

Delphi-Consensus Weights for Ischemic and Bleeding Events to Be Included in a Composite Outcome for RCTs in Thrombosis Prevention

Agnes Dechartres^{1,2,3,4}, Pierre Albaladejo^{5,6}, Jean Mantz^{7,8}, Charles Marc Samama^{3,9}, Jean-Philippe Collet^{10,11}, Philippe Gabriel Steg^{8,12,13}, Philippe Ravaud^{1,2,3,4}, Florence Tubach^{1,2,8,14,15*}

1 UMR-S 738, INSERM, Paris, France, **2** UMR-S 738, Université Paris Diderot, Paris, France, **3** Université Paris Descartes, Paris, France, **4** Centre d'Epidémiologie Clinique, Hôpital Hôtel-Dieu, APHP, Paris, France, **5** Pôle d'Anesthésie et de Réanimation, Centre Hospitalier Universitaire, Grenoble, France, **6** Université Joseph Fourier, Grenoble, France, **7** Service d'anesthésie réanimation et SMUR, Hôpital Beaujon, APHP, Clichy, France, **8** Université Paris Diderot, Paris, France, **9** Service d'anesthésie réanimation, Hôpital Hôtel-Dieu, APHP, Paris, France, **10** Service de cardiologie, Hôpital Pitié-Salpêtrière, APHP, Paris, France, **11** Université Pierre et Marie Curie, France, **12** INSERM U-698, Paris, France, **13** Service de cardiologie, Hôpital Bichat-Claude Bernard, APHP, Paris, France, **14** CIE 801, INSERM, Paris, France, **15** Département d'Epidémiologie, Biostatistique et Recherche Clinique, Hôpital Bichat-Claude Bernard, APHP, Paris, France

Abstract

Background and Objectives: To weight ischemic and bleeding events according to their severity to be used in a composite outcome in RCTs in the field of thrombosis prevention.

Method: Using a Delphi consensus method, a panel of anaesthesiology and cardiology experts rated the severity of thrombotic and bleeding clinical events. The ratings were expressed on a 10-point scale. The median and quartiles of the ratings of each item were returned to the experts. Then, the panel members evaluated the events a second time with knowledge of the group responses from the first round. Cronbach's α was used as a measure of homogeneity for the ratings. The final rating for each event corresponded to the median rating obtained at the last Delphi round.

Results: Of 70 experts invited, 32 (46%) accepted to participate. Consensus was reached at the second round as indicated by Cronbach's α value (0.99 (95% CI 0.98-1.00)) so the Delphi was stopped. Severity ranged from under-popliteal venous thrombosis (median = 3, Q1 = 2; Q3 = 3) to ischemic stroke or intracerebral hemorrhage with severe disability at 7 days and massive pulmonary embolism (median = 9, Q1 = 9; Q3 = 9). Ratings did not differ according to the medical specialty of experts.

Conclusions: These ratings could be used to weight ischemic and bleeding events of various severity comprising a composite outcome in the field of thrombosis prevention.

Citation: Dechartres A, Albaladejo P, Mantz J, Samama CM, Collet J-P, et al. (2011) Delphi-Consensus Weights for Ischemic and Bleeding Events to Be Included in a Composite Outcome for RCTs in Thrombosis Prevention. PLoS ONE 6(4): e18461. doi:10.1371/journal.pone.0018461

Editor: Rochelle E. Tractenberg, Georgetown University Medical Center, United States of America

Received: September 7, 2010; **Accepted:** March 8, 2011; **Published:** April 7, 2011

Copyright: © 2011 Dechartres et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The authors have no support or funding to report.

