

Philosophies of Surgical Care Are Embedded in Outcome Studies: An Illustrative Reanalysis of the Cartiva MOTION Trial

Gregory P. Guyton, MD¹ 

Foot & Ankle International®
1–6

© The Author(s) 2022

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/10711007221112928

journals.sagepub.com/home/fai

Abstract

Background: Subjective assumptions on the definition of surgical success are inherent to the design of clinical trials with a categorical outcome. The current study used reasonable alternative assumptions about surgical care to reassess data for the randomized controlled Cartiva trial (MOTION).

Methods: Data from the published study were augmented by publicly accessible internal US Food and Drug Administration documents. As in the published report, 1-sided lower bound 95% CIs (LBCI95) for the difference of proportions were calculated for a series of alternative scenarios in which the assumptions underlying what constitutes surgical success were altered.

Results: Using a noninferiority margin of –15%, the MOTION trial reported success based on a 1-sided LBCI95 of –10.9%. Each of the 3 independent alternative scenarios analyzed yielded results that altered the primary outcome of the trial: (1) eliminating failures based solely upon radiographs findings, thereby considering a painless pseudarthrosis as a success (1-sided LBCI95 of –15.9%), (2) considering only major surgical revision as a failure and discounting isolated hardware removal (1-sided LBCI95 of –15.1%), and (3) using a visual analog scale (VAS) pain threshold of <30 as the success criterion rather than a 30% reduction in VAS pain score (1-sided LBCI95 of –15.8%).

Conclusion: In this reanalysis, applying any of 3 reasonable alternative assumptions about the definition of surgical success to the data resulted in failure to prove noninferiority of Cartiva over arthrodesis, a reversal of the reported trial result. These results highlight the effect of subjective assumptions in the design of clinical trials with a categorical outcome and illustrate how differing philosophies about what constitutes surgical success can be pivotal in determining the final result.

Level of Evidence: Level II, prospective comparative study.

Keywords: noninferiority trial, statistical analysis, randomized controlled trial, RCT, foot

Introduction

All trials in medicine require that clinical success be defined. This definition is often subjective, particularly when assessing elective surgeries designed to improve quality of life. For a variety of legitimate reasons, investigators may reduce a large volume of categorical and continuous data into a single binary outcome of “success” or “failure” for each subject in a trial. The choice of variables, margins, and criteria used in these studies inherently reflects the designers’ philosophy about the meaning of surgical success.

The MOTION study is a well-known multicenter, open label, randomized controlled noninferiority trial that found a novel motion-preserving cylindrical polyvinyl first metatarsophalangeal (MTP) implant (Cartiva) noninferior to

arthrodesis for treatment of first metatarsophalangeal joint degeneration.⁴ The company-sponsored trial was used for US Food and Drug Administration (FDA) approval of the device. The primary endpoint of the MOTION trial evaluated the categorical success or failure of study subjects using a binary definition of surgical success that required satisfaction of all criteria in a checklist (Table 1). Although the

¹Department of Orthopaedic Surgery, MedStar Union Memorial Hospital, Baltimore, MD, USA

Corresponding Author:

Gregory P. Guyton, MD, c/o Lyn Jones, MA, ELS, Department of Orthopaedic Surgery, MedStar Union Memorial Hospital, 3333 North Calvert Street, Suite 400, Baltimore, MD 21218, USA.

Email: lyn.m.jones@medstar.net

Table 1. Criteria for Surgical Success in the MOTION Trial.¹⁷

Criterion No. ^a	Criterion
1	Improvement on the visual analog scale (VAS) pain score of 30% or greater.
2	Absence of worsening beyond the minimum clinically important difference (MCID) of the sports subscale of the Foot and Ankle Ability Measure (FAAM). This was later modified to the activities of daily living subscale of the FAAM at the request of the FDA.
3	Absence of any predetermined safety events.
4	Absence of predetermined radiographic failure criteria including nonunion in the case of arthrodesis and component migration in the case of the Cartiva implant.

^aSuccess required satisfaction of all criteria.

