Survey of Composite Endpoints in Therapeutic Premarket Approval (PMA) submissions

Natasha Sahr[1], Ja-An Lin[2], Rajesh Nair[2]

[1]Biostatistics Department, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105

[2]Center for Devices and Radiological Health, US Food and Drug Administration, 10903 New Hampshire Ave, Silver Spring, MD 20993

Abstract

In many clinical studies for therapeutic medical devices, a single clinical outcome may not capture the main benefits and risks of intervention. A composite clinical endpoint combines the most relevant clinical outcomes for assessing the effectiveness and safety of a device and combines them into a single endpoint. Composite endpoints can lead to smaller and shorter clinical trials which has resulted in an ever–increasing number of device trials utilizing a composite primary endpoint. However, composite endpoints often lead to difficulties in the interpretation of results. Here we report the results of a systematic review of clinical studies from Premarket Approval (PMA) submissions for therapeutic medical devices approved between 2007 and 2016 that had a primary composite endpoint.

Key Words: Types of endpoints, composite endpoints, survey, medical devices

1. Introduction

A composite endpoint is a clinically relevant endpoint that is constructed from combinations of two or more other clinically relevant outcomes, termed component endpoints. Composite endpoints are typically used in therapeutic areas where a single clinical endpoint does not adequately capture the benefits and risks of treatment. There are many ways in which component endpoints can be combined to form the composite endpoints. In addition, the composite endpoint and its components are most commonly binary, continuous, or time--to--event outcomes.

For studies where component outcomes are equally important, a composite endpoint provides a single summary measure of treatment effect. If multiple significance tests were to be conducted for each of the component outcomes, the rate of false positive findings could be greatly inflated without the use of an adjustment for multiple testing. Additional advantages supporting the use of a composite outcome are that it increases statistical efficiency because of higher event rates which reduces sample size requirements, costs, and time; helps investigators avoid an arbitrary choice between several important outcomes that refer to the same disease process to estimate the clinical benefit; and, assess multiple clinical endpoints that may address both safety and effectiveness simultaneously (Cordoba et a, BMJ, 2010).

Unfortunately, making treatment decisions based on composite outcomes can be challenging. The interpretation can be difficult when the endpoints are not of equal importance or the treatment effect is in opposite directions for some of the components, especially for a comparative multiple arm study. For example, suppose a device leads to a substantial reduction on a composite outcome of ``death, stroke, or myocardial infarction.'' This finding could mean that the device resulted in fewer deaths, stokes, and myocardial infarctions. But it is also possible that the composite was driven entirely by a reduction in myocardial infarction with no change, or even an increase, in death and stroke rates. In many clinical studies the effects often vary, and in some cases, the effect is biggest for the less important component outcomes and smallest or opposite for the most important components.

The goal of our survey was to summarize usage and reporting of composite endpoints in pivotal medical device clinical trials submitted to Center for Devices and Radiological Health (CDRH) to support Premarket Approval (PMA) for therapeutic devices.

2. Survey Methodology

2.1 Methods

Clinical studies are usually required for PMA of therapeutic medical devices. For approved devices, the results of clinical studies are summarized in the FDA Summary of Safety and Effectiveness Data (SSED) which is publicly available. We performed a systematic review of clinical studies from PMA submissions approved between 2007 and 2016 that had a primary composite endpoint. We excluded studies where the composite endpoint was a secondary outcome measure.

2.2 Search Strategy

We extracted Original and Panel Track PMA submissions from 2007–2016 from the Premarket Approval (PMA) database publicly available on the US Food and Drug Administration (FDA) website (https://www.accessdata.fda.gov). The landing page for the PMA database can be seen in Figure 1. Submissions were extracted by utilizing the Supplement Type and Decision Date fields from the PMA database. The Supplement Type field was limited to “Original” and “Panel Track”. The Decision Date was limited to “1/1/2007” to “12/31/2016”. The search resulted in 390 PMA submissions.