Competing Interests: P. Albaladejo: Research grants (to the institution) from Bristol Myers Squibb, Pfizer, Daiichi Sankyo, Eli-Lilly, Boehringer Ingelheim, LFB, consulting or lecture fees from Bayer, Bristol-Myers Squibb, Pfizer, Daiichi-Sankyo, Eli-Lilly, Boehringer Ingelheim, Sanofi-Aventis, LFB, and Vifor. J. Mantz: Consulting fees from Baxter, Pall Medical and Orionpharma. C.M. Samama: Research grants (to the institution) from LFB, NovoNordisk, CSL Behring, SanofiAventis, Bayer, consulting or lecture fees from AstraZeneca, Bayer, Bristol-Myers Squibb, Daiichi-Sankyo, Eli-Lilly, Sanofi-Aventis Group, Pfizer, Boehringer-Ingelheim, and GSK. J.P. Collet: Research grants (to the institution) from Abbott Vascular, Bristol-Myers Squibb, Boston Scientific, Cordis, Eli-Lilly, Fédération Française de Cardiologie, Fondation de France, Guerbet Medical, INSERM, Medtronic, Sanofi-Aventis Group, Société Française de Cardiologie, consulting or lecture fees from Zsra-Zeneca, Bayer, Bristol-Myers Squibb, Daiichi-Sankyo, Eli-Lilly, Sanofi-Aventis Group, and Servier. P.G. Steg: Hamilton Health Services Astellas, AstraZeneca, Boehringer-Ingelheim, Bristol-Myers Squibb, GSK, Daiichi-Sankyo/Eli-Lilly Alliance, Medtronic, Merck Sharpe Dohme, Roche, Sanofi-Aventis, Servier, The Medicines Company, and Aterovax. P. Ravaud: None. F. Tubach: Research grants (to the institution) from Abbott, AstraZeneca, Pfizer, and Schering Plough.

* E-mail: florence.tubach@bch.aphp.fr

Introduction

A composite outcome consists of two or more component outcomes. Patients who have experienced any one of the events specified by the components are considered to have experienced the composite outcome[1,2]. The use of composite outcomes in RCTs is common, particularly in cardiology[3] having the advantage of reducing sample size requirement, costs and time because of higher event rates. Composite outcomes estimate the net clinical benefit of treatment and enable to avoid an arbitrary

choice between a number of important outcomes[2,4–7] so they may be used to summarize the risk/benefit profile of an intervention[8,9]. In the field of thrombosis prevention where treatments aim to decrease the rate of ischemic events but may cause hemorrhagic side effects of various severity, using composite outcomes including both ischemic and hemorrhagic events may be particularly appropriate to capture the net clinical benefit. Many authors have argued that all components of a composite outcome should be of similar importance to adequately interpret treatment effect[1–4,6–8,10–12] which is not frequently the case. Cordoba

showed that the components were not of similar importance in 70% of RCTs reporting a binary composite outcome[1]. Choosing individual components of the same importance might also be irrelevant if the aim is to capture the overall impact of treatment. This is why some authors have proposed to assign each component a weight reflecting severity[8,12–14]. Since weighting may be somewhat arbitrary, it should be subjected to consensus panel[12,13,15].

STRATAGEM is a multicenter, randomized, double-blind, placebo-controlled trial whose objective was to compare low-dose aspirin therapy versus placebo (stopping anti platelet therapy) in the perioperative period in patients treated with antiplatelet therapy as secondary prevention (with documented symptomatic stable atherothrombotic disease) who undergo non-coronary surgery (registration number: NCT00190307, IRB authorization from the “Comité Consultatif de Protection des Personnes se prêtant à la Recherche Biomédicale (CCPPRB) de Paris Bichat” (Ref 2004/18, authorization obtained the 10th of Novembre 2004). The composite outcome took into account the balance of risk and benefit associated with maintaining antiplatelet therapy in the peri-operative period including both ischemic events (e.g., ischemic stroke, non-fatal myocardial infarction, acute limb ischemia, clinical deep venous thrombosis) and bleeding events (e.g., life-threatening bleeding or conducive to revision, or redo surgery, cerebral hemorrhage, intra- or retroperitoneal bleeding, bleeding requiring the transfusion of more than 3 units of packed red blood cells) in addition to overall mortality within one month following surgery. Since the individual components of this composite outcome clearly do not have the same value and severity, the aim of the present project was to attribute consensus-driven weights to ischemic and bleeding events according to their severity to be used in a composite outcome in RCTs in the field of thrombosis prevention.

