 2 abnormality, leukoencephalopathy was treated as a binary variable (present or absent). There were no significant differences in the cumulative dosages of known neurotoxic chemotherapeutic agents between patients with or without persistent leukoencephalopathy (On-line Table 1).

After correcting for the false discovery rate, there were significant associations between age-adjusted DTI parameters in mul-

Statistical Analysis

For demographic and clinical data, means, SDs, medians, and interquartile ranges were calculated for continuous variables and frequencies were reported for categoric variables. The Mann-Whitney \(U\) test (continuous variables) and the \(\chi^2\) test (categoric variables) were used to evaluate whether demographic and clinical characteristics were different between survivors with and without persistent leukoencephalopathy. Associations between persistent leukoencephalopathy and white matter integrity (fractional anisotropy and mean diffusivity) were examined using general linear modeling, adjusting for current age. \(P\) values were adjusted by controlling for the false discovery rate within the global white matter tracts. Only tracts that were significantly associated with leukoencephalopathy were tested for associations with methotrexate exposure using general linear modeling. Neurocognitive scores were transformed into age-adjusted \(z\) scores \((\mu = 0, \sigma = 1.0)\) using nationally representative norms. One-sample \(t\) tests were applied to compare the group average with the expected population value \((\mu = 0)\) for the specific tests. Impairment on an individual neurocognitive test was defined as a score falling below the tenth percentile of the norm \((z \text{ score} \leq -1.286)\). Global neurocognitive impairment was defined as having \(\geq 2\) neurocognitive test scores that fall \(>1.5\) SDs or 1 score that falls \(>2\) SDs below the mean. Associations between neurocognitive scores and persistent leukoencephalopathy were evaluated using general linear modeling, adjusting for age at evaluation. Associations between global neurocognitive impairment and white matter integrity were assessed using general linear modeling, adjusted for age at evaluation. All analyses were conducted in SAS (SAS 9.4; SAS Institute, Cary, North Carolina). Statistical significance was defined as a \(P\) value < .05, and all statistical tests were 2-sided.
Multiple fiber tracts and the presence of persistent leukoencephalopathy (Fig 2 and On-line Table 2). Lower fractional anisotropy and higher mean diffusivity were demonstrated in the corpus callosum, corona radiata, superior longitudinal fasciculi, and superior fronto-occipital fasciculi; higher mean diffusivity was detected in the posterior thalamic radiations in survivors with leukoencephalopathy. The total number of intrathecal administrations of methotrexate, hydrocortisone, and cytarabine was associated with higher mean diffusivity values in several fiber tracts (Table 2). High-dose intravenous methotrexate was not associated with DTI parameters.

Survivors had lower performance scores than population norms on measures of executive function, memory, processing speed, and global intelligence (almost all P < .001). There were no significant associations between neurocognitive performance measures and the presence of persistent leukoencephalopathy (On-line Table 3). However, higher mean diffusivity was associated with overall neurocognitive impairment (Fig 3 and On-line Table 4).

**DISCUSSION**

To our knowledge, this is the largest study of persistent leukoencephalopathy and tractography-based examination of white matter integrity in long-term survivors of childhood ALL treated with chemotherapy only. With serial imaging of the brain conducted during therapy and 5 years following diagnosis in 173 survivors, we identified the prevalence of persistent leukoencephalopathy to be 23.7%, with 78.8% of survivors who developed acute leukoencephalopathy continuing to demonstrate leukoencephalopathy at long-term follow-up. Persistent leukoencephalopathy was associated with microstructural indices of poor white matter integrity on DTI, though only the DTI parameters were associated with clinically relevant neurocognitive outcomes. Such findings suggest that DTI may be a better method for evaluation of white matter pathology in patients and survivors of childhood ALL.

A study of 54 patients treated on 2 Pediatric Oncology Group ALL protocols reported a prevalence of leukoencephalopathy of 22% for one protocol and 68% for the second at roughly 7 years following diagnosis.26 The investigators attributed the higher prevalence in the second treatment group to the greater use of triple intrathecal therapy, increased frequency of intrathecal therapy, and lack of leucovorin rescue compared with the first group. The approximately 24% prevalence of leukoencephalopathy observed in the present investigation is similar to that reported in survivors in the first group of the Pediatric Oncology Group study. We detected no significant associations between specific chemotherapy cumulative doses and leukoencephalopathy.

The association between poor DTI parameters in multiple supratentorial white matter tracts and the presence of leukoencephalopathy in our survivors confirms the persistence of white matter injury well into long-term follow-up. While our DTI results are similar to those previously reported by others,13,27-33 our novel linkage of DTI to leukoencephalopathy suggests that white matter integrity is negatively impacted during the early course of chemotherapy treatment.
FIG 3. Association between white matter integrity and global neurocognitive impairment. Black circles represent survivors with global neurocognitive impairment. White circles represent survivors without global neurocognitive impairment. Global neurocognitive impairment is defined as having ≥2 neurocognitive tests (listed in On-line Table 3) that fall >1.5 SDs or 1 test that falls >2 SDs below the age-adjusted population normative data. The y-axis represents mean diffusivity. 

Table 2: Association between methotrexate and white matter integritya

<table>
<thead>
<tr>
<th>Tracts</th>
<th>Mean Diffusivity</th>
<th>Fractional Anisotropy</th>
<th>Est. SE</th>
<th>P</th>
<th>Est. SE</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intrathecal MHAb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corpus callosum</td>
<td>0.0017</td>
<td>0.0008</td>
<td>.40</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Genu</td>
<td>0.0016</td>
<td>0.0009</td>
<td>.07</td>
<td>—</td>
<td>0.0015</td>
<td>.21</td>
</tr>
<tr>
<td>Body</td>
<td>0.0008</td>
<td>0.0010</td>
<td>.45</td>
<td>—</td>
<td>0.0000</td>
<td>.99</td>
</tr>
<tr>
<td>Splenium</td>
<td>0.0022</td>
<td>0.0011</td>
<td>.04</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Corona radiata</td>
<td>0.0014</td>
<td>0.0007</td>
<td>.40</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Anterior (left)</td>
<td>0.0017</td>
<td>0.0009</td>
<td>.06</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Anterior (right)</td>
<td>0.0011</td>
<td>0.0009</td>
<td>.19</td>
<td>—</td>
<td>0.0005</td>
<td>.41</td>
</tr>
<tr>
<td>Superior (left)</td>
<td>0.0013</td>
<td>0.0005</td>
<td>.02</td>
<td>—</td>
<td>0.0003</td>
<td>.70</td>
</tr>
<tr>
<td>Superior (right)</td>
<td>0.0015</td>
<td>0.0005</td>
<td>.05</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Posterior (left)</td>
<td>0.0009</td>
<td>0.0007</td>
<td>.23</td>
<td>—</td>
<td>0.0003</td>
<td>.65</td>
</tr>
<tr>
<td>Posterior (right)</td>
<td>0.0016</td>
<td>0.0007</td>
<td>.02</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Posterior thalamic radiation</td>
<td>0.0015</td>
<td>0.0007</td>
<td>.03</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Left</td>
<td>0.0011</td>
<td>0.0010</td>
<td>.29</td>
<td>0.0000</td>
<td>0.0008</td>
<td>.98</td>
</tr>
<tr>
<td>Right</td>
<td>0.0018</td>
<td>0.0006</td>
<td>.03</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Superior longitudinal fasciculus</td>
<td>0.0007</td>
<td>0.0005</td>
<td>.16</td>
<td>0.0006</td>
<td>0.0005</td>
<td>.25</td>
</tr>
<tr>
<td>Left</td>
<td>0.0006</td>
<td>0.0005</td>
<td>.22</td>
<td>0.0008</td>
<td>0.0006</td>
<td>.16</td>
</tr>
<tr>
<td>Right</td>
<td>0.0007</td>
<td>0.0005</td>
<td>.13</td>
<td>0.0005</td>
<td>0.0006</td>
<td>.41</td>
</tr>
<tr>
<td>Superior fronto-occipital fasciculus</td>
<td>0.0020</td>
<td>0.0009</td>
<td>.02</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Left</td>
<td>0.0021</td>
<td>0.0010</td>
<td>.04</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Right</td>
<td>0.0019</td>
<td>0.0007</td>
<td>.01</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Note: — indicates parameter estimate; SE, standard error; MHA, methotrexate plus hydrocortisone plus cytarabine.
a General linear modeling was applied for the test of strength of association between treatment variables with mean diffusivity and fractional anisotropy for each tract, adjusted for age at evaluation. Higher mean diffusivity and lower fractional anisotropy are indicative of worse white matter integrity. 
b Intrathecal MHA was defined as the number of intrathecal injections of methotrexate plus hydrocortisone plus cytarabine; high-dose IV methotrexate was defined as a daily dose of >1 g/m², presented as cumulative doses (g/m²).

et al26 reported a significant association between leukoencephalopathy and poorer performance on tests of attention in a group of 43 ALL survivors treated with chemotherapy only. Although there were no significant differences in the mean full-scale intelligence quotient, verbal intelligence quotient, or performance intelligence quotient between patients with or without leukoencephalopathy, survivors with leukoencephalopathy composed 80% of those scoring >1 SD below the mean for verbal intelligence quotient; 64%, for performance intelligence quotient; and 89%, for full-scale intelligence quotient.26 In contrast, we did not see similar trends in >150 survivors who completed both MR imaging examinations and neurocognitive testing. However, our DTI measures were associated with neurocognitive function. 

Mean diffusivity in the genu of the corpus callosum, corona radiata, and the superior fronto-occipital fasciculus was associated with global neurocognitive impairment. These tracts all involve the frontal lobes, which play a role in executive function14 and which demonstrate prolonged development during childhood, potentially placing them...
at extended risk for neurotoxicity. Other investigators have also found associations between abnormal DTI parameters and deficits in neurocognitive performance in survivors of ALL treated with chemotherapy only.\(^{13,27,28,32}\) That only mean diffusivity and not fractional anisotropy was associated with neurocognitive deficits suggests that the amount of diffusion, rather than the degree of anisotropy, in our survivors’ white matter had a greater effect on their neurocognitive abilities. The larger number of associations between mean diffusivity in multiple fiber tracts and the total number of intrathecal administrations of chemotherapeutic agents, compared with fractional anisotropy, also indicates that chemotherapy has a greater effect on the amount of diffusion in those tracts than the degree to which it is anisotropic. This is an area that requires further investigation.

Limitations to our study include distinguishing leukoencephalopathy from terminal zones of myelination, which can be difficult in some cases. The grades of leukoencephalopathy were compared over multiple MR imaging examinations to try to standardize the visual thresholds used across the studies; however, it is impossible, on the basis of visual inspection alone, to be certain that some cases of leukoencephalopathy were not classified as terminal zones or that terminal zones may have been incorrectly considered leukoencephalopathy. Given the size of our cohort, though, the instances of misclassification likely did not have a large effect on our results. In addition, the strong correlation between the presence of leukoencephalopathy and DTI parameters suggests that misclassification did not significantly affect our data. Another limitation is that this investigation lacked a control group for comparison with the DTI findings. The differences in the DTI parameters between survivors with and without leukoencephalopathy and the linkage of the DTI measures to standardized neurocognitive testing, however, demonstrate the importance of our DTI results. The limited research DTI sequence used in this study was designed and implemented 8 years ago and was not optimal for tracking all possible connections. However, the connections reported were reproducible and reliable. Future studies will likely include higher spatial and angular resolution acquisitions. We had no pretreatment DTI of the brain, which prevented us from determining whether DTI parameters were already abnormal due to disease. Because treatment typically begins within 24 hours of diagnosis for ALL and because the imaging in this study was acquired solely for research, we were not able to obtain pretreatment imaging. Last, no neurocognitive testing was performed before active therapy; thus, some neurocognitive deficits in our subjects may have existed before treatment. The lack of a baseline study, however, is not likely to account for most of the problems in neurocognitive performance we detected.

Children successfully treated for ALL with chemotherapy alone are at risk for persistent leukoencephalopathy. Although leukoencephalopathy is an MR imaging biomarker of white matter abnormality, it was not associated with neurocognitive deficits in our survivors. Differences in DTI parameters, a more precise measure of microstructural white matter integrity, however, were associated with global neurocognitive impairment, suggesting that DTI may serve as a more sensitive imaging measure of clinically consequential loss of white matter integrity in ALL survivors than the presence of leukoencephalopathy on standard MR imaging sequences. While the precise significance of the loss of structural integrity that leukoencephalopathy represents is unclear, continued refinement of chemotherapy treatment regimens for ALL should nevertheless aim to reduce the occurrence of white matter abnormalities. Potential treatments and interventions to enhance white matter development and function also need to be explored.

CONCLUSIONS

Leukoencephalopathy that develops in patients treated for childhood ALL and long-term neurocognitive impairment are associated with microstructural white matter integrity. DTI may be more sensitive than standard MR imaging sequences for identifying clinically consequential white matter abnormalities in this group of cancer survivors.

ACKNOWLEDGMENTS

We would like to acknowledge Cara Kimber, PhD, of Psychological Care Associates, Woburn, Massachusetts; and Ms Cynthia Jones, Ms Deborah Stewart, and Ms Adrienne Studaway of the Department of Epidemiology and Cancer Control, St. Jude Children’s Research Hospital, for administering the neurocognitive tests; and Ms Joycelynn Butler, of the Department of Epidemiology and Cancer Control, St. Jude Children’s Research Hospital, for extracting and cleaning the data.


REFERENCES

23. Conners CK. Conners’ Continuous Performance Test II. North Tonawanda: Multi-Health Systems; 2001
24. Meyers JE, Meyers KR. Rey Complex Figure and Recognition Trial. Odessa: Psychological Assessment Resources; 1996
Apparent Diffusion Coefficient Levels and Neurodevelopmental Outcome in Fetuses with Brain MR Imaging White Matter Hyperintense Signal

E. Katorza, G. Strauss, R. Cohen, M. Berkenstadt, C. Hoffmann, R. Achiron, E. Barzilay, and O. Bar-Yosef

ABSTRACT

BACKGROUND AND PURPOSE: One of the perplexing findings of fetal brain MR imaging is white matter T2 hyperintense signal. The aims of our study were initially to determine the main etiologies associated with white matter T2 hyperintense signal, then to examine whether the different etiologies have different ADC values, and, last, to assess the association of white matter T2 hyperintense signal with developmental outcome.

MATERIALS AND METHODS: This was a prospective cohort study of 44 MR imaging scans of fetal brains obtained for suspected brain pathologies at a tertiary medical center during 2011–2015. Clinical data were collected from electronic medical charts. ADC values were measured and averaged in the frontal, parietal, occipital, and temporal lobes. Neurodevelopmental assessments were performed with the Vineland Adaptive Behavior Scales II.

RESULTS: Half of the cases of MRI hyperintense T2 signal of the fetal brain were associated with congenital cytomegalovirus infection. The other half were mainly idiopathic. Thus, the study group was divided to subgroups positive and negative for cytomegalovirus. Both groups had hyperintense signal in the temporal lobe. The group positive for cytomegalovirus had involvement of the parietal lobe. Only this group had increased ADC values in the temporal and parietal lobes. There was no association between the neurodevelopment outcome and the etiologies or ADC values.

CONCLUSIONS: T2 hyperintense signal in fetal brain MRI associated with positive cytomegalovirus infection has increased ADC values in the temporal and parietal lobes, suggestive of brain edema in these areas. However, the association between this finding and neurodevelopment outcome requires further evaluation.

ABBREVIATIONS: CMV = cytomegalovirus; fbMRI = fetal brain MRI; ICC = interclass correlation coefficient; VABS = Vineland Adaptive Behavior Scales II; WMHS = white matter hyperintense signal

Fetal brain MR imaging (fbMRI) has been increasingly used in recent years as a means of tracking normal and pathologic fetal brain maturation. One of the perplexing findings of fbMRI is white matter T2 hyperintense signal (WMHS). On the one hand, it has been associated with in utero brain pathologies, such as ischemia and cytomegalovirus (CMV) infection. On the other hand, the validity and relevance of this finding have been questioned.

In recent years, DWI and its ADC metric have become a quantitative method for evaluation of fetal brain maturation. Previous studies showed that ADC values of the developing fetal brain correlate with fetal brain maturation. Deviation from normal ADC values has been shown to be associated with brain pathologies such as ischemia, CMV infection, and ventriculomegaly. Postmortem studies of animal and human fetuses with hypoxic-ischemic brain injury have demonstrated a transition from low ADC values after the injury to increased values days after the injury. This transition was associated with histologic findings changing from initial cytotoxic edema and swollen astrocytes to vasogenic edema, astrogliosis, and abundance of macrophages. Thus, ADC and its association with histopathology could be used to test the validity and meaning of T2 hyperintensity.

The aims of our study were initially to determine the main
etologies associated with WMHS, then to examine whether the different etiologies have different ADC values, and, last, to assess the association of WMHS with developmental outcome.

**MATERIALS AND METHODS**

**Subjects**

This was a prospective cohort study of women who were referred for fetal brain MR imaging to our tertiary medical center, Sheba Medical Center, Ramat Gan, Israel, between 2011 and 2015. The cohorts for this study were chosen on the basis of identification of a hyperintense signal on the T2-weighted sequences. Demographic and clinical data were collected from the electronic records of each patient.

Data obtained from the records included maternal history, prenatal screening tests, imaging results from anatomic sonography and MR imaging, maternal CMV status, and perinatal history. Fetal cytomegalovirus infection was confirmed by either amniocentesis performed during pregnancy or by the presence of CMV DNA in neonatal urine or saliva.

The study was approved by Sheba Medical Center, Ramat Gan, Israel, institutional ethics board. Informed consent was obtained from each participant in the study prospectively.

**MR Imaging Scans**

Fetal brain MR imaging was performed using a 1.5T system (Optima MR450w with GEM Suite; GE Healthcare, Milwaukee, Wisconsin). Single-shot fast spin-echo T2-weighted sequences in 3 orthogonal planes were performed using a half-Fourier technique (NEX = 0.53) with the following parameters: section thickness, 3 or 4 mm; no gap; flexible coil (8-channel cardiac coil). FOV was determined by the size of the fetal head with a range of 24 × 24 to 30 × 30 cm; acquisition time was between 40 and 45 seconds (matrix, 320 × 224; TE, 90 ms; TR, 1298 ms; pixel bandwidth, 122 Hz/pixel; specific absorption rate values, 1.1–1.7 W/kg). A DWI sequence in 1, 2, or 3 orthogonal planes was then performed, with a 40-cm FOV, b-values of 0 and 700 ms, and a slice thickness of 4 mm with no gap. All MR images were obtained by the same protocol at our institution and assessed by a specialist in fetal sonography and a neuroradiologist expert in MR imaging as previously published.15

**Target Variables**

ADC calculation was performed on 8 circular ROIs: 2 on the white matter of both frontal (1, 2), parietal (3, 4), temporal (5, 6), and occipital lobes (7, 8). A circular ROI was placed over the desired anatomic area, ranging from 75 to 98 mm².