MOTION trial demonstrated noninferiority for the prosthesis vs fusion, the results of subsequent independent reports following FDA approval have been mixed.^{1,5,12} There have been subsequent positive reports as well, but these represented long-term follow-up of subsets of the original trial patients in whom the prosthesis had not been explanted.^{6,7}

These disparities may arise in part from reasonable but differing underlying philosophies of surgical outcomes. This reanalysis evaluates the outcome of the MOTION trial using reasonable but differing alternative criteria for success to determine what effect several subjective choices about surgical success embedded in the trial design might have had on the statistical outcome.

The MOTION study was performed using a noninferiority design. Noninferiority studies were originally developed for use in pharmaceutical trials when an existing medication was already established that clearly provided a patient benefit and it would be unethical to perform a placebo-controlled trial. In these trials, a new medication that is thought to provide the original benefit but also have some ancillary advantage (such as a reduced side-effect profile, better dosing regimen, or reduced cost) is tested against the original medication to ensure that it is noninferior with regard to the main effect.

Noninferiority studies use the same statistical metrics as standard comparative trials: statistical significance (α) and power (β). In addition, they require that a third metric be determined before the trial begins: the noninferiority margin, M . This margin is the degree to which the new treatment could be worse than the original on the main treatment effect and still be considered noninferior. It is typically set to equal the minimum clinically important difference (MCID) of the primary outcome of the study.

Many noninferiority trials involving new surgical implants use a categorical design. Each participant is recorded as being a success or failure based on his or her outcome in each of a series of categories that may include the absence of complications, pain outcomes, improvement on patient-reported outcomes, and radiographic findings. In a categorical noninferiority trial, success/failure proportions are calculated for the original treatment and the new treatment, and a 95% CI for the difference of those proportions is calculated. Two procedures that have exactly the same outcome will have a

difference of success/failure proportions equal to 0, but the mean is not the value of interest in a noninferiority trial. The degree to which the outcome of the new procedure could be worse than the outcome of the new one (with 95% confidence) determines whether noninferiority has been established. The lower boundary of the 95% CI is compared to the predetermined noninferiority margin to make that determination.

There is an important caveat to this approach. Noninferiority trials that use a composite success/failure checklist have essentially created a new ad hoc rating scale for which an MCID has not been established. Without an established MCID, the choice of the noninferiority margin is arbitrary, and no established consensus margin exists for orthopaedic trials of this type. The MOTION trial utilized a -15% noninferiority margin with a 1-sided 95% CI for the difference of success/failure proportions, a combination that is less statistically rigorous than the large majority of noninferiority trials in the orthopaedic literature.⁸ It is not the purpose of this reanalysis to judge whether or not those standards were justified and, as in the original MOTION trial, they are used for all calculations.

Methods

This reanalysis used the data from the peer-reviewed publication of the MOTION trial supplemented by more detailed trial results documented for the FDA approval process.¹⁶⁻¹⁸

Individual cases were reassigned to success or failure based on each alternative criterion evaluated. The lower-bound 95% CI (LBCI95) for the difference of the success/failure proportions was calculated. One-sided 95% CIs were calculated as used in the original trial.⁸

Three alternative criteria for surgical success were analyzed independently:

1. Consider as success any patient with clinical success regardless of the radiographic interpretation.
2. Consider only major surgical revision as a failure and discount isolated hardware removal or scar release.
3. Use a visual analog scale (VAS) pain threshold of <30 as the success criterion instead of a 30% improvement from baseline.

Results

The results of these reanalyses are summarized in Figure 1.

Use of Radiographs

The MOTION trial paper did not directly indicate that radiographic outcomes were used as a separate criterion to determine success or failure.⁴ The FDA documentation, however, clearly indicates that radiographic interpretation was used in this way.¹⁶⁻¹⁸ The analysis disproportionately affected the arthrodesis group by categorizing as failures any patients with radiographic anomalies but no significant pain. Three patients in the arthrodesis group were considered failures based solely on radiographic criteria: 1 with broken hardware and 2 with radiographic pseudarthroses.¹⁶ All would have been considered successes based on pain status, functional status, and absence of secondary surgical interventions. If the 3 painless pseudarthrosis patients in the

arthrodesis arm were reclassified as successes, the primary study endpoint changed to a 1-sided LBCI95 of -15.9%, outside the range chosen to establish noninferiority of Cartiva.

