Review Article

AO SPINE

Global Spine Journal
2020, Vol. 10(3) 332-345

DOI: 10.1177/2192568219841046
journals.sagepub.com/home/gsj
SAGE

Superiority Claims for Spinal Devices: A Systematic Review of Randomized Controlled Trials

S. Raymond Golish, MD, PhD, MBA[1], Michael W. Groff, MD[2], Ali Araghi, DO[3], and Jason A. Inzana, PhD[4]

Abstract

Study Design: Systematic review.

Objectives: Superiority claims for medical devices are commonly derived from noninferiority trials, but interpretation of such claims can be challenging. This study aimed to (*a*) establish the prevalence of noninferiority and superiority designs among spinal device trials, (*b*) assess the frequency of post hoc superiority claims from noninferiority studies, and (*c*) critically evaluate the risk of bias in claims that could translate to misleading conclusions.

Methods: Study bias was assessed using the Cochrane Risk of Bias Tool. The risk of bias for the superiority claim was established based on post hoc hypothesis specification, analysis of the intention-to-treat population, post hoc modification of a priori primary outcomes, and sensitivity analyses.

Results: Forty-one studies were identified from 1895 records. Nineteen (46%) were noninferiority trials. Fifteen more (37%) were noninferiority trials with a secondary superiority hypothesis specified a priori. Seven (17%) were superiority trials. Of the 34 noninferiority trials, 14 (41%) made superiority claims. A medium or high risk of bias was related to the superiority claim in 9 of those trials (64%), which was due to the analyzed population, lacking sensitivity analyses, claims not being robust during sensitivity analyses, post hoc hypotheses, or modified endpoints. Only 4 of the 14 (29%) noninferiority studies provided low bias in the superiority claim, compared with 3 of the 5 (60%) superiority trials.

Conclusions: Health care decision makers should carefully evaluate the risk of bias in each superiority claim and weigh their conclusions appropriately.

Keywords

superiority, noninferiority, randomized controlled trial, claim bias, clinical trial design, spinal device

Introduction

Randomized controlled trials (RCTs) are pivotal in establishing the safety and efficacy of novel spinal devices. Spinal device trials are designed either as noninferiority (NI) or superiority trials. In NI trials, the aim is to demonstrate that an investigational device is similar to an accepted surgical procedure or device by showing that the investigational device is not worse (by a small margin). In superiority trials, the goal is to show that the investigational device is superior to a control treatment, which may be nonsurgical care or a gold standard surgical procedure.[1]

In the United States, most investigational device exemption (IDE) studies of novel spinal devices are designed as NI trials because of effect size, secondary benefits, and ethical considerations.[2,3] Many of these NI trials also test for superiority of the investigational device (NI + S), since sponsors are under pressure from physicians and payers to show improvements in safety, efficacy, and cost-effectiveness. The nuances associated

[1] Jupiter Medical Center, Palm Beach, FL, USA
[2] Brigham and Women's Hospital, Boston, MA, USA
[3] The CORE Institute, Phoenix, AZ, USA
[4] Telos Partners, LLC, Denver, CO, USA

Corresponding Author:
Jason A. Inzana, Telos Partners, LLC, Lafayette, CO 80026, USA.
Email: jinzana@telospartnersllc.com

with post hoc tests of superiority in this setting can make interpretation of such superiority claims challenging and potentially misleading.[4] The aims of NI versus superiority trials differ substantially, so the methodology associated with design, analysis, and interpretation is also different. For example, it is conservative to analyze the intention-to-treat (ITT) population for superiority analyses, but it is not conservative for NI analyses since any confounding events will drive the result toward equivalence.[1,5-7] Additionally, post hoc specification of hypotheses must be avoided in confirmatory trials,[4] which requires that superiority analyses are well-defined in the statistical plan a priori. Finally, it is critical to address not only the statistical superiority but also the clinical significance of the differences observed. This is particularly true when a NI margin is imposed for the primary analysis, so that interpretation can be symmetric with less potential for bias.[8]

The purpose of this study was to review the literature for reports of randomized controlled trials of spinal devices from the year 2000 to present. For each report, the primary study design was classified as NI, superiority, or NI with an additional predefined superiority analysis (NI + S). For each trial, superiority claims were identified and were assessed for potential sources of bias by multiple reviewers using a standardized tool. The hypothesis was that NI trials would predominate, and that superiority claims derived from NI trials would have a greater risk of potential bias and less reliability.

Methods

Study Selection

This systematic review was performed according to the guidelines provided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement.[9] Search criteria were developed to identify RCTs of medical devices or biologics for the spine through PubMed/MEDLINE, Embase, ClinicalTrials.gov, the World Health Organization's International Clinical Trials Registry Platform (ICTRP), as well as the Food and Drug Administration's (FDA) databases on premarket approvals (PMA), postapproval studies (PAS), and proceedings from FDA advisory committee meetings of the Orthopaedic and Rehabilitation Devices Panel. Search filters included information available in the English language and publication since the year 2000 to focus on more recent trends in trial design, analysis, and interpretation. Search terms and inclusion/exclusion criteria for record screening are summarized in Table 1 for the PubMed and Embase searches while further details for these and the other databases are provided in Appendix A (see Supplementary Material available in the online version of the article). Two independent researchers screened the identified records for inclusion and exclusion. The final search of each database was completed between May 15 and June 15, 2018. When relevant studies were identified through one database, the other databases were further queried to identify protocols or reports that may provide supplemental study information for data extraction. Only the primary endpoint and primary outcomes were evaluated in this review, considering those were the basis for trial design.

Table 1. Search Terms and Screening Criteria Used for the PubMed and Embase Databases.

Anatomical Terms	Inclusion Criteria
1. Spine	• Spinal devices or biologics
2. Lumbar	• Randomized controlled trials (RCTs) with description of trial design (noninferiority or superiority)
3. Lumbosacral	• Reports on the primary study endpoint
4. Thoracolumbar	
5. Cervical	
6. Intervertebral disc	
Device Terms	Exclusion Criteria
7. Device	• Drugs, diagnostics, or nonsurgical management methods
8. Instrumentation	• Surgical techniques or other non–device-based interventions (navigation and robot-assisted surgery were inclusive)
9. Spacer	• Nonclinical studies (eg, preclinical research or economics analyses)
10. Cage	• Non-RCT studies (eg, single-arm or nonrandomized studies)
11. Interbody fusion	• Insufficient description of study design (eg, noninferiority or superiority)
12. Annular (anular) closure	• Information from multiple publications about a single clinical study were merged, but only for the primary endpoint
13. Fixation	
14. Stimulator	
15. Prosthesis	
16. Replacement	
17. Arthroplasty	
18. Biomaterial	
19. Biologic	
Study Design Terms	Search Filters
20. Investigational device exemption (IDE)	• Language: English
21. Non-inferior(-ity)	• Years: 2000 to Present
22. Noninferior(-ity)	• Study type: Randomized Controlled Trial
23. Superiority	
24. Randomized controlled trial (RCT)	
25. Meta-analysis	
26. Review	
Search Term Combination Strategy	
[1-6]/or AND {([7-19]/or AND [24]) or [20-24]/or} NOT [25-26]/or	

Data Extraction

Relevant data was extracted from each included study by 2 independent researchers. Discrepancies in identifying the study design were resolved through discussion and identification of additional, clarifying documentation in five cases. The data of interest for this review included the study objective, hypotheses for primary endpoints (NI or superiority), margins or effect size used in trial design, primary outcomes and endpoints, sample sizes, conclusions or claims made in the report (NI or superiority), treatment effects of superior devices, and any statistical or clinical significance considerations relating to the superiority claims. When multiple articles reported on the same study (eg, outcomes at different time points), each article was screened for the data of interest related to the a priori study design and primary endpoint.