**White Matter T2 Hyperintense Signal**

WMHS is a subjective interpretation of the signal from the fetal brain white matter on T2 sequences, made by the radiologist. This diagnosis is established when specific areas of the brain white matter appear hyperintense in comparison with other areas or with that expected according to the gestational age (Fig 2).16

**Interobserver Validity of ADC Measurements**

To validate the consistency of measurements and reliability of results, 2 observers evaluated 10 fetuses. Interobserver variability was assessed by the interclass correlation coefficient (ICC). We considered an ICC value of ≥0.8 as excellent agreement.

**Neurodevelopmental and Hearing Outcome**

Children were assessed by the Vineland Adaptive Behavior Scales, 2nd edition (VABS), which is a structured parent interview assessing 4 different domains of behavior: communication, daily living skills, socialization, and motor skills. All 4 domains are included in an adaptive composite score.17,18 VABS assessment was conducted by a phone interview by 2 medical students trained and supervised by a pediatric neurologist and child development expert experienced in conducting VABS (O.B.-Y.). Validation of the phone interviews were assessed by correlation of the ICC to the VABS scores of 15 children (divided into 7 and 8 children between the students) evaluated by the pediatric neurologist (O.B.-Y.). The ICC was >0.83 for the 4 VABS domains and the composite score. Scores of children were considered abnormal if the standard score was <70.

Hearing outcome was assessed 2–3 days after delivery by Transient Evoked Otoacoustic Emissions. Neonates positive for CMV were tested routinely before 1 month of age by brain stem auditory evoked potential, then at 3, 6, and 12 months by Transient Evoked Otoacoustic Emissions and behavioral assessment.

**Statistical Analysis**

Categoric variables were expressed as number and percentage. Distribution of continuous variables was assessed using a histo-
gram and Q-Q plot. Continuous variables were described using median and interquartile range or mean and SD as appropriate. Categoric variables were compared using the χ² test, Fisher exact test, or McNemar test as appropriate. Continuous variables were compared using the Student t test or Mann-Whitney test as appropriate. A 2-tailed P < .05 was considered statistically significant. Analyses were performed with SPSS (Version 24.0, 2016; IBM, Armonk, New York).

RESULTS

Demographic and Clinical Characteristics of the Study Population

The study subjects comprised 43 singleton pregnant women and 1 single twins pregnant woman who underwent fb-MRI scans. All 44 fetuses included in this study were in their third trimester of pregnancy at the time of the fetal MR imaging, with a median gestational age of 32 weeks (interquartile range, 28.25–35 weeks). The average maternal age was 33 years (interquartile range, 25–32 years). The indications for MR imaging included the following: maternal cy-

![FIG 2. T2 MR images (single-shot fast spin-echo T2-weighted sequences in 3 orthogonal planes using a half-Fourier technique, NEX = 0.53) of 2 fetuses at 33 weeks of gestational age. The CMV-positive fetus has diffuse WMHS (white arrowheads), unilateral ventriculomegaly, and intraventricular adhesions (black asterisks), suggesting ventriculitis (A–C). The CMV-negative fetus has WMHS located in the white matter in the temporal lobes (white arrowhead) (D–F).](image-url)
In the CMV-positive group, there were significantly less early and late anatomic scan findings ($P = .02$, $P = .009$, respectively) (Table 1).

**Distribution of Hyperintense Signal and Additional MR Imaging Findings**

Both groups had similar rates of WMHS in the temporal lobe ($P > .999$). WMHS was not depicted in the occipital lobe in both groups. However, the CMV-positive group had statistically higher rates of parietal hyperintense signal ($P = .002$) and a trend toward higher rates in the frontal lobe ($P = .067$). The CMV-positive group had more extended WMHS involving all 3 lobes ($P = .004$). The CMV-negative fetuses had statistically significantly higher rates ($P = .046$) of minor additional findings on imaging, including subarachnoid cyst ($n = 2$), slightly enlarged subarachnoid space ($n = 1$), and lateral ventricle asymmetry ($n = 4$) (Table 2 and Fig 2).

**Interobserver Validity of ADC Measurements**

ADC measurements showed excellent interobserver agreement for all regions, with the ICC ranging between 0.81 and 0.97.

**ADC Value Measurements**

ADC values for each group were compared with the normal ADC values as published by Hoffmann et al. Fetuses in the CMV-positive group were found to have statistically higher ADC values in the temporal lobe bilaterally ($P = .002$ and $P < .001$) and the left parietal lobe ($P = .033$) and a trend in the right parietal lobe ($P = .057$) (Table 3).

ADC values were compared for each region between the 2 study groups. Fetuses in the CMV-positive group were found to have higher ADC values in the left frontal lobe ($P = .026$), the parietal lobe bilaterally (right, $P = .001$; left, $P = .002$), and the temporal lobe bilaterally (right, $P = .011$; left, $P = .002$) (Table 4).

**Delivery Data, Hearing, and Neurodevelopmental Assessment**

Neurodevelopmental assessment (VABS) and delivery data were analyzed for 20 children in the CMV-positive group and 18 children in the CMV-negative group (Table 5). Six patients (2 from the CMV-positive group and 4 from the CMV-negative group) refused to participate in the neurodevelopmental assessment. Delivery data and age at developmental assessment were not statistically different between the groups. Comparison of the VABS score for the 4 domains and the adaptive composite score showed only

---

**Table 2: Radiologic MR imaging findings**

<table>
<thead>
<tr>
<th>Lobe</th>
<th>CMV-Positive</th>
<th>CMV-Negative</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frontal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parietal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temporal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occipital</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**Table 3: Comparison of ADC values of the study group with the control group**

<table>
<thead>
<tr>
<th>Lobe, Side</th>
<th>Study Group</th>
<th>ADC</th>
<th>ADC Control</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frontal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parietal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temporal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occipital</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**Note:**—CMV + indicates CMV-positive; CMV −. CMV-negative.

*Data are presented as median (interquartile range) for continuous variables or number (percentage) for categoric variables.

---

tomegalovirus infection ($n = 31$), family history of central nervous system pathology ($n = 3$), and abnormal findings during fetal sonography ($n = 10$). The abnormal sonographic findings included the following: lateral ventricular asymmetry ($n = 4$), mega cisterna magna ($n = 2$), small head circumference ($n = 2$), fisted hands ($n = 1$), and hyperechogenic bowel ($n = 1$). Twenty-two (50%) of the fetuses were found to be positive for CMV by amniocentesis or by saliva or urinary CMV DNA testing after birth. Thus, to examine the association between ADC values and etiology, we divided the main study group into a CMV-positive group consisting of 22 fetuses with WMHS from an unknown etiology.

These 2 distinct patient populations are described separately and compared by their demographic, clinical, imaging, and neurodevelopmental characteristics.

The median gestational age, pregnancy history, and sex did not differ statistically between the groups. All mothers in the CMV-positive group had normal medical background, compared with 6 (27%) mothers in the CMV-negative group who had significant abnormal medical background including thrombophilia, Raynaud disease, and a history of cerebellar infarction ($P = .02$).
a trend for motor skills ($P = .07$). Four children in the CMV-positive group and 1 child in the CMV-negative group had at least 1 VABS score <70; however, this was not statistically significant. Only 1 child had a hearing deficit; this child was CMV-positive with a low VABS score.

**DISCUSSION**

WMHS is a puzzling finding in fetal MR imaging, and its significance is not completely understood. The aims of our study were initially to explore the etiologies associated with this finding, then to examine whether the different etiologies are characterized by different ADC values, and, last, to investigate the association of WMHS with neurodevelopmental outcome.

Diverse etiologies were found to be associated with WMHS in this study. The major one (half of the study group), in accordance with fetal MR imaging literature, is congenital CMV infection. The precise etiologies of the second half of the group are not completely clear. Ten fetuses had further imaging findings; thus, the WMHS might be a different aspect of the overall brain pathology.

Comparison between the 2 groups demonstrated a difference in the distribution of the WMHS. Both groups had the same prevalence of WMHS in the temporal lobes, but the CMV-positive group had a higher rate in the parietal lobes and a trend in the frontal lobes. In general, the CMV-positive group had more diffuse WMHS.

The median ADC values were higher in the CMV-positive group than in the CMV-negative group in the temporal, parietal, and left frontal lobes. The ADC values of the CMV-negative group did not differ from those of the control group. Thus, although the incidences of WMHS were similar in the temporal lobes, the median ADC values were different. Because ADC is a quantitative measure and not a subjective interpretation, it is possible that some WMHS of the CMV-negative group is actually overinterpretation.

As for the CMV-positive group, the specific involvement of the temporal lobes was previously described in fMRIs. WMHS is part of a spectrum of findings localized to this area, including cysts, dilation of the temporal horns, and reduced temporal lobe volume. The reason for the specific vulnerability of the temporal lobe is unclear. Most interesting, in addition to the presence of WMHS in the temporal lobe, most of the fMRIs in our study depicted a similar signal in the parietal lobe. This finding was previously described in children with congenital CMV.

Thus, the combination of temporal and parietal hyperintense T2 signal is potentially a sign of CMV infection involving the brain. Further involvement of WMHS including the frontal lobes may serve as an additional sign of congenital CMV infection.

The combination of WMHS and high ADC values in the CMV-positive group suggests a possible etiology for this finding. Hyperintense signal was associated at postmortem examination with astrogliosis and extracellular brain edema in the subacute stage of hypoxic-ischemic injury. Although hypoxic-ischemic injury pathogenesis is different from CMV infection, the simplest explanation of the T2 finding is an increase in extracellular fluid content in the white matter. Kotovich et al. found low ADC values in fetuses with congenital CMV infection without T2 hyperintense signal. In a previous study of the same group, Yaniv et al. described postmortem histology consisting of edema and cellular infiltration of plasma cells, lymphocytes, and microglia. The difference in T2 hyperintense signal and ADC values between the 2 latter studies and ours might reflect a different ratio between the edema and cellular component in the brain tissue. When the microgliosis and edema component are more prominent, the ADC values are lower than those of controls; whereas when the component of cellular degeneration and immune cell infiltration is more prominent, the ADC values are lower than those of controls.

The association between WMHS and abnormal neurodevelopmental outcome was not found in this study. In the CMV-positive group, 4 children (20%) had abnormal neurodevelopmental findings or hearing loss. Based on neurodevelopmental data from previous studies, the expected range of these abnormal findings in children with asymptomatic congenital CMV (normal prenatal imaging findings) is 10%–15%. Thus, it is difficult to assess whether the WMHS and the increased ADC values predict worse prognosis.

Combining the results from our study and others investigating the association of ADC and brain pathologies suggests the following practical advice for clinicians: The combination of WMHS and normal ADC

---

**Table 4: Comparison of ADC values between fetuses with CMV-positive infection and fetuses with isolated white matter hyperintense signal**

<table>
<thead>
<tr>
<th>Lobe, Side</th>
<th>CMV-Positive</th>
<th>CMV-Negative</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frontal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td>1872 (160)</td>
<td>1793 (192)</td>
<td>.347</td>
</tr>
<tr>
<td>Left</td>
<td>1858 (193)</td>
<td>1741 (141)</td>
<td>.026</td>
</tr>
<tr>
<td>Parietal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td>1840 (140)</td>
<td>1659 (200)</td>
<td>.001</td>
</tr>
<tr>
<td>Left</td>
<td>1852 (138)</td>
<td>1690 (188)</td>
<td>.002</td>
</tr>
<tr>
<td>Temporal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td>1835 (161)</td>
<td>1702 (169)</td>
<td>.011</td>
</tr>
<tr>
<td>Left</td>
<td>1846 (122)</td>
<td>1689 (193)</td>
<td>.002</td>
</tr>
<tr>
<td>Occipital</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td>1723 (129)</td>
<td>1659 (224)</td>
<td>.259</td>
</tr>
<tr>
<td>Left</td>
<td>1725 (111)</td>
<td>1675 (169)</td>
<td>.254</td>
</tr>
</tbody>
</table>

*Data are presented as median (interquartile range) for continuous variables or as number (percentage) for categoric variables.

**Table 5: Delivery data, hearing, and VABS assessment**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>CMV-Positive (n = 20)</th>
<th>CMV-Negative (n = 18)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birth week</td>
<td>39.2 (38.5–39.9)</td>
<td>37.9 (37.9–40.6)</td>
<td>.55</td>
</tr>
<tr>
<td>Birth weight (percentile)</td>
<td>49 (31–73)</td>
<td>49 (31–73)</td>
<td>.53</td>
</tr>
<tr>
<td>Apgar score at 5 min $\geq$9</td>
<td>20 (100%)</td>
<td>18 (100%)</td>
<td>&gt;.99</td>
</tr>
<tr>
<td>Hearing deficit</td>
<td>1 (5%)</td>
<td>0 (0%)</td>
<td>&gt;.99</td>
</tr>
<tr>
<td>Age (mo) at VABS test</td>
<td>30 (17–49)</td>
<td>27.5 (17.2–28.5)</td>
<td>.75</td>
</tr>
<tr>
<td>VABS motor skills</td>
<td>92.5 (81.75–104)</td>
<td>103 (95.25–108)</td>
<td>.07</td>
</tr>
<tr>
<td>VABS daily living skills</td>
<td>103 (99.25–117)</td>
<td>109 (99.25–117)</td>
<td>.28</td>
</tr>
<tr>
<td>VABS socialization</td>
<td>101 (98.5–105.75)</td>
<td>100 (96.75–115.75)</td>
<td>.89</td>
</tr>
<tr>
<td>VABS communication</td>
<td>102.5 (100.25–113)</td>
<td>108 (99.25–117)</td>
<td>.84</td>
</tr>
<tr>
<td>VABS adaptive score comp</td>
<td>102 (90–18)</td>
<td>107 (98–115)</td>
<td>.16</td>
</tr>
<tr>
<td>Child with any VABS score</td>
<td>4 (20%)</td>
<td>1 (5.5%)</td>
<td>.34</td>
</tr>
</tbody>
</table>

*Data are presented as median (interquartile range) for continuous variables or as number (percentage) for categoric variables. Hearing deficit was assessed by either transient evoked otoacoustic emissions or brain stem evoked potential.
suggests that the source of WMHS is most likely without clinical significance. As for cases of abnormal ADC either below or above the expected value for gestational age, they raise the suspicion of brain pathology such as CMV infection, ventriculomegaly, or ischemia. However, the association of abnormal ADC with neurodevelopmental outcome is unclear.

Our study is limited by the size of the study group. A larger number of patients could be used to correlate the distribution of T2 hyperintense signal or ADC values with neurodevelopment. On the other hand, the small number of children in each group strengthens the significance of the difference in T2 hyperintense signal and ADC values between the CMV-positive and -negative groups.

**CONCLUSIONS**

ADC measurements supported the validity of T2 hyperintense signal in fbMRI only in fetuses with CMV infection. The association of abnormal ADC with neurodevelopmental outcome requires further investigation.

**REFERENCES**

Longitudinal Findings of MRI and PET in West Syndrome with Subtle Focal Cortical Dysplasia


ABSTRACT

BACKGROUND AND PURPOSE: Despite the development of neuroimaging, identification of focal cortical dysplasia remains challenging. The purpose of this study was to show the longitudinal changes of MR imaging and FDG-PET in patients with West syndrome and subtle focal cortical dysplasia.

MATERIALS AND METHODS: Among 52 consecutive patients with West syndrome, 4 were diagnosed with subtle focal cortical dysplasia on 3T MR imaging. MR imaging and PET findings were evaluated longitudinally at onset and at 12 and 24 months of age.

RESULTS: At the onset of West syndrome, MR imaging demonstrated focal signal abnormalities of the subcortical white matter in 2 patients. In the other 2 patients, focal subcortical high-intensity signals became visible on follow-up T2WI as myelination progressed. PET at onset showed focal cortical hypometabolism in 3 patients, with 1 of these patients also having focal hypermetabolism and 1 having normal findings. On PET at 24 months, hypometabolism persisted in 2 patients and disappeared in 1, and hypermetabolism disappeared in 1. In 1 patient with normal MR imaging and PET findings at onset, focal hyperintensity and hypometabolism first appeared at 24 months of age. The findings on MR imaging and PET in these patients evolved differently with brain maturation and the clinical course.

CONCLUSIONS: Subtle focal cortical dysplasia can be undetectable on MR imaging at the onset of West syndrome and is not always accompanied by hypometabolism or hypermetabolism on PET. Longitudinal MR imaging and PET studies may be useful for detecting such lesions. Even in West syndrome with a congenital structural abnormality, PET findings evolve differently with brain maturation and the clinical condition.

ABBREVIATIONS: EEG = electroencephalography; FCD = focal cortical dysplasia; WS = West syndrome

West syndrome (WS) is an age-dependent epileptic encephalopathy characterized by a triad of epileptic spasms, hypsarrhythmia on electroencephalography (EEG), and neurodevelopmental regression. WS is attributed to various etiologies, but the cause is undetermined in 20% of patients.1 The presumed etiology in such cases includes genetic abnormalities and focal cortical dysplasia (FCD).

Despite recent developments in neuroimaging techniques, identification of FCD remains challenging. Subtle FCD can be missed on MR imaging at the onset of WS.2 It has also been reported that MR imaging findings of subtle FCD can disappear with white matter maturation.3 FDG-PET may suggest FCD even if initial MR imaging findings are normal.4 However, the findings of PET at onset evolve along with the patient’s epilepsy.5-9 On MR imaging evaluation of infants, the status of myelination affects the findings, and MR imaging findings of FCD cannot be detected or differ from those in older patients.7 Thus, longitudinal MR imaging and PET studies may be useful to diagnose FCD accurately.

We hypothesized that serial evaluation of MR imaging and PET from the onset to 2 years of age shows the evolution of subtle FCD findings and leads to the accurate diagnosis and location of the lesions in patients with WS. To verify this hy-
pothesis, we reviewed our series of patients with WS in whom MR imaging and PET were longitudinally performed during 2007 and 2013.

MATERIALS AND METHODS
This study was approved by the Research Ethics Committee at Nagoya University Graduate School of Medicine. Informed consent was obtained from patients who participated in clinical investigations.