Stratifying Revision Surgery

The MOTION trial treated all secondary surgical events as categorical failures, an approach that inherently equates isolated hardware removal with revision fusion. A reasonable alternative assumption would be to assume that only major secondary procedures have clinical significance. For this reanalysis, the secondary surgical procedures were divided into major procedures that involved bony fusion, osteotomy, or repositioning of the prosthesis and minor procedures that involved hardware removal or scar release only.

The Cartiva group contained 12 major events: 10 cases of explantation and conversion to fusion, 1 Moberg osteotomy, and 1 revision of the prosthesis with repositioning.

			Lower Bound of 1-sided 95% Confidence Interval for Difference of Proportions	Outcome Using a -15% Margin
As reported		Cartiva		
	Success	104	40	
	Failure	26	10	
			-10.9%	Demonstrated noninferiority
Excluding failures based solely on radiographs – clinical outcome only		Cartiva		
	Success	104	43	
	Failure	26	7	
			-15.9%	Fails to demonstrate noninferiority
Counting only major second surgeries as failure – hardware removal and scar release excluded		Cartiva		
	Success	105	43	
	Failure	25	7	
			-15.1%	Fails to demonstrate noninferiority
Using VAS pain threshold of ≤ 30 – not a relative improvement of 30%		Cartiva		
	Success	98	40	
	Failure	32	10	
			-15.8%	Fails to demonstrate noninferiority

Figure 1. The statistical results of the MOTION trial as originally reported and as modified using the 3 alternative conditions explored in this reanalysis.

It contained 1 minor event: a manipulation and scar release. The arthrodesis group contained 3 major events: 3 cases of revision arthrodesis. It contained 3 minor events: 3 patients who underwent hardware removal.¹⁶⁻¹⁸ One of the revision arthrodesis patients underwent an initial hardware removal followed by a later revision arthrodesis.⁴ This patient was therefore counted once as one of the 3 major events.

Among those patients who had secondary surgeries, the average 24-month VAS pain scores were 4 for arthrodesis and 12 for Cartiva.¹⁶ Residual pain and dysfunction were not disqualifying factors for success among the patients who underwent minor secondary procedures.

Considering only major surgical procedures as failures, 1 Cartiva patient and 3 arthrodesis patients were reclassified as successes. This changed the primary study endpoint to a 1-sided LBCI95 of -15.1%, outside the range chosen to establish noninferiority of Cartiva.

Using a Threshold Value for VAS Pain

The MOTION trial used a 30% improvement of VAS pain scores as a criterion for success. The value of 30% change as the MCID for a VAS pain scale is among many possibilities that have been supported in the literature,^{9,10} and the article indicates that this criterion was chosen on that basis.

The MCID of a psychometric scale is the interval change that determines whether any clinically meaningful effect has occurred. However, recent studies have indicated that patient satisfaction across a wide range of chronic diseases is less closely correlated with whether a given subject can detect any level of improvement than with whether they achieve a desired absolute threshold of symptomatic control.^{11,13,15} Subjectively, this principle might be best stated as "It is more important to feel good than to feel better."

Technically, this paradigm of a Patient Acceptable Symptomatic State (PASS) score for a psychometric scale should be used only after a threshold has been determined through separate validation using global anchor questions. It may be reasonable, however, to consider the inherently intuitive VAS pain scale to be an exception to this requirement. When formal PASS scores for VAS pain have been set in orthopaedic conditions, they have been found to range from 21.1 to 36.4, corresponding to a more familiar "I would like my pain to be less than 3 out of 10."^{9-11,15}

Precedent exists in the foot and ankle literature for using the PASS concept for VAS pain without a formal threshold determination. Baumhauer et al³ reported the percentage of patients with residual VAS pain above 20 of 100 following autograft harvest. The cited reference for that value was one of many papers that have calculated an absolute value for the MCID for VAS pain, not a formally determined PASS threshold.¹⁴

If an absolute threshold requiring a final VAS pain score of less than 30 is used as the criterion for success in lieu of

a 30% improvement relative to the preoperative value, a total of 6 patients in the Cartiva group and 0 patients in the arthrodesis group change from successes to failures. This implies those patients now categorized as failures had residual VAS pain scores above 30 despite some improvement, a fact also noted by the FDA statisticians.¹⁶⁻¹⁸ The primary outcome of the 1-sided LBCI95 changes to -15.8%, outside the range chosen to establish noninferiority of Cartiva.