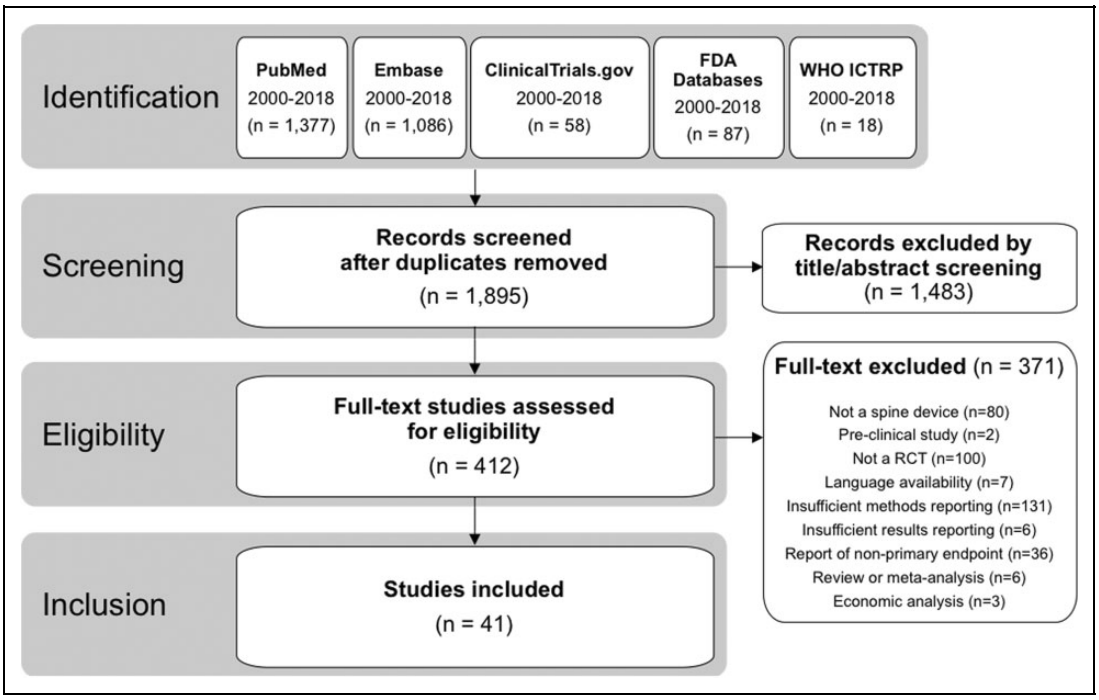

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram demonstrating flow of records from identification through inclusion.

Risk of Bias Assessment

The general risk of bias was evaluated for each study using the Cochrane Risk of Bias Tool for Randomized Controlled Trials[10] and interpreted according to the key domains described by Pavon et al[11] (Supplementary Table S1, Appendix B). The reporting of any financial disclosures, or lack thereof, was also noted but was not considered in the overall risk of bias evaluation. Additionally, the risk of bias specifically related to superiority claims was assessed. The criteria for this assessment included analyses that were not specified a priori, analysis of the ITT population, post hoc modification of primary outcomes, and any sensitivity analyses performed on the analysis population or missing value imputation (Supplementary Table S2, Appendix B). The reporting of confidence intervals for superiority claims was also noted but was not considered in the overall risk of bias for the superiority claim.

Results

Overview of Included Studies

Across all 7 databases, 1895 unique records were identified, and 41 unique studies met the inclusion/exclusion criteria (Figure 1). Among these 41 studies, the most common investigational spinal devices were cervical disc replacements (9/41; 22%), followed by interspinous/interlaminar spacers (7/41; 17%), biologics used to support spinal fusion (7/41; 17%), lumbar disc replacements (6/41; 15%), vertebroplasty materials (2/41; 5%), spinal cord stimulators (2/41; 5%), interbody fusion cages (2/41; 5%), and 1 each (2%) of a dural sealant, an adhesion barrier gel, an annular closure device, a sacroiliac joint fusion device, a dynamic posterolateral pedicle screw system, and a surgical robot used for pedicle screw placement (Table 2).

There were 19 (46%) studies designed as NI trials, 15 (37%) studies designed as NI + S trials, and 7 (17%) studies designed as superiority trials. Five of the 7 superiority trials were reported within the past 3 years (Figure 2). A composite clinical success (CCS) criterion was the most common primary outcome measure and was typically defined as: an improvement in a patient reported outcome greater than a clinically relevant threshold; the absence of secondary surgical interventions or procedures; the absence of neurologic deterioration; the absence of device and/or procedure related serious adverse events; and possibly radiographic findings.[12]

Sample size calculations were often performed using the methods described by Blackwelder et al[13,14] with a NI margin of 10% (Table 3). A few studies assumed other margins for power calculations, but data was also analyzed with a 10% margin at the request of the FDA.[15-18] No study estimated the NI margin from a prior superiority study that measured the effect size compared with sham or placebo. Three of the superiority studies used Bayesian methods for sample size.[19-21] Two assumed superiority effect sizes of 9%[22] and 23%.[23] One did not describe its power analysis[24] and one assumed a medium effect size (Cohen's d=0.4) for differences in disability scores.[25] Three superiority studies compared with nonsurgical management,[20,21,23] while the rest of the studies used an active surgical control. The active surgical controls represented a standard treatment technique for the respective condition (eg,

Table 2. Summary of Included Studies.

