Methodologies for pragmatic and efficient assessment of benefits and harms: application to the SOCRATES trial

Scott R Evans[1], Mikael Knutsson[2], Pierre Amarenco[3], Gregory W. Albers[4], Philip M Bath[5], Hans Denison[2], Per Ladenvall[2], Jenny Jonasson[2], J. Donald Easton[6], Kazuo Minematsu[7], Carlos A. Molina[8], Yongjun Wang[9], K.S. Lawrence Wong[10], and S. Claiborne Johnston[11], for the SOCRATES Steering Committee and Investigators

[1]Biostatistics Center, George Washington University, Washington, DC, USA; [2]AstraZeneca, Research and Development, Gothenburg, Sweden; [3]Department of Neurology and Stroke Centre, Bichat Hospital, Paris University, Paris, France; [4]Stanford Stroke Center, Stanford University, Stanford, CA, USA; [5]Stroke Trials Unit, Division of Clinical Neuroscience, University of Nottingham, Nottingham, UK; [6]Department of Neurology, University of California, San Francisco, CA, USA; [7]National Cerebral and Cardiovascular Center, Suita, Osaka, Japan; [8]Stroke Unit, Hospital Vall d'Hebron, Barcelona, Spain; [9]Department of Neurology, Beijing Tiantan Hospital, Beijing, China; [10]Department of Medicine & Therapeutics, Chinese University of Hong Kong, Shatin, Hong Kong; [11]Dean's Office, Dell Medical School, University of Texas at Austin, Austin, TX, USA.

Corresponding author: Scott Evans, Ph.D., M.S., Director, Biostatistics Center, Professor of Epidemiology and Biostatistics, Milken Institute School of Public Health, George Washington University, 6110 Executive Boulevard, Suite 750, Rockville, MD 20852-3943, USA.

Phone: (301) 881-9260; Fax: (301) 881-3742; Email: sevans@bsc.gwu.edu

Running head (limit ≤40 characters): Pragmatic benefit-risk assessment

Trial registration number/trial register: NCT01994720: ClinicalTrials.gov

Word count: ~~3874~~ 3882 words

Abstract

Background/Aims: Standard approaches to trial design and analyses can be inefficient and non-pragmatic. Failure to consider a range of outcomes impedes evidence-based interpretation and reduces power. Traditional approaches synthesizing information obtained from separate analysis of each outcome fail to incorporate associations between outcomes and recognize the cumulative nature of outcomes in individual patients, suffer from competing risk complexities during interpretation, and since efficacy and safety analyses are often conducted on different populations, generalizability is unclear. Pragmatic and efficient approaches to trial design and analyses are needed.

Methods: Approaches providing pragmatic assessment of benefits and harms of interventions, summarizing outcomes experienced by patients, and providing sample size efficiencies are described. Ordinal outcomes recognize finer gradations of patient responses. Desirability-of-outcome ranking (DOOR) is an ordinal outcome combining benefits and harms within patient. Analysis of DOOR can be based on rank-based methodologies including the DOOR probability, the win ratio, and the proportion in favor of treatment. Partial credit analyses, involving grading the levels of the DOOR outcome similar to an academic test, provides an alternative approach. The methodologies are demonstrated using the acute stroke or transient ischemic attack treated with aspirin or ticagrelor and patient outcomes (SOCRATES; NCT01994720), a randomized clinical trial.

Results: Two five-level ordinal outcomes were developed for SOCRATES. The first was based on a Modified Rankin scale. The odds ratio is 0.86 (95% CI 0.75, 0.99; p = 0.04) indicating that the odds of worse stroke categorization for a trial participant assigned to ticagrelor is 0.86 times that of a trial participant assigned to aspirin. The 5-level DOOR outcome incorporated and prioritized survival; the number of strokes, myocardial infarction, and major bleeding events; and whether a stroke event was disabling. The DOOR probability and win ratio are 0.504 (95% CI 0.499, 0.508; p = 0.10) and 1.11 (95% CI 0.98, 1.26; p = 0.10) respectively, implying that the probability of a more desirable result with ticagrelor is

50.4% and that a more desirable result occurs 1.11 times more frequently on ticagrelor vs. aspirin.

Conclusion: Ordinal outcomes can improve efficiency though require pre-specification, careful construction and analyses. Greater pragmatism can be obtained by composing outcomes within-patient. DOOR provides a global assessment of the benefits and harms that more closely reflect the experience of patients. The DOOR probability, the proportion in favor of treatment, the win ratio, and partial credit can more optimally inform patient treatment, enhance the understanding of the totality of intervention effects on patients, and potentially provide efficiencies over standard analyses. The methods provide the infrastructure for incorporating patient values and estimating personalized effects.