Patient Selection
Between 2007 and 2013, fifty-two children admitted to Nagoya University Hospital were newly diagnosed as having WS. In a systematic review of these 52 patients, 23 were diagnosed as having an unknown etiology at the onset of WS, meeting the following criteria (previously termed “cryptogenic”): 1) normal birth and absence of any etiologic factors related to WS; 2) normal development before onset and absence of neurologic abnormalities at onset; 3) the occurrence of clusters of spasms without any other types of seizures before the onset of spasms; and 4) normal laboratory and MR imaging findings at onset. Among the 23 patients with an unknown etiology at onset, 2 patients were later diagnosed as having FCD on MR imaging. In the remaining 29 patients with genetic, structural, or metabolic etiologies, 2 patients had subtle focal abnormalities of FCD on 3T MR imaging at onset. Two other patients had diffuse or bilateral FCD that was clearly recognized on MR imaging. These 2 patients were not included in this study because the focus was on patients with focal, subtle lesions that can be missed on conventional MR imaging. Thus, the present study included 4 patients in whom subtle FCD was seen at onset or during the follow-up period.

Neuroimaging Protocol
MR Imaging. MR imaging was performed longitudinally at the diagnosis of WS before adrenocorticotropic hormone therapy and at 12 and 24 months of age. Additional MR imaging was performed according to clinical need. MR imaging was performed using a 3T scanner (Magnetom Trio, a Tim System; Siemens, Erlangen, Germany) with a 32-channel phased array head coil. All patients were sedated with oral chloral hydrate (80 mg/kg) before the examination. When patients did not appear sufficiently sedated after chloral hydrate intake, intravenous midazolam or ketamine was administered. We acquired the following images: axial T1WI (TR/TE, 400/11 or 700/6.8 ms; slice thickness, 5 mm); T2WI (TR/TE, 5210/72 or 5200/131 ms; slice thickness, 5 mm); FLAIR images (TR/TE, 9000/139 ms; TI, 2500 ms; slice thickness, 5 mm); sagittal T1WI (TR/TE, 500/6.8 ms; slice thickness, 5 mm); coronal T2WI (TR/TE, 5210/72 or 5210/131 ms; slice thickness, 5 mm); and T1-weighted sagittal MPRAGE (TR/TE, 1570/2.2 ms; slice thickness, 1 mm) with reconstructed axial and coronal images.

FDG-PET. FDG-PET was performed as part of the clinical routine to search for underlying pathologies using a Headtome V scanner (Shimadzu, Kyoto, Japan). All patients were sedated with a chloral hydrate suppository (50 mg/kg) before the PET examination. When patients did not appear sufficiently sedated after chloral hydrate, intravenous midazolam or ketamine was administered 45 minutes after the FDG injection. Patients were scanned for 60 minutes after intravenous administration of FDG. EEGs were not monitored during the PET scans. Instead, pediatric neurologists observed all patients throughout the scan and confirmed that there were no seizure manifestations during the examination. Thirty-four axial images (4-mm-thick) were obtained, and the images from 45 to 60 minutes after FDG injection were used for the evaluation.

Assessment of MR Imaging and PET Findings
MR imaging and PET findings were assessed with visual inspection by 3 pediatric neurologists (J.N., Y.S., and H. Kidokoro) and 2 radiologists (T. Nakane for MR imaging and K.K. for PET). One of the radiologists was a pediatric neuroradiologist (T. Nakane) regularly involved in the MR imaging evaluation of patients with epilepsy. Another radiologist (K.K.) specialized in nuclear medicine and evaluating PET findings of patients with epilepsy. MR imaging was independently evaluated by 3 pediatric neurologists (J.N., Y.S., and H. Kidokoro) and 1 neuroradiologist (T. Nakane). PET was evaluated independently by 3 pediatric neurologists (J.N., Y.S., and H. Kidokoro) and 1 radiologist specialized in nuclear medicine (K.K.). Four reviewers agreed on the presence or absence of abnormal findings on MR imaging or PET, and they decided on the distribution of the abnormalities by discussion. On MR imaging, whether there were structural abnormalities, abnormal high or low intensity signals, and delayed myelination was evaluated. On PET, focal hypometabolism or hypermetabolism was defined as a regional decrease or increase, respectively, in FDG accumulation in ≥2 gyri on ≥2 slices. To assess localization of PET abnormalities, we performed an MR imaging–PET coregistration technique using SPM12b software (http://www.fil.ion.ucl.ac.uk/spm/software/spm12) and the image visualization software Register (Brain Imaging Center, Montreal Neurological Institute, Montreal, Quebec, Canada).

Clinical Data
Other clinical variables were collected from medical records. EEG recordings were performed at onset, during adrenocorticotropic hormone therapy, and at least at 12 and 24 months of age. The EEG records were assessed by 3 pediatric neurologists (J.N., Y.S., and H. Kidokoro). All patients with WS were followed by pediatric neurologists at our outpatient clinics until 5–7 years of age. The developmental quotient or intelligence quotient was assessed at the last follow-up using the Kinder Infant Development Scale, a developmental test standardized in Japan in 1989. If the patient was older than 7 years of age, the Wechsler Intelligence Scale for Children, Fourth Edition, was administered because the Kinder Infant Development Scale was appropriate for children younger than 7 years of age. Developmental status was considered as follows: 1) normal, when patients had a developmental quotient of ≥80; 2) borderline, when patients had a developmental quotient between 70 and 79; 3) mild delay, when patients had a developmental quotient between 50 and 69; and 4) severe delay, when patients had a developmental quotient of <50.
RESULTS

The profiles of the 4 patients are summarized in On-line Table 1. The age at onset of spasms ranged from 2 to 18 months. The age at diagnosis of WS ranged from 4 to 19 months. All patients had normal development before the onset of spasms and normal neurologic examination findings at onset. All patients received adrenocorticotropic hormone therapy. After adrenocorticotropic hormone therapy, 3 patients were free of seizures. Another one (patient 1) had a relapse of spasms at 17 months of age, and oral antiepileptic drugs were administered. The patient became free of seizures from 2 years of age. The developmental quotient or intelligence quotient at the last follow-up was normal in 3 patients and mildly delayed in 1 patient.

MR Imaging and PET Findings

MR imaging, PET, and EEG findings are summarized in On-line Table 2. The initial PET scans at the onset of WS showed focal cortical hypometabolism in 3 patients, and one of them (patient 3) also had focal hypermetabolism. The longitudinal findings in each patient are detailed below (Figure and On-line Figs 1–4).

Patient 1. This patient began having epileptic spasms at 7 months of age. Her initial MR imaging at 11 months of age showed a focal hyperintense lesion in the subcortical white matter of the left frontal lobe on axial MR imaging at onset (arrow, A). PET hypometabolism is seen in the left frontotemporal lobe at 11 months of age (arrows, B), in the right frontotemporal lobe at 24 months of age (arrows, C), and in the left frontal lobe at 36 months of age (arrow, D). PET at 24 months of age also shows possible hypermetabolism in the left temporal lobe (arrowheads, C). A coregistered coronal image of PET on MR imaging at 36 months of age shows hypometabolism in the left orbitofrontal area that corresponds to the MR imaging lesion (arrow, E). Multifocal hypometabolism in the frontal and temporal lobes is seen on PET at onset (arrows, G) that disappears at 24 months of age (H). No PET hypometabolism at 24 months of age (H) is seen. Hypometabolism in the left frontal lobe corresponds to the MR imaging lesion on a coronal MR imaging–PET coregistration image (arrows, I). Hyperintense subcortical white matter is seen in the right precentral area on MR imaging at 12 months of age (arrow, K). Left parietotemporal hypometabolism and right frontal hypometabolism on PET is seen at onset (arrow, L). Right frontal hypermetabolism disappears at 12 months of age (M). Left temporal hypometabolism is seen on PET at 24 months of age, whereas MR imaging shows no abnormality (arrow, N). Patient 4: both MR imaging (O) and PET (Q) findings are normal at onset. A hyperintense area in the right orbitofrontal white matter is seen on axial MR imaging at 24 months of age (arrow, P). Hypometabolism in the right orbitofrontal lobe corresponds to the MR imaging lesion (arrow, R and S) and is seen in the left temporal lobe on PET (arrows, R).
the MR imaging lesion (Fig 1D, -E). The EEG at 24 months of age showed repetitive focal spikes, polyspikes, and slow waves in the left frontal area, concordant with the MR imaging lesion. The paroxysmal epileptic discharges were less frequent at 36 months of age. At the last follow-up at 5 years of age, the patient was free of seizures with antiepileptic drugs (levetiracetam and topiramate). She had mild a developmental delay. Her developmental quotient by the Kinder Infant Development Scale at 5 years of age was 57.

**Patient 2.** The patient had late-onset spasms at 18 months of age, with otherwise no previous medical history of neurological abnormalities. T2WI at onset showed slight hyperintensity in the left frontal white matter (Fig 1F). PET at onset showed multifocal hypometabolism in the left frontal, left temporal, and right frontal areas (Fig 1G, -H). At 24 months of age, PET showed no abnormality, even at the site of the MR imaging lesion (Fig 1H). EEG at 24 months of age showed sporadic spikes, polyspikes, and sharp waves in the left frontal area, concordant with the MR imaging lesion. At the last follow-up at 5 years of age, the patient was free of seizures without antiepileptic drugs, except for febrile seizures. Her psychomotor development was normal. Her developmental quotient by the Kinder Infant Development Scale at 5 years of age was 95.

**Patient 3.** The patient had onset of spasms at 2 months of age. MR imaging findings at onset were normal (Fig 1I), but MR imaging at 12 months of age showed a hyperintense area in the right frontal subcortical white matter (Fig 1J). The hyperintense lesion was more clearly recognized at 24 months of age with progress of myelination in the surrounding white matter. PET at 4 months of age showed left parietotemporal hypometabolism and right frontal hypermetabolism (Fig 1L). At 12 and 24 months of age, hypermetabolism in the right frontal lobe disappeared (Fig 1M), while PET showed left temporal hypometabolism and MR imaging showed no abnormality (Fig 1N). The EEG at 24 months of age showed sporadic polyspikes in the left temporal lobe that were concordant with PET hypometabolism. At the last follow-up at 7 years of age, the patient was free of seizures without antiepileptic drugs. His intelligence quotient by the Wechsler Intelligence Scale for Children, Fourth Edition, was 84. His intelligence level was within normal limits, but he had an attention deficit/hyperactivity disorder.

**Patient 4.** The patient began having spasms at 3 months of age. She had no abnormalities on both MR imaging and PET at 4 and 12 months of age (Fig 1O, -Q). The right orbitofrontal MR imaging lesion and hypometabolism became evident at 24 months of age (Fig 1P, -R, -S). PET at 24 months of age showed hypometabolism in the left temporal lobe at a different site from the MR imaging lesion (Fig 1R). EEG at 24 months of age showed sharp waves in the right frontal and left temporal areas that were concordant with 2 areas of PET hypometabolism. At the last follow-up at 5 years of age, the patient was free of seizures without antiepileptic drugs. She had normal psychomotor development. Her developmental quotient by the Kinder Infant Development Scale was 91 at 5 years of age.

**DISCUSSION**

Recent developments in neuroimaging technology have made it possible to understand the causes of WS, such as FCD. However, it is still difficult to accurately diagnose microscopic lesions.

In the present study, longitudinal MR imaging and PET findings of patients with WS and subtle MR imaging lesions of FCD are presented. In 2 cases (patients 3 and 4), no MR imaging abnormality was visible in the early infantile period, but an MR imaging abnormality became visible later with progress of myelination in the surrounding white matter. PET showed additional regional hypometabolism remote from the MR imaging lesions. In patient 3, PET at the onset of WS also showed hypermetabolism in the MR imaging lesion that was visible later at 12 months of age.

Two of the 4 patients had negative MR imaging findings in the early infantile period, and they were regarded as having WS of unknown etiology at onset. Abnormal white matter hyperintensity became visible on the T2-weighted follow-up MR imaging. Signal change of the white matter on MR imaging occurs during the first 2 years of life as a result of myelination. Myelination is recognized as a high signal on T1WI and a low signal on T2WI. In general, myelination progresses from caudal to cephalad and from dorsal to ventral. Therefore, the subcortical white matter matured last (other than the in calcarine and Rolandic areas), proceeding from the occipital region anteriorly to the anterior frontal and temporal lobes. Changes of myelination on T2WI appear at 14–18 months of age in the midfrontal subcortical white matter and at 24–30 months of age in the anterior frontal subcortical white matter. In the present study, the lesion in the precentral area (patient 3) became visible at 12 months of age, and the lesion in the orbitofrontal area (patient 4) became visible at 24 months of age. The age at detection of the lesion on MR imaging matched the period of myelination. Before the maturational change of subcortical white matter, it is difficult to detect the characteristic findings of FCD, such as blurring of the gray-white matter junction and signal changes in subcortical white matter. Thus, all patients with WS of unknown etiology at diagnosis should be re-evaluated by MR imaging at 18–24 months of age.

Concerning FDG-PET findings, it has been reported that maturational change occurs in regional cerebral glucose metabolism, especially during the infantile period. During this period, the cerebral glucose uptake pattern progresses from the subcortical gray matter to the cortex and from posterior-to-anterior cortical areas. In early infancy, PET shows lower uptake in the frontal and temporal lobes than in the parietal and occipital lobes, and frontaltemporal pathologic hypometabolism can be missed in this period. In the present study, the frontal hypometabolism at onset in patient 4 was unclear, possibly due to the nature of the glucose metabolism pattern in the infantile period.

FDG-PET is useful for detecting FCD in WS. A previous study showed that PET was more sensitive than MR imaging for patients with a mild degree of cortical dysplasia, though MR imaging was performed using a 1.5T scanner. More recent reports with 1.5T or 3T MR imaging scanners have also shown the usefulness of PET in patients with MR imaging negative for FCD, especially with PET/MR imaging coregistration. Chugani et al showed that
PET hypometabolism was consistent with EEG localization of epileptogenic foci in patients with MR imaging negative for WS. Asano et al. also reported that the ictal electrocorticographic findings associated with spasms originated within a PET hypometabolic region. In the present study, PET showed hypometabolism or hypermetabolism that corresponded to MR imaging lesions, but also hypometabolism in areas remote from the MR imaging lesions. In 2 patients (patients 3 and 4), the remote hypometabolic areas were accompanied by focal independent epileptiform discharges on the interictal EEG. The remote hypometabolic areas with focal epileptiform discharges may indicate latent dysplastic lesions or secondary epileptogenic foci.

Even in the region of structural abnormalities such as FCD, PET findings evolve with time from hypometabolism to normal, hypermetabolism to normal, or normal to hypometabolism during the first years of life. Previous studies have demonstrated that FDG-PET showed hypometabolism in the region of FCD, usually corresponding to, but somewhat more extensive than, the MR imaging lesion. In contrast, cases with FCD lesions showing hypermetabolism have also been reported. While subclinical seizures during the scan possibly cause hypermetabolism on PET, it has been reported that hypermetabolism can be seen in regions with frequent epileptiform discharges, even in the interictal state, especially with FCD. While several studies have shown that PET has an advantage for detecting subtle FCD, Sankar et al. reported that PET was not as sensitive as follow-up MR imaging in patients with WS. The present observation of normal PET findings at the site of FCD at some points is consistent with the results of Sankar et al. Conversely, this serial PET study showed hypometabolism or hypermetabolism at the site of FCD on PET scans at least at 1 point in all patients. Evaluation of serial PET with follow-up MR imaging has a complementary effect in detecting subtle lesions.

Generally, patients with WS with FCD have poor seizure control by antiepileptic drugs and are candidates for resective surgery. However, all patients in the present study had favorable seizure outcomes without surgery, at least until preschool age, and the developmental outcome was also favorable in 3 patients. This difference may be explained by the different characteristics of the MR imaging and PET findings of FCD in the present cases. Typical MR imaging findings of FCD are a focal area of cortical thickness and blurring of the gray-white matter junction, with or without signal change in subcortical white matter on T2WI and FLAIR. All patients in the present study, however, showed only a hypointense signal on T2WI in the subcortical white matter, without cortical thickness or blurring of the gray-white matter junction. Additionally, PET hypometabolism was not consistently observed at the MR imaging lesion of FCD. The different neuroimaging characteristics and, probably, prompt initiation of adrenocorticotropic hormone with a short treatment lag after onset might be associated with the favorable outcomes in the present study.

The present study has some limitations. First, the diagnosis of FCD was not confirmed pathologically because the patients showed favorable seizure outcomes. The differential diagnosis of the MR imaging lesion may include benign tumor and polymicrogyria. However, none of the patients showed less cortical thickening and intravenous contrast enhancement than observed in brain tumors or an abnormal gyral pattern, increased cortical thickness, or irregularity of the cortical–white matter junction as seen in polymicrogyria. It has been reported that FCD has a variety of features on MR imaging studies and does not always show the characteristic signs, such as cortical thickening, blurring of the gray-to-white matter surface, and signal changes in the underlying white matter. From these observations, the MR imaging findings in the present study suggest FCD even without confirmation by histopathology. Second, a strong limitation of the present study is that the sample size was very small. Studies with a large number of patients are needed to draw conclusions on the usefulness of MR imaging and PET. Third, MR imaging and PET were assessed by visual inspection without using quantitative or statistical techniques (such as volumetric MR imaging analysis or quantification of glucose metabolism). Using these techniques may increase sensitivity and objectivity.

CONCLUSIONS

A case series of patients with WS with FCD who showed subtle or unrecognized MR imaging abnormalities at diagnosis was presented. The findings on MR imaging and PET in these patients evolved differently with brain maturation and the clinical course of WS. Their seizure outcomes were consistently favorable compared with previous reports. Serial MR imaging and PET studies may be useful to accurately diagnose the etiology and make the appropriate prediction of seizure outcomes in WS.

REFERENCES

1. Riikinson R. Epidemiological data of West syndrome in Finland. Brain Dev 2001;23:539–41 CrossRef Medline
12. Ballesteros MC, Hansen PE, Soila K. MR imaging of the developing
ABSTRACT

BACKGROUND AND PURPOSE: We present the largest case series to date on basiocciput abnormalities in CHARGE syndrome (Coloboma of the eye, Heart defects, Atresia of the choanae, Retardation of growth and/or development, Genital and/or urinary abnormalities, and Ear abnormalities and/or deafness). We aimed to show that basiocciput abnormalities are common and may aid in diagnosis. We furthermore explored whether clivus size correlates with the type of chromodomain-helicase-DNA binding protein 7 gene (CHD7) mutation, which causes CHARGE syndrome, and with clinical criteria according to Blake et al and Verloes.