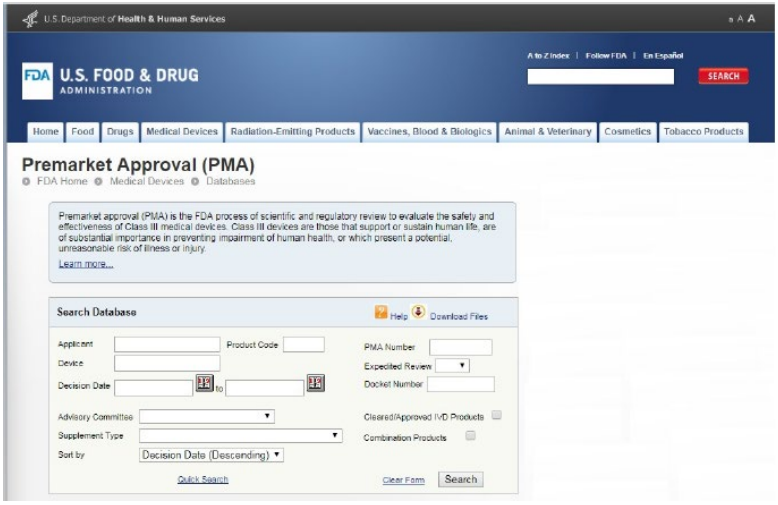


Figure 1: Landing page for the Premarket Approval (PMA) database publicly available on the US Food and Drug Administration website.

2.3 Study Selection and Data Extraction

For the 390 Original and Panel Track PMA submissions, the SSED was reviewed to identify pivotal studies with primary composite endpoint. We identified 85 PMA submissions containing composite primary outcomes in 90 clinical studies. A standard form was created to extract data on approval year, panel, randomization status, masking status, number of study arms, country where study was conducted (US/Outside US), enrolled and analyzed sample size, statistical methodology (frequentist/Bayesian), analysis type (performance goal/superiority hypothesis/non–inferiority hypothesis/descriptive), number of components, component combination method (any/all/sum), type of composite endpoint (binary/continuous/time–to–event), composite result (achieve/fail to achieve/descriptive), number of failed components, component labeling claims (yes/no), pre–specified missing data plans (yes/no), number of missing observations, sample size calculations provided (yes/no), and components considered in sample size calculation (yes/no).

3. Survey Results

3.1 Use of Composite Endpoints by Filing Year

For PMA submissions filed between 2007–2011, 26 PMA's had at least one clinical study where the primary endpoint was a composite outcome. From 2012–2016, 59 PMA submissions had at least one clinical study where the primary endpoint was a composite outcome. The proportion of PMA submissions with composite primary endpoint varied from 15-30% per year (Figure 2). From 2012-2016, about 25% of the submissions utilized a composite endpoint as the primary endpoint.