Methods

Study design

The Delphi method was used to synthesize expert opinion [16,17]. It is a well-recognized method to reach consensus, relying on the following principles: anonymity, iteration, controlled feedback, and statistical aggregation of group responses [18–20].

Staff

A steering committee was initiated to perform this study and included all authors. The committee was responsible for the selection of events to be evaluated and experts, the analysis of the responses and the presentation of results.

Selection of experts

Experts were recruited from clinical disciplines involved in the management of patients with atherothrombotic disease in the perioperative period. In France, both cardiologists and anesthesiologists are involved in this field. Experienced academic experts were identified from different centers all over the country within national organizations such as the French Society of Anesthesia and Intensive Care or the French Society of Cardiology. The selected experts had also to be involved in design, execution and evaluation of clinical trials. Thirty cardiologists and 40 anaesthesiologists were invited to participate in the study. The experts were sent a standardized information package containing a synopsis of the study and a description of the Delphi process. The experts were informed that the consensus-driven ratings would be used as weights in a composite outcome.

Selection of events to be evaluated

Events to be evaluated were identified from the Common Terminology Criteria for Adverse Events (CTCAE) v3.0[21] which is a descriptive terminology that can be used for Adverse Event (AE) reporting. A grading (severity) scale is provided for each AE term. One author (F.T.) identified 28 ischemic and bleeding events that were then submitted to the steering committee for validation to enter the first Delphi round. They covered all the fields addressed by the STRATAGEM composite endpoint, in a more detailed way (for instance myocardial infarction was addressed by 3 different events corresponding to 3 different levels of severity in accordance with the CTCAE). We did not include death among the events to be assessed since the steering committee decided to attribute it automatically the worse rating (i.e., 10). The items involved in the Delphi process are reported in table 1.

Delphi consensus

The steering committee planned to perform at least two Delphi rounds. If consensus was not reached after 2 rounds, it was planned to perform additional rounds until a consensus was reached. The consensus process was conducted via email. Two reminders were sent at each round in case of non response.

In the first Delphi round, each member of the panel evaluated the severity of each of the 28 events on a 10-point scale. For each event, the experts were asked to answer the following question: “According to you, how severe is this event?”. A 10-point scale with the anchors “not severe at all” at 0 and “extremely severe” at 9 was used to record the responses. The experts had the possibility to suggest events that were missing. They were added at the following round provided that they were not redundant with the other events. The median rating (1st quartile–3rd quartile (Q1–Q3)) for the whole group was established for each individual event.

In the second round, the experts considered the same event, and were also informed of each event rating at the first round by reporting of the median ((Q1–Q3)) rating on the scale for each event. The experts were asked to rate each event again in light of the responses at the first round.

Analysis

For each event, the experts’ ratings were summarized as median (Q1–Q3). We applied a Last Observation Carried Forward (LOCF) strategy for missing data after the first round that is to say that, if an expert did not answer the second round, we considered his answers at the first round.

The concept of consensus within a group was defined as homogeneity or consistency opinion among the experts. Assuming that each event was characterized by a constant but unknown severity, the ratings of the experts could be considered as multiple measures of this characteristic. We used Cronbach’s α to measure internal consistency among the experts for the set of events reflects the extent of consensus within the group for the severity of the set of events. When Cronbach’s α is close to 1.0, it can be argued that there is consistency in the responses of the index panel, suggesting consensus. According to the recommendation of Bland and Altman [22], we considered that a consensus would be reached for a Cronbach’s α value of 0.95. We also calculated intra-class correlation coefficient as a measure of the overall agreement between experts [23]. Ninety five percent confidence intervals for both Cronbach’s α and intra-class correlation coefficient were calculated with bootstraps (1000 simulations). We planned to stop the Delphi consensus after the second round if the Cronbach’s α value was superior to 0.95. The final weight for each event was the median rating obtained at the last Delphi round.

All analyses were performed on R version 2.10.0[24].

Table 1. Delphi panel events.