MATERIALS AND METHODS: We retrospectively analyzed the clivus of 23 patients with CHARGE syndrome with CHD7 mutations on MR imaging or CT. We recorded the size of the clivus, the Welcher angle, basilar invagination, and Chiari I malformations. We compared the clival size and Welcher angle of patients with CHARGE syndrome with those of 72 age-matched controls. Additionally, we tested for correlations between clivus size and mutation type or clinical criteria.

RESULTS: Eighty-seven percent of the patients with CHARGE syndrome had an abnormal clivus; 61% had a clivus >2.5 SD smaller than that of age-matched controls. An abnormally large Welcher angle was observed in 35%. Basiocciput hypoplasia was found in 70%, and basilar invagination, in 29%. None of the patients had a Chiari I malformation. At the group level, patients with CHARGE syndrome had a smaller clivus and larger Welcher angle than controls. No significant correlation between clivus size and mutation type or clinical criteria was found.

CONCLUSIONS: Most patients with CHARGE syndrome have an abnormal clivus. This suggests that clivus abnormalities may be used as an additional diagnostic tool. Our results provide evidence that CHD7, which is expressed in the presomitic mesoderm during somitogenesis, plays an important role in the formation of the clivus.

ABBREVIATIONS: Ba-Xs = exosphenobasion; Ba-Es = endosphenobasion; CHARGE = Coloboma of the eye, Heart defects, Atresia of the choanae, Retardation of growth and/or development, Genital and/or urinary abnormalities, and Ear abnormalities and deafness; CHD7 = chromodomain-helicase-DNA binding protein 7 gene. The animal homologue is Chd7. Non-italicized CHD7 (human) and Chd7 (animal) refer to the protein.

Coloboma of the eye, Heart defects, Atresia of the choanae, Retardation of growth and/or development, Genital and/or urinary abnormalities, and Ear abnormalities and deafness (CHARGE) syndrome is a complex disorder with multiple congenital anomalies that occurs in approximately 6 in 100,000 live births. First described independently by Hall and Hittner et al in 1979, the acronym CHARGE was coined by Pagon et al in 1981. Many more features are associated with the syndrome, such as semicircular canal dysplasia, facial nerve palsy, anosmia with or without olfactory bulb hypoplasia, delayed puberty, and cleft lip/palate. Clinical criteria have been published by Blake et al in 1998, Verloes in 2005, and Hale et al in 2015, which aid in the clinical diagnosis (On-line Table 1). Guidelines for cranial imaging were published by de Geus et al in 2017. In 2004, CHARGE syndrome was found to be caused by mutations or deletions of the chromodomain-helicase-DNA binding protein 7 gene (CHD7) gene, and molecular confirmation currently plays a pivotal role in the diagnosis. Pathogenic mutations in the CHD7 gene usually occur de novo, though familial occurrence has been described.
complicates early diagnosis. CT and MR imaging in CHARGE may play an important role in the diagnosis by demonstrating congenital abnormalities of the labyrinth, which are present in almost all patients and can be assessed on CT and MR imaging.\(^\text{11}\) Olfactory bulb hypoplasia, cerebellar dysplasia, and other congenital brain abnormalities may be demonstrated on MR imaging, but they are not invariably present in all patients.\(^\text{12,13}\) Morphologic changes of the clivus in CHARGE syndrome were first described on neuroimaging by Fujita et al in 2009.\(^\text{14}\) A smaller size, a malformed shape, platybasia, basilar invagination, and Chiari I malformations have since been described by several authors.\(^\text{15-18}\) Normally, the body of the sphenoid occupies the upper portion of the clivus and is joined to the basilar occipital bone to form the complete clivus (Fig 1A).\(^\text{19}\) Steepness (of the clivus) may be quantified by the Welcher angle formed by a line through the frontal skull base and a line along the dorsal clivus (Fig 2A).

In the present study, we have elaborated on the previously published basiocciput findings by evaluating a large cohort of patients with molecularly proved CHARGE syndrome in comparison with age-matched controls. We further attempted to correlate clival size with mutation type and clinical criteria.

**MATERIALS AND METHODS**

**Patients**

The diagnosis of CHARGE syndrome was molecularly confirmed in all patients. \(\text{CHD7}\) nonsense and frameshift mutations and larger deletions were categorized as truncating (ie, mutations leading to a nonfunctional protein or no protein at all). \(\text{CHD7}\) missense mutations were categorized as nontruncating (ie, mutations leading to production of an altered protein that may have residual function). \(\text{CHD7}\) splice site mutations may have truncating or nontruncating effects, and these mutations could therefore not be further categorized. All patients were scored using the Blake et al\(^\text{5}\) and Verloes\(^\text{6}\) criteria (On-line Table 1).

Neuroimaging was performed in the authors’ hospital in 12 patients on a 1.5T scanner (Siemens, Erlangen, Germany) using a regular head coil. The remainder of the patients were scanned at other hospitals using different scanners (1–1.5T) of different brands using different protocols. Only patients with sagittal T1 2D TSE, 3D T1 MPRAGE, or a sagittal 2D T2 TSE imaging were included in this study. Three patients assessed at an external hospital had a CT scan with the possibility of sagittal reconstruction of a transverse CT scan of either a head scan or a mastoid scan.

All neuroimaging studies were assessed and measurements were made by an experienced pediatric neuroradiologist (L.C.M.).

**Controls**

Age-matched controls from 6 age groups (0–3 months and 1, 2, 6, 10, and 16 years of age) had been scanned for various neurologic and endocrine indications, not suspicious for CHARGE syn-

---

**FIG 1.** Clival abnormalities in CHARGE syndrome. A, Sagittal T1 scan of a 4.5-year-old boy without CHARGE syndrome. White lines show the measurement of the Ba-Es and Ba-Xs. B, A 22-month-old boy with CHARGE syndrome (patient 14). He has a hypoplastic sclerotome of the clivus (arrow) with a large Welcher angle. Although the top of the odontoid process of the dens does not extend cranially to the Chamberlain line, there is a slight angulation of the medulla oblongata without impression. C and D, Clival size versus age. White dots show length of the Ba-Es (C) and Ba-Xs (D) of the individual controls; black dots show the same parameters of individual patients with CHARGE syndrome.

**FIG 2.** Platybasia in patients with CHARGE syndrome. A, Sagittal T1 scan of a 4.5-year-old boy without CHARGE syndrome. The Welcher line is shown in white. B, A 22-month-old boy with CHARGE syndrome (patient 14). Note the large Welcher angle on the midsagittal T1 scan. C, Welcher angle versus age. White dots show the Welcher angle of the individual controls. Black dots show the Welcher angles of individual patients with CHARGE syndrome.
drome or skull base abnormalities, on a 1.5T MR imaging system (Siemens, Erlangen, Germany), in the authors’ hospital between 2002 and 2014. All controls had sagittal 2D or 3D T1-weighted MR imaging included in the scanning protocol. The brain scan findings had been assessed as normal by various experienced neuroradiologists, and at selection, this assessment was confirmed by an experienced pediatric neuroradiologist (L.C.M.).

Radiologic Analysis of the Clivus
Anatomic definitions and measurements of the clivus were used as described by Fujita et al.14 Clivus size was quantified by measuring the exosphenobasion (Ba-Xs) and endosphenobasion (Ba-Es, Fig 1A). This was done by measuring from the basion, the point of the clivus at the midpoint on the anterior margin of the foramen magnum, to the ventral (Ba-Xs) and dorsal (Ba-Es) margins of the visible synchondrosis.14 The Welcher angle is formed by the intersection between the nasion-tuberculum line and the tuberculum-basion line.20 Basioccipital hypoplasia has been defined as hypoplasia of ≥1 of the 5 clival segments (sclerotomes) of the clivus (simplified from Fujita et al14). Basilar invagination is commonly defined as cranial displacement of >5 mm of the tip of the odontoid above to the Chamberlain line. This line extends from the posterior margin of the foramen magnum anteriorly along the hard palate. Type 1 Chiari malformation is defined as herniation of at least 1 cerebellar tonsil ≥5 mm below the foramen magnum.

The exosphenobasion, the endosphenobasion, the Welcher angle, basilar invagination, and type I Chiari malformation were measured.

Statistical Analysis
The scans of the patients with CHARGE syndrome were compared with the findings on sagittal T1-weighted MR images of 72 controls in 6 age groups: 0–3 months and 1, 2, 6, 10, and 16 years of age.

Measurements of Ba-Es, Ba-Xs, and Welcher angle in controls were used to calculate age-specific mean and SD values. Patient measurements were compared with age-specific mean control values. A clivus was determined to be abnormally small if it was >2.5 SDs below its age-specific control. The Welcher angle was determined to be abnormally large if it was >2.5 SDs above its age-dependent control value.

Because the Ba-Es, Ba-Xs, and the Welcher angle were highly correlated (On-line Table 2), we orthogonalized the data with factor analysis. The first factor was then modeled with nonlinear regression on the normal data. To control for bias due to the difference in age distribution of patients with CHARGE syndrome and controls, we computed the observed minus the predicted values, which were then tested with a 2-sample t test (for additional statistical methods, see the On-line Appendix). A 2-sided Fisher exact test was performed to examine correlations between the size of the clivus and the type of mutation (truncating versus nontruncating), the presence of choanal atresia or coloboma, and satisfaction of the criteria of Verloes6 or Blake et al17 (listed in On-line Table 1). No t test was performed for the criteria of Hale et al17 because all patients satisfied these criteria.

RESULTS
In total, 23 patients with an age range of 3 days to 16 years (median age, 20 months) were included in this study. Table 1 summarizes the patients’ clinical criteria according to Blake et al, Verloes, and Hale et al and their type of CHD7 mutation. The full spectrum was represented in patients with both clinically typical and atypical CHARGE syndrome.

Figure 1C, -D summarizes the lengths of the Ba-Es and Ba-Xs for patients and controls with increasing age. For 2 patients, the presence of basilar invagination could not be reliably determined (patients 13 and 20 in On-line Table 3). Most patients with CHARGE syndrome (87%, 20/23) had a clivus that was small or had abnormal morphology or both. Fourteen patients had a clivus of >2.5 SDs smaller than in their age-matched controls (61%, 14/23). In 8 patients (35%, 8/23), an extra synchondrosis was seen. Just more than two-thirds (70%, 16/23) of the patients with CHARGE syndrome showed very short clivi with loss of the normal triangular shape (basiocciput hypoplasia), which was further illustrated by 9 of them having a Welcher angle of ≥2.5 SDs above that in controls, indicating platybasia (39%, 9/23; Fig 2C). The Welcher angle varied between 124° and 176° (mean 140°, SD 11.2°).

The results for the comparison at the group level are shown in On-line Tables 2 and 4 and the On-line Figure. Most of the patients with CHARGE syndrome showed a reduced value in factor 1. Factor 1 had a positive correlation with the Ba-Es and Ba-Xs and a negative correlation with the Welcher angle. Despite some patients with CHARGE syndrome having clearly normal values, as a group they had significantly lower values (P = 6 × 10^-6), which correspond with a smaller Ba-Es and Ba-Xs and larger Welcher angles.

Six (29%, 6/21, missing data n = 2) patients with CHARGE syndrome showed basilar invagination. In 2 of these patients, a minor impression of the craniovertebral junction on the medulla
oblongata was suggested (see Fig 1B for an example). None of the patients exhibited herniation of the cerebellar tonsils.

Table 2 shows the correlation among clivus size, clinical criteria, and type of mutation. No significant correlation was found between the size of the clivus in the patients with CHARGE syndrome and the mutation type (truncating versus nontruncating), occurrence of facial clefts, ocular coloboma, atresia of the choanae, or satisfaction of the Verloes or Blake et al criteria. DISCUSSION

In this study, the presence of clival abnormalities was assessed in a large group of patients with CHARGE syndrome and compared with that in healthy controls.

Clival abnormalities in CHARGE syndrome have been published previously. Fujita et al14 were the first to publish examples of basioccipital hypoplasia in 7/8 patients (88%) and associated basilar invagination in 5 (63%). Furthermore, one of their patients exhibited a Chiari I malformation and syringomyelia. Nattung et al15 described a case with a short clivus, fused cervical vertebrae, occipitalization of the atlas, and basilar invagination. Hoch et al16 found skull base hypoplasia in 9/10 and a dorsally angulated clivus in 7/10 patients. Mahdi and Whitehead17,18 described a child with a coronal clival cleft in 2017 and recently published a study consisting of 15 genetically and clinically confirmed CHARGE cases, in which they reported a coronal cleft in 13 (87%) patients and clival hypoplasia without a cleft in the remaining 2 patients.

Although we did not particularly assess the coronal cleft, the prevalence of clival abnormalities and/or skull base hypoplasia in these articles is similar to our numbers. We found that 20/23 patients with CHARGE syndrome (87%) had an abnormal clivus, either in morphology or size. At the group level, patients with CHARGE syndrome had a smaller clivus and larger Welcher angle (On-line Figure). However, only 14/23 (61%) had an abnormally small clivus, defined as >2.5 SDs smaller than that of age-matched controls (On-line Figure). Basilar invagination was seen in only 6 patients (29%, 6/21, missing data in 2 patients), with only 2 of these patients showing a minimal impression on the ventral medulla oblongata. None of the patients in the present study had a Chiari I malformation. If we combined our study and the above-mentioned 2 case series, 51 of 56 patients with CHARGE syndrome (91%) had clival abnormalities and/or skull base hypoplasia, underscoring the potential of this feature as a diagnostic tool in CHARGE syndrome.

However, because all of these case series represent nonrandomly selected samples, there is a danger of overestimation. All samples may be biased toward the more severe end of the clinical spectrum because more severely affected patients may be more likely to undergo cerebral imaging. Our data do not show correlation between the severity of the disorder (satisfaction of clinical criteria) and the presence of clival abnormalities. In fact, in our case series, a large number of patients were atypical on the Verloes criteria (5/17) or negative on the Blake et al criteria (8/20), yet clival abnormalities were found in most. This finding underscores the importance of clival abnormalities on imaging in supporting the diagnosis, especially in mildly affected patients.

Basilar invagination in patients with CHARGE syndrome may be of clinical importance because it may cause compression of the medulla with ensuing clinical symptoms. No obvious neurologic symptoms that could be attributed to the basilar invagination were reported in the clinical data of the 2 patients who showed possible involvement of the medulla (patients 14 and 15 in On-line Table 3). Only 1 of the patients in the series of Fujita et al14 had neurologic sequelae, but she was reported to have syringomyelia in addition to basilar impression.

In our study, hypoplasia of the clivus was suggested as an all-or-nothing event: If hypoplasia was present, the degree of hypoplasia was severe (Ba-Es of −4 to −10 SDs smaller than in controls, Fig 1 and On-line Table 3). The high variability and incomplete penetrance of specific features are well-known aspects of CHARGE syndrome.21 CHARGE syndrome is exclusively caused by mutations in CHD7. The chromodomain helicase DNA-binding protein 7 (protein CHD7) is essential in embryologic development, and mutations in CHD7 result in a wide range of features with incomplete penetrance. Much of the clinical variability is still unexplained, though there is some correlation between clinical severity and mutation type: Patients with a missense CHD7 mutation generally have a milder presentation of clinical features.22 In our cohort, no correlations were found between clivus length and mutation type, satisfaction of all CHARGE criteria, or specific symptoms. Milder missense mutations are, however, fairly rare in CHARGE syndrome. This is reflected in our study because only 2/23 patients had a missense mutation.

The precise function of the CHD7 protein in the formation of

<table>
<thead>
<tr>
<th>Variable</th>
<th>Size of Clivus</th>
<th>&lt;2.5 SDs Compared with Controls</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Truncating mutation (total n = 19)</td>
<td>Normal</td>
<td>5</td>
<td>12</td>
</tr>
<tr>
<td>+</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Presence of choanal atresia</td>
<td></td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>+</td>
<td>8</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Presence of coloboma (total n = 22)</td>
<td></td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>+</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Presence of cleft (total n = 22)</td>
<td></td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>+</td>
<td>6</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Verloes criteria satisfied</td>
<td>+</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>-</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Blake et al criteria satisfied</td>
<td>+</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>-</td>
<td>4</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

Note: + indicates yes; −, no. a Four patients had a splice site mutation and could not be classified as either truncating or nontruncating. b For 1 patient, no data regarding presence of coloboma were available. For 1 patient, no information regarding cleft lip/palate was available. c See also On-line Table 1.
the skull base is unknown. CHD7 is expressed in the presomitic mesoderm during somitogenesis,23 in which it plays a role in controlling left-right symmetry. Somitogenesis is an important process in the formation of the clivus, which is formed from 4 occipital somites through a complicated process. In this process, fusion of the first 3 somites creates the rostral basiocciput.24 After formation of a transient sclerotome called the proatlas, parts of the fourth somite then form the basion. Chd7 knockout zebrafish exhibit irregularly shaped vertebral arches,23,25 supporting the role of CHD7 in somitogenesis. The altered anatomy of the clivus in many patients with CHARGE syndrome may therefore reflect errors in somitogenesis due to faulty CHD7 signaling.

This study has several limitations. It was based on retrospective assessment of MR imaging and CT scans obtained at different hospitals using different scanning protocols and image parameters. Nevertheless, a sagittal MR imaging scan or a sagittal CT reconstruction, on which the clivus could be assessed, was always available. However, in 2 cases, the clivus was difficult to assess. An altered anatomy of the remainder of the skull base, also described by Natung et al.,19 made the definition of clival borders difficult in several cases. Extreme clival hypoplasia also limited accurate measurements.

CONCLUSIONS
This is the largest case-control series on clivus abnormalities in CHARGE syndrome, to our knowledge. Although the clinical relevance of clival hypoplasia and platybasia in CHARGE syndrome is not yet clear, the results of this study confirm the suggestion by Mahdi and Whitehead18 that besides the well-known labyrinthe anomalies and hypo- or aplasia of the olfactory bulb,26 clival abnormalities may provide an important extra clue for the diagnosis of CHARGE syndrome in neuroimaging studies. We hypothesize that CHD7 may play an important role in the development of clival hypoplasia.

ACKNOWLEDGMENTS
We thank J.Senior for editing the manuscript and G.J. te Meer- man, biostatistician, for his valuable help with the statistical analysis.