Discussion

Noninferiority trials using categorical success/failure checklists have become particularly popular in FDA trials for new surgical implants. Regulatory authorities must look at the totality of the data and reach a binary decision to either approve or deny the application. More nuanced approaches to how those implants are used are ultimately left to individual surgeons and patients. An FDA panel, like a granting agency, is charged with generating a yes or no recommendation. This goal is well served by the reductionist methodology of a categorical noninferiority trial.

The philosophical choices made by the designers of a noninferiority trial are embedded in the trial design and may not be immediately apparent. The MOTION trial provides a particularly interesting case study primarily because the statistical outcome was quite close, and subtle changes in the criteria used for success would have had profound effects on the final result. How patients and surgeons view those criteria varies both between individuals and in the same individual over the course of their life. The reanalysis presented here confirmed that any of 3 independent scenarios would have changed the outcome of the trial if translated into statistics:

1. A surgeon and patient may not consider an unplanned painless pseudoarthrosis to be a bad outcome for arthrodesis when compared to an implant procedure that itself is designed to create a painless pseudoarthrosis.
2. A surgeon and patient may not consider an isolated hardware removal with successful resolution of symptoms to be an important enough adverse surgical event to consider the outcome a failure. They may consider a revision fusion to be a much more morbid procedure that should be treated differently.
3. Patients may feel it is more important to reach an absolute level of pain control than to achieve a percentage reduction in pain that leaves them with significant residual symptoms.

Subjective choices are also involved in setting the noninferiority margin and determining whether a 1- or 2-sided 95% CI is used, a complex topic previously reviewed in connection with the MOTION trial.⁸ The -15% margin

with a 1-sided interval chosen for this trial was a lenient statistical standard in the overall context of noninferiority trials in the orthopaedic literature.⁸

The MOTION trial was created using a common template for pivotal FDA device trials: a noninferiority design, a checklist for success, and a categorical outcome variable. There is a strong reason for this choice of design. Pivotal device trials need to include patient safety events, which are inherently categorical. This reanalysis is designed to highlight subjective choices made within the framework of a categorical checklist noninferiority trial, but it is also instructive to ask what outcomes and ambiguities might have arisen had a simple comparative design been used.

Figure 2 is a graphical representation of the VAS pain scores for arthrodesis and Cartiva reported over time in the MOTION trial. If a simple comparative study for pain outcomes had been performed, 3 conflicting arguments for interpreting this graph could be made, all of which would be correct:

1. In favor of arthrodesis: At all late time points, arthrodesis resulted in a better mean result for VAS pain than Cartiva and that difference was statistically significant.
2. In favor of Cartiva: Although statistically significant, the differences between mean VAS pain for arthrodesis and Cartiva were below most of the commonly accepted values for the MCID of the VAS pain scale. The difference in mean pain outcomes between the procedures was therefore clinically insignificant, although Cartiva offers the advantage of some retained toe motion.²
3. In favor of arthrodesis: Despite the relatively similar means, the SD for VAS pain in the Cartiva group was much larger at all time points and extended above the value of 30 mm, whereas that for arthrodesis did not after 6 weeks. Only the Cartiva group had a substantial number of outliers with unacceptable levels of pain if a VAS pain score of 30 mm is used as the threshold for “unacceptable.”

These variable interpretations illustrate that subjectivity is not confined to the noninferiority design. Even with a simple comparative study, one can emphasize the presence of statistical significance, the failure of the difference of means to exceed the MCID, or the presence of outliers above a threshold value in only 1 treatment arm. All are defensible and correct.