Clinical events	Definition
Ischemic events	
Transient ischemic attack	Transient ischemic event q 24 hrs duration
Ischemic stroke with no symptom at 7 days	Defined by a modified Rankin scale of 0–1
Ischemic stroke with slight disability at 7 days	Defined by a modified Rankin scale of 2
Ischemic stroke with moderate disability at 7 days	Defined by a modified Rankin scale of 3
Ischemic stroke with severe disability at 7 days	Defined by a modified Rankin scale of 4–5
Limb ischemia not requiring heparin or intervention	Brief (q 24 hrs) episode of ischemia managed non surgically and without permanent deficit
Limb ischemia requiring heparin or intervention	Recurring or prolonged (> 24 hrs) requiring medical or surgical intervention
Limb ischemia requiring amputation*	Life-threatening, disabling limb ischemia requiring end organ damage (i.e., limb loss)
Increased level of troponin	Without new Q wave or heart failure
Non-fatal myocardial infarction without heart failure	Killip 1
Non-fatal myocardial infarction with heart failure	Killip ≥ 2
Under-popliteal deep venous thrombosis	Under-popliteal deep venous thrombosis
Deep venous thrombosis with iliac extension	Deep venous thrombosis with iliac extension
Venous thrombosis of the pectoral limb	Venous thrombosis of the pectoral limb
Venous thrombosis other	Venous thrombosis other
Pulmonary embolism	Pulmonary embolism
Massive pulmonary embolism	Clinical or echographical acute pulmonary heart and/or impact on hepatic biology and/or more than half obstruction at angiography
Hemorrhagic events	
Intracerebral hemorrhage with no symptom at 7 days	Defined by a modified Rankin scale of 0–1
Intracerebral hemorrhage with slight disability at 7 days	Defined by a modified Rankin scale of 2
Intracerebral hemorrhage with moderate disability at 7 days	Defined by a modified Rankin scale of 3
Intracerebral hemorrhage with severe disability at 7 days	Defined by a modified Rankin scale of 4–5
Bleeding with increased length of stay	Mild bleeding not requiring intervention other than iron supplements
Bleeding requiring redo surgery or endoscopic sclerosis	
Bleeding requiring both redo surgery and interventions to maintain cardiac output	Bleeding with life-threatening consequences requiring urgent and major interventions
Bleeding requiring transfusion of 3 U or more packed red blood cells	Bleeding requiring transfusion of 3 U or more packed red blood cells
Bleeding requiring both transfusion of 3 U or more packed red blood cells and interventions to increase cardiac output	Bleeding with life-threatening consequences requiring urgent and major interventions
Intra or retroperitoneal bleeding	Intra or retroperitoneal bleeding
Intra or retroperitoneal bleeding requiring interventions to maintain cardiac output	Bleeding with life-threatening consequences requiring urgent and major interventions

Mesenteric ischaemia is considered as peripheral ischemia and then classified as “Life-threatening, disabling limb ischemia requiring end organ damage (i.e., limb loss)” and thus with limb ischemia requiring amputation.

doi:10.1371/journal.pone.0018461.t001

Results

Delphi process

Of the 70 experts invited (30 cardiologists and 40 anaesthesiologists), 32 (46%) accepted to participate in the survey and completed the first round (9 cardiologists (30%) and 23 anesthesiologists (57%)). Twenty five experts (78%) completed the second round (6 cardiologists and 19 anesthesiologists). One event suggested by an expert was added at the second round.

At the second round, Cronbach's α was 0.99 (95% CI 0.98–1.00) showing a high internal consistency indicating consensus between the experts and therefore the end of the Delphi process. Overall agreement between experts was good with an intra-class correlation coefficient at 0.72 (95% CI: 0.59–0.80).