Disclosures: Christa M. de Geus—RELATED: Grant: CHARGE Syndrome Foundation pilot grant.* Grant: University Medical Center Groningen GrAGIKO grant. Jonieke E. H. Bergman—RELATED: Grant: Netherlands Organization for Health Research and Development, Comments: grant no. 92003460 to J.E.H. Bergman for a PhD (the PhD was obtained between 2007 and 2011).* *Money paid to the institution.

REFERENCES
Multiple Brain Developmental Venous Anomalies as a Marker for Constitutional Mismatch Repair Deficiency Syndrome

S.I. Shiran, L. Ben-Sira, R. Elhasid, J. Roth, U. Tabori, M. Yalon, S. Constantini, and R. Dvir

ABSTRACT

BACKGROUND AND PURPOSE: Biallelic constitutional mutations in DNA mismatch repair genes cause a distinct syndrome, constitutional mismatch repair deficiency syndrome (CMMRD), characterized by cancers from multiple organs, most commonly brain tumors, during childhood. Surveillance protocols include total and brain MR imaging among other modalities to enable early detection of tumors. Brain surveillance scans revealed prominent brain developmental venous anomalies (DVAs) in some patients. DVAs are benign vascular anomalies, and their incidence in the general population is 2.6%–6.4%. Most developmental venous anomalies are asymptomatic and are found incidentally. Our purpose was to assess the prevalence of DVAs in CMMRD patients and describe their phenotype.

MATERIALS AND METHODS: A retrospective descriptive analysis of brain MR imaging studies from 10 patients from 3 families with CMMRD was performed. Analysis included the number of developmental venous anomalies, location, draining vessels, and associated vascular anomalies (ie, cavernomas), with clinical correlation of symptoms and tumors.

RESULTS: All 10 patients had ≥2 developmental venous anomalies, and 2 had, in addition, non-therapy-induced cavernomas. There was no clinically symptomatic intracranial bleeding from developmental venous anomalies. Six patients had malignant brain tumors. The location of brain tumors was not adjacent to the developmental venous anomalies. No new developmental venous anomalies developed during follow-up.

CONCLUSIONS: The occurrence of multiple developmental venous anomalies in all our patients with CMMRD suggests that developmental venous anomalies may be a characteristic of this syndrome that has not been previously described. If confirmed, this quantifiable feature can be added to the current scoring system and could result in early implementation of genetic testing and surveillance protocols, which can be life-saving for these patients.

ABBREVIATIONS: CMMRD = constitutional mismatch repair deficiency syndrome; DVA = developmental venous anomaly

Constitutional mismatch repair deficiency syndrome (CMMRD) is a cancer-predisposition syndrome characterized mainly by a high risk for developing cancer in childhood and young adulthood as well as nonmalignant features. The most common malignancies are brain tumors (predominantly malignant gliomas, though other tumors are reported), lymphoid malignancies (most commonly non-Hodgkins lymphomas and leukemia), and gastrointestinal cancers (Lynch syndrome–associated tumors, especially colorectal cancer). Nonmalignant manifestations of the syndrome include benign tumors such as adenomas and neurofibromas, features of neurofibromatosis type 1, predominantly café au lait macules, other hypo- and hyperpigmented skin alterations, pilomatrixomas, and other features that were included in the suggested diagnostic criteria by the European Consortium. Because CMMRD is a cancer-predisposition syndrome with a very high penetrance and most reported individuals are affected during childhood, surveillance protocols were developed aiming at early detection and interventions of these cancers. MR imaging...
of the brain, every 6 months, starting at diagnosis or birth, is part of the surveillance protocol recommended for these children by the European Consortium for CMMR-D and the International Biallelic Mismatch Repair Deficiency Consortium. It is important to arrive at a definite diagnosis of CMMR-D in a pediatric or young adult patient with cancer as early as possible to allow the recommended surveillance for the patient and genetic counseling for family members. To facilitate CMMR-D testing in patients with cancer, the European Consortium Care for CMMR-D suggested a score based on diagnostic criteria. In the suggested score, several features of the syndrome are assigned points; a patient with a 3-point score or above is referred for genetic counseling. In Tel Aviv Sourasky medical center, we noted prominent developmental venous anomalies (DVAs) on routine brain MR imaging performed as a part of our surveillance protocol in patients with CMMR-D.

DVA is the most frequently encountered cerebral vascular malformation, with an incidence of 2.6%–6.4% in different studies. A DVA is characterized by a cluster of venous radicles that converge into a collecting vein, resulting in the typical caput medusa appearance of the DVA. The collecting vein crosses a variable length of brain parenchyma to join either the superficial or deep venous system. Two or more DVAs coexisting in separate regions of the brain were observed in 7%–16% of described patients with DVAs. Most DVAs are asymptomatic and are found incidentally. However, DVAs have been documented as a rare cause of cerebrovascular bleeding and ischemic events. There is a known association between DVAs and cavernomas.

Our aim was to assess the prevalence of DVAs in children with CMMR-D and to describe their phenotype to elucidate possible distinct features associated with CMMR-D.

MATERIALS AND METHODS

We performed detailed clinical and imaging analysis of all children from 3 different families that are carriers of DNA mismatch repair genes, each family with a distinct type of mutation. The first family has a PMS2 mutation (c.2458dupA), with 5 affected homozygous members. The second family has a MSH6 mutation (c.2314c>T (p.Arg727Trp)), with 2 affected homozygous members. The third family has 2 different pathogenic mutations in MSH6 (c.3984_3987dupGTCA/c.3959_3962delCAAG), with 3 affected members who are compound heterozygotes.

All children had surveillance brain MR imaging during the past 10 years. This retrospective study was approved by Tel Aviv Sourasky medical center institutional review board. A retrospective descriptive analysis of their brain MR imaging studies included the number of DVAs, location, draining vessels (peripheral or central), length and diameter of the collecting vein, associated parenchymal changes (as reflected by increased T2WI signal intensity), and the presence of associated cavernomas. Correlation with coexisting brain tumor presence and location was also performed. Descriptive statistics were performed with SPSS (IBM, Armonk, New York).

RESULTS

Brain MR imaging studies of 10 children with CMMR-D were evaluated. The complete patient information is presented in the On-line Table. There were 6 boys and 4 girls. The age at the earliest study available per child ranged from 1.0 to 12.0 years (mean, 6.5 years). Seven children were treated for CNS tumors (4 glioblastoma multiforme, 1 medulloblastoma, 1 gliomatosis cerebri, and 1 cord primitive neuroectodermal tumor), 1 child was treated for acute myeloid leukemia (AML), and 2 children were treated for lymphoma. Colon polyposis was present in 6 children, and adenocarcinoma of the colon, in 1 child. DVAs were noted in all children, constituting 100% prevalence. Five children had 2 DVAs, 1 child had 3 DVAs (Fig 1), 3 children had 5 DVAs, and 1 child had 7 DVAs. In total, 35 DVAs were observed; 83% were in the supratentorial brain and 64% drained to peripheral cortical veins. The diameters of the collecting veins ranged from a minimum of 0.8 mm to a maximum of 3.0 mm (average, 1.9 mm; median, 1.8 mm; mode, 2.2 mm). The length of the collecting vein ranged from a minimum of 5.0 mm to a maximum of 39.0 mm (average, 24.9 mm; median, 27.0 mm; mode, 30.0 mm). There were T2 signal changes in the brain parenchyma adjacent to the DVA in 26% of DVAs (Fig 2).

In children with brain tumors, there was no correlation between the tumor locations and the location of the DVA. Cavernomas were present in 2 patients. One patient with a brain tumor had a single small cavernoma adjacent to the DVA before therapy; on follow-up studies, after tumor treatment with radiation therapy, he developed focal bleeding in the known cavernoma and bleeding in additional therapy-induced multiple cavernomas that were not apparent on the initial study. A second patient had 2 non-therapy-induced cavernomas present but distant from the DVA. For 9 children, follow-up studies were available; the time between first and last available studies ranged from 0.2 to 7.5 years, and there was no change in the number of DVAs on follow-up studies.

DISCUSSION

In this study, we observed 100% prevalence of multiple DVAs in children with CMMR-D. This is strikingly different from the prevalence of DVAs in the healthy population, which was reported as 2.7% in a study based on postmortem evaluation, and 6.4% in a more recent study based on modern MR imaging. Gökçe et al evaluated 1165 patients with brain MR imaging studies and found DVAs in 75 patients; of those, 10 had multiple DVAs, which is 0.9% of their total evaluated population and 13.3% of the group of patients who had DVAs. The phenotypic characteristics of the DVAs in their study included 73% supratentorial location, 50% drainage to peripheral veins, and a collecting vein diameter range of 1.0–4.3 mm (median, 2.0 mm). These findings are not significantly different from ours with 83% supratentorial location, 64% drainage to peripheral veins, and a collecting vein diameter range of 0.8–3.0 mm (median, 1.8 mm). We found associated T2 signal changes in the parenchyma adjacent to the DVA in 26% of the DVAs, which is within the range of those reported in different studies: 7.8%, 11.6%, and 28%. The phenotypic similarities of the DVAs in our study group compared with those reported in the general population exclude morphologic unique features of this vascular malformation in CMMR-D patients.

An increased prevalence of DVAs was reported in association with extensive head and neck venolymphatic malformations. In a
series of 40 patients with facial venous malformations, 20% had associated DVAs and 12.5% had multiple DVAs. In a series of 33 patients with orbital and periorbital lymphatic or venolymphatic malformations, up to 60% of patients had associated DVAs. We did not encounter associated head and neck vascular malformations in our study population, and we did not find an association between CMMRD and vascular malformations in the literature. It has been suggested that DVAs may be part of intracranial manifestations of neurocutaneous disorders because there are case reports describing an association between DVAs and blue rubber bleb nevus syndrome. Patients with CMMRD have not been reported to have any association with blue rubber bleb nevus syndrome, and a family history of malignancies in young relatives, especially brain tumors, lymphomas, or leukemias. Indeed, these and additional clinical manifestations were recently suggested to serve as criteria for a clinical diagnosis of CMMRD. Our data suggest that the prevalence of multiple DVAs in an MR imaging of the brain with an associated brain tumor or other CMMRD-related malignancy can be another clue to the diagnosis. This marker can be especially helpful to determine the clinical diagnosis of CMMRD in individuals who are unaffected by cancer and in whom the genetic tests are inconclusive. Most interesting, in 2 patients, in whom there was clinical suspicion of CMMRD, the lack of DVAs was associated with negative molecular evidence of CMMRD (unpublished data).

The main limitation of this study is the small group of patients.
Further assessment of larger patient cohorts from the international consortia will further delineate the exact prevalence of DVAs and the characteristics of DVAs in this patient population. However, the presence of multiple DVAs in all patients in this study has no parallel description in any other pathology, supporting the important role of DVAs for the clinical diagnosis of CMMRD.

CONCLUSIONS

Our cohort had 100% prevalence of multiple CNS DVAs in patients with CMMRD. If confirmed, this quantifiable feature can be added to the current scoring system and result in early implementation of genetic testing and surveillance protocols, which can be life-saving for these patients and families as well as allowing targeted tumor management.²⁰,²¹

ACKNOWLEDGMENTS

We thank the International Biallelic Mismatch Repair Deficiency Consortium for their support in this work. This research was further supported by Meagan’s Walk, the CIHR Joint Canada-Israel Health Research Program, and an American Association for Cancer Research Stand Up 2 Cancer Catalyst TM Research Grant.

REFERENCES

Spinal Imaging Findings of Open Spinal Dysraphisms on Fetal and Postnatal MRI


ABSTRACT

BACKGROUND AND PURPOSE: Fetal MRI has become a valuable tool in the evaluation of open spinal dysraphisms making studies comparing prenatal and postnatal MRI findings increasingly important. Our aim was to determine the accuracy of predicting the level of the spinal dysraphic defect of open spinal dysraphisms on fetal MR imaging and to report additional findings observed when comparing fetal and postnatal MR imaging of the spine in this population.

MATERIALS AND METHODS: A single-center retrospective analysis was performed of fetal MRIs with open spinal dysraphisms from 2004 through 2016 with available diagnostic postnatal spine MR imaging. Images were reviewed by 2 board-certified fellowship-trained pediatric neuroradiologists. Corresponding clinical/operative reports were reviewed.

RESULTS: One hundred nineteen fetal MRIs of open spinal dysraphisms were included. The level of the osseous defect between fetal and postnatal MR imaging was concordant in 42.9% (51/119) of cases and was 1 level different in 39% (47/119) of cases. On postnatal MR imaging, type II split cord malformation was seen in 8.4% (10/119) of cases, with only 50% (5/10) of these cases identified prospectively on fetal MR imaging. Syrinx was noted in 3% (4/119) of prenatal studies, all cervical, all confirmed on postnatal MR imaging.

CONCLUSIONS: Fetal MR imaging is accurate in detecting the level of the spinal dysraphic defect, which has an impact on prenatal counseling, neurologic outcomes, and eligibility for fetal surgery. In addition, fetal MR imaging is limited in its ability to detect split cord malformations in patients with open spinal dysraphisms. Although rare, fetal MR imaging has a high specificity for detection of cervical spinal cord syrinx.

ABBREVIATION: OSD = open spinal dysraphism

Fetal MR imaging has been well-established as a powerful tool in the prenatal evaluation of the neuroaxis and continues to play an increasing role in prenatal diagnosis, management, and counseling. The Management of Myelomeningocele (MOMS) trial remains the sentinel work driving more centers across the country to offer prenatal repair of open spinal dysraphisms (OSDs), for which fetal MR imaging has become an essential part of the work-up, guiding clinical management.

Despite the heavy reliance on fetal MR imaging in the evaluation of open spinal dysraphisms and its use in the selection of candidates for fetal surgery, the scientific literature examining MR imaging of the fetal spine is limited, and there are very few studies that compare pre- and postnatal MR imaging findings.

Determining the level of the defect is a key component of the inclusion criteria for fetal surgery and provides valuable information for prognosis and prenatal counseling.

Our aim was to determine the accuracy of predicting the upper level of the spinal dysraphic defect of open spinal dysraphisms on fetal MR imaging and to report additional findings observed when comparing fetal and postnatal MR imaging of the spine.

MATERIALS AND METHODS

Study Design

This study is a single-center, retrospective chart review. The case list was manually compiled from all the fetal MRIs performed at Cincinnati Children’s Hospital Medical Center in Cincinnati,
Ohio, between 2004 and 2016. Inclusion criteria were fetuses with diagnostic-quality fetal MRIs for OSD (either myelomeningocele or myelocole). Only fetuses with adequate available postnatal neuroimaging and clinical/neurosurgical follow-up were included. Criteria for adequate postnatal neuroimaging for this study included a diagnostic-quality MR imaging of the spine within the first 3 months of life, which was used as the gold standard in this study. Determination of diagnostic-quality imaging was made at the neuroradiologists’ discretion. The images were viewed in the PACS. A chart review was performed to obtain relevant clinical data. This study was compliant with the Health Insurance Portability and Accountability Act and was approved by the institutional review board. The requirement for informed consent was waived.

Scanning Parameters
All fetuses included in our study were scanned prenatally on a 1.5T magnet at Cincinnati Children’s Hospital Medical Center by using a Ingenia 1.5T (Philips Healthcare, Best, the Netherlands) or 1.5T Signa HDxt (GE Healthcare, Milwaukee, Wisconsin) system. The spinal dysraphism protocol included axial, sagittal, and coronal T2-single-shot FSE and balanced fast-field echo/FIESTA images of the neuroaxis at 3- to 4-mm slice thickness with no skip. Although this imaging protocol did not change during the study period, the TR and TE varied on each scanner and were changed at times of scanner upgrades to optimize image quality. At least 2 stacks of images in each plane were obtained to the radiologists’ satisfaction. The smallest FOV possible was used. Postnatal MR imaging was performed on 1 of 6 inpatient clinical magnets with evolving imaging protocols across the years. By 2014, T2-single-shot FSE images were largely replaced by T2-FSE images, and by 2015, true fast imaging with steady-state precession/FIESTA images became a routine part of the postnatal spinal dysraphism protocol.

Image Interpretation
All images were reviewed by 2 board-certified radiologists (U.D.N., B.M.K.-F.), both with added qualifications in pediatric radiology and fellowship training in pediatric neuroradiology. The readers were blinded to the pre- and postnatal imaging findings and reported fetal sonographic findings at the time of interpretation. Differences were resolved by consensus.

Multiple imaging parameters were evaluated on fetal MR imaging of the spine. The superior level of the spinal dysraphic defect was determined on fetal MR imaging by establishing the most caudal spinal hyperintense disc space as L5–S1 and the lowest horizontal vertebral body as L5 and by counting vertebral bodies superior to the highest level of the absence of the posterior elements at the bone/skin defect. On postnatal MR imaging, the defect level was determined superior to inferior by examining the entire spine, with the odontoid being labeled C2 and assuming 7 cervical vertebral bodies and 12 thoracic vertebral bodies. The presence of an arachnoid cyst, defined as an intradural extramedullary thin-walled CSF-intensity fluid collection, was documented. The presence or absence of a measurable postoperative fluid collection, visible spinal cord syrinx, or type II split cord malformation was also documented.

Additional intracranial findings on fetal MR imaging were recorded, including the degree of posterior fossa hindbrain herniation by Chiari grades 1–3, with grade 1 having a patent fourth ventricle and cisterna magna, grade 2 having effacement of the fourth ventricle with a patent cisterna magna, and grade 3 having effacement of both the cisterna magna and the fourth ventricle. Lateral ventricular size, by measuring the transverse atrial diameter on an axial or coronal image, and third ventricular size, by measuring the transverse diameter on a coronal image, were also recorded. The presence or absence of a clubfoot and, when present, whether it was unilateral or bilateral were also routinely reported.

Statistical Analysis
Descriptive analyses were performed to demonstrate the distribution of the imaging findings. Continuous variables were presented as means ± SDs, and categoric variables were presented as number (percentage). A 2-sample t test or 1-way ANOVA was used to detect the differences in continuous variables among different groups. The correlation among categoric variables was assessed by the χ² or Fisher exact test when appropriate. All analyses were performed using the Statistical Analysis System, Version 9.4 (SAS Institute, Cary, North Carolina). A P value < .05 was considered statistically significant.

RESULTS
Description of Cohort
A total of 119 fetuses (52 male, 67 female) met the criteria and were included in this analysis. The average gestational age at fetal MR imaging was 23.9 ± 3.6 weeks in the cohort as a whole. While 31.9% (38/119) of fetuses underwent open in utero repair of OSD, the remaining 68.1% (81/119) underwent postnatal repair. The average age at postnatal spine MR imaging was 19.9 ± 20.2 days. These and other descriptors of the cohort are summarized in Table 1.