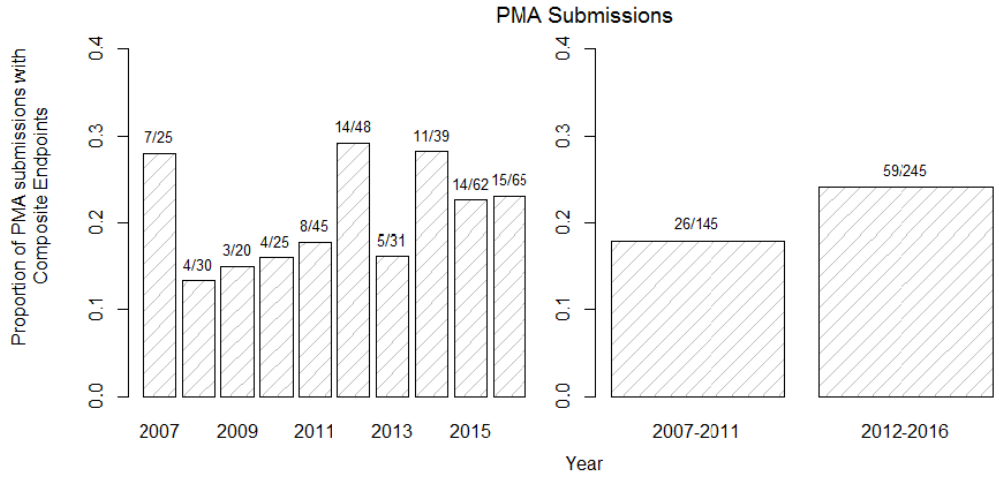
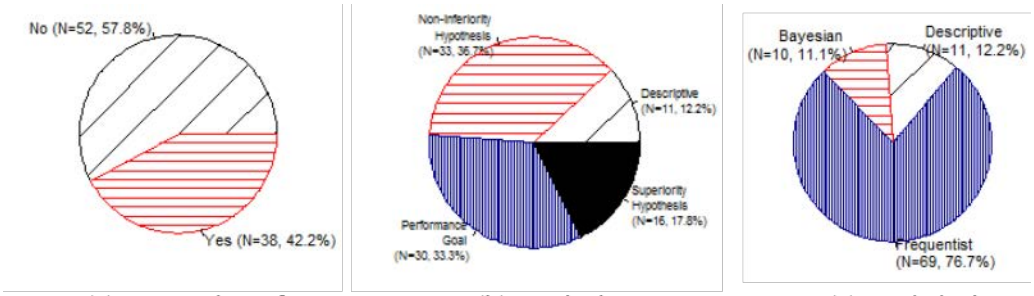


Figure 2: Usage of composite endpoints by year in PMA submissions for therapeutic devices.

3.2 Study Characteristics

Most device studies were conducted entirely in the US (72%). As can be seen from Figure 3(a), 42% of the studies were randomized trials. Non-inferiority hypothesis testing was utilized in 37% of the studies (Figure 3(b)). This was followed by single-arm studies utilizing a performance goal approach (33%). Most of the trials (77%) used a frequentist approach for inference (Figure 3(c)). Bayesian approaches were used in 11.1% of the trials.



(a) Proportion of randomized studies

(b) Analysis type

(c) Statistical methodology

Figure 3: Study characteristics of therapeutic PMA submissions utilizing composite endpoints.

Most clinical studies with composite endpoints had 3 components in the composite outcome (40%), followed by 4 components (24%) and 2 components (20%). The component endpoints for most studies were categorical (82%). Drastically fewer studies had time–to–event or continuous outcomes. The categorical endpoints in these studies were always binary. Two methods were commonly used to combine the component endpoints into a composite (Figure 4): (a) a patient is considered to be a composite success if success was achieved on any component, and (b) a patient is considered to be a composite success only if success was achieved on all components. Two clinical studies had continuous component outcomes that were combined via summation. An important factor to consider with the combination method is the studies that had time–to–event component endpoints

always used the combination method in which the composite outcome was recorded as an event if a patient experienced any component event. Fourteen (14) studies utilized a time-to-event composite endpoint. Eliminating the time–to–event and continuous outcomes from consideration resulted in 37 and 36 clinical studies that required success on any or all components, respectively.

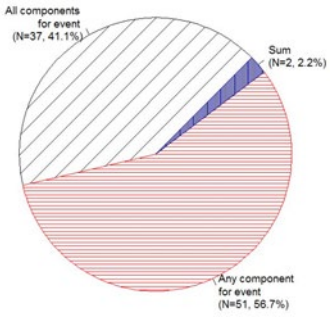


Figure 4: Summary of therapeutic PMA submissions by component endpoint combination method.

3.3 Reporting of Composite Endpoints

For studies where the composite was defined as success on any component, 80% of the studies achieved significance on the composite endpoint. This is surprisingly close to the 80% nominal power for most device studies. However, for studies where the composite endpoint is defined as success on all components, only 68% of the studies achieved significance. This could be due to the planned power for these studies not being achieved due to various reasons like the bigger impact of missing data (see discussion in section 3.3).