Consensus

A summary of experts' rating for each event and for each Delphi round is presented in Table 2. The ranking of the events slightly changed between the 1st and 2nd round. Events with the lowest rating of severity were: increased Troponin level (median = 3, Q1 = 3; Q3 = 4) and infra-popliteal venous thrombosis (median = 3, Q1 = 2; Q3 = 3). Events with the highest rating of importance were: ischemic stroke with severe disability at 7 days (median = 9, Q1 = 9; Q3 = 9), non-fatal myocardial infarction with heart failure (median = 9, Q1 = 8; Q3 = 9), massive pulmonary embolism (median = 9, Q1 = 9; Q3 = 9) and intra-cerebral hemorrhage with severe disability at 7 days (median = 9, Q1 = 9; Q3 = 9). Delphi-consensus weights are presented in Table 3. Ratings did not differ according to the specialty of experts (Appendix S1). Ratings at the first Delphi round did not differ

Table 2. Summary of experts' rating at each Delphi round for the assessment of severity on a 10-point scale of events deriving from individual components of a composite outcome.

Events	Median (Q1–Q3) 1 st Delphi round	Median (Q1–Q3) 2 nd Delphi round
Thrombotic events:		
Transient ischemic attack	5 (3–6)	5 (4–5)
Ischemic stroke with no symptom at 7 days	6 (5–6)	6 (5–6)
Ischemic stroke with slight disability at 7 days	7 (6–7)	7 (6–7)
Ischemic stroke with moderate disability at 7 days	8 (7–8)	8 (8–8)
Ischemic stroke with severe disability at 7 days	9 (9–9)	9 (9–9)
Limb ischemia not requiring heparin or intervention	5 (4–6)	5 (4–5)
Limb ischemia requiring heparin or intervention	7 (6–7)	6 (6–7)
Limb ischemia requiring amputation	9 (8–9)	9 (8–9)
Increased level of troponin	4 (3–5)	4 (3–4)
Non-fatal myocardial infarction without heart failure	7 (6–8)	7 (6–7)
Non-fatal myocardial infarction with heart failure	8 (8–9)	9 (8–9)
Under-popliteal deep venous thrombosis	3 (2–4)	3 (2–3)
Deep venous thrombosis with iliac extension	6 (5–6)	6 (6–6)
Venous thrombosis of the pectoral limb	5 (4–6)	5 (4–5)
Venous thrombosis other	7 (6–8)	7 (6–7)
Pulmonary embolism	7 (6–8)	8 (7–8)
Massive pulmonary embolism	9 (8–9)	9 (9–9)
Hemorrhagic events:		
Intracerebral hemorrhage with no symptom at 7 days	6 (5–6)	6 (5–6)
Intracerebral hemorrhage with slight disability at 7 days	7 (7–7)	7 (7–7)
Intracerebral hemorrhage with moderate disability at 7 days	8 (8–8)	8 (8–8)
Intracerebral hemorrhage with severe disability at 7 days	9 (9–9)	9 (9–9)
Bleeding with increased length of stay		3 (2–4)
Bleeding requiring redo surgery or endoscopic sclerosis	5 (4–6)	5 (5–6)
Bleeding requiring both redosurgery and interventions to maintain cardiac output	7 (6–8)	7 (7–8)
Bleeding requiring transfusion of 3 U or more packed red blood cells	5 (4–7)	5 (4–6)
Bleeding requiring both transfusion of 3 U or more packed red blood cells and interventions to increase cardiac output	7 (6–8)	7 (7–8)
Intra or retroperitoneal bleeding	6 (4–7)	6 (5–7)
Intra or retroperitoneal bleeding requiring interventions to maintain cardiac output	7 (6–8)	8 (7–8)

doi:10.1371/journal.pone.0018461.t002

between experts who responded at the second Delphi round and those who did not respond (Appendix S2).

Discussion

Before introducing a new treatment or strategy to common practice, or in comparative effectiveness research, capturing the

Table 3. Delphi-consensus weights for ischemic and bleeding events comprising a composite outcome in the field of thrombosis prevention.