Imaging Findings
Level of the Defect. The upper level of the osseous defect on fetal MR imaging ranged from as high as T10 to as low as S3. Most patients had lumbar defects, 68.9% (82/119), with the largest per-
percentage being at L4 (23.5%, 28/119); 13.4% (16/119) had a defect at S2 or lower (Fig 1); 42.9% (51/119) had the same defect level interpreted on postnatal spinal MR imaging compared with the fetal MR imaging; 39.5% (47/119) were 1-level discrepant; 11.8% (14/119) were discrepant at 2 levels; and 5.9% (7/119) were 3-levels discrepant (Fig 2). Of the discrepant levels, 42.6% (29/68) were interpreted as lower than on the postnatal MR imaging and 57.4% (39/68) were higher on postnatal MR imaging; 13.4% (16/119) of patients were interpreted as having a defect at S2 or S3, which would preclude them from fetal surgery; 62.5% (10/16) of these patients were concordant to the exact level on postnatal MR imaging and 25% (4/16) were higher and 12.5% (2/16) were lower.

There was a significant correlation between the level of the defect and lateral ventricular size. There was no significant correlation between defect level and Chiari grade, third ventricular size, and the presence of a clubfoot (Table 2).

**Split Cord Malformation and Syrinx.**

On fetal MR imaging, the possibility of a type II split cord malformation was raised in 13% (16/119) of patients; however, only 31% (5/16) were confirmed on postnatal MR imaging. On postnatal MR imaging, split cord malformation was seen in 8.4% (10/119) of cases, with only of 50% (5/10) of these cases identified prospectively on fetal MR imaging (Fig 3). Syrinx was noted in 3% (4/119) of prenatal studies, all cervical, all confirmed on postnatal MR imaging; however, fetal MR imaging was performed after the second trimester in 75% (3/4) of these patients (Fig 4).

**Postnatal Spine MR Imaging.** Postnatal MR imaging was performed in all patients at an average age of 19.9 ± 20.2 days; 31.9% (38/119) of patients had undergone prenatal repair; 65.5% (78/119) had undergone postnatal repair; 2.5% (3/119) of patients had spine MR imaging before postnatal repair; 39% (47/119) of postnatal spine MRIs had evidence of spinal cord syrinx, and of these subjects 29.8% (14/47) had undergone prenatal repair, 68% (32/47) had undergone postnatal repair, and 2% (1/47) was imaged before postnatal repair. Measurable postoperative extraspinal fluid collections were seen in 32.8% (38/116) of postnatal spine MRIs, 5.3% (2/38) of these patients had undergone prenatal repair; 3.4% (4/119) of patients had intraspinal arachnoid cysts on postnatal spine MR imaging, 50% (2/4) of whom had undergone prenatal repair; 36.2% (42/116) of patients did not have any evidence of syrinx, fluid collection, or arachnoid cyst on postnatal spine MR imaging, 50% (21/42) of whom had undergone prenatal repair (Table 3).

**DISCUSSION**

We describe our experience with spine imaging findings of OSD on fetal and postnatal MR imaging and have made several observations: Concordance within 1 level of the osseous defect was seen in 82% of patients between pre- and postnatal MR imaging. There was a significant correlation between the level of the defect and lateral ventricular size (p = 0.001), with a higher defect level correlating with larger lateral ventricular size. There was no significant correlation between defect level and Chiari grade, third ventricular size, and the presence of a clubfoot (Table 2).

**FIG 1.** Distribution of defect levels on fetal MR imaging.

**FIG 2.** Sagittal balanced fast-field echo/FIESTA image from fetal MR imaging at 24 weeks’ 5 days’ gestational age (A) demonstrates findings of a lumbosacral myelomeningocele with the upper level of the spinal dysraphic defect beginning at L4. Sagittal T2-FSE image from postnatal MR imaging of the spine in the same patient at 20 days of age (B) shows postoperative changes from OSD closure, with the upper level of the defect again beginning at L4 (same level), with L3 having an intact neural arch (arrow).
There is literature describing the correlation of the prenatal defect level with the postnatal neurologic function, making the identification of the defect level prenatally of potential importance in counseling.\textsuperscript{20} Despite previous sonographic literature describing no significant relationship between defect level and ventricular size, our study illustrates how a higher spinal defect is associated with larger lateral ventricular size on fetal MR imaging in a larger sample size.\textsuperscript{21} This finding is of potential clinical significance because larger ventricles on prenatal imaging have been shown to be associated with an increased need for shunting in those undergoing fetal surgery.\textsuperscript{22} On the contrary, despite previous literature describing a higher incidence of foot deformity with higher lesion levels, we did not observe a significant correlation between the fetal spinal defect level and the presence of clubfoot deformity in our cohort.\textsuperscript{23} This finding suggests that the pathophysiology of this disease process is complex and influenced by multiple factors, and further studies may be helpful in improving prenatal counseling. Our study also adds the observation that there was no significant correlation between defect level and Chiari grade or third ventricle size.

Prenatal diagnosis of split cord malformation, also known as diastematomyelia, has been described in a few isolated cases; however, the diagnostic reliability of fetal MR imaging in a larger population is still unclear.\textsuperscript{24,25} Our series adds to the literature by demonstrating a relatively limited ability of fetal MR imaging to identify type II split cord malformation. The lack of our ability to consistently identify split cord malformation prospectively on fetal MR imaging can likely be explained by the very small size of the fetus being imaged. However, it is less clear why cases in which split cord malformation was questioned prenatally were not confirmed on postnatal MR imaging. It is possible that clumped nerve roots from the ventral neural placode or a prominent anterior median fissure of the spinal cord was mistaken for split cord malformation on fetal MR imaging.\textsuperscript{26} Another possibility is that postoperative changes from OSD closure obscure the underlying split cord malformation on postnatal MR imaging. While the association between type II split cord malformation and open spinal dysraphisms is known and does not preclude fetal surgery, knowing that our ability to diagnose this prenatally is limited may be helpful in counseling. There are notable clinical implications of split cord malformation in this population: Not only do these patients often require more extensive surgical untethering early in life or at the time of primary repair, but they are also at increased risk of tethered cord syndrome and progressive scoliosis postnatally.\textsuperscript{28-31}

Our study also adds to the literature a description of imaging findings on postnatal spine MR imaging in patients with OSD after both prenatal and postnatal repair. Most notably, we observed that syrinx can be seen in association with open spinal dysraphism on fetal MR imaging.\textsuperscript{32} We describe its incidence in a relatively large sample of fetuses with open spinal dysraphism (3.4%, 4/119) and demonstrate a high specificity of fetal MR imaging, particularly in the third trimester, because all fetuses with a prenatally diagnosed syrinx in our series had the same findings on
postnatal MR imaging. The increased incidence of syrinx on postnatal MR imaging (39.5%, 47/119) compared with prenatal MR imaging (3.4%, 4/119) is likely explained by ongoing associated abnormalities of CSF flow dynamics in this population, including hindbrain herniation, hydrocephalus, and tethered cord rather than a missed congenital syrinx.\textsuperscript{13} Although postoperative fluid collections were more frequently seen in the postnatal repair group (46.2%, 36/78) compared with the prenatal repair group (5.3%, 2/38), this finding can likely be explained by the relative time of imaging after the operation. We also add our observed incidence of intraspinal arachnoid cysts on postnatal spine MR imaging in this patient population (3.4%, 4/119), which was seen equally in those that underwent prenatal versus postnatal repair. Our study has some limitations. First, this is a retrospective study, which limits its internal validity. Also, given that this is a single-institution study performed within a certain timeframe, its external validity may be limited as well. Along those same lines, this study is also likely subject to some degree of selection bias because the data were collected from one of the largest referral centers in the country for prenatal repair of open spinal dysraphisms. Also, given that imaging studies were acquired during a 12-year period on multiple different clinical magnets with periodic upgrades, the heterogeneity of scanning parameters may affect our results as well. In clinical practice, sonography is often used in conjunction with fetal MR imaging to determine the defect level of an open spinal dysraphism, and in our practice, we emphasize the sonographic findings for defect level in fetal counseling because some believe that sonography can better delineate the osseous structures of the fetal spine than fetal MR imaging.\textsuperscript{14} However, given that sonography is highly operator-dependent, which is problematic in clinical practice and in the context of a retrospective analysis, studies in fetal MR imaging of the spine are becoming increasingly important.

**CONCLUSIONS**

We describe several observations at our institution on fetal and postnatal MR imaging of the spine in patients with OSD and add to the existing literature several important findings. First, fetal MR imaging is accurate in detecting the level of the dysraphic defect, which has an impact on prenatal counseling, neurologic outcomes, and eligibility for fetal surgery. We also found that a higher defect level correlated with increased fetal ventricular size. Second, we describe additional findings of associated split cord malformation and spinal cord syrinx on fetal MR imaging in this population. Fetal MR imaging was limited in its ability to identify split cord malformation in our series. Finally, although cervical spinal cord syrinx is uncommon, we found fetal MR imaging to be a useful tool for detecting it in patients with OSD.

**Table 3: Postnatal spine MRI findings**

<table>
<thead>
<tr>
<th>Percentage of</th>
<th>Percentage of</th>
<th>Percentage of</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total (N = 119)</td>
<td>Prenatal Repair Group (n = 38)</td>
<td>Prenatal Repair Group (n = 78)</td>
</tr>
<tr>
<td>Extradural postoperative fluid collection</td>
<td>32.8% (38/119)</td>
<td>5.3% (2/38)</td>
</tr>
<tr>
<td>Arachnoid cyst</td>
<td>3.4% (4/119)</td>
<td>5.3% (2/38)</td>
</tr>
<tr>
<td>Syrinx</td>
<td>39.5% (47/119)</td>
<td>36.8% (14/38)</td>
</tr>
<tr>
<td>No syrinx, no fluid collection, and no arachnoid cyst</td>
<td>36.2% (42/116)</td>
<td>55.3% (21/38)</td>
</tr>
</tbody>
</table>

**REFERENCES**


Dilated Vein of the Filum Terminale on MRI: A Marker for Deep Lumbar and Sacral Dural and Epidural Arteriovenous Fistulas


ABSTRACT

BACKGROUND AND PURPOSE: Conventional MR imaging can provide important clues regarding the location of a spinal vascular malformation. We hypothesized that a dilated vein of the filum terminale, identified as a curvilinear flow void on T2WI, could be an imaging marker for a lower lumbar (L3–L5) or sacral fistula.

MATERIALS AND METHODS: We retrospectively identified all spinal dural and spinal epidural arteriovenous fistulas from 2 large tertiary referral centers from 2005 to 2018. All patients had a lumbar spinal MR imaging and a conventional spinal angiography. Images were reviewed by 2 neuroradiologists who categorized the level of the arterial feeder to the fistula and the presence or absence of a dilated vein of the filum terminale on T2WI and T1 postcontrast images. We calculated the sensitivity, specificity, positive predictive value, and negative predictive value of the presence of a dilated filum terminale vein for a deep lumbar or sacral fistula.

RESULTS: One hundred sixty-two patients were included. An enlarged filum terminale vein was identified in 39 patients. Sensitivity, specificity, positive predictive value, and negative predictive value of the presence of a dilated filum terminale vein for a deep lumbar or sacral fistula were 86%, 98.3%, 94.9%, and 95.1%, respectively.

CONCLUSIONS: The presence of a dilated vein of the filum terminale can accurately localize a spinal dural arteriovenous fistula/spinal epidural arteriovenous fistula to the lower lumbar or sacral spine in patients being evaluated for such lesions. This finding can be used to facilitate both noninvasive and conventional spinal angiography.

ABBREVIATIONS: SDAVF = spinal dural arteriovenous fistula; SEDAVF = spinal epidural arteriovenous fistula; VFT = vein of the filum terminale

Received April 13, 2018; accepted after revision July 17.

From the Departments of Radiology (W.B., J.M.M., H.C., G.L.) and Neurosurgery (W.B., G.L.), Mayo Clinic, Rochester, Minnesota; and Joint Department of Medical Imaging (W.B., C.A.H., A.A.D., V.M.P., T.K.), Toronto Western Hospital, Toronto, Ontario, Canada.

Timo Krings and Giuseppe Lanzino contributed equally as senior supervising authors for this work.

Please address correspondence to Waleed Brinjikji, MD, Mayo Clinic, 200 1st St SW, Rochester, MN 55905; e-mail: brinjikji.waleed@mayo.edu; @wbrinjikji

http://dx.doi.org/10.3174/ajnr.A5784

Spinal dural arteriovenous fistulas (SDAVFs) and spinal epidural arteriovenous fistulas (SEDAVFs) are the 2 most common types of spinal vascular malformations. These lesions can result in considerable morbidity from congestive myelopathy secondary to chronic venous hypertension. Delay in the identification of spinal vascular malformations has been shown to result in high rates of irreversible morbidity due to the natural history of the disease, unnecessary surgeries including spinal cord biopsies and laminectomies, costly-yet-ineffective medical interventions including intravenous immunoglobulin therapy, and substantial costs from unnecessary serologic and imaging investigations.

Once the imaging findings suggestive of a spinal vascular malformation are properly identified, further investigations are required to characterize the angioarchitecture and location of the dominant arterial feeders to the lesion. This is important for planning surgical and/or endovascular therapies. In many centers, a spinal CT angiogram or time-resolved large-FOV MR angiogram is obtained to get a sense of the location of the arterial feeders so that spinal angiography can be focused on a few levels. This is important because a complete spinal angiography can be time-consuming and result in high radiation exposure for the patient and operator. A complete catheter spinal angiography also requires high volumes of iodinated contrast material, which can result in renal toxicity in some patients.

While noninvasive spinal angiography can be highly accurate in identifying the location of the fistula, it has limitations related to low spatial resolution (imaging typically requires a large FOV) and poor temporal resolution. While it was previously believed that conventional MR imaging findings of spi-
Dilated Vein of the Filum Terminale and Fistula Location

Outcomes and Statistical Analysis

The primary outcomes of this study were sensitivity, specificity, and accuracy of the presence of an enlarged VFT on conventional lumbar spine MR imaging in identifying lower lumbar and sacral fistulas. Lower lumbar and sacral fistulas were defined as fistulas located at the L3–S5 levels—the levels of the filum terminale. All statistical values are reported with their associated 95% confidence intervals. A $\kappa$ statistic measuring interobserver agreement for the presence of a VFT was calculated as well. All statistical analyses were performed using JMP13.0 (SAS Institute, Cary, North Carolina).

RESULTS

Patient Population and Baseline Characteristics

One hundred sixty-two patients were included in this study. Mean patient age was 64.4 ± 12.3 years; 78.4% (127/162) of patients were men. There were 119 SDAVFs and 43 SEDAVFs. Forty-three fistulas (26.5%) were deep lumbar or sacral, and 119 fistulas (73.4%) were upper lumbar or thoracic. Of the 43 deep lumbar or sacral fistulas, 20 fistulas were sacral fistulas and 23 were deep lumbar fistulas. All patients had noncontrast lumbar spine MRIs with T2-weighted imaging in the sagittal plane, and 102 patients (63%) had T1 contrast enhanced lumbar spine MR imaging in the sagittal plane.

MATERIALS AND METHODS

Patient Selection

Following institutional review board approval at 2 institutions, we retrospectively identified all patients with SDAVFs and SEDAVFs who underwent lumbar spinal MR imaging and conventional spinal angiography from 2005 to 2018. All patients had an angiographically confirmed SDAVF or SEDAVF. Children and patients who did not consent to the use of their charts for retrospective research were excluded. Patients with intramedullary, perimedullary, and metameric-type vascular malformations were excluded.

Imaging Review

Images were reviewed by 2 diagnostic and interventional neuroradiologists. MR images were reviewed for the presence of a dilated VFT, which was defined as a curvilinear flow void coursing along the filum terminale on sagittal T2-weighted images and/or a curvilinear enhancing VFT coursing along the filum on contrast-enhanced T1-weighted images. The neuroradiologists were blinded to the level and type of fistula at the time of MR imaging interpretation. Following interpretation of the MR images, the 2 neuroradiologists reviewed the conventional spinal angiograms and identified the location of the arterial feeder to the spinal vascular malformation as well as the type of spinal vascular malformation (SEDAVF versus SDAVF).

Dilated Vein of the Filum Terminale and Fistula Location

Thirty-nine patients had a dilated VFT. Interobserver agreement for identifying the presence of a dilated VFT was excellent ($\kappa = 0.84$; 95% CI, 0.81–0.87). In all cases, a dilated VFT was present on both T2 and T1 contrast enhanced imaging.

The sensitivity of a dilated VFT in identifying a lower lumbar or sacral fistula was 86% (37/43; 95% CI, 72.1%–94.7%). The specificity of a dilated VFT in identifying a lower lumbar or sacral fistula was 98.3% (117/119; 95% CI, 94.1%–99.8%). The positive predictive value of a dilated VFT in identifying a lower lumbar or sacral fistula was 94.9% (37/39; 95% CI, 82.3%–98.7%). The negative predictive value of a dilated VFT in identifying a lower lumbar or sacral fistula was 95.1% (117/123; 95% CI, 90.3%–97.6%). The accuracy of a dilated VFT in identifying a lower lumbar or sacral fistula was 95.1% (154/162; 95% CI, 90.5%–97.8%). The area under the curve of a dilated VFT in identifying a lower lumbar or sacral fistula was 0.921. Ninety-five percent (19/20) of sacral SEDAVFs and SDAVFs had a dilated VFT, and 78.3% (18/23) of lower lumbar SEDAVFs and SDAVFs had a dilated VFT. Age, sex, and fistula type were not associated with the presence of a dilated VFT. Examples of dilated VFTs are provided in Figs 1–3.

FIG 1. A 63-year-old woman with a sacral dural arteriovenous fistula at S1. A, Sagittal T2-weighted MR imaging shows a dilated vein of the filum terminale (white arrow). B, Contrast-enhanced T1-weighted MR imaging again demonstrates the dilated vein of the filum terminale (white arrow). C, Rotational digital subtraction angiogram reformatted in the coronal plane following injection into the right internal iliac artery demonstrates a fistula at S1 (arrowhead), with a dilated vein of the filum terminale.
DISCUSSION

Our large retrospective study found that the presence of a dilated VFT on conventional T2WI is a reliable indicator for the presence of a lower lumbar or sacral spinal vascular malformation. Sensitivity, specificity, accuracy, negative predictive value, and positive predictive value were all high, and interobserver agreement was excellent. Ninety-five percent of sacral spinal vascular malformations were characterized by the presence of a dilated VFT, and >80% of lower lumbar spinal vascular malformations had a dilated VFT. These findings are important because the identification of a dilated VFT can help in focusing noninvasive angiographic imaging and conventional spinal angiography on the lumbosacral levels, thus allowing more efficient identification of arterial feeders to these lesions.