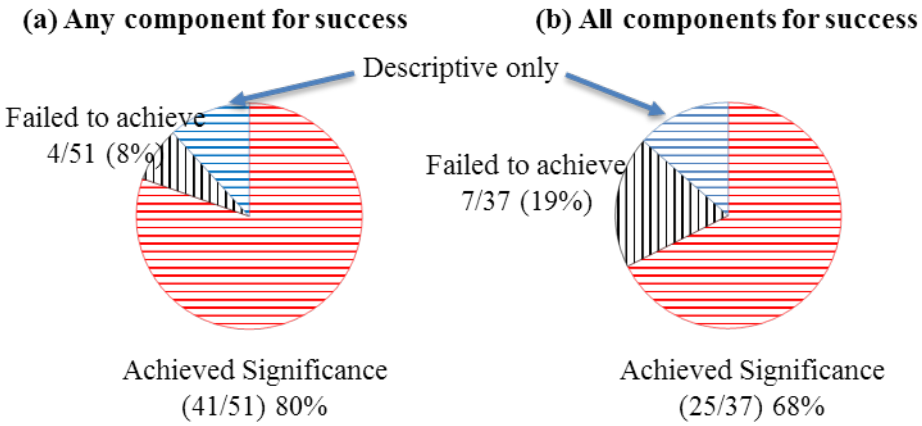


Figure 5: Summary of studies which achieved significance on the composite endpoint.

Since the component endpoints are typically not of equal importance for most studies, results for component endpoints are extremely important for interpreting study results. While most of the trials at least provided descriptive results for the component endpoints, 13% of the trials did not report any results for the for the component endpoints. Table 1 provides an example of a study for which detailed results for the component endpoints were

provided in the SSED. Bayesian methods were used to analyze the study results and the posterior mean and 95% Highest Posterior Density (HPD) credible intervals were reported for each component endpoint. Non-inferiority and superiority hypothesis testing were also pre-specified for the component endpoints.

Table 1: Example of Study with Detailed Reporting of Component Endpoints in the SSED.

Primary Outcome Variable	24-Month posterior mean (95% HPD credible interval)		24-Month posterior probabilities	
	Investigational	Control	Non-inferiority	Superiority
Neck Disability Index	85.0% (159) (79.7%, 89.9%)	76.2% (140) (69.7%, 82.6%)	~100%	98.0%
Neurological	92.4% (159) (88.4%, 96.1%)	90.9% (140) (86.4%, 95.3%)	~100%	69.2%
Free from SAE	97.6% (160) (95.5%, 99.4%)	95.2% (140) (92.1%,98.1%)	~100%	89.8%
Re-intervention	97.9% (160) (95.9%,99.5%)	96.1% (140) (93.2%,98.7%)	~100%	85.1%
Overall Success	80.1% (160) (74.3%, 85.8%)	71.8% (140) (65.0%, 78.9%)	~100%	96.9%

Table 2 provides an example of a study for which results for the component endpoints were not provided in the SSED. The primary composite endpoint consisted of three (3) components: 1) heel pain upon first step of the day, 2) heel pain while doing daily activities, and 3) heel pain after application of the Dolorimeter. Results were only provided for the composite endpoint.

Table 2: Example of Study where Descriptive Statistics was not provided for Component Endpoints.

	Investigational (n=125)	Control (n=118)	Effect size	P-value one sided
Composite VAS score: Percent change from Baseline at Visit 7				
Mean (SD) Median	-56.0 (39.31) -72.1	-44.1 (41.81) -44.7	0.5753	0.0220

To further examine the reporting of results for component endpoints, we focused on clinical studies that employed a superiority or non–inferiority hypothesis test for the primary composite endpoint and eliminated clinical studies that evaluated performance goals or solely provided descriptive statistics. This resulted in a pool of 49 clinical studies with primary composite endpoints from 46 PMA submissions. Of these 49 studies, 40 achieved

significance on the primary composite endpoint. Of the 14 superiority studies, 21% achieved statistical significance on all components (Figure 6(a)). No hypothesis testing was conducted for the component endpoints in 50% (Figure 6(a)) of the superiority studies. Of the 26 non-inferiority studies, 42% achieved statistical significance on all components (Figure 6(b)). No hypothesis testing was conducted for the component endpoints in 38% (Figure 6(b)) of the non-inferiority studies.