Device Type	Device	Study Identifier	Citation	Study Comparators	Sample Size	Risk of Study Bias
Cervical disc replacement	BRYAN	NCT00437190 PMA P060023	Heller et al, 2009[33]	BRYAN vs ACDF at 2 years	Test: 242 Control: 221	L
	Prestige ST	NCT00642876 PMA P060018	Mummaneni et al, 2007[15]	Prestige ST vs ACDF at 2 years	Test: 276 Control: 265	M
	Prestige LP	NCT00637156 PMA P090029	Gornet et al, 2017[37]	2-level Prestige vs ACDF at 2 years	Test: 209 Control: 188	L
	ProDisc-C	NCT00291018 PMA P070001	Murrey et al, 2009[40]	ProDisc-C vs ACDF at 2 years	Test: 103 Control: 106	M
	Kineflex|C	NCT00374413	Coric et al, 2011[42]	Kineflex|C vs ACDF at 2 years	Test: 136 Control: 133	M
	Mobi-C	NCT00389597 PMA P110002 PMA P110009	Hisey et al, 2014[51]	1-level Mobi-C vs ACDF at 2 years	Test: 164 Control: 81	L
			Davis et al, 2013[35]	2-level Mobi-C vs ACDF at 2 years	Test: 225 Control: 105	L
	Secure-C	NCT00882661 PMA P100003	Vaccaro et al, 2013[16]	SECURE-C vs ACDF at 2 years	Test: 240 Control: 140	M
	PCM Cervical Disc	NCT00578812 PMA P100012	Phillips et al, 2013[41]	PCM vs ACDF at 2 years	Test: 189 Control: 153	M
Lumbar disc replacement	Charite	NCT00215306 PMA P040006	Blumenthal et al, 2005[52]	Charite vs ALIF at 2 years	Test: 205 Control: 99	M
	ProDisc-L	IDE #G010133 PMA P050010	Zigler et al, 2007[43]	1-level ProDisc-L vs fusion at 2 years	Test: 161 Control: 75	L
		NCT00295009	Delamarter et al, 2011[53]	2-level ProDisc-L vs fusion at 2 years	Test: 165 Control: 72	M
	MAVERICK	NCT00635843	Gornet et al, 2011[17]	MAVERICK vs ALIF at 2 years	Test: 405 Control: 172	L
	Kineflex	NCT00292292	Pettine et al, 2011[26]	Kineflex vs Charite at 2 years	Test: 33 Control: 31	M
	activL	NCT00589797 PMA P120024	Garcia et al, 2015[18]	activL vs ProDisc-L or Charite at 2 years	Test: 218 Control: 106	L
Biologic: Fusion	OP-1 Putty	NCT00677950	Vaccaro et al, 2008[54]	OP-1 vs autograft (noninstrumented fusion) at 2 years	Test: 208 Control: 87	L
		ISRCTN43648350	Delawi et al, 2016[55]	OP-1 vs autograft (instrumented fusion) at 1 year	Test: 60 Control: 59	L
	Novosis (rhBMP-2)	NCT01764906	Cho et al, 2017[56]	Novosis vs autograft at 6 months	Test: 42 Control: 51	L
	Bonion	NCT01615328	Yi et al, 2015[57]	Bonion vs. βTCP/HA allografts at 2 years	Test: 38 Control: 39	M
	i-Factor	NCT00310440 PMA P140019	Arnold et al, 2016[58]	i-Factor vs autograft at 1 year	Test: 165 Control: 154	M
	INFUSE with LT-cage	PMA P000058	Burkus et al, 2002[59] and 2003[60]	INFUSE vs autograft at 2 years	Test: 143 Control: 136	L
	AMPLIFY (rhBMP-2)	PMA P050036	FDA Executive Summary[61]	AMPLIFY vs autograft at 2 years	Test: 239 Control: 224	L

(continued)

Table 2. (continued)

Device Type	Device	Study Identifier	Citation	Study Comparators	Sample Size	Risk of Study Bias
Adhesion barrier	Oxiplex/SP	PMA P070023	Rhyne et al, 2012[62] and FDA Executive Summary[24]	Surgery + Oxiplex vs surgery alone at 6 months	Test: 177 Control: 175	L
Vertebroplasty	Cortoss	NCT00290862	Bae et al, 2012[63]	Cortoss vs PMMA at 2 years	Test: 162 Control: 94	L
	Kiva	NCT01123512	Tutton et al, 2015[64]	Kiva vs. balloon kyphoplasty at 1 year	Test: 147 Control: 153	L
Dural Sealant	Adherus	NCT01158378 PMA P130014	Strong et al, 2017[27]	Adherus vs DuraSeal at 4 months	Test: 124 Control: 126	L
Interlaminar/ Spinous device	Superion	NCT00692276 PMA P140004	Patel et al, 2015[28]	Superion vs X-Stop at 2 years	Test: 190 Control: 201	M
	Coflex	NCT00534235 PMA P110008	Davis et al, 2013[46]	Coflex vs fusion at 1-2 levels at 2 years	Test: 230 Control: 114	L
		NCT01316211	Schmidt et al, 2018[25]	Coflex vs decompressive surgery at 2 years	Test: 115 Control: 115	L
	Aperius	NCT00905359	Meyer et al, 2018[47]	Aperius vs decompressive surgery at 2 years	Test: 82 Control: 81	M
	X-Stop	N/A	Strömqvist et al, 2013[48]	X-Stop vs decompressive surgery at 2 years	Test: 50 Control: 50	M
		PMA P040001	FDA SSED[44] and memo[23]	X-Stop vs nonsurgical management at 2 years	Test: 100 Control: 91	M
	DIAM	IDE G050025 PMA P140007	FDA Executive Summary[21]	DIAM vs nonsurgical management at 1 year	Test: 181 Control: 100	H
Annular closure device	Barricaid	NCT01283438	Thome et al, 2018[19]	Barricaid vs discectomy only at 2 years	Test: 276 Control: 278	L
Sacroiliac joint fusion	iFuse	NCT01640353	Whang et al, 2015[20]	iFuse vs nonsurgical management at 6 months	Test: 102 Control: 46	M
Interbody fusion cage	Novomax	N/A	Lee et al, 2016[29]	NovoMax vs titanium cage at 1 year	Test: 41 Control: 39	L
	BAK/C	PMA P980048	Hacker et al, 2000[65] and FDA SSED[66]	BAK/C vs bone grafting at 2 years	Test: 164 Control: 134	M
Dynamic stabilization	Dynesys	NCT00759057 PMA P070031	FDA Executive Summary[67]	Dynesys vs posterolateral fusion at 2 years	Test: 253 Control: 114	M
Spinal cord stimulator	Senza	NCT01609972 PMA P130022	Kapural et al, 2016[30]	Senza vs low-frequency stimulation at 2 years	Test: 101 Control: 97	L
	Axium	NCT01923285 PMA P150004	FDA SSED[31]	Axium vs marketed control device at 3 months	Test: 76 Control: 76	M
Surgical robot	Renaissance	NCT02121249	Kim et al, 2017[22]	Robot-assisted pedicle screw accuracy vs freehand	Test: 37 Control: 41	L

Abbreviations: L, low; M, medium; H, high; N/A, not applicable; FDA, US Food and Drug Administration; SSED, Summary of Safety and Effectiveness Data.

fusion as a control for disc replacements and autograft for biologics). Seven of the studies compared with devices of the same class that were already available on the market.[18,26-31]

The overall risk of study bias was low in 22 of the 41 studies (54%), medium in 18 (44%), and high in 1 (2%) of the studies (Table 2; Supplementary Table S1). Medium risk ratings were attributable to potential attrition bias, an unclear blinding of outcome assessors, or potential limitations in randomization. The study with a high risk of bias suffered from a high rate of crossover subjects in the primary analysis dataset, a lack of

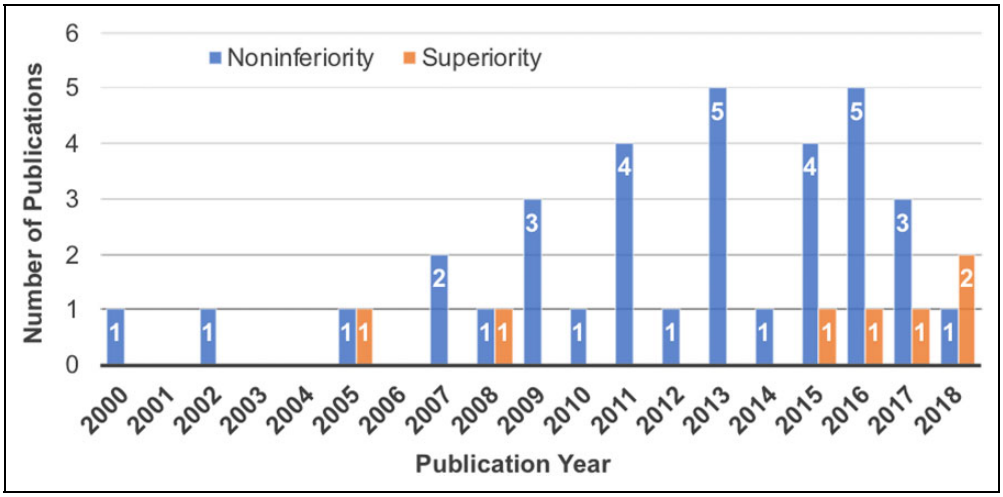

Figure 2. Trends in noninferiority versus superiority designs for randomized controlled trials of spinal devices that were included in this review since the year 2000.

sensitivity analyses, and uncertainty of concurrent interventions that could confound outcomes. The use of independent assessors, such as radiologists who were blinded to other outcomes, was considered an appropriate substitute for investigator blinding. Financial disclosure statements were only provided in 23 (56%) of the reports.

