Number Needed to Treat with Vertebral Augmentation to Save a Life

J.A. Hirsch, R.V. Chandra, N.S. Carter, D. Beall, M. Frohbergh, and K. Ong

ABSTRACT

BACKGROUND AND PURPOSE: Evidence from randomized controlled trials for the efficacy of vertebral augmentation in vertebral compression fractures has been mixed. However, claims-based analyses from national registries or insurance datasets have demonstrated a significant mortality benefit for patients with vertebral compression fractures who receive vertebral augmentation. The purpose of this study was to calculate the number needed to treat to save 1 life at 1 year and up to 5 years after vertebral augmentation.

MATERIALS AND METHODS: A 10-year sample of the 100% US Medicare data base was used to identify patients with vertebral compression fractures treated with nonsurgical management, balloon kyphoplasty, and vertebroplasty. The number needed to treat was calculated between augmentation and nonsurgical management groups from years 1–5 following a vertebral compression fracture diagnosis, using survival probabilities for each management approach.

RESULTS: The adjusted number needed to treat to save 1 life for nonsurgical management versus kyphoplasty ranged from 14.8 at year 1 to 11.9 at year 5. The adjusted number needed to treat for nonsurgical management versus vertebroplasty ranged from 22.8 at year 1 to 23.8 at year 5.

CONCLUSIONS: Both augmentation modalities conferred a prominent mortality benefit over nonsurgical management in this analysis of the US Medicare registry, with a low number needed to treat. The calculations based on this data base resulted in a low number needed to treat to save 1 life at 1 year and at 5 years.

ABBREVIATIONS: BKP = balloon kyphoplasty; NNT = number needed to treat; NSM = nonsurgical management; VCF = vertebral compression fracture; VP = vertebroplasty

Vertebral augmentation techniques became popular in the late 1990s and 2000s due to their efficacy in reducing pain and disability in patients with painful vertebral compression fractures (VCFs). Early data from large observational studies and open-label randomized controlled trials comparing vertebral augmentation with nonsurgical therapy supported the use of augmentation. However, since that time, evidence for the effectiveness of vertebroplasty has been made controversial by 2 randomized sham trials. These high-profile randomized controlled trials in 2009 did not demonstrate that vertebroplasty conferred benefit over active control sham. These findings considerably reduced the use of vertebral augmentation procedures. In 2016, in an attempt to control for a number of perceived methodologic limitations from those earlier trials, the authors of the Vertebroplasty for Acute Painful Osteoporotic Fractures (VAPOUR) trial selected a group of patients with severe pain from a recent fracture using advanced imaging and modified the sham procedure to a subcutaneous injection of local anesthetic. It became the first placebo-controlled randomized controlled trial of vertebroplasty to demonstrate positive results. The evidentiary landscape again changed with the 2018 publication of A Randomised Sham Controlled Trial of Vertebroplasty for Painful Chronic Osteoporotic Vertebral Fractures (VERTOS IV), another active control sham randomized controlled trial, which again failed in its primary end point to show benefit for use of vertebroplasty, primarily due to the high level of pain relief seen in the active sham group and a statistically highly controversial method of comparing the difference in pain reduction between the 2 treatment groups.
Multiple analyses of larger datasets from national registries or insurance-based claims data have demonstrated a significant survival benefit for patients with VCFs who receive vertebral augmentation, taking research in a new direction because the sham trials were not intended to assess mortality risk.\textsuperscript{10,11} Most recently, Ong et al\textsuperscript{7} studied a cohort that included 2,077,944 patients with VCF, analyzing vertebral augmentation use and mortality risk in Medicare patients from 2005 to 2014. The use of kyphoplasty conferred a $>50\%$ 1-year mortality benefit and reduced the 10-year mortality by up to 24\% compared with nonsurgical management (NSM).\textsuperscript{10} This mortality benefit has also been observed in an analysis of claims data in the German population, which found a 43\% reduction in 5-year mortality in those treated by vertebral augmentation compared with NSM.\textsuperscript{11} The discovery of this survival benefit that spans countries, cultures, and races would seem an important informational component on which to base treatment decisions in the population of patients with VCF. On the basis of this information and on the entire Medicare data base for a decade, our aim in this study was to calculate the number needed to treat (NNT) to save 1 life at 1 and up to 5 years after treatment.\textsuperscript{12}