Keywords

Benefit-risk, composite outcome, desirability of outcome ranking (DOOR), ordinal outcome, partial credit, win ratio

Introduction

Randomized clinical trials are the gold standard for evaluating the benefits and risks of interventions. However, these studies often fail to provide the necessary evidence to fully inform practical medical decision-making. The important implications of these deficiencies are largely absent from discourse in medical research communities.[1] Contributing factors to these deficiencies include the over-reliance on dichotomized outcomes and the manner in which benefits and harms are integrated and analyzed.

Outcomes in clinical trials are often dichotomized. However, dichotomized outcomes may fail to convey and recognize finer gradations of important patient responses. This potential limitation may result in reduced sensitivity and statistical power for detecting important effects, resulting in larger sample sizes than necessary.

Typical analyses of clinical trials involve separate intervention comparisons for each of efficacy, safety, and quality-of-life outcomes; for example, Johnston et al, 2016.[2] Outcome-specific effects are tabulated and potentially systematically or unsystematically combined in benefit:risk analyses with the belief that such analyses inform of the totality of effects on patients. However, such approaches do not incorporate associations between outcomes of interest, fail to summarize the cumulative nature of different outcomes on individual patients, suffer from competing risk challenges when interpreting individual outcomes, and, since efficacy and safety analyses are frequently conducted on different analysis populations, the population to which these benefit:risk analyses apply, is unclear.

We present and apply recently proposed methodologies for benefit:risk assessment in clinical trials. The methods have greater pragmatism and efficiency than traditional methods in that they recognize finer gradations of important patient responses and provide a global assessment that more closely reflects the totality of the patient experience and status. We

apply the methods in a post-hoc analysis of the data from SOCRATES, a large international clinical trial.[2]

Methods

Methodologies that can improve efficiency and provide greater pragmatism include use of ordinal outcomes, composing outcomes within a patient, and new approaches to statistical analyses.

Ordinal outcomes

Frequently utilized dichotomized outcomes may fail to elicit finer gradations of important responses. This approach results in reduced sensitivity and power for detecting important effects. Ordinal outcomes offer the opportunity to improve efficiency by recognizing finer gradations of patient responses.

Composing outcomes within a patient

Ordinal outcomes represent an incremental step to greater pragmatism. The next step is to appropriately compose the outcomes. Recently proposed methodologies for the evaluation of multiple outcomes provide the opportunity to obtain global assessments of the benefits and harms of interventions that more closely reflect the experience of patients. The general concept is based on using the outcomes to globally analyze the patient rather than using patient data for segmented evaluation of each outcome, and allowing for gradations in outcomes to be considered in assessing overall performance of an intervention through the use of a composite ordinal outcome. The methodologies synthesize multiple outcomes as they are experienced by the patients, consider the relative importance of outcomes, and avoid many competing risk issues that plague analyses and interpretation of single outcomes.

Analysis methods based on pairwise comparisons

Methodologies have recently been proposed to compare two treatments based on multiple outcomes using the concept of pairwise patient comparisons. All possible pairwise comparisons of the outcomes from patients in one treatment arm to the outcomes from patients in the other treatment arm are conducted. For example, if one treatment arm has N1 patients and the other treatment arm has N2 patients then there are N1*N2 possible pairwise comparisons. When comparing a specific patient's results from one treatment group to a patient from the other treatment group, a more desirable ('win'), less desirable ('loss'), or equally desirable (tie) result will be observed.

One strategy for making comparisons based on multiple outcomes is via prioritization of outcomes.[3] For example, suppose two outcomes are considered: survival and whether an adverse event occurred. Further, suppose that survival is prioritized over the adverse event. When comparing two patients, if one survived but the other did not then the patient that survived had the most desirable outcome. If both patients survived, they would then be compared with regard to their adverse event status.

Comparisons can also be made by combining multiple outcomes into a single ordinal composite outcome. For example, the desirability of outcome ranking (DOOR) is a concept whereby patients are classified into an ordinal global outcome based on overall outcome desirability.[4] The DOOR is often constructed using the major endpoints of efficacy, safety and quality-of-life outcomes of interest. Table 1 represents a simple example of a DOOR outcome using survival and adverse events.

When patients have been classified then the *DOOR probability* (i.e. the probability of a more desirable result [adjusted for tied desirability]) in one treatment relative to another treatment, the proportion in favor of treatment,[5] and the *win ratio*[3] (i.e. the relative frequency

by which one treatment has a more desirable result than another) can be calculated by tabulating the pairwise comparison results.

$$\text{DOOR probability} = (\#\text{wins} + 0.5[\#\text{ties}])/(N1*N2)$$

$$\text{Proportion in favor of treatment (net benefit)} = (\#\text{wins} - \#\text{losses})/(N1*N2)$$

$$\text{Win ratio} = \#\ \text{wins}/\#\text{losses}$$

The DOOR probability (similar to the Mann-Whitney probability) and the proportion in favor of treatment (similar to the Somers' D[C|R]) can be viewed as absolute measures whereas the win ratio can be considered as a relative measure. Presentation of the absolute measures is recommended. Consider a comparison of two treatments for which pairwise comparisons result in two more desirable comparisons, one less desirable comparison, and 97 ties resulting in a win ratio of 2. A comparison of two treatments for which pairwise comparisons result in two more desirable comparisons, one less desirable comparison, and 9997 ties also results in a win ratio of 2. These two scenarios are importantly different and would be readily illuminated using the DOOR probability or the proportion in favor of treatment.