Evaluation of Superiority Claims

Among the 19 NI studies, 4 (21%) made post hoc superiority claims. Ten of the 15 (67%) NI + S studies and 5 of the 7 (71%) superiority trials satisfied their a priori superiority hypothesis (Table 3). All 19 superiority conclusions were based on statistical analyses with a superiority margin equal to zero. Although none of the studies discussed the superiority margin, the difference in proportions of treatment success (CCS) exceeded a +10% margin in 16 of the 19 studies (Table 3). However, the lower bound of the 95% confidence interval did not exceed +10% in most of the studies reporting that information.

The superiority claims in 4 of the 10 NI + S studies were found to be at a high risk of bias and 1 was at a medium risk of bias (Table 3; Supplementary Table S2). The NI + S study with a medium risk of bias for the superiority claim did not describe the analysis population and the FDA panel recommended only allowing NI claims.[15,32] In 1 NI + S study at a high risk of bias, the superiority claim was not robust to the sensitivity analyses of imputed values or the per-protocol analysis.[33,34] Two other NI+S studies at high risk of bias performed the superiority analysis on the as-treated population rather than the ITT population and did not describe any sensitivity analyses for the population or missing value imputations.[35-38] This is particularly important when up to 18% of patients did not receive the assigned treatment, which could compromise the efficacy of randomization.[37] The fourth NI + S study at high risk of bias only reported the safety analysis to be a predefined superiority analysis, which failed to meet statistical superiority; yet, overall success rates were claimed to be superior.[39]

All 4 superiority claims from NI studies were rated to be at a high risk of bias due to the apparent post hoc specification of the superiority hypothesis and lack of multiplicity adjustment.[40-43] Furthermore, the analysis population was either not described[40,42] or the per-protocol population was used[41,43] in each of these studies. One NI study claimed superiority based solely on a post hoc modified CCS outcome since only NI could be claimed with the original primary endpoint.[40]

Three of the 5 superiority trials were rated as a low risk of bias for the superiority claim,[19,23,25,44] 1 was rated with a medium risk of bias due to the lack of reporting on the analysis population or sensitivity analyses,[20] and 1 was at a high risk of bias due to a high rate of crossover in the primary analysis dataset, no sensitivity analyses, and potentially confounding concurrent interventions.[21] The study at high risk of bias did not lead to FDA approval of the investigational device. Among these 5 superiority trials, only 2 described sensitivity analyses (the conclusions were robust to the alternate analyses).[19,23] Although it was not considered in the risk of bias evaluation, 12 of the 19 (63%) studies claiming superiority reported the associated confidence intervals, which are useful for understanding the effect size. These 12 studies were comprised of 1 of the 4 NI studies, 7 of the 10 NI + S studies, and 4 of the 5 superiority studies.

Discussion

The majority of RCTs for spinal device trials are designed primarily as NI trials based on effect size, secondary benefits, or ethical considerations; however, sponsors frequently attempt to establish post hoc superiority claims. The present study demonstrates that post hoc superiority claims derived from NI trials often suffer from a high to medium risk of bias due to analyzing the per-protocol or as-treated populations without sensitivity

Table 3. Summary of Study Design Characteristics and Conclusions.

Device Type	Device	Citation	Hypotheses	Primary Outcome	Margin	Report Claim	Risk of Superiority Claim Bias	Superiority Outcomes and Considerations
Cervical disc replacement	BRYAN	Heller et al, 2009[33]	Primary: NI Secondary: S	CCS	NI: 10%	S	H	Test CCS: 83% (95% CI: 77%-87%) Control CCS: 73% (95% CI: 66%-79%) $P = .01$, one-sided FET; S claim not supported by PP analysis or sensitivity analyses of missing values
	Prestige ST	Mummaneni et al, 2007[15]	Primary: NI Secondary: S	CCS	NI: 10%	S	M	Test CCS: 79% Control CCS: 68% $P = .005$, one-sided FET; Analysis population not described; FDA panel recommended limit of NI claim only
	Prestige LP	Gornet et al. 2017[37]	Primary: NI Secondary: S	CCS	NI: 10%	S	H	Test CCS: 81% Control CCS: 69% BPP = 0.99; As-treated analysis, 8%-18% of patients received different treatment; no sensitivity analysis
	ProDisc-C	Murrey et al, 2009[40]	NI	CCS	NI: 10%	CCS: NI mCCS: S	H	Test mCCS: 74% Control mCCS: 61% $P = .047$, one-sided FET; Original CCS confirmed NI, post hoc mCCS used for superiority test; analysis population not described
	Kineflex C	Coric et al. 2011[42]	NI	CCS	NI: 10%	S	H	Test CCS: 85% Control CCS: 71% $P = .05$, 2-sided FET; Superiority testing not prespecified; analyzed population not described
	Mobi-C	Hisey et al, 2014[51]	Primary: NI Secondary: S	CCS	NI: 10%	NI	N/A	N/A
		Davis et al, 2013a[35]	Primary: NI Secondary: S	CCS	NI: 10%	S	H	Test CCS: 70% Control CCS: 37% $P < .0001$, Farrington-Manning test; Analysis of as-treated population; Sensitivity analyses only described to support NI
	Secure-C	Vaccaro et al, 2013[16]	Primary: NI Secondary: S	CCS	NI: 10%	S	L	Test CCS: 84% Control CCS: 73% BCI: 0.6%-20%, BPP = 0.98; Sensitivity analysis not performed for S claim
	PCM	Phillips et al, 2013[41]	NI	CCS	NI: 12.5%	S	H	Test CCS: 75% (95% CI: 69%-81%) Control CCS: 65% (95% CI: 57%-73%) $P = .02$, one-sided Z-test; S analysis not defined a priori, no ITT analysis
	Charite	Blumenthal et al, 2005[52]	NI	CCS	NI: 15%	NI	N/A	N/A

(continued)

Table 3. (continued)