\section*{MATERIALS AND METHODS}

\subsection*{Patient Population}

Inpatient/outpatient US Medicare Fee-For-Service claims data from 2005 to 2014 were used to identify incident patients with VCFs (International Classification of Diseases, Ninth Revision, Clinical Modification [ICD-9-CM] codes 733.13, 805.0, 805.2, 805.4, 805.6, and 805.8). The first VCF diagnosed in the study period was used as the incident fracture. The patients were required to have at least a 12-month claims history before the VCF diagnosis to confirm a VCF-free period. Those without 12 months of claims history before the VCF diagnosis were excluded due to potential incompleteness in their claims history. Patients undergoing vertebral augmentation in the 12 months before the index VCF were excluded. Those younger than 65 years of age were also excluded due to potential confounding factors from their Medicare eligibility, including certain disabilities such as permanent kidney failure, amyotrophic lateral sclerosis, and other significant medical conditions that may result in Medicare enrollment. Patients enrolled in a Health Maintenance Organization plan and not residing in the 50 states were also excluded.\textsuperscript{10}

Patients were stratified into NSM, balloon kyphoplasty (BKP), and vertebroplasty (VP) cohorts. Surgical treatment meant that patients underwent vertebral augmentation within the first year of the VCF diagnosis. Those who underwent fusion surgery between the VCF diagnosis and BKP/VP were excluded. The NSM cohort comprised patients who did not undergo augmentation or fusion during the study period and those who underwent augmentation or fusion only $>$1 year after the index VCF diagnosis. Balloon kyphoplasty was identified using ICD-9-CM code 81.66 or Current Procedural Terminology codes 22289 and 22523–22525.\textsuperscript{13}

This study was based on publicly available datasets, did not use private health identifiable information, and did not represent human subject research; therefore, it did not require oversight by our institutional review boards.

\subsection*{Calculation of NNT}

The NNT for survival between BKP and NSM was evaluated from years 1 to 5 following a VCF diagnosis via a time-to-event approach. The input for this calculation was the Kaplan-Meier survival probability as an estimate of the hazard ratio, as described by Altman and Anderson\textsuperscript{14} and Bowry et al.\textsuperscript{15} To estimate the unadjusted NNT for each year following a VCF diagnosis, we first obtained the corresponding unadjusted NSM survival probabilities via the Kaplan-Meier approach from previously published data on survival of patients with VCF. These data had been stratified into groups by NSM, BKP, and VP. Using the adjusted hazard ratio for mortality risk of BKP obtained from published data from a mortality study of BKP and NSM,\textsuperscript{6} we calculated the survival probability for BKP at each year as the survival probability for NSM with the hazard ratio as the exponent (as per Equation 2 in Bowry et al). The unadjusted NNT was then determined from the inverse of the difference in the BKP and NSM survival probabilities. The 95\% confidence interval for the NNT was obtained by replacing the hazard ratio with the 2 limits of the 95\% confidence interval for the hazard ratio. To estimate the adjusted NNT, we replaced the unadjusted survival probability for NSM with the corresponding adjusted survival probability. The NNT for survival between VP and NSM used a similar approach but relied on the adjusted hazard ratio for mortality risk of VP relative to NSM, instead of BKP relative to NSM. Similarly, for the NNT between BKP and VP, the VP results were used in place of the NSM results from the previous BKP-NSM analysis.

\section*{RESULTS}

\subsection*{BKP versus NSM}

The patients with NSM had an unadjusted survival probability of 76.8\% at 1 year post-VCF diagnosis, which decreased to 42.5\% at 5 years post-VCF diagnosis, while the corresponding survival probabilities for patients with BKP were 84.2\% and 50.9\% respectively, after accounting for the relative hazard ratios between BKP and NSM (Table 1). The adjusted survival probabilities decreased from 79.1\% to 41.9\% at years 1–5 for patients with NSM and from 85.9\% to 50.3\% at years 1–5 for those with BKP. The unadjusted NNT for BKP versus NSM ranged from 13.5 patients (95\% CI, 13.1–13.9 patients) at year 1 to 12.0 patients (95\% CI, 11.4–12.6 patients) at year 5 (Table 2). The adjusted NNT ranged from 14.8 (95\% CI, 14.4–15.2) at year 1 to 11.9 (95\% CI, 11.3–12.6) at year 5.

\subsection*{VP versus NSM}

The survival probabilities for patients with VP were 81.6\% at year 1 and 46.7\% at year 5, after accounting for the relative hazard ratios between VP and NSM (Table 1). The adjusted survival probabilities decreased from 83.5\% to 46.1\% at years 1–5 for patients with VP. The unadjusted NNT for VP versus NSM ranged from 20.8 patients (95\% CI, 19.9–21.8 patients) at year 1 to 23.8 patients (95\% CI, 21.8–26.3 patients) at year 5 (Table 2).
The adjusted NNT ranged from 22.8 (95% CI, 21.8–23.9) at year 1 to 23.8 (95% CI, 21.7–26.3) at year 5.