Event	Rating
Death	10
Ischemic stroke with severe disability at 7 days	9
Limb ischemia requiring amputation	9
Non-fatal myocardial infarction with heart failure	9
Massive pulmonary embolism	9
Intracerebral hemorrhage with severe disability at 7 days	9
Ischemic stroke with moderate disability at 7 days	8
Pulmonary embolism	8
Intracerebral hemorrhage with moderate disability at 7 days	8
Intra or retroperitoneal bleeding requiring interventions to maintain cardiac output	8
Ischemic stroke with slight disability at 7 days	7
Non-fatal myocardial infarction without heart failure	7
Venous thrombosis other	7
Intracerebral hemorrhage with slight disability at 7 days	7
Bleeding requiring both redosurgery and interventions to maintain cardiac output	7
Bleeding requiring both transfusion of 3 U or more packed red blood cells and interventions to increase cardiac output	7
Ischemic stroke with no symptom at 7 days	6
Limb ischemia requiring heparin or intervention	6
Deep venous thrombosis with iliac extension	6
Intracerebral hemorrhage with no symptom at 7 days	6
Intra or retroperitoneal bleeding	6
Transient ischemic attack	5
Limb ischemia not requiring heparin or intervention	5
Venous thrombosis of the pectoral limb	5
Bleeding requiring redo surgery or endoscopic sclerosis	5
Bleeding requiring transfusion of 3 U or more packed red blood cells	5
Increased level of troponin	4
Under-popliteal deep venous thrombosis	3
Bleeding with increased length of stay	3
No event	0

doi:10.1371/journal.pone.0018461.t003

overall impact of a therapeutic strategy in term of benefit and risk is important[25]. This is a well-recognized advantage of composite outcomes, but their use relies on the underlying assumption that patients will attach similar importance to each component [5]. However, this is rarely true. As outlined by Ferreira-Gonzalez[4] and cordoba[1], most composite end points showed either a large or moderate gradient in importance to patients. Weighting composite outcomes according to severity or importance to patients has been suggested to deal with this issue[8,12–14]. This approach is possible only if a consensus can be reached on the importance of each individual component[15]. We report in this study how consensus-driven severity ratings were obtained for a wide range of ischemic and bleeding events comprising a composite outcome. The Delphi method was used to assign each individual component of the composite outcome a rating reflecting its severity. This well-recognized method to reach consensus in

health care research[18–20] presents major advantages : it can be conducted via mail or email which improves feasibility and lowers costs and it can be completely anonymous which limits the influence of a single expert. Experts presented a high level of agreement so the Delphi was stopped at the second round.

All individual components of the composite outcome were ranked from the most (i.e., death) to the least severe (i.e., absence of event) considering the final median rating attributed by the experts for each event. There are several possibilities to deal with the fact that a single patient may present several events of interest during the follow-up period. As proposed by Braunwald[13], the score for each patient may represent the score of the most serious event encountered by this patient regardless of the number of events having occurred what we planned to do in this study. Another possibility could be to use the sum of the ratings for all outcomes encountered[14]. We believe that presenting both a transient ischemic attack (weight = 5) and increased level of troponin (weight = 4) during the follow-up period is not equivalent to ischemic stroke with severe disability at 7 days (weight = 9). Furthermore, we believe that death from myocardial infarction should not account for a higher rating than death from unknown cause occurring at home, which might also be due to myocardial infarction. Rating multiple events was not possible in our study given the number of possible combinations so the consensus was limited to severity ratings for each event and did not relate to their combination.

Felker proposed an alternative method[26]: all patients who met the worst event (i.e., death) during the follow-up would be assigned the worst ranks, in order to their time to event (e.g., the patient who died first would have the worst rank, the second patient who died the second worst rank). Patient not dying during study follow-up would be evaluated for the second worst endpoint and ranked above those who died, using the same methodology. Those patients not experiencing any of the event components during follow-up would be ranked according to quality of life scores from baseline to last follow-up. After all study subjects are ranked, the comparative efficacy of the 2 treatments is evaluated by comparing the ranks between the 2 groups.