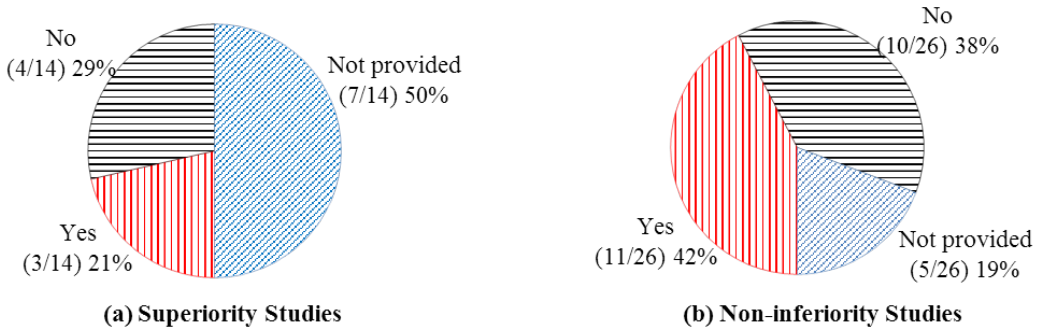


Figure 6: Summary of superiority/non-inferiority studies which achieved statistical significance on all component endpoints.

Some or all of the component endpoints were specified as secondary endpoints in 43 of the 90 studies with composite primary endpoint. Of these, 22 (51%) submissions had labeling claims for the component endpoints.

3.4 Missing Data in Composite Endpoint Studies

A detailed account of the survey results for missing data can be found in Figure 6. Some limitations regarding missing data reported in the SSED include inconsistent definitions of enrolled sample size in different studies. Moreover, many SSED reports do not explicitly specify the number of missing observations for the composite and component outcomes. No missing data was reported in 63% (32/51) studies where the composite endpoint is defined as success on any component (Figure 6). Only 20% (10/51) of these studies reported greater than 5% missing data. In contrast, studies where the composite endpoint is defined as success on all components had substantially more missing data. This is primarily because for this component endpoint combination method, missing data in any component results in missing value for the composite outcome. Whereas for studies where the composite is defined as success on any component, the composite outcome is missing only if there are no events (success) in any of the component endpoints and at least one component is missing outcome data. On account of this, no missing data was reported in only 19% (7/37) of studies where the composite is defined as success on all components (Figure 6). Almost half the studies (49%) with this component endpoint combination method reported greater than 5% missing data. The survey also indicated a pattern of increased missing data with increasing number of component endpoints.