**BKP versus VP**

Patients with VP had an unadjusted survival probability of 81.4% at 1 year post-VCF diagnosis, which decreased to 44.3% at 5 years post-VCF diagnosis, while the corresponding survival probabilities for patients with BKP were 84.1% and 48.5%, respectively, after accounting for the relative hazard ratios between BKP and VP (Table 1). The adjusted survival probabilities decreased from 79.7% to 43.0% at years 1–5 for patients with VP and from 82.6% to 47.2% at years 1–5 for patients with BKP. The unadjusted NNT for BKP versus VP ranged from 36.7 (95% CI, 34.5–39.2) at year 1 to 23.9 (95% CI, 21.8–26.4) at year 5.

**DISCUSSION**

There is a discordance between the findings of 3 sham-controlled trials of vertebroplasty and the mortality advantage suggested by claims-based studies. There are several possible reasons for this divergence. First, NSM is not risk-free therapy. The combination of anagelsic therapy with a period of bed rest and limitation of daily activities can be counterproductive in the geriatric population. Moreover, NSM has long included opioid medications, which are increasingly understood to be problematic in this population. The nonsurgical arm in the VAPOUR trial had more significant complications than the vertebroplasty cohort, including paralysis related to vertebral body collapse in 2 patients several weeks after enrollment in the trial. It has proved difficult to design a methodologically sound trial that compares vertebral augmentation against a viable alternative while still avoiding the active placebo effect. While the impact of a placebo (and nocebo) is real in pain trials, it is difficult to replicate or harness these positive effects in routine clinical practice. Furthermore, the sham trials that showed a robust placebo response used needle docking in the periosteum and periosteal injection of local anesthetic, a technique that is known to produce pain relief, as opposed to the VAPOUR trial, which used a subcutaneous injection of local anesthetic.

In the 2010 VERTOS II trial, an open-label comparison of vertebroplasty and nonsurgical management, the positive clinical outcomes observed in those who had vertebroplasty were numerous and statistically robust. The 2 most recent sham-controlled blinded studies, VERTOS IV and VAPOUR, yielded different results regarding pain outcomes. This difference might, in part, relate to the design of the sham procedure, in particular the use of an active control (periosteal numbing) in VERTOS IV compared with a truer sham (subcutaneous numbing) in VAPOUR, because the amount of pain decrease seen immediately after the sham in VERTOS IV was a dramatic 3.1 point reduction on the numeric rating scale compared with the 1.8 point decrease seen in the VAPOUR trial. Understanding the potential for a control to have a treatment effect is critical for interpreting pain trials.

Vertebroplasty practitioners, and thus trials of vertebroplasty, have focused on palliation of pain and improvement in functional status. In both VAPOUR and VERTOS IV, serial follow-up studies for the long term have demonstrated that augmented vertebral bodies preserve height more readily than those that are not augmented, and collapsed vertebral bodies have resulted in severe adverse events with loss of neurologic function. The preservation of sagittal alignment and vertebral body height also has intuitive benefits that perhaps contribute to the observed mortality advantage. The elevated risk of mortality for older patients with hyperkyphotic posture, specifically due to atherosclerosis, has been reported in a prospective study of >1300 patients.

Claims-based data have tended to demonstrate relative mortality advantages of kyphoplasty over vertebroplasty. While no blinded trial of kyphoplasty exists, primarily due to the ethical challenges of such a trial along with the difficulty in adequately designing a sham procedure, the findings of the 2009 vertebroplasty trials also ultimately cast some doubt on kyphoplasty.

Table 1: Survival probability for BKP versus NSM, VP versus NSM, and BKP versus VP