Events rated by the experts to be included in the final composite outcome can be considered as patient important outcomes (which was previously defined as death, morbidity or, patient reported outcomes[27]). Nevertheless, a potential limitation of this study is the absence of involvement of patients to assess the severity of clinical events which may be differently perceived than by physicians. We believed that explaining clearly all events with their possible consequences to make the judgment of patients possible would have been difficult.

Whatever the way to use the ratings to build the composite outcome, there is no evidence that such a composite outcome represents a clinically meaningful endpoint. A validation study should be undertaken with comparison of the different strategies for integrating the ratings. Important questions may be also raised about which between-arm difference will be relevant, with implications for interpretation of results and sample size calculation. Calculating sample size is generally difficult for

composite outcomes since information for the control group may be available for one or several components separately but rarely for the overall outcome. The most important problem pertains to the interpretation of results, which is not intuitive using this approach. Which between-arm difference for the final composite outcome corresponds to a clinically relevant difference is an issue.

It has to be noted that the severity ratings were ordinal and not true interval so the composite outcome should not theoretically be treated as a continuous variable. We also made the assumption that the experts not responding at the second round would have had identical answers in the second round and applied a LOCF strategy. We compared the ratings at the first round between the experts having responded at the second round and those who did not and checked that there was no difference in the ratings (appendix S2). Third, we made the assumption that cardiologists and anesthesiologists would be consistent in their ratings, which we verified by comparing their ratings (appendix S1).

In conclusion, the consensus-driven ratings that were obtained could be used to weight ischemic and bleeding events of various severity comprising a composite outcome in the field of thrombosis prevention. This approach could be reproduced for other types of treatment and medical areas.

Supporting Information

Appendix S1 Summary of experts' rating for the assessment of importance on a 10-point scale of events deriving from individual components of a composite outcome at the second Delphi round according to the specialty of experts. (DOC)

Appendix S2 Comparison of summary of experts' rating at the first Delphi round between experts who responded at the second round and those who did not. (DOC)

Acknowledgments

We would like to thank all experts who participated in this study: Michel Bertrand, Emmanuel Teiger, Emile Ferrari, Yves Juillièrre, Marie-Claude Morice, Pierre Aubry, Gérald Roul, Jeanne Barré, Vincent Piriou, Bernard Cholley, Souhayl Dahmani, Jean-Paul Depoix, Hawa Keita, Gilles Lebuffe, Hervé Dupont, Benoit Plaud, Jean-François Payen, Paul Zufferey, Benjamin Tremey, Philippe Vanderlinden, Sylvain Belisle, Jean François Hardy, Samir Jaber, Jean Luc Fellahi, Gérard Janvier, Vincent Minville, Karim Asehnoune, Sylvie Schlumberger.

Author Contributions

Conceived and designed the experiments: AD PA JM J-PC CMS PGS PR FT. Performed the experiments: AD JM CS J-PC PGS FT. Analyzed the data: AD FT. Contributed reagents/materials/analysis tools: FT JM PA CS J-PC PGS PR. Wrote the paper: AD FT. Drafting the manuscript: AD FT. Critical revision of the manuscript for important intellectual content: PA JM CS J-PC PGS PR.

References

- Cordoba G, Schwartz L, Woloshin S, Bae H, Gotzsche PC (2010) Definition, reporting, and interpretation of composite outcomes in clinical trials: systematic review. *BMJ* 341: e3920.
- Ferreira-Gonzalez I, Permanyer-Miralda G, Busse JW, Bryant DM, Montori VM, et al. (2007) Methodologic discussions for using and interpreting composite endpoints are limited, but still identify major concerns. *J Clin Epidemiol* 60: 651-657; discussion 658-662.
- Lim E, Brown A, Helmy A, Mussa S, Altman DG (2008) Composite outcomes in cardiovascular research: a survey of randomized trials. *Ann Intern Med* 149: 612-617.
- Ferreira-Gonzalez I, Busse JW, Heels-Ansdell D, Montori VM, Akl EA, et al. (2007) Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. *BMJ* 334: 786.
- Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C (2003) Composite outcomes in randomized trials: greater precision but with greater uncertainty? *JAMA* 289: 2554-2559.
- Freemantle N, Calvert MJ (2010) Interpreting composite outcomes in trials. *BMJ* 341: c3529.
- Ross S (2007) Composite outcomes in randomized clinical trials: arguments for and against. *Am J Obstet Gynecol* 196: 119 e111-116.