Device Type	Device	Citation	Hypotheses	Primary Outcome	Margin	Report Claim	Risk of Superiority Claim Bias	Superiority Outcomes and Considerations
Lumbar disc replacement	ProDisc-L	Zigler et al, 2007[43]	NI	CCS	NI: 12.5%	S	H	Test CCS: 53% Control CCS: 41% $P = .044$, Stat method not specified; S analysis not defined a priori, ITT population not analyzed, 50 nonrandomized patients included
		Delamarter et al, 2011[53]	NI	CCS	NI: 12.5%	NI	N/A	N/A
	MAVERICK	Gornet et al, 2011[17]	Primary: NI Secondary: S	CCS	NI: 10%	S	L	Test CCS: 74% Control CCS: 55% $P < .001$, one-sided FET
	Kineflex	Pettine et al, 2011[26]	NI	ODI and VAS scores	NI: 10pt ODI and 18pt VAS	NI	N/A	N/A
	activL	Garcia et al, 2015[18]	Primary: NI Secondary: S	CCS	NI: 10%	S	L	Treatment difference ~ 14% $P = .02$, O'Brien-Fleming sequential spending function
Biologic: Fusion	OP-1 Putty	Vaccaro et al, 2008[54] Delawi et al, 2016[55]	NI NI	CCS CCS	NR NI: 15%	NI Inferior	N/A N/A	N/A N/A
	Novosis (rhBMP-2)	Cho et al, 2017[56]	NI	Fusion rate	NI: 10%	NI	N/A	N/A
	Bonion	Yi et al, 2015[57]	NI	Fusion rate	NI: 15%	NI	N/A	N/A
	i-Factor	Arnold et al, 2016[58]	Efficacy: NI Safety: S	CCS	NI: 10-15%	S	H	Test CCS: 69% Control CCS: 57% $P = .038$, Wald asymptotic approach; Failed safety superiority endpoint; S claim based on CCS that was not predefined in statistical plan
	INFUSE with LT-cage	Burkus et al, 2002[59] and 2003[60]	Primary: NI Secondary: S	Fusion rate	NR	NI	N/A	N/A
	AMPLIFY (rhBMP-2)	FDA Executive Summary[61]	Primary: NI Secondary: S	CCS	NI: 10%	NI	N/A	N/A
Adhesion barrier	Oxiplex/SP	FDA Executive Summary[24]	S	LSOQ	S: 0	Not S	N/A	N/A
Vertobroplasty	Cortoss	Bae et al, 2012[63]	NI	VAS and ODI	\NI: 12.5%	NI	N/A	N/A
	Kiva	Tutton et al, 2015[64]	Primary: NI Secondary: S	CCS	NI: 12.5%	NI	N/A	N/A
Dural sealant	Adherus	Strong et al, 2017[27]	NI	CCS	NI: 10%	NI	N/A	N/A

(continued)

Table 3. (continued)

Device Type	Device	Citation	Hypotheses	Primary Outcome	Margin	Report Claim	Risk of Superiority Claim Bias	Superiority Outcomes and Considerations
	Superion	Patel et al, 2015[28]	NI	CCS	NI: 10%	NI	N/A	N/A
	Coflex	Davis et al, 2013b[46]	Primary: NI Secondary: S	CCS	NI: 10%	NI	N/A	N/A
		Schmidt et al, 2018[25]	S	CCS	S: 0	S	L	Test CCS: 58% Control CCS: 42% 95% CI of difference: 3%-30% $P = .017$; No sensitivity analysis
	Aperius	Meyer et al, 2018[47]	NI	PF of ZCQ	NI: 10%	NI	N/A	N/A
	X-Stop	Strömqvist et al, 2013[48]	NI	ZCQ	NR	NI	N/A	N/A
		FDA SSED[44] and memo[23]	S	CCS	S: 0	S	L	Test CCS: 46% Control CCS: 5% $P < .001$, 2-sided FET Nonoperative care as control
Interlaminar/Spinous spacer	DIAM	FDA Executive Summary[21]	S	CCS	S: 0	S	H	Test CCS: 64% Control CCS: 15% BPP = 1.0 Nonoperative care as control High rate (>40%) of crossover in primary analysis dataset; No sensitivity analysis discussed; Not approved by FDA
Annular closure device	Barricaid	Thomé et al, 2018[19]	S	CCS & recurrence	S: 0	S	L	Test mCCS: 76% Control mCCS: 66% 95% CI of difference: 2%-18%, $P < .02$ Test recurrence: 50% Control recurrence: 70% 95% CI of difference: −12% to −28%, $P < .001$
Sacroiliac joint fusion	iFuse	Whang et al, 2015[20]	S	CCS	S: 0	S	M	Test CCS: 81% (72%-88%) Control CCS: 24% (13%-39%) BPP > 0.999; Analysis population and sensitivity analysis not reported; Nonoperative care as control
Interbody fusion spacer	Novomax	Lee et al, 2016[29]	NI	Fusion rate	NI: 15%	NI	N/A	N/A
	BAK/C	Hacker et al, 2000[65] and FDA SSED[66]	NI	CCS	NR	NI	N/A	N/A

(continued)

Table 3. (continued)

Device Type	Device	Citation	Hypotheses	Primary Outcome	Margin	Report Claim	Risk of Superiority Claim Bias	Superiority Outcomes and Considerations
Dynamic stabilization	Dynesys	FDA Executive Summary[67]	NI	CCS	NI: 10%	NI	N/A	N/A
Spinal cord stimulator	Senza	Kapural et al, 2016[30]	Primary: NI Secondary: S	≥50% pain reduction	NI: 10%	S	L	Test back pain response: 77% Control back pain response: 49% 95% CI of difference: 10%-42%, $P < .001$ Test leg pain response: 73% Control leg pain response: 49% 95% CI of difference: 6%-39%, $P = .003$
	Axium	PMA P150004[31]	Primary: NI Secondary: S	CCS	NI: 10%	S	L	Test CCS: 81% (95% CI: 70%-90%) Control CCS: 56% (95% CI: 43%-68%) $P = .0004$
Surgical robot	Renaissance	Kim et al, 2017[22]	S	Screw accuracy	S: 9%	Not S	N/A	N/A

Abbreviations: NI, noninferiority; S, superiority; NR, not reported; N/A, not applicable; H, high; M, medium; L, low; PP, per-protocol; ITT, intention to treat; FDA, US Food and Drug Administration; BPP, Bayesian posterior probability; BCI, Bayesian credible interval; FET, Fisher's exact test; CB, confidence bound; CCS, composite clinical success (reports often referred to this as "overall success"); mCCS, modified CCS; PF of ZCQ, physical function component of Zurich Claudication Questionnaire.

analysis, the claims not being robust during sensitivity analysis, or the claim being based on post hoc modified endpoints. This is important, since sponsors are under pressure from physicians, payers, and health care systems to demonstrate improvements in safety, efficacy, and cost-effectiveness. By claiming superiority in some aspects of safety and effectiveness, the sponsor can argue an improved value proposition. The current study suggests that such post hoc claims may be valid in some instances but should be scrutinized closely by the intended audiences.

The strengths of the present study include the use of multiple databases, the inclusion of important governmental databases in addition to indices of journal articles, a rigorous query methodology, and the use of multiple reviewers to eliminate false positives and combine duplicates from the query results. However, there are several shortcomings to the results. No set of databases or queries can assure complete capture of relevant results. Also, many RCTs have multiple published reports at multiple follow-up timepoints. We focused on the timepoint for the trial's primary endpoint; however, it is possible that additional superiority claims were made at later timepoints. Published protocols that provided adequate details of the a priori study plans were usually unavailable. A published protocol was only identified for 1 study.[19,45] Another limitation was that important details were sometimes not reported, which resulted in an "Unclear" rating for the bias assessments. Similarly, the disclosure of potential conflicts-of-interest was not consistent and could not be meaningfully collected and analyzed. While regulatory bodies and payers may receive additional, nonpublic details of the trials from the sponsor, other researchers must rely on publicly available data. Finally, only superiority claims related to primary outcomes at the primary endpoint were evaluated in this review; however, analyses of secondary outcomes specified a priori can be important for determining the utility of a new device, particularly for NI trials.