<table>
<thead>
<tr>
<th>Year</th>
<th>Unadjusted</th>
<th>Adjusted</th>
<th>Unadjusted</th>
<th>Adjusted</th>
<th>Unadjusted</th>
<th>Adjusted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NSM</td>
<td>BKP</td>
<td>NSM</td>
<td>BKP</td>
<td>NSM</td>
<td>VP</td>
</tr>
<tr>
<td>1</td>
<td>76.8%</td>
<td>84.2%</td>
<td>79.1%</td>
<td>85.9%</td>
<td>76.8%</td>
<td>81.6%</td>
</tr>
<tr>
<td>2</td>
<td>66.4%</td>
<td>74.2%</td>
<td>67.9%</td>
<td>75.4%</td>
<td>66.4%</td>
<td>70.9%</td>
</tr>
<tr>
<td>3</td>
<td>57.4%</td>
<td>65.6%</td>
<td>58.3%</td>
<td>66.4%</td>
<td>57.4%</td>
<td>62.0%</td>
</tr>
<tr>
<td>4</td>
<td>49.5%</td>
<td>57.8%</td>
<td>49.6%</td>
<td>57.9%</td>
<td>49.5%</td>
<td>53.9%</td>
</tr>
<tr>
<td>5</td>
<td>42.5%</td>
<td>50.9%</td>
<td>41.9%</td>
<td>50.3%</td>
<td>42.5%</td>
<td>46.7%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>Unadjusted</th>
<th>Adjusted</th>
<th>Unadjusted</th>
<th>Adjusted</th>
<th>Unadjusted</th>
<th>Adjusted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NSM</td>
<td>BKP</td>
<td>NSM</td>
<td>BKP</td>
<td>NSM</td>
<td>VP</td>
</tr>
<tr>
<td>1</td>
<td>13.5 (12.1–13.9)</td>
<td>14.8 (14.4–15.2)</td>
<td>20.8 (19.9–21.8)</td>
<td>22.8 (21.8–23.9)</td>
<td>36.7 (34.5–39.2)</td>
<td>33.9 (31.9–36.2)</td>
</tr>
<tr>
<td>2</td>
<td>12.9 (12.6–13.3)</td>
<td>13.4 (13.1–13.6)</td>
<td>22.2 (20.9–23.8)</td>
<td>23.0 (21.6–24.6)</td>
<td>30.3 (28.3–32.9)</td>
<td>29.1 (27.0–31.6)</td>
</tr>
<tr>
<td>3</td>
<td>12.2 (12.0–12.5)</td>
<td>12.4 (12.1–12.7)</td>
<td>21.6 (20.1–23.3)</td>
<td>21.9 (20.3–23.6)</td>
<td>26.7 (24.6–29.2)</td>
<td>26.1 (24.0–28.5)</td>
</tr>
<tr>
<td>4</td>
<td>12.1 (11.8–12.4)</td>
<td>12.1 (11.8–12.4)</td>
<td>22.9 (21.1–25.1)</td>
<td>23.0 (21.1–25.1)</td>
<td>25.8 (23.6–28.5)</td>
<td>25.4 (23.2–28.3)</td>
</tr>
<tr>
<td>5</td>
<td>12.0 (11.4–12.6)</td>
<td>11.9 (11.3–12.6)</td>
<td>23.8 (21.8–26.3)</td>
<td>23.8 (21.7–26.3)</td>
<td>241 (22.0–26.6)</td>
<td>23.9 (21.8–26.4)</td>
</tr>
</tbody>
</table>

Table 2: Number needed to treat for BKP versus NSM, VP versus NSM, and BKP versus VP

<table>
<thead>
<tr>
<th>Year</th>
<th>Unadjusted (95% CI)</th>
<th>Adjusted (95% CI)</th>
<th>Unadjusted (95% CI)</th>
<th>Adjusted (95% CI)</th>
<th>Unadjusted (95% CI)</th>
<th>Adjusted (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NSM</td>
<td>BKP</td>
<td>NSM</td>
<td>BKP</td>
<td>NSM</td>
<td>VP</td>
</tr>
<tr>
<td>1</td>
<td>13.5 (12.1–13.9)</td>
<td>14.8 (14.4–15.2)</td>
<td>20.8 (19.9–21.8)</td>
<td>22.8 (21.8–23.9)</td>
<td>36.7 (34.5–39.2)</td>
<td>33.9 (31.9–36.2)</td>
</tr>
<tr>
<td>2</td>
<td>12.9 (12.6–13.3)</td>
<td>13.4 (13.1–13.6)</td>
<td>22.2 (20.9–23.8)</td>
<td>23.0 (21.6–24.6)</td>
<td>30.3 (28.3–32.9)</td>
<td>29.1 (27.0–31.6)</td>
</tr>
<tr>
<td>3</td>
<td>12.2 (12.0–12.5)</td>
<td>12.4 (12.1–12.7)</td>
<td>21.6 (20.1–23.3)</td>
<td>21.9 (20.3–23.6)</td>
<td>26.7 (24.6–29.2)</td>
<td>26.1 (24.0–28.5)</td>
</tr>
<tr>
<td>4</td>
<td>12.1 (11.8–12.4)</td>
<td>12.1 (11.8–12.4)</td>
<td>22.9 (21.1–25.1)</td>
<td>23.0 (21.1–25.1)</td>
<td>25.8 (23.6–28.5)</td>
<td>25.4 (23.2–28.3)</td>
</tr>
<tr>
<td>5</td>
<td>12.0 (11.4–12.6)</td>
<td>11.9 (11.3–12.6)</td>
<td>23.8 (21.8–26.3)</td>
<td>23.8 (21.7–26.3)</td>
<td>241 (22.0–26.6)</td>
<td>23.9 (21.8–26.4)</td>
</tr>
</tbody>
</table>
have shown that there is no number one can treat in 5 years of the use of stents for coronary artery disease, of which studies et al.7 used propensity score matching to best account for patient heterogeneous population being analyzed retrospectively. Ong with future studies including it as a prespecified end point.