This reanalysis is purely illustrative. It is not intended to make a definitive statement about the use of the Cartiva implant. The alternative scenarios presented here were specifically chosen both because they had some potential to affect the statistical outcome and because they may be reasonable to some physicians and patients. Just as the choices

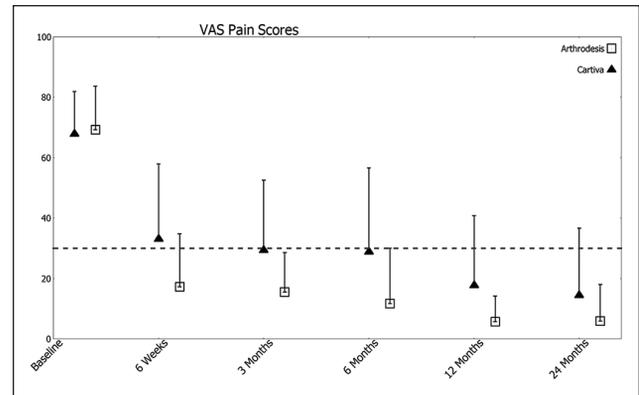


Figure 2. This graph demonstrates VAS pain scores over time in the MOTION study demonstrating positive standard errors and a 30-mm threshold value. The graph represents the data in table 11 from the original article.⁴ In a subsequent communication, the investigators noted that *t* test analysis of the aggregate outcomes for Cartiva and arthrodesis showed a statistically significant advantage for arthrodesis that was considered clinically unimportant because it was within the MCID for VAS pain.² However, 6 Cartiva patients and 0 arthrodesis patients with improvement of VAS pain of 30% or more still had residual VAS pain scores above 30 mm at 24 months.¹⁸

made by the trial investigators may not represent universal philosophies of surgical success, the choices made in this analysis are also wholly subjective and are not universally accepted themselves.

Importantly, only 3 potential alternative scenarios were analyzed. This article is not intended to be an exhaustive examination of every possible way to look at the data that could be considered reasonable. Although the 3 examined here would have reversed the trial result, other alternative scenarios can easily be envisioned that would have further favored the implant.

Conclusion

In this reanalysis, applying any 1 of 3 reasonable alternative assumptions to the MOTION trial data resulted in failure to prove noninferiority of Cartiva compared with arthrodesis, a reversal of the reported trial result. These results highlight the subjective nature inherent in design of clinical trials with a categorical outcome and illustrate that underlying assumptions about how surgical success is measured can be statistically pivotal. Reductionist methodologies such as noninferiority studies have a role and are particularly valuable in a regulatory setting where yes or no decisions must be made. Carefully constructed randomized controlled trials are critical to advancing the field and will always require that subjective choices be made. The MOTION trial remains a seminal work, and the skill with which it was constructed and implemented by its authors is a model for future

investigators. Nevertheless, physician philosophies of care and patient desires are not universal, and they may not match those reflected in even the most carefully designed investigations. This reanalysis and discussion is meant to encourage the clinician to carefully review the separate component parts of complex trials in the context of their own individual philosophies of surgical care.

Acknowledgments

I thank Lyn Jones, MA, ELS, of our department for editorial support.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. ICMJE forms for all authors are available online.

Funding

The author received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Gregory P. Guyton, MD,  <https://orcid.org/0000-0002-1238-3673>