8. Pogue J, Thabane L, Devereaux PJ, Yusuf S (2010) Testing for heterogeneity among the components of a binary composite outcome in a clinical trial. *BMC Med Res Methodol* 10: 49.
9. Tugwell P, Judd MG, Fries JF, Singh G, Wells GA (2005) Powering our way to the elusive side effect: a composite outcome 'basket' of predefined designated endpoints in each organ system should be included in all controlled trials. *J Clin Epidemiol* 58: 785–790.
10. Ferreira-Gonzalez I, Permyer-Miralda G, Busse JW, Devereaux PJ, Guyatt GH, et al. (2009) Composite outcomes can distort the nature and magnitude of treatment benefits in clinical trials. *Ann Intern Med* 150: 566–567.
11. Montori VM, Permyer-Miralda G, Ferreira-Gonzalez I, Busse JW, Pacheco-Huergo V, et al. (2005) Validity of composite end points in clinical trials. *BMJ* 330: 594–596.
12. Neaton JD, Gray G, Zuckerman BD, Konstam MA (2005) Key issues in end point selection for heart failure trials: composite end points. *J Card Fail* 11: 567–575.
13. Braunwald E, Cannon CP, McCabe CH (1992) An approach to evaluating thrombolytic therapy in acute myocardial infarction. The 'unsatisfactory outcome' end point. *Circulation* 86: 683–687.
14. Sampson UK, Metcalfe C, Pfeffer MA, Solomon SD, Zou KH (2010) Composite outcomes: weighting component events according to severity assisted interpretation but reduced statistical power. *J Clin Epidemiol* 63: 1156–1158.
15. Califf RM, Harrelson-Woodlief L, Topol EJ (1990) Left ventricular ejection fraction may not be useful as an end point of thrombolytic therapy comparative trials. *Circulation* 82: 1847–1853.
16. Dalkey NC, Helmer O (1963) An experimental application of the Delphi method to the use of experts. *Manage Sci*.
17. Delbecq AL, van de Ven AH, Gustafson DH (1975) Group techniques for program planning: a guide to nominal group and Delphi processes. Glenview: Scott, Foresman.
18. Dalkey NC (1969) The delphi method: an experimental study of group opinion. Santa MonicaCA: RAND Corporation.
19. Fink A, Kosecoff J, Chassin M, Brook RH (1984) Consensus methods: characteristics and guidelines for use. *Am J Public Health* 74: 979–983.
20. Pill J (1971) The Delphi method: substance, context, a critique and an annotated bibliography. *Socio-Econ Plan Sci* 5: 57–71.
21. Program CTE (2006) Common Terminology Criteria for Adverse Events, Version 3.0, DCTD, NCI, NIH, DHHS. March 31, 2003 (<http://ctep.cancer.gov>).
22. Bland JM, Altman DG (1997) Cronbach's alpha. *BMJ* 314: 572.
23. Bravo G, Potvin L (1991) Estimating the reliability of continuous measures with Cronbach's alpha or the intraclass correlation coefficient: toward the integration of two traditions. *J Clin Epidemiol* 44: 381–390.
24. R Development Core Team (2009) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria ISBN 3-900051-07-0, URL <http://www.R-project.org>.
25. Kraemer HC, Frank E (2010) Evaluation of comparative treatment trials: assessing clinical benefits and risks for patients, rather than statistical effects on measures. *JAMA* 304: 683–684.
26. Felker GM, Anstrom KJ, Rogers JG (2008) A global ranking approach to end points in trials of mechanical circulatory support devices. *J Card Fail* 14: 368–372.
27. Rahimi K, Malhotra A, Banning AP, Jenkinson C (2010) Outcome selection and role of patient reported outcomes in contemporary cardiovascular trials: systematic review. *BMJ* 341: c5707.