Further study should address the risks of vertebral body collapse and the importance of preserving vertebral body height, and it could be conjectured that balloon expansion leads to better sagittal reconstruction. Confounding factors may be involved in this difference in benefit. Most notably, patients with severe comorbidities and anesthetic risk may be offered vertebroplasty over kyphoplasty; hence, a difference in outcome may be attributed to underlying medical conditions.

The original use of NNT was an epidemiologic measure designed to present data in which 2 different treatments are compared with respect to incidence rates of an unfavorable event.21 In practice, it facilitates understanding the clinical significance of an intervention. Although the NNT was, at first, designed to be used to indicate treatment impact in randomized controlled trials, it has also been applied to observational studies. In the case of vertebral augmentation, the very salient unfavorable event is death as demonstrated in the single largest claims-based study of VCFs.7

To put these numbers into practical terms, we can make comparisons with known interventions. The European Cooperative Acute Stroke Study III (ECASS-III) trial studied the use of IV-tPA administered from 3 to 4.5 hours following ischemic stroke. It was found that 15 people with acute stroke symptoms needed to be treated to achieve a single favorable outcome.22 The US Preventive Services Task Force (USPSTF) trial, pooling data from 15 trials with >70,000 patients at low risk for cardiovascular disease, found that 0.4% fewer patients taking a statin died than patients taking a placebo. This equated to an NNT of 250.23 In another study, for those taking aspirin for 1 year to prevent a first heart attack or stroke, a cardiovascular event was prevented for 1 person in a patient population of 1667, compared with 1 in 3000 for stroke.24 Each of these interventions are more favorable than the use of stents for coronary artery disease, of which studies have shown that there is no number one can treat in 5 years of follow-up to achieve any benefit.25

Following the publication of the 2009 sham trials of vertebroplasty, the controversy and debate were so substantial that opportunities for learning from their findings were lost. As Firanescu et al.9 argued, clinical care pathways that focus on improvement to NSM, moving it away from the scourge of high-dose opioid anesthesia, should be considered.26 The positive outcomes demonstrated in the sham-controlled groups could indicate the potential for open-label peristioelal numbing and/or medial branch blocks to be used as treatment in patients thought to be at low risk of subsequent vertebral body collapse. Further study should address the risks of vertebral body collapse and the importance of preserving vertebral body height, with future studies including it as a prespecified end point.

Using large claims-based datasets inherently equates to a heterogeneous population being analyzed retrospectively. Ong et al.7 used propensity score matching to best account for patient covariates and reduce the bias of confounding variables. Recent advances in vertebral augmentation practices have led to discussion of the suitability of particular augmentation procedures for specific patient subgroups (eg, traumatic fractures, neoplastic fractures, patients with significant comorbidities).27 Thus, there is scope for future research to analyze the utility of different techniques in different subgroups on the basis of their unique clinical conditions.28 Despite these new considerations coming into focus in the discussion of vertebral augmentation, there still remains debate regarding the utility of augmentation over non-operative management.29 Thus, the low NNT to save a life when offering augmentation is an important consideration in today’s evidentiary landscape.

CONCLUSIONS

This NNT analysis of >2,000,000 patients with VCF reveals that only 15 patients need to be treated to save 1 life at 1 year. This is an obvious clinically significant impact, and given that all augmentation clinical trials are underpowered to detect a mortality benefit, this large dataset analysis reveals that vertebral augmentation provides a significant mortality benefit over nonsurgical management with a low NNT.

REFERENCES


Altman DG, Andersen PK. Calculating the number needed to treat for trials where the outcome is time to an event. *BMJ* 1999;319:1492–95 CrossRef Medline


Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. *BMJ* 1995;310:542–54 CrossRef Medline