The vascular supply to the filum terminale has been well-described in both radiographic and cadaveric studies. The filum terminale has a single artery arising from the termination of the anterior spinal artery, which travels along the anterior aspect of the filum into the sacral canal. The vein of the filum terminale, which is continuous with the anterior spinal vein, travels in front of the filum but behind the artery and is of uniform caliber along its course. There are no veins along the dorsal aspect of the filum. Because the VFT is the only intradural venous structure below the L2 vertebral body level, it is the only longitudinal collecting vein below this level. Thus, all dural and epidural AVFs with intradural venous drainage must drain through the VFT. The VFT can drain in 2 directions: 1) descending to the sacral venous plexus and hypogastric vein, and 2) ascending toward the anterior spinal vein along the spinal cord. It is thought that the ascending route is generally the preferred route. The propensity for the VFT to drain superiorly could explain the propensity for deep lumbar and sacral fistulas to result in a dilated VFT and the low rate of more superiorly located fistulas presenting with dilated VFTs. A small number of patients with deep lumbar and sacral fistulas did not have a dilated VFT on MR imaging. Angiographically, these lesions drained into the VFT; however, the VFT was likely not engorged enough to be easily seen on MR imaging.

Prior studies and case reports have shown high rates of VFT enlargement in the setting of deep lumbar and sacral fistulas. One of the first case reports demonstrating such an association was published by Chen and Hsu in 2002. The authors reported a sacral dural AVF, which presented with an enlarged VFT on sagittal T2-weighted MR imaging. The presence of such a vein prompted the investigators to pursue pelvic angiography before thoracolumbar angiography, resulting in prompt identification of the fistulous point. In a more recent series of sacral fistulas by Gioppo et al, the authors found that all 15 sacral fistulas had an enlarged VFT on MR imaging. Meanwhile, a recently published series of spinal vascular malformations at the L5 level or below found that 60% of fistulas had a dilated filum terminale vein, further reinforcing the association between a dilated VFT and the presence of a deep lumbar or sacral fistula. Our study differs from these prior studies because we are able to confirm that a dilated VFT identified using conventional

FIG 2. A 75-year-old man with a sacral dural arteriovenous fistula at S1. A, Sagittal T2-weighted MR imaging shows a dilated vein of the filum terminale (white arrow). B and C, Contrast-enhanced MRIs demonstrate marked dilation of the vein of the filum terminale (white arrow). D, Conventional angiogram in the anteroposterior plane following injection into the right internal iliac artery demonstrates a fistula at S1 with a dilated vein of the filum terminale (black arrow).

FIG 3. An 80-year-old man with a sacral dural arteriovenous fistula at S1. A, Sagittal T2-weighted MR imaging shows a dilated vein of the filum terminale (white arrows). B, Contrast-enhanced T1-weighted MR imaging again demonstrates the dilated vein of the filum terminale (white arrows). C, Conventional angiogram in the anteroposterior plane following injection into the right internal iliac artery demonstrates a fistula at S1 with a dilated vein of the filum terminale (black arrows).
MR imaging techniques is a sensitive and specific sign for localization of SDAVF/SEDAVFs to the lower lumbar or sacral spine in patients being evaluated for such lesions.

Numerous studies have demonstrated the utility of various MR and CT angiographic techniques in identifying the feeding artery and angioarchitecture of spinal vascular malformations. These techniques have proved instrumental in focusing spinal angiography to make conventional spinal angiography more efficient and reduce operative time, radiation exposure, and contrast dose. However, these techniques have their limitations. For gadolinium bolus spinal angiography, some centers lack the capability and scanner time to image the entire spine in patients with suspected spinal vascular malformations. Thus, such centers will often perform a gadolinium bolus MRA focused on the thoracic and upper lumbar spine and will miss deep lumbar and sacral fistulas on initial evaluation. This can be mitigated by performing 2 gadolinium-bolus MRAs; however, it is preferable to avoid this situation. In the setting of noninvasive 4D MR imaging, there is often a trade-off among spatial, temporal, and contrast resolution. Thus, the ability to narrow the FOV to a smaller region in the setting of a dilated VFT could allow improved noninvasive characterization of spinal vascular malformations. Our study suggests that when a dilated VFT is identified, radiologists should focus their noninvasive angiographic investigations in the lumbosacral region, including branches of the iliac arteries. We have found that lack of adequate angiographic investigation of the iliac branches is one of the most common reasons for missing spinal vascular malformations on false-negative spinal angiographic studies.

Limitations
Our study has limitations. First, given its retrospective nature, there is a propensity for selection bias. We included only patients with SDAVFs and SEDAVFs; thus, our results do not apply to fistulas of the filum terminale, pial AVFs, and nidus-type AVMs. A fistula of the filum terminale will, by definition, have a dilated VFT. Thus, there is the potential for an overreliance on this sign to cause clinicians to miss such a lesion. However, these lesions are also characterized by a markedly enlarged anterior spinal artery. The patients included in this study were imaged during a 13-year period using various MR imaging scanners, MR imaging protocols, and various field strengths (1.5T versus 3T). This feature can limit the generalizability of our results.

CONCLUSIONS
Our retrospective study including >160 patients with SDAVFs and SEDAVFs and lumbar spine imaging found that the presence of a dilated VFT is sensitive and specific for localization of SEDAVF/SEDAVFs to the lower lumbar or sacral spine in patients being evaluated for such lesions. Our findings suggest that the presence of an enlarged VFT should prompt the neuroangiographer to initially catheterize the lower lumbar and internal iliac vessels when attempting to identify the location and arterial feeders of a SEDAVF/SEDAVF. Our results are also valuable in planning noninvasive spinal angiography by allowing imagers to narrow their FOVs or center on the lumbosacral region.

REFERENCES
Spinal Instrumentation Rescue with Cement Augmentation

A.C. Cianfoni, M. Giamundo, M. Pileggi, K. Huscher, M. Shapiro, M. Isalberti, D. Kuhlen, and P. Scarone

ABSTRACT

BACKGROUND AND PURPOSE: Altered biomechanics or bone fragility or both contribute to spine instrumentation failure. Although revision surgery is frequently required, minimally invasive alternatives may be feasible. We report the largest to-date series of percutaneous fluoroscopically guided vertebral cement augmentation procedures to address feasibility, safety, results and a variety of spinal instrumentation failure conditions.

MATERIALS AND METHODS: A consecutive series of 31 fluoroscopically guided vertebral augmentation procedures in 29 patients were performed to address screw loosening (42 screws), cage subsidence (7 cages), and fracture within (12 cases) or adjacent to (11 cases) the instrumented segment. Instrumentation failure was deemed clinically relevant when resulting in pain or jeopardizing spinal biomechanical stability. The main study end point was the rate of revision surgery avoidance; feasibility and safety were assessed by prospective recording of periprocedural technical and clinical complications; and clinical effect was measured at 1 month with the Patient Global Impression of Change score.

RESULTS: All except 1 procedure was technically feasible. No periprocedural complications occurred. Clinical and radiologic follow-up was available in 28 patients (median, 16 months) and 30 procedures. Revision surgery was avoided in 23/28 (82%) patients, and a global clinical benefit (Patient Global Impression of Change, 5–7) was reported in 26/30 (87%) cases at 1-month follow-up, while no substantial change (Patient Global Impression of Change, 4) was reported in 3/30 (10%), and worsening status (Patient Global Impression of Change, 3), in 1/30 (3%).

CONCLUSIONS: Our experience supports the feasibility of percutaneous vertebral augmentation in the treatment of several clinically relevant spinal instrumentation failure conditions, with excellent safety and efficacy profiles, both in avoidance of revision surgery and for pain palliation.

ABBREVIATIONS: PGIC = Patient Global Impression of Change; PMMA = polymethylmethacrylate

Spinal instrumentation is widely used in the treatment of degenerative, traumatic, and neoplastic conditions. Altered biomechanics and/or bone fragility may lead to instrumentation failure, bone resorption, or new fractures with consequent instability and recurrent or progressive pain.1,2 The most commonly encountered types of instrumentation failure are implant fracture or disassembly, bone resorption around the screws, or impaction fracture adjacent to an implanted cage.1 In addition, vertebral insufficiency fractures can occur within the instrumented segment or at adjacent levels (junctional fractures). In many instances described above, revision surgery is performed,3 with the potential of further morbidity, increased cost, and reduced patient satisfaction. Re-operation is often an unattractive option in elderly, medically complex, and fragile patients.

Minimally invasive options would be desirable to address instrumentation failure. Vertebral cement augmentation is used in the treatment of painful osteoporotic and tumor-related compression fractures.4–7 Numerous reports also document the utility of polymethylmethacrylate (PMMA) to augment pedicular screws at the time of insertion.8–12

To date, several small series have described the use of cement augmentation in implant failure, including junctional fractures and screw loosening, both in osteoporotic13–15 and neoplastic settings.16–17

Received May 21, 2018; accepted after revision July 23.
From the Departments of Neuroradiology (A.C., M.P., M.I.) and Neurosurgery (M.G., D.K., P.S.), Neurocenter of Southern Switzerland, Ente Ospedaliero Cantonale, Lugano, Switzerland; Department of Neurosurgery (K.H.), Hôpital du Valais, Sion, Switzerland; Department of Radiology (M.S.), New York University Langone Medical Center, New York, New York; and Department of Neuroradiology (A.C.), Inselspital, Bern, Switzerland.

Please address correspondences to Alessandro Cianfoni, PD, MD, Department of Neuroradiology, Neurocenter of Southern Switzerland, Ospedale Regionale di Lugano, Via Tesserete 46, 6900 Lugano, Switzerland; Department of Neuroradiology, Inselspital, Freiburgstrasse 18, 3010 Bern, Switzerland; e-mail: alessandro.cianfoni@eoc.ch

Indicates article with supplemental on-line table.

http://dx.doi.org/10.3174/ajnr.A5795

We report the largest and most encompassing series to date of percutaneous fluoroscopically guided vertebral cement augmentation in a variety of clinically significant instrumentation failure conditions, including screw loosening, cage subsidence, and vertebral fractures within or adjacent to the instrumented segment.

MATERIALS AND METHODS

A retrospective analysis of 31 consecutive procedures in 29 patients (male/female, 10:19; mean age, 71.6 years; range, 50–82 years) with instrumentation failure treated by percutaneous cement augmentation between May 2013 and October 2016 was performed. Informed consent was obtained from all patients, and the study was approved by the Comitato Etico Cantonale del Ticino institutional ethics committee.

Indications for primary spinal instrumentation included a variety of traumatic, degenerative, and neoplastic conditions. Indications for cement rescue treatment were the following: clinically relevant screw loosening with bone resorption, cage subsidence, and vertebral fracture within or adjacent to the instrumented segment. Instrumentation failure was deemed clinically relevant if accompanied by new or recurrent pain correlated with imaging findings or deemed to be a threat to spinal biomechanical stability. Determination of implant failure, the decision to offer treatment, and the treatment approach were based on case review by the Spine Unit staff of our institution, composed of neurosurgeons, neurologists, neuroradiologists, pain physicians, and physical medicine and rehabilitation physicians. All patients underwent preprocedural noncontrast CT in addition to pre-existing imaging studies. Patients with uncorrectable coagulopathy and local or systemic infection were excluded.

Cement-Augmentation Procedure

Procedures were performed with the patient under moderate sedation and local anesthesia. A single dose of prophylactic intravenous antibiotic (cefazolin, 2 g) was administered 1 hour before the procedure. All procedures were performed under fluoroscopic guidance in a monoor bi-plane angiosuite. A variety of approaches to the target vertebral body were used (Fig 1) to overcome access constraints posed by the presence of implants. Vertebroplasty 15-ga bevel-tip trocars were inserted via a transpedicularr approach to reach the desired target location, and high-viscosity PMMA cement (VertaPlex HV; Stryker, Kalamazoo, Michigan) was injected under real-time fluoroscopic control. Postprocedural target-level CT was performed. Patients were mobilized immediately following recovery from moderate sedation. Most patients were discharged the same day.

Vertebral Body Access

Transpedicular Approach. An access trajectory parallel to the screw was used for treatment of screw loosening. An en face view of the screw was obtained as a “bull’s eye” projection, the fluoroscopic camera was then angled slightly toward the side of the
desired modified access, and the proximal or midportion of the screw was set as the target. With use of a 15-cm 15-ga beveled-tip vertebroplasty needle, contact with the midshaft of the screw was obtained and the flexible needle was pushed to slide with the bevel along the distal shaft of the screw, maintaining the needle tip position inside the bone resorption halo. Alternatively, the tip of the screw was set as the fluoroscopic target on the en face view and aligned with a transpedicular access route a few millimeters away from the screw.

For access to instrumented level fractures (rather than loosening), a trajectory crossing from lateral to medial and from cranial to caudal relative to pedicle screw was usually chosen. In this case, the fluoroscope was angled from the bull’s eye view of the screw slightly laterally and cranially. Whenever the needle contacted a screw, appropriately turning the bevel would allow the needle to slide past the screw and continue forward progress to reach the desired target in the vertebral body.

Extrapediclar Access. In general, we thought that small-caliber, straight anteroposterior, and steep craniocaudal orientation of mid and upper thoracic pedicles favored an extrapedicular approach between the pedicle and the rib head. The access trajectory could be slightly lateral and cranial to the screw and then parallel to the screw, exploiting the bevel design and flexibility of the access needle. For loose screw indications, a similar final needle position was targeted as discussed in the transpedicular section above. For instrumented-level fractures, the approach was to cross the path of the screw at the junction between the pedicle and vertebral body, from lateral to medial, and cranial or caudal to the screw, depending on target fracture location (Fig 3). In case of access to the vertebral body without pedicular screws, a standard approach was used, with only necessary adjustments to avoid the vertical rods.

Injection of high-viscosity PMMA cement was performed under real-time high-resolution fluoroscopic control, predominantly in the lateral view, with intermittent anteroposterior or oblique control views. In case of screw loosening (Fig 2), cement injection was aimed at filling the halo of bone resorption around the screw, to simulate screw oversizing, and reducing or nulling screw micromobility. Whenever possible, the adjacent trabecular network was also filled with cement in an effort to achieve a more stable anchoring cast between the screw and the vertebral body. In case of vertebral fracture, cement injection was aimed at filling, with a trabecular interdigitation pattern, the anterior two-thirds of the vertebral body, from superior to inferior endplates on both sides of the midline. In case of cage subsidence, cement injection was aimed at augmenting the bone-metal interface and arresting bone compaction (Fig 3).

Follow-Up
Patients were followed 1 month after the procedure with a clinical visit and standing plain films; when deemed necessary, CT or MR imaging was performed. Clinical and imaging follow-up was continued at variable intervals, depending on the clinical situation and the referring physician’s preference.

Study End Points
Study end points were feasibility, safety, and efficacy. Feasibility and safety were assessed by review of chart records and periprocedural complications; efficacy was based on the rate of revision surgery, while the clinical effect on pain was measured at 1-month follow-up. The Patient Global Impression of Change (PGIC) 7-point response was scored as follows: 1, extremely worse; 2, much worse; 3, a little worse; 4, no change; 5, a little better; 6, much better; and 7, extremely better.

RESULTS
The On-line Table summarizes the results.
In 31 procedures, performed in 29 patients (2 patients underwent 2 procedures), 42 loose screws, 7 levels with cage subsidence,
12 vertebral body fractures at levels within the instrumented segment, and 11 fractures at levels adjacent to the instrumented segment (junctional fractures) were treated. In addition, prophylactic cement augmentation was performed at nonfractured adjacent levels in 9 cases with the intent of preventing subsequent junctional fractures in patients with osteoporosis when focal kyphosis was noted at the junctional level. In some patients, during the same procedure, distant vertebral or sacral fractures were also treated with cement augmentation. Prophylactic and distant site augmentation was not included in analysis.

In all 32 thoracic, 24 lumbar, 13 sacral, and 3 pelvic (iliac screws) targets, a total of 72 targets, were treated. The mean interval between the last spinal instrumentation and the procedure was 14 months (range, 1 week to 11 years; median, 4.5 months).

All except 1 procedure was technically feasible and successfully accomplished (ie, satisfactory fluoroscopic target visibility, needle placement, and cement injection). One procedure was aborted due to access failure caused by implant-related inability to visualize fluoroscopic landmarks. No periprocedural minor or major complications occurred; specifically, there were no neurologic complications or clinically significant PMMA leaks. Clinical and radiologic follow-up at 1 month was available in 28 patients; extended follow-up ranged from 2 to 54 months (median, 16 months; interquartile range, 14.7 months). Specifically, follow-up was available at 1 month in 28 patients, at 3 months in 26, at 6 months in 24, and beyond 12 months in 18. One patient was lost to follow-up.

During the follow-up period, 5/28 (18%) patients required revision surgery: The patient whose procedure was aborted due to access failure underwent revision surgery with replacement of a loose screw; another patient developed re-fracture and kyphosis despite cement augmentation of a junctional fracture requiring extension of instrumentation. Three patients underwent implant removal. In 1 of these 3 patients, instrumentation failure was ultimately due to a low-grade chronic infection.

In the 28 patients (30 procedures) with available follow-up, global clinical benefit (PGIC 5–7) was reported in 26/30 (87%) cases at 1 month after the procedure (PGIC 7 in 10/26, PGIC 6 in 9/26, PGIC 5 in 7/26), while no significant changes in status (PGIC 4) were reported in 3/30 (10%) and worsening status (PGIC 3) occurred in 1/30 (3%) in the absence of obvious procedural complications.

FIG 3. Cage subsidence/fracture and multiple targets. A and B, CT images of an L1 fracture treated with corpectomy, cage grafting, and T11–L3 posterior stabilization in a patient with osteoporosis. Due to bone compaction/fracture cranial and caudal to the cage, there is cage subsidence and focal kyphosis (arrows in B). Another fracture is noted at T11 (arrowhead in B), and there is bone resorption and screw loosening at L3 (not shown), with initial screw pullout. C and D, Anteroposterior and lateral fluoroscopy views after placement of multiple needles to perform cement augmentation at the cranial and caudal bone-metal cage interface in T12 and L2 (arrowheads), in the T11 fracture (arrow), in L3 (arrow) to augment the screw osseous purchase, and in T10 to perform prophylactic augmentation (arrow). E, Postprocedural sagittal MIP CT image demonstrates satisfactory cement filling of the target levels. F, Standing plain film at 12-month follow-up, with stable results.