Regression approaches for these methods are available to incorporate other factors, for example, stratification variables.[6] These regression approaches have different assumptions compared to ordinal logistic regression but have the advantage of the ability to deal with censoring.[7]

One concern with the methods based on pairwise comparisons is that a decrement in a very important component could be offset by a large advantage in a component outcome of lesser importance despite appropriate prioritization. In the case of the ordinal DOOR

outcome in Table 1, the step between ‘survival without adverse event’ and ‘survival with adverse event’ may be viewed as smaller than the step between ‘survival with adverse event’ and ‘death’. Researchers may wish to directly account for such perspectives during analyses.

Partial credit

Partial credit analyses[8] can be conducted to directly address the concerns with pairwise comparison methodologies. Partial credit analyses involve grading the levels of the ordinal DOOR outcome similar to an academic test, i.e. from 0% to 100%. If the patient experiences the most desirable overall outcome then they receive a score of 100%. If the patient has the least desirable result (e.g. death) then they receive a score of 0. Partial credit is given for intermediate categories (Table 1) directly accounting for cases in which steps between categories are viewed as unequal. Partial credit can be informed from patients using quality-of-life instruments or from a survey of expert clinicians. Treatment comparisons can then be made by comparing mean partial credit scores. The advantage of the partial credit scoring approach is that it strategically scores the DOOR categories to account for non-uniform steps between categories. A disadvantage of the partial credit approach is that it is more challenging to score outcomes than to rank or prioritize them.

Although partial credit scoring can be pre-specified for transparency, the treatment contrast can be displayed as partial credit assignment varies. This approach gives providers and patients the freedom to choose a treatment based on how they value the intermediate categories.

Trial design using new methodologies

When designing a trial with an ordinal outcome, the outcome must be clearly defined and pre-specified for transparency as with traditional outcomes. The calculation of sample size for a trial using a DOOR or other ordinal outcome can be conducted in three ways. The first option involves testing the null hypothesis that a patient randomly assigned to a new therapy will have an X% chance of a better DOOR than one assigned to the control. For X, 50% is often selected noting that >50% implies superiority of the new therapy to the control. Other values of X can be selected. The alternative hypothesis is that a patient randomly assigned to a new therapy will have a Y% chance of a better DOOR than one assigned to the control. Here Y is selected so that Y is greater than X with the improvement based upon assessment of clinical relevance. Using this paradigm, the required sample size can be calculated with standard software using the Mann-Whitney U Test when X = 50%. Simulation can be used with X ≠ 50%.

A second, more-desirable option involves obtaining information on the proportions for the respective categories of the ordinal response for the two groups. The proportions in the categories for the control group may represent the expected responses when treated with the control. The proportions in the categories for the experimental group would be constructed by the amount of improvement in desirability that is important to detect. Once proportions have been identified then sample size can be calculated using standard software when X = 50% and simulation when X ≠ 50%.

The third option involves testing the null hypothesis that the treatment-group means are the same when implementing a pre-specified partial credit scoring strategy. Though analyses can be conducted varying the scoring strategy, a specific scoring strategy can be pre-specified and informed by, for example, surveys of patient perspectives or treatment experts. Given assumed DOOR distributions for the experimental and control strategies, treatment-specific means and standard deviations can be obtained, and the trial can be sized using t-tests and standard software.

Example: SOCRATES Trial

The acute stroke or transient ischemic attack treated with aspirin or ticagrelor and patient outcomes (SOCRATES; NCT01994720) study was a multicenter, randomized, double-blind, double-dummy, parallel-group trial conducted at 674 sites in 33 countries between January 7, 2014 and October 29, 2015.[2,9] The objective of SOCRATES was to compare ticagrelor (180 mg on day 1 followed by 90 mg twice daily on days 2–90) with aspirin (300 mg on day 1, followed by 100 mg daily on days 2–90) for the prevention of major vascular events (a composite of stroke, myocardial infarction [MI], or death) over a period of 90 days in patients with acute ischemic stroke or TIA. Eligible patients were aged ≥40 years, had an acute ischemic stroke with a National Institutes of Health Stroke Scale score of ≤5 or had high-risk transient ischemic attack (ABCD^2 stroke risk score of ≥4 [scores assessing the risk of stroke on the basis of age, blood pressure, clinical features, duration of transient ischemic attack, and presence or absence of diabetes mellitus] or symptomatic intracranial or extracranial arterial stenosis), and could undergo randomization within 24 hours after symptom onset.[9]