Only 17% of the reviewed RCTs of spinal devices since 2000 were designed as superiority trials. Major categories of NI trials included disc replacements (15 studies), biologics for fusion (7 studies), and interspinous/interlaminar spacers (7 studies). Most disc replacement studies compared with fusion, offering the secondary advantage of retaining range of motion. While some of these studies included radiographic measures of motion or fusion, they still used a NI design for the primary endpoint. Biologics studies had the secondary advantage of avoiding donor site morbidity compared with autologous iliac bone grafts, but this was not articulated as a superiority hypothesis and was only indirectly captured in patient reported outcomes described in the NI hypothesis. Interlaminar and interspinous process spacers are promoted as less invasive surgery, but only 1 trial compared an interlaminar device directly to fusion in order to justify the implication that reduced operating room time and blood loss resulted in a net benefit.[46] Other reports comparing interspinous process spacers to decompressive surgery alone referred to improving patient satisfaction, complication rates, and reducing subsequent surgical interventions for the potential advantages of the new devices.[47,48] Such comparisons would be most appropriate as a superiority trial with adverse events included in the CCS, as exemplified by Schmidt et al[25] for an interspinous process spacer versus decompression and analogously by Thomé et al[19] for an annular closure device compared with discectomy alone. Updates to the Consolidated Standards of Reporting Trials (CONSORT) statement were proposed in 2006[49] and incorporated in 2012,[50] which suggest that studies should report the rationale for adopting a NI design and the associated NI margin. Most NI or NI + S studies published after these updates did not specifically discuss rationale for NI vs. superiority designs. Furthermore, only 4 reports provided any rationale for the NI margin, referring to requirements by the FDA.

Using well-rounded CCS measures as the primary endpoint may reduce the options for secondary benefits of the device beyond possible economic advantages. Among the reviewed RCTs on disc replacement, the primary endpoint CCS rates in the control group (fusion) ranged from 37% to 73%[15,16,33,35,37,40-42,51] for cervical discs and from 41% to 55%[17,43,52,53] for lumbar discs, suggesting that a ceiling effect should not be a concern in those studies. Yet each disc replacement was evaluated with NI as the primary hypothesis and superiority as the secondary hypothesis. By focusing on appropriate endpoints, at-risk populations, and CCS criteria that demand well-rounded device success, the ceiling effect can be diminished and areas for improvement can be elucidated.

This review observed four superiority claims made through post hoc analyses of NI trials. These superiority claims were inherently at a high risk of bias due to post hoc hypothesis specification in a confirmatory trial.[4] Furthermore, 50% of the superiority claims from NI+S studies were observed to be at a medium or high risk of bias due to inappropriate methodology for analysis or interpretation of the superiority hypothesis. This was usually attributable to analyzing the as-treated or per-protocol population without consideration of the ITT dataset. Relying solely on as-treated or per-protocol analyses could bias the conclusions, particularly if a significant number of patients did not receive the assigned treatment, there was missing follow-up data, or significant attrition.[7] Overall, such deficiencies were apparent in 64% (9/14) of the NI or NI + S studies making superiority claims, which demonstrates the challenge of ensuring high fidelity conclusions when the superiority hypothesis is secondary to the NI design.

Based on the studies reviewed herein alongside the theoretical considerations of trial design and interpretation, superiority claims derived from NI trials may have a greater likelihood of confounding by methodological mistakes, ambiguities or sources of bias compared to claims derived from superiority trials. However, RCTs with an NI + S design can indeed be rigorous and present superiority claims with high levels of confidence. A few of the reviewed NI+S studies had a low risk of bias in the superiority conclusion because of the meticulous nature of the analysis and reporting, which included sensitivity analyses of both the population dataset and missing value imputations along with confidence intervals that demonstrated substantial margins.[17,18,30,31] The rationale for conducting

these studies as NI + S trials rather than focusing on superiority was unclear. Regulatory or commercial considerations may provide a possible explanation.

Conclusions

Spine studies rarely employ superiority designs for confirmatory trials. NI studies can sometimes yield reliable superiority claims, but meticulous study conduct, analysis, reporting, and interpretation is paramount. Considering the singular goal of superiority trials and the standard methodology of such designs, greater confidence may be derived more readily from the resulting superiority claims. Investigators and sponsors are encouraged to consider superiority trial designs when evaluating novel technologies against a standard of care when feasible. Readers are encouraged to carefully evaluate the risk of bias behind each superiority claim by examining the methodology of the study and associated analyses.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: SRG reports personal fees from Intrinsic Therapeutics, during the conduct of the study; personal fees from US FDA, other from AAOS, personal fees from Paradigm Spine, and personal fees from Wright Medical outside of the submitted work. MWG reports royalty payments from Depuy Spine and Biomet Spine outside of the submitted work. AA reports personal fees from Intrinsic Therapeutics outside of the submitted work. JAI is a salaried employee of Telos Partners, LLC, which received consulting fees from Intrinsic Therapeutics during the conduct of this study.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Telos Partners, LLC received funding from Intrinsic Therapeutics in support of the systematic literature review.

ORCID iD

Jason A. Inzana, PhD https://orcid.org/0000-0002-5221-273X

Supplemental Material

The supplemental material is available in the online version of the article.

References

1. Christensen E. Methodology of superiority vs. equivalence trials and non-inferiority trials. *J Hepatol*. 2007;46:947-954.
2. Golish SR. Pivotal trials of orthopedic surgical devices in the United States: predominance of two-arm non-inferiority designs. *Trials*. 2017;18:348.
3. Golish SR, Reed ML. Spinal devices in the United States-investigational device exemption trials and premarket approval of class III devices. *Spine J*. 2017;17:150-157.
4. Ng TH. Issues of simultaneous tests for noninferiority and superiority. *J Biopharm Stat*. 2003;13:629-662.
5. Committee for Proprietary Medicinal Products. Points to consider on switching between superiority and non-inferiority. *Br J Clin Pharmacol*. 2001;52:223-228.
6. International Council for Harmonisation. Guidance for industry. E9: statistical principals for clinical trials. https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm073137.pdf. Published September 1998. Accessed August 6, 2018.
7. International Council for Harmonisation of Technical Requirements for pharmaceuticals for human use. ICH Harmonised Guideline. Estimands and sensitivity analysis in clinical trials. Draft E9(R1). http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/E9-R1EWG_Step2_Guideline_2017_0616.pdf. Accessed August 8, 2018.
8. Ganju J, Rom D. Non-inferiority versus superiority drug claims: the (not so) subtle distinction. *Trials*. 2017;18:278.
9. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6: e1000097.
10. Higgins JP, Altman DG, Gotzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011;343:d5928.
11. Pavon JM, Williams JW Jr, Adam SS, et al. *Effectiveness of Intermittent Pneumatic Compression Devices for Venous Thromboembolism Prophylaxis in High-Risk Surgical and Medical Patients*. Washington, DC: Health Services Research & Development Service; 2015. https://www.hsrd.research.va.gov/publications/esp/VenousThromboembolism-REPORT.pdf.
12. Golish S, Kurtz SM, Guyer RD. Clinical trials for cervical disk replacement. https://aaos.org/AAOSNow/2018/Jun/Research/research02/? ssopc=1. Accessed September 6, 2018.
13. Blackwelder WC. "Proving the null hypothesis" in clinical trials. *Control Clin Trials*. 1982;3:345-353.
14. Blackwelder WC, Chang MA. Sample size graphs for "proving the null hypothesis." *Control Clin Trials*. 1984;5:97-105.
15. Mummaneni PV, Burkus JK, Haid RW, Traynelis VC, Zdeblick TA. Clinical and radiographic analysis of cervical disc arthroplasty compared with allograft fusion: a randomized controlled clinical trial. *J Neurosurg Spine*. 2007;6:198-209.
16. Vaccaro A, Beutler W, Peppelman W, et al. Clinical outcomes with selectively constrained SECURE-C cervical disc arthroplasty: two-year results from a prospective, randomized, controlled, multicenter investigational device exemption study. *Spine (Phila Pa 1976)*. 2013;38:2227-2239.
17. Gornet MF, Burkus JK, Dryer RF, Peloza JH. Lumbar disc arthroplasty with Maverick disc versus stand-alone interbody fusion: a prospective, randomized, controlled, multicenter investigational device exemption trial. *Spine (Phila Pa 1976)*. 2011;36: E1600-E1611.
18. Garcia R Jr, Yue JJ, Blumenthal S, et al. Lumbar total disc replacement for discogenic low back pain: two-year outcomes of the activL Multicenter Randomized Controlled IDE Clinical Trial. *Spine (Phila Pa 1976)*. 2015;40:1873-1881.
19. Thomé C, Klassen PD, Bouma GJ, et al; Annular Closure RCT Study Group. Annular closure in lumbar microdiscectomy for