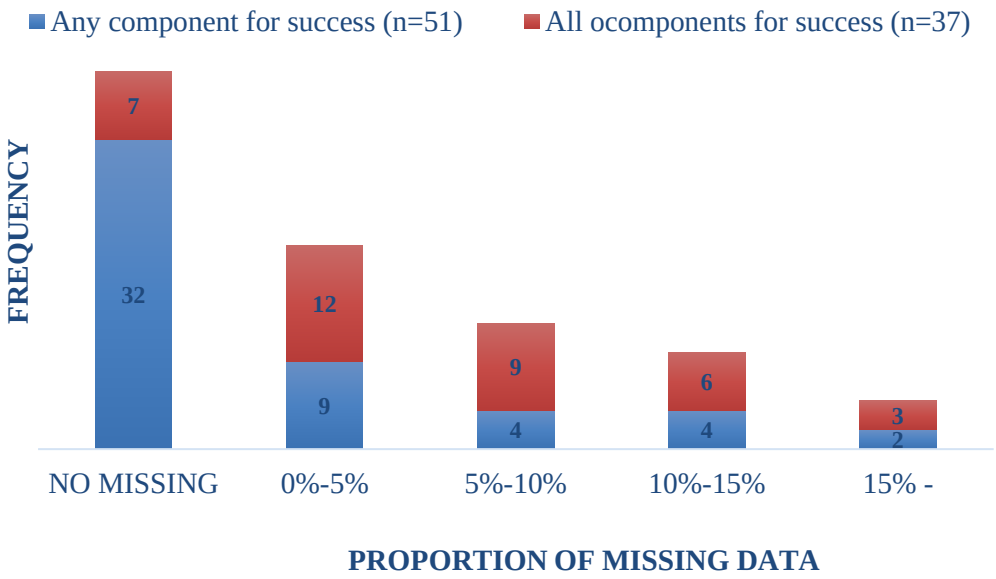
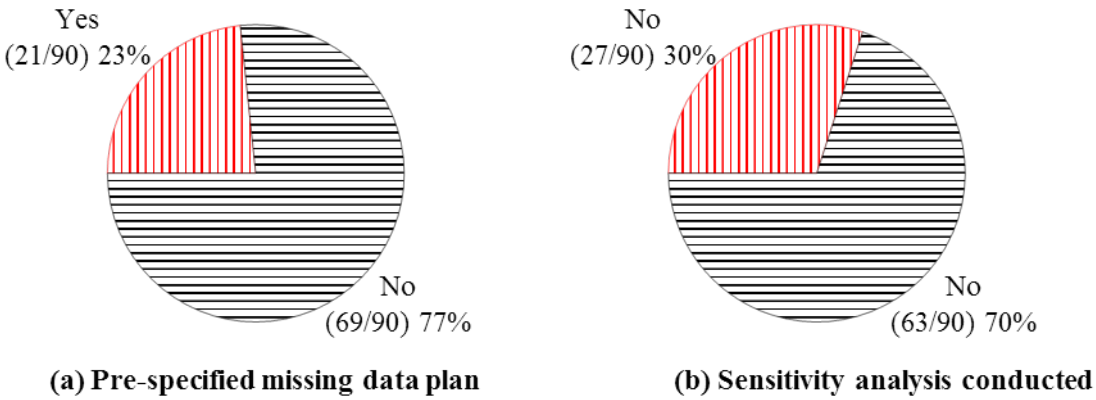


Figure 6: Summary of missing data in composite endpoint studies by component endpoint combination method.

It would be a reasonable assumption to require a pre–specified missing data plan and/or sensitivity analysis for PMA submissions containing composite primary endpoints. However, the majority of studies did not have a pre-specified missing data analysis plan (Figure 7(a)).



(a) Pre-specified missing data plan **(b) Sensitivity analysis conducted**

Figure 7: Summary of missing data analysis in composite endpoint studies.

4. Concluding Remarks

A survey of the approved PMA submissions from the US Food and Drug Administration Premarket Approval (PMA) database with a composite primary endpoint has resulted in many interesting findings. The survey identified 90 clinical trials using composite primary endpoint during a 10-year period from 2007-2016. Among the findings include some difficulty in interpretation of composite outcomes. Such difficulty in interpretation results from limitations surrounding component versus composite significance, missing data,

component labeling claims, sample size, and power. The component endpoints for most device studies were categorical. The proportion of studies which achieved significance on the primary composite endpoint depended on the method used to combine the component endpoints into a composite endpoint. Results for the component endpoints were reported at least descriptively for most studies. Almost 50% of the studies specified one or more component endpoints as secondary endpoints. Most studies did not have labeling claims for the component endpoints. The amount of missing data on the composite endpoint also depended on the method used to combine the component endpoints into a composite endpoint. Most studies did not have a pre-specified missing data plan or conduct sensitivity analyses.

References

G. Cordoba, L. Schwartz, S. Woloshin, H. Bae, and P. Gøtzsche, “Definition, reporting, and interpretation of composite endpoints in clinical trials: systemic review,” BMJ, vol. 341, pp. 1–7, 2010.