In SOCRATES, 13,199 patients were randomized. Demographics and baseline characteristics were balanced between treatment groups.[2] Ticagrelor was not found to be superior to aspirin in reducing the rate of stroke, MI, or death at 90 days. There was a 6.7% event rate in the ticagrelor arm versus 7.5% rate in the aspirin arm resulting in an estimated hazard ratio of 0.89 (95% confidence interval [CI]: 0.78, 1.01; p = 0.07).[2] PLATO major bleeds occurred in 31 patients (0.5%) on ticagrelor and 38 patients (0.6%) on aspirin (hazard ratio, 0.83; 95% CI: 0.52, 1.34) with few intracranial hemorrhages.[10]

Traditional evaluations in cardiovascular-event prevention trials, such as SOCRATES, are limited to evaluating time-to-the-first event (stroke, MI, or death within 90 days). One

paradoxical property to such analyses is that a patient with an MI at 40 days is considered to be a worse outcome than a patient who dies at 60 days, as the MI occurred first and the focus is on the event time. This approach further ignores potentially valuable information that can be derived from the differential importance of the events of interest, the cumulative nature of multiple events, the association between events, and complexities induced by competing risks, for example, death informatively censoring the time to stroke. Typical benefit:risk assessment conducted by separately estimating an effect for each important outcome, for example, death, MI, stroke, and bleeding, and then combining the effects on these outcomes in some way, is difficult to interpret. Since these events are not mutually exclusive (e.g. fatal bleeding event), events can be double-counted.[11] Ordinal outcome and DOOR analyses provide an opportunity to address these issues.

Results

Separate ordinal outcomes

The results of the SOCRATES trial for an ordinal outcome based on a Modified Rankin scale using ordinal logistic regression analyses[12,13] are shown in Table 2. Advantages of this approach are: (1) recognition of important finer gradations of patient response, more closely reflecting the differential impact of therapies and (2) greater sensitivity providing efficiencies.

The results of the efficacy stroke analyses suggest that ticagrelor is superior to aspirin (odds ratio = 0.862; 95% CI 0.747, 0.995; p = 0.04) indicating that the odds of worse stroke categorization for a trial participant assigned to ticagrelor is 0.862 times that of a trial participant assigned to aspirin. These results are similar to the time-to~~dichotomized all~~ stroke analyses (estimated hazard ratio of 0.86; 95% CI 0.75, 0.00; p = 0.03) and the ~~dichotomized~~ primary endpoint i.e. the time-to event analyses of the composite of stroke, MI, or death (estimated hazard ratio of 0.89 [95% CI 0.78, 1.01; p = 0.07]).[2]

The results of the safety analyses of hemorrhage suggest a significant increase in bleeding for ticagrelor versus aspirin (p<0.0001), a result mainly driven by no event vs. any event analyses. These ~~though~~ analyses are difficult to interpret ~~, mainly driven by no event versus any event~~ since there are low percentages for the three most severe types of events and aspirin had more PLATO major bleeding events. Given the tradeoff of efficacy and safety, careful synthesis of these evaluations through benefit:risk evaluation is required.

Limitations of the separate ordinal outcome analyses approach include: (1) MIs are not included; (2) ordinal logistic regression requires that the proportional odds assumption hold to ensure validity and has difficulty dealing with censoring; (3) difficulty synthesizing separate relative measures for efficacy and safety, for example, an odds ratio of 2 could imply

increase of risk of 1 in 10 to 2 in 10 indicating a high level of importance, or alternatively imply an increase of risk from 1 in 10,000 to 2 in 10,000 representing a much lower level of importance; (4) associations between efficacy and safety not evaluated; and (5) the efficacy and safety analyses conducted on different analysis populations.

DOOR

DOOR analyses were conducted building upon the advantages of the ordinal outcome. DOOR addresses the limitations of the separate ordinal outcome evaluation described above by: (1) including MIs; (2) using the DOOR probability and win ratio, which do not require the proportional odds assumption; (3) focusing on a composed absolute measure (i.e. DOOR probability); (4) examining the association between the various component outcomes by synthesizing the benefits and harms within patient, and (5) conducting harmonized analyses on a single analysis population (intent-to-treat).

DOOR was constructed as a global composite ordinal outcome incorporating major events (death, stroke, MI, and major bleeding). Construction of the DOOR prioritized death over non-fatal events, disabling over non-disabling events, and more events over fewer events. The DOOR consisted of five levels (from most to