DISCUSSION
In this series, minimally invasive fluoroscopically guided percutaneous cement augmentation was associated with successful avoidance of revision surgery in 82% of patients with screw loosening, cage subsidence, or vertebral fractures within or adjacent to the instrumented segment. Although limited by the retrospective study design, we believe that implant failure would have otherwise led to revision surgery, as discussed in the Spine Unit multidisciplinary review. Technical feasibility was highly satisfactory, with 30/31 procedures executed with the desired technical results. Safety was excellent, with no periprocedural complications, and the procedure resulted in significant pain amelioration in 87% of cases.

Spinal instrumentation performed for degenerative, traumatic, or oncologic diseases that may cause deformity or instability aims at providing stability while osseous fusion develops. Adequate bone quality is of primary importance in this success. Inadequate fixation and subsequent segmental microinstability may put the implant at risk of ultimate failure. A combination of
factors such as initial or subsequent poor bone density/quality, instrumentation-induced stress shielding and subsequent disuse osteopenia, plastic deformation at the bone-metal interface, or high static stress combined with cyclic loading also contribute to instrumentation failure. These processes are often accompanied by new or recurrent pain, altered biomechanics, and deformity. Implant failure can occur at variable time intervals after instrumentation from weeks to years. To interrupt this vicious spiral, spinal anatomy can be altered by laminectomies and bony plants, and the operator is therefore forced to seek other trajectories within narrow anatomic windows. Finally, after the operation, spinal anatomy can be altered by laminectomies and bony fusion masses.

Both CT and fluoroscopic guidance were described in prior reports. We invariably used fluoroscopic guidance based on careful planning and mandatory preoperative CT. Transpedicular access was favored in the lumbar and lower thoracic spine, while an extrapedicular approach was preferably adopted in the mid and upper thoracic spine. A small-caliber 15-ga beveled-tip vertebroplasty needle was used, allowing steerability and flexibility to precisely reach targets.

Precise needle positioning, use of high-viscosity cement, and use of high-quality fluoroscopic imaging equipment might have contributed to the excellent safety profile of the procedures in our series. Although not used in this series, intraoperative CT, O-arm Multidimensional Surgical Imaging System (Medtronic, Minneapolis, Minnesota), or conebeam CT guidance could be implemented to aid needle insertion. Moreover, when necessary, injection of cement can be directed by the use of a coaxial curved cannula.

**Clinical Results**

The main clinical objective of this study was to assess the rate of revision surgery avoidance. In a retrospective study, it is not possible to rule out that in some circumstances, a conservative approach would have been sufficient. Pain palliation was also assessed; however, back pain in this kind of complex patient cohort is an elusive target: It is indeed rather difficult to definitively attribute new or recurrent symptoms to imaging findings of implant failure. We preferred to use the PGIC scale rather than a Visual Analog Scale for pain assessment because, in our opinion, it yields a more global and general assessment of the patient’s perception of treatment effectiveness. Among patients with available follow-up, the clinical benefit of the procedure was reported in 26/30 cases. The 1 patient in whom the procedure was aborted due to access failure reported unchanged clinical status and ultimately underwent revision surgery. Two more patients reporting a PGIC 4 and 1 patient with PGIC 3 required revision surgery, while 1 patient reporting initial clinical benefit (PGIC 6) after augmentation of 2 loose screws and a junctional fracture had a delayed recurrent collapse of the junctional level and underwent instrumentation extension. Finally, 1 patient was lost to follow-up and remained unreachable; this patient might have undergone revision surgery at another institution.

In our series of 29 patients with implant failure, it was necessary to treat a total of 72 targets, including loose screws, cage subsidence, and vertebral fractures within or adjacent to the instrumented segment. Our data suggest that an important cause of implant failure is poor bone quality, a systemic problem leading to multilevel breakdown (Fig. 3). We therefore stress the extreme importance of aggressive osteoporosis management in this patient population.

Infrequently, bone resorption around implants can be caused by an infectious process. In case of clinical-radiologic suspicion, biopsies and cultures are recommended before proceeding to cement augmentation. Despite these measures, 1 patient in this series, following revision surgery with implant removal, was diagnosed with a low-grade infection, despite biopsy performed 2 weeks before cement augmentation resulting in cultures negative for infection. This infection was likely present before cement augmentation. Suboptimal sensitivity of spine biopsies for low-grade infections remains a challenge in such cases.

**Limitations**

Major limitations include the retrospective study design and an intrinsically subjective definition of clinically relevant instrumentation failure. Also recognized is the subjective nature of patient self-assessment. Although this is by far the largest series of its kind to date, some conditions such as cage subsidence are infrequent and considerably more experience is necessary. Given the frequency of spinal instrumentation surgery, it is likely that percutaneous salvage options are currently underused; therefore, greater awareness and prospective investigations are necessary.

**CONCLUSIONS**

This series supports the feasibility of safe, efficacious, minimally invasive percutaneous vertebral cement augmentation in the treatment of clinically relevant instrumentation failure, with an
excellent safety profile and efficacious clinical results in terms of pain palliation and avoidance of revision surgery. A larger prospective cohort will be necessary to determine optimal candidates for this treatment and to provide more generalizable outcome data.

Disclosures: Dominique Kühlen—UNRELATED: Board Membership: DePuySynthes Advisory Board.* Pietro Scarone—RELATED: Consulting Fee or Honorarium: Depuy Synthes. Comments: I am a consultant spine surgeon for this company, from which my institution receives fees.* Money paid to the institution.

REFERENCES


We read with interest the letter “Common Origin of Brachiocephalic and Left Common Carotid Arteries: Proposal of New Terminology.”1 The authors describe the variations in the origins of arch vessels, including the common origin of the brachiocephalic trunk and left common carotid artery, and propose a new terminology.

The most common variation in aortic arch branching seen is the common origin of brachiocephalic trunk and left common carotid artery commonly called the “bovine arch.” The bovine arch (in cattle) bears no resemblance to this branching pattern described in humans.2 The aortic arch in cattle gives rise to a large trunk, which gives off a bicarotid trunk and both subclavian arteries. The embryologic basis of variations in aortic arch branching in humans can be explained by the aberrations in the 6 embryonic aortic arches. The true bovine aortic arch in humans, which presents a common trunk further branching into the bilateral carotid and subclavian arteries, can be explained by the persistence of the embryonic fifth aortic arch along with involution of the embryonic fourth aortic arch distal to the left subclavian artery.3,4 The common origin of the brachiocephalic trunk and the left common carotid artery (erroneously called the “false bovine aortic arch”) results from the involution of the embryonic fourth aortic arch between the left common carotid artery and the left subclavian artery with a persistent embryonic fifth aortic arch.

The second most common variation of aortic arch branching in our Indian population is a separate origin of the left vertebral artery from the aortic arch, which can also be described as a 4-vessel origin from the aortic arch.

The reverse pattern of a common origin of the brachiocephalic trunk and left common carotid artery may be seen in patients with a right aortic arch in whom there may be a common origin of left brachiocephalic trunk and right common carotid artery. This is a commonly seen association in patients with tetralogy of Fallot with a right aortic arch. The bicarotid trunk is basically a common origin of the brachiocephalic trunk and left common carotid artery combined with a variation in the development of the right subclavian artery, in which the right subclavian artery, instead of originating from the brachiocephalic trunk, takes its origin from a bulbous dilation from the proximal descending thoracic aorta. This is due to the involution of the embryonic right fourth aortic arch and persistence of the embryonic proximal right dorsal aorta and seventh intersegmental artery.2 The bulbous dilation of the embryonic proximal right dorsal aortic component is called the Kommerell diverticulum.

Finally, the authors propose a new terminology for the common origin: “brachio-bicephalic trunk” for its brevity and anatomic correctness. The term “bicephalic” in Latin means having 2 heads.6 Hence, we think that the term “brachio-bicarotid trunk” may be more appropriate for describing this pattern of aortic arch branching.7

REFERENCES


R. Rajagopal
S. Sharma
Department of Cardiovascular Radiology and Endovascular Interventions
All India Institute of Medical Sciences
New Delhi, India
We would like to thank Drs Rajagopal and Sharma for their interest in our letter “Common Origin of Brachiocephalic and Left Common Carotid Arteries: Proposal of New Terminology.”1 We are glad they agree that the term “bovine” is a misnomer and should no longer be used to describe a very common variation in the normal human anatomy. However, there are a few points in their comment that are worth further discussion.

First, in their attempt to explain the embryology of the brachio-bicephalic trunk, they included a theory based on the persistence of the fifth embryonic arch and the involution of the fourth arch between the left common carotid and left subclavian arteries. This theory is not widely accepted on the basis of the rarity of the persistent fifth arch in contrast to the common occurrence of the brachio-bicephalic trunk.2 The embryologic development of the brachio-bicephalic trunk was not in the scope of our original letter; however, there are other simpler and more reasonable theories based on slower growth of the ventral aortic roots between the third and fourth arches or arrested bifurcation of the aortic sac allowing fusion between the brachiocephalic and left common carotid arteries.3,4

Second, they described the “bi-carotid” trunk as a brachio-bicephalic trunk in which the right subclavian artery has an aberrant origin from a bulbous dilation from the proximal descending thoracic aorta (the Kommerell diverticulum). We would like to further explain that the bi-carotid trunk denotes a common trunk for both carotid arteries and separate origins of the subclavian arteries. Thus, the right subclavian artery may arise in its original location as the first aortic arch branch5 or arise as the last branch from the aorta as an aberrant right subclavian artery. Moreover, the current understanding is that an aberrant right subclavian artery is not necessarily associated with a Kommerell diverticulum.

Finally, the prefix “bi-” has, unsurprisingly, more than 1 meaning.6 In our suggested terminology, the intended meaning is “both sides”; hence, the term brachio-bicephalic trunk refers to a trunk for a brachial artery and 2 arteries for both sides of the head. Moreover, the term is short and is similar to the original brachiocephalic term.

REFERENCES
2. Gupta SK, Bamforth SD, Anderson RH. How frequent is the fifth arch artery? Cardiol Young 2015;25:628 – 46 CrossRef Medline

A.M. Tawfik
D.M. Sobh
N.M. Batouty
Department of Diagnostic and Interventional Radiology
Faculty of Medicine, Mansoura University
Mansoura, Egypt
Quality-Control Assessment to Improve the Accuracy of Dynamic Contrast-Enhanced MR Imaging Perfusion

We read with much interest the article by Dr Morales and colleagues published in the May 2018 issue of American Journal of Neuroradiology in which they showed that T1-weighted dynamic contrast-enhanced (DCE) MR imaging might help differentiate atypical hemangiomas from metastatic vertebral lesions.

The authors used a dynamic 2-step approach with a visual assessment of T1 curves and a modelization of the DCE MRI using the extended Tofts 2-compartment pharmacokinetic model, providing quantitative perfusion parameters such as the plasma volume (Vp) and the volume transfer constant (Ktrans). This approach could have been strengthened by the integration of a quality-control assessment of the DCE based on the individual arterial input function (AIF) curves. When an individual AIF is near the mean AIF of the population, it might be integrated into the simplified visual assessment of the curves. Alternatively, when an individual’s AIF is far from the mean AIF of the population, 2 approaches might be used to provide more accurate quantitative data: the first one using only the pharmacokinetic model approach to analyze the quantitative parameters; the second correcting data with a B-spline-based model-independent deconvolution to obtain a renormalized tissular kinetic. This latter approach gathers more accurate quantitative parameters.

The authors used the Ktrans to demonstrate preservation of permeability in cases of atypical hemangiomas with an abnormally elevated Vp. This approach should be used cautiously because the Ktrans method of calculation includes both perfusion and permeability-related phenomena, leading to a possible misinterpretation of the parameters, even when data are accurately fitted. The use of the extravascular and extracellular volume fraction (Ve) provided by the extended Tofts model, or, even better, the use of a more comprehensive and complex model such as the 2-compartment exchange or 2CX model can better quantify the permeability measurement.

It would be very interesting to perform a second segmentation by a second reader, if possible an inexperienced one, to evaluate the inter- and intraobserver variability and to reinforce the findings of the study and support its use in clinical practice.

Besides this feedback, our colleagues’ very interesting work could considerably help radiologists before and during diagnosis and lead to changes in patient management, reducing biopsies, additional imaging, and patient anxiety in persons with atypical hemangiomas.


REFERENCES

http://dx.doi.org/10.3174/ajnr.A5787
REPLY:

I appreciate the comments in the letter to the Editor and the interest in our publication. The parameters used in our project best fit our dynamic contrast-enhanced (DCE) model with the FDA-approved software. I find the comments interesting and truly hope that the commentators use their suggestions on their own spine DCE data, and I look forward to their publications in this area along with a discussion of histopathology of hemangiomas.

S. Karimi
Department of Radiology
Memorial Sloan-Kettering Cancer Center
New York, New York

http://dx.doi.org/10.3174/ajnr.A5826
The review of the latest data concerning vertebroplasty and kyphoplasty for osteoporotic vertebral fractures seemed a desperate attempt to cleanse the unconscionable.1

The first vertebroplasty case report appeared in an obscure low-Impact Factor (0.8) journal, from a team who subsequently published dozens of articles but never a trial.2 Since 1987, although thousands of articles have been published, it is exceptional to find an adequately designed trial (eg, blinded, compared with a sham procedure, adequate number of patients, large patient cohorts, long clinical follow-up using relevant outcomes such as patients’ quality of life). In 2016, the Vertebroplasty for Acute Painful Osteoporotic Fractures (VAPOUR) trial showed improved pain relief but with an absurdly short 14-day follow-up.3 The most recent trial on vertebroplasty showed robust evidence that it did not result in benefit for patients with acute osteoporotic vertebral compression fractures.4

Sadly, health care systems are resistant to learning from error. The 2009 Angioplasty and Stenting for Renal Artery Lesions (ASTRAL) trial showed that dilation and stent placement in renal arteries provided no benefit over drug treatments but only increased serious harm. However, the procedure had spread like wildfire since the 1980s with 45,000 procedures annually in the United States.5 Surgeons and radiologists are not alone because the issue also concerns medicines. Encainide and flecainide have long been the standard of care after myocardial infarction to suppress ventricular premature complexes, until the Cardiac Arrhythmia Suppression Trial (CAST) investigators showed it increased mortality.6 Presently, most approvals of anticancer medicines are based on flimsy or untested surrogate end points, and most drugs offer marginal benefits that may be lost in the real world of heterogeneous patients in whom only harms appear.7

The solution is a challenge. In 1660, Pascal warned, “People almost invariably arrive at their beliefs not on the basis of proof but on the basis of what they find attractive” (The Art of Persuasion).

Pragmatically, Chandra et al1 should have pledged that all new patients be included in randomized controlled trials and, if that were not possible, in registries for monitoring adverse effects.

At the system level, training in history and humility should be an integral part of the medical curriculum. Editors must restrain researchers with preliminary observations from “scienzationalism” (sensationalism in science).8 Terms such as “therapeutic revolution” should be unacceptable.9 Professional societies must scrutinize their recommendations and grade them according to the evidence.

Disclosures: Susan Bewley—UNRELATED: Chair of HealthWatch, a charity that stands for better understanding by the public and the media of the importance of application of evidence from robust clinical trials [https://www.healthwatch-uk.org/].

REFERENCES

http://dx.doi.org/10.3174/ajnr.A5762

Vertebroplasty: Expectation or Evidence-Based Interventional Radiology?
We thank Braillon and Bewley for their interest in our article, “Vertebroplasty and Kyphoplasty for Osteoporotic Vertebral Fractures: What Are the Latest Data?” The stated aim of this article was to provide an update to clinicians on the evolution of evidence for reduction in pain and disability from vertebroplasty and kyphoplasty for osteoporotic vertebral fractures and, in particular, highlight the limitations of the various prospective randomized controlled trials (RCTs). The most recent trial, A Randomised Sham-Controlled Trial of Vertebralplasty for Painful Acute Osteoporotic Vertebral Fractures (VERTOS IV) had not been published at the time this review was conducted. Notably, there are important differences in the enrolled patient cohorts between A Controlled Trial of Vertebralplasty for Acute Painful Osteoporotic Fractures (VAPOUR) and VERTOS IV, which again highlight the challenge in the interpretation of outcomes from these procedures.²

Apart from the RCTs, there are also strong signals of benefit from large national or insurance-based-claims datasets from Germany, Sweden, France, Taiwan, and the United States.³-⁹ In one of the largest analyses of more than 2 million patients during 10 years from the US Medicare dataset, there was a strong signal of reduced mortality after vertebral augmentation compared with medical management.⁸ This signal of survival benefit has been replicated in further analysis of German⁵ and Taiwanese⁷ health insurance datasets. In addition, various national and medical societies have varied in their interpretation of the evidence, depending on when they examined the literature.¹⁰-¹² Most notably, the National Institute for Health and Care Excellence, which provides evidence-based guidance and advice to the National Health Service in the United Kingdom, recommends vertebroplasty and kyphoplasty as treatment options for patients with severe pain after a recent osteoporotic vertebral compression fracture and concluded that it was reasonable to assume that the procedures reduce mortality.¹³

Akin to many other areas in medicine, clinicians must integrate their clinical expertise with patient values and interpretation of the research evidence to provide optimized and meaningful care. For years, the results from the various RCTs have shown that vertebroplasty and kyphoplasty are best considered for patients with severe pain and disability and only after rigorous clinical and advanced imaging selection. Moreover, earlier treatment (potentially <3 weeks from fracture onset) may provide the best chance of benefit. Important questions remain unanswered; for example, what are the implications of the progressive height loss evident in untreated-versus-cemented levels in VERTOS IV and VAPOUR? Does this prevention of height loss help explain the mortality benefit observed in almost all claims-based studies? We concur with Braillon and Bewley that new patients should be included in further RCTs to clarify the role of these procedures or included in large registries in which data can be pooled and additional meaningful conclusions reached.

REFERENCES

http://dx.doi.org/10.3174/ajnr.A5781


R.V. Chandra
Interventional Neuroradiology Unit, Monash Imaging
Monash Health
Melbourne, Victoria, Australia
Faculty of Medicine, Nursing and Health Sciences
Monash University
Melbourne, Victoria, Australia

J. Maingard
Interventional Neuroradiology Service, Department of Radiology
Austin Hospital
Melbourne, Victoria, Australia
School of Medicine
Faculty of Health, Deakin University
Warrn Ponds, Victoria, Australia