prevention of reherniation: a randomized clinical trial. *Spine J*. 2018;18:2278-287.
20. Whang P, Cher D, Polly D, et al. Sacroiliac joint fusion using triangular titanium implants vs. non-surgical management: six-month outcomes from a prospective randomized controlled trial. *Int J Spine Surg*. 2015;9:6.
21. US Food and Drug Administration. FDA executive summary: DIAM Spinal Stabilization System (P140007). https://wayback.archive-it.org/7993/20170723181343/https://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/MedicalDevices/MedicalDevicesAdvisoryCommittee/OrthopaedicandRehabilitationDevicesPanel/UCM486413.pdf. Accessed September 4, 2018.
22. Kim HJ, Jung WI, Chang BS, Lee CK, Kang KT, Yeom JS. A prospective, randomized, controlled trial of robot-assisted vs free-hand pedicle screw fixation in spine surgery. *Int J Med Robot*. 2017;13(3). doi:10.1002/rcs.1779
23. Office of Surveillance and Biometrics. Final statistical review of PMA P040001, St. Francis Medical's X-STOPÒ Interspinous Process Distraction System. https://wayback.archive-it.org/7993/20170405120856/https://www.fda.gov/ohrms/dockets/ac/04/briefing/2004-4064b1_04_statistics%20memo-print%20version.pdf. Accessed September 4, 2018.
24. US Food and Drug Administration. FDA PMA P070023 executive summary. OxiplexÒ/SP Gel. https://wayback.archive-it.org/7993/20170405042451/https://www.fda.gov/ohrms/dockets/ac/08/briefing/2008-4374b1-07.pdf. Published July 15, 2008. Accessed September 4, 2018.
25. Schmidt S, Franke J, Rauschmann M, Adelt D, Bonsanto MM, Sola S. Prospective, randomized, multicenter study with 2-year follow-up to compare the performance of decompression with and without interlaminar stabilization. *J Neurosurg Spine*. 2018;28:406-415.
26. Pettine K, Hersh A. Kineflex lumbar artificial disc versus Charite lumbar total disc replacement for the treatment of degenerative disc disease: a randomized non-inferiority trial with minimum of 2 years' follow-up. *SAS J*. 2011;5:108-113.
27. Strong MJ, West GA, Woo H, et al. A pivotal randomized clinical trial evaluating the safety and effectiveness of a novel Hydrogel Dural Sealant as an adjunct to dural repair. *Oper Neurosurg (Hagerstown)*. 2017;13:204-212.
28. Patel VV, Whang PG, Haley TR, et al. Superion interspinous process spacer for intermittent neurogenic claudication secondary to moderate lumbar spinal stenosis: two-year results from a randomized controlled FDA-IDE pivotal trial. *Spine (Phila Pa 1976)*. 2015;40:275-282.
29. Lee JH, Kong CB, Yang JJ, et al. Comparison of fusion rate and clinical results between CaO-SiO_2-P_2O_5-B_2O_3 bioactive glass ceramics spacer with titanium cages in posterior lumbar interbody fusion. *Spine J*. 2016;16:1367-1376.
30. Kapural L, Yu C, Doust MW, et al. Comparison of 10-kHz high-frequency and traditional low-frequency spinal cord stimulation for the treatment of chronic back and leg pain: 24-month results from a multicenter, randomized, controlled pivotal trial. *Neurosurgery*. 2016;79:667-677.
31. US Food and Drug Administration. Premarket approval (P150004). https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfPMA/pma.cfm? id=P150004. Accessed June 14, 2018.
32. US Food and Drug Administration. Premarket approval (P060018). https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpma/pma.cfm? id=P060018. Accessed August 6, 2018.
33. Heller JG, Sasso RC, Papadopoulos SM, et al. Comparison of BRYAN cervical disc arthroplasty with anterior cervical decompression and fusion: clinical and radiographic results of a randomized, controlled, clinical trial. *Spine (Phila Pa 1976)*. 2009;34:101-107.
34. US Food and Drug Administration. Premarket approval (P060023). https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpma/pma.cfm? id=P060023. Accessed August 7, 2018.
35. Davis RJ, Kim KD, Hisey MS, et al. Cervical total disc replacement with the Mobi-C cervical artificial disc compared with anterior discectomy and fusion for treatment of 2-level symptomatic degenerative disc disease: a prospective, randomized, controlled multicenter clinical trial: clinical article. *J Neurosurg Spine*. 2013;19:532-545.
36. US Food and Drug Administration. Premarket approval (P110009). https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpma/pma.cfm? id=P110009. Accessed August 8, 2018.
37. Gornet MF, Lanman TH, Burkus JK, et al. Cervical disc arthroplasty with the Prestige LP disc versus anterior cervical discectomy and fusion, at 2 levels: results of a prospective, multicenter randomized controlled clinical trial at 24 months. *J Neurosurg Spine*. 2017;26:653-667.
38. US Food and Drug Administration. Premarket approval (P090029). https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpma/pma.cfm? id=P090029. Accessed August 7, 2018.
39. Arnold PM, Sasso RC, Janssen ME, et al. Efficacy of i-Factor Bone Graft versus autograft in anterior cervical discectomy and fusion: results of the prospective, randomized, single-blinded Food and Drug Administration Investigational Device Exemption Study. *Spine (Phila Pa 1976)*. 2016;41:1075-1083.
40. Murrey D, Janssen M, Delamarter R, et al. Results of the prospective, randomized, controlled multicenter Food and Drug Administration Investigational Device Exemption Study of the ProDisc-C total disc replacement versus anterior discectomy and fusion for the treatment of 1-level symptomatic cervical disc disease. *Spine J*. 2009;9:275-286.
41. Phillips FM, Lee JY, Geisler FH, et al. A prospective, randomized, controlled clinical investigation comparing PCM cervical disc arthroplasty with anterior cervical discectomy and fusion. 2-year results from the US FDA IDE clinical trial. *Spine (Phila Pa 1976)*. 2013;38:E907-E918.
42. Coric D, Nunley PD, Guyer RD, et al. Prospective, randomized, multicenter study of cervical arthroplasty: 269 patients from the Kineflex|C artificial disc investigational device exemption study with a minimum 2-year follow-up: clinical article. *J Neurosurg Spine*. 2011;15:348-358.
43. Zigler J, Delamarter R, Spivak JM, et al. Results of the prospective, randomized, multicenter Food and Drug Administration Investigational Device Exemption Study of the ProDisc-L total disc replacement versus circumferential fusion for the treatment of 1-level degenerative disc disease. *Spine (Phila Pa 1976)*. 2007;32:1155-1163.