References

1. An TW, Cassinelli S, Charlton TP, Pfeffer GB, Thordarson DB. Radiographic and magnetic resonance imaging of the symptomatic synthetic cartilage implant. *Foot Ankle Int.* 2020;41(1):25-30. doi:10.1177/1071100719877147
2. Baumhauer J. Response to "Letter regarding: prospective, randomized, multi-centered clinical trial assessing safety and efficacy of a synthetic cartilage implant versus first metatarsophalangeal arthrodesis in advanced hallux rigidus." *Foot Ankle Int.* 2017;38(1):108-110. doi:10.1177/1071100716681632
3. Baumhauer JF, Glazebrook M, Younger A, et al. Long-term autograft harvest site pain after ankle and hindfoot arthrodesis. *Foot Ankle Int.* 2020;41(8):911-915. doi:10.1177/1071100720920846
4. Baumhauer JF, Singh D, Glazebrook M, et al. Prospective, randomized, multi-centered clinical trial assessing safety and efficacy of a synthetic cartilage implant versus first metatarsophalangeal arthrodesis in advanced hallux rigidus. *Foot Ankle Int.* 2016;37(5):457-469. doi:10.1177/1071100716635560
5. Cassinelli SJ, Chen S, Charlton TP, Thordarson DB. Early outcomes and complications of synthetic cartilage implant for treatment of hallux rigidus in the United States. *Foot Ankle Int.* 2019;40(10):1140-1148. doi:10.1177/1071100719855049
6. Daniels TR, Younger AS, Penner MJ, et al. Midterm outcomes of polyvinyl alcohol hydrogel hemiarthroplasty of the first metatarsophalangeal joint in advanced hallux rigidus. *Foot Ankle Int.* 2017;38(3):243-247. doi:10.1177/1071100716679979
7. Glazebrook M, Blundell CM, O'Dowd D, et al. Midterm outcomes of a synthetic cartilage implant for the first metatarsophalangeal joint in advanced hallux rigidus. *Foot Ankle Int.* 2019;40(4):374-383. doi:10.1177/1071100718815469
8. Guyton GP. Standards for noninferiority trials in orthopaedic surgery would be arbitrary: it's time we had some. *J Bone Joint Surg Am.* 2021;103(17):e69. doi:10.2106/JBJS.20.01949
9. Olsen MF, Bjerre E, Hansen MD, et al. Pain relief that matters to patients: systematic review of empirical studies assessing the minimum clinically important difference in acute pain. *BMC Med.* 2017;15(1):35. doi:10.1186/s12916-016-0775-3
10. Olsen MF, Bjerre E, Hansen MD, Tendal B, Hilden J, Hrobjartsson A. Minimum clinically important differences in chronic pain vary considerably by baseline pain and methodological factors: systematic review of empirical studies. *J Clin Epidemiol.* 2018;101:87-106.e2. doi:10.1016/j.jclinepi.2018.05.007
11. Paulsen A, Roos EM, Pedersen AB, Overgaard S. Minimal clinically important improvement (MCII) and patient-acceptable symptom state (PASS) in total hip arthroplasty (THA) patients 1 year postoperatively. *Acta Orthop.* 2014;85(1):39-48. doi:10.3109/17453674.2013.867782
12. Shimozone Y, Hurley ET, Kennedy JG. Early failures of polyvinyl alcohol hydrogel implant for the treatment of hallux rigidus. *Foot Ankle Int.* 2021;42(3):340-346. doi:10.1177/1071100720962482
13. Tashjian RZ, Deloach J, Porucznik CA, Powell AP. Minimal clinically important differences (MCID) and patient acceptable symptomatic state (PASS) for visual analog scales (VAS) measuring pain in patients treated for rotator cuff disease. *J Shoulder Elbow Surg.* 2009;18(6):927-932. doi:10.1016/j.jse.2009.03.021
14. Todd KH, Funk JP. The minimum clinically important difference in physician-assigned visual analog pain scores. *Acad Emerg Med.* 1996;3(2):142-146. doi:10.1111/j.1553-2712.1996.tb03402.x
15. Tubach F, Ravaud P, Martin-Mola E, et al. Minimum clinically important improvement and patient acceptable symptom state in pain and function in rheumatoid arthritis, ankylosing spondylitis, chronic back pain, hand osteoarthritis, and hip and knee osteoarthritis: results from a prospective multinational study. *Arthritis Care Res (Hoboken).* 2012;64(11):1699-1707. doi:10.1002/acr.21747
16. US Food and Drug Administration. Cartiva synthetic implant (P150017): FDA introduction. Published 2016. Accessed September 22, 2020. <https://www.fda.gov/media/97335/download>
17. US Food and Drug Administration. FDA executive summary: Cartiva Synthetic Cartilage Implant. Published 2016. Accessed February 7, 2020. <https://www.fda.gov/media/96916/download>
18. US Food and Drug Administration. Premarket approval (PMA). Summary of safety and effectiveness data. Published 2016. Accessed January 10, 2020. https://www.accessdata.fda.gov/cdrh_docs/pdf15/P150017B.pdf