44. US Food and Drug Administration. Summary of safety and effectiveness. X-STOPÒ premarket approval (P040001). https://www.accessdata.fda.gov/cdrh_docs/pdf4/P040001B.pdf. Accessed September 4, 2018.
45. Klassen PD, Hes R, Bouma GJ, et al. A multicenter, prospective, randomized study protocol to demonstrate the superiority of a bone-anchored prosthesis for anular closure used in conjunction with limited discectomy to limited discectomy alone for primary lumbar disc herniation. *Int J Clin Trials*. 2016;3:120-131.
46. Davis RJ, Errico TJ, Bae H, Auerbach JD. Decompression and Coflex interlaminar stabilization compared with decompression and instrumented spinal fusion for spinal stenosis and low-grade degenerative spondylolisthesis: two-year results from the prospective, randomized, multicenter, Food and Drug Administration Investigational Device Exemption trial. *Spine (Phila Pa 1976)*. 2013;38:1529-1539.
47. Meyer B, Baranto A, Schils F, et al. Percutaneous interspinous spacer vs decompression in patients with neurogenic claudication: an alternative in selected patients? *Neurosurgery*. 2018;82:621-629.
48. Stromqvist BH, Berg S, Gerdhem P, et al. X-stop versus decompressive surgery for lumbar neurogenic intermittent claudication: randomized controlled trial with 2-year follow-up. *Spine (Phila Pa 1976)*. 2013;38:1436-1442.
49. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ; CONSORT Group. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA*. 2006;295:1152-1160.
50. Piaggio G, Elbourne DR, Pocock SJ, Evans SJ, Altman DG; CONSORT Group. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. *JAMA*. 2012;308:2594-2604.
51. Hisey MS, Zigler JE, Jackson R, et al. Prospective, randomized comparison of one-level Mobi-C cervical total disc replacement vs. anterior cervical discectomy and fusion: results at 5-year follow-up. *Int J Spine Surg*. 2016;10:10.
52. Blumenthal S, McAfee PC, Guyer RD, et al. A prospective, randomized, multicenter Food and Drug Administration Investigational Device Exemptions Study of lumbar total disc replacement with the CHARITE artificial disc versus lumbar fusion: part I: evaluation of clinical outcomes. *Spine (Phila Pa 1976)*. 2005;30:1565-1591.
53. Delamarter R, Zigler JE, Balderston RA, Cammisa FP, Goldstein JA, Spivak JM. Prospective, randomized, multicenter Food and Drug Administration Investigational Device Exemption Study of the ProDisc-L total disc replacement compared with circumferential arthrodesis for the treatment of two-level lumbar degenerative disc disease: results at twenty-four months. *J Bone Joint Surg Am*. 2011;93:705-715.
54. Vaccaro AR, Lawrence JP, Patel T, et al. The safety and efficacy of OP-1 (rhBMP-7) as a replacement for iliac crest autograft in posterolateral lumbar arthrodesis: a long-term (>4 years) pivotal study. *Spine (Phila Pa 1976)*. 2008;33:2850-2862.
55. Delawi D, Jacobs W, van Susante JL, et al. OP-1 compared with iliac crest autograft in instrumented posterolateral fusion: a randomized, multicenter non-inferiority trial. *J Bone Joint Surg Am*. 2016;98:441-448.
56. Cho JH, Lee JH, Yeom JS, et al. Efficacy of *Escherichia coli*-derived recombinant human bone morphogenetic protein-2 in posterolateral lumbar fusion: an open, active-controlled, randomized, multicenter trial. *Spine J*. 2017;17:1866-1874.
57. Yi J, Lee GW, Nam WD, et al. A prospective randomized clinical trial comparing bone union rate following anterior cervical discectomy and fusion using a polyetheretherketone cage: hydroxyapatite/B-tricalcium phosphate mixture versus hydroxyapatite/demineralized bone matrix mixture. *Asian Spine J*. 2015;9:30-38.
58. Arnold PM, Anderson KK, Selim A, Dryer RF, Burkus JK. Heterotopic ossification following single-level anterior cervical discectomy and fusion: results from the prospective, multicenter, historically controlled trial comparing allograft to an optimized dose of rhBMP-2. *J Neurosurg Spine*. 2016;25:292-302.
59. Burkus JK, Gornet MF, Dickman CA, Zdeblick TA. Anterior lumbar interbody fusion using rhBMP-2 with tapered interbody cages. *J Spinal Disord Tech*. 2002;15:337-349.
60. Burkus JK, Heim SE, Gornet MF, Zdeblick TA. Is INFUSE bone graft superior to autograft bone? An integrated analysis of clinical trials using the LT-CAGE lumbar tapered fusion device. *J Spinal Disord Tech*. 2003;16:113-122.
61. US Food and Drug Administration. Executive summary for P050036 Medtronic's AMPLIFY™ rhBMP-2 matrix. https://wayback.archive-it.org/7993/20170114052355/http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/MedicalDevices/MedicalDevicesAdvisoryCommittee/OrthopaedicandRehabilitationDevicesPanel/UCM220079.pdf. Published July 27, 2010. Accessed September 4, 2018.
62. Rhyne AL, Blumenthal SL, Frank EH, et al; Oxiplex Clinical Study Group. Oxiplex reduces leg pain, back pain, and associated symptoms after lumbar discectomy. *Spine (Phila Pa 1976)*. 2012;37:631-641.
63. Bae H, Hatten HP Jr, Linovitz R, et al. A prospective randomized FDA-IDE trial comparing Cortoss with PMMA for vertebroplasty: a comparative effectiveness research study with 24-month follow-up. *Spine (Phila Pa 1976)*. 2012;37:544-550.
64. Tutton SM, Pflugmacher R, Davidian M, Beall DP, Facchini FR, Garfin SR. KAST study: the Kiva system as a vertebral augmentation treatment-a safety and effectiveness trial: a randomized, noninferiority trial comparing the Kiva system with balloon kyphoplasty in treatment of osteoporotic vertebral compression fractures. *Spine (Phila Pa 1976)*. 2015;40:865-875.
65. Hacker RJ, Cauthen JC, Gilbert TJ, Griffith SL. A prospective randomized multicenter clinical evaluation of an anterior cervical fusion cage. *Spine (Phila Pa 1976)*. 2000;25:2646-2655.
66. US Food and Drug Administration. Summary of safety and effectiveness (P980048). https://www.accessdata.fda.gov/cdrh_docs/pdf/P980048B.pdf. Accessed September 5, 2018.
67. US Food and Drug Administration. FDA executive summary for Zimmer Spine's Dynesys Spinal System (P070031). https://wayback.archive-it.org/7993/20170405192046/https://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/MedicalDevices/MedicalDevicesAdvisoryCommittee/OrthopaedicandRehabilitationDevicesPanel/UCM188734.pdf. Published November 4, 2009. Accessed September 4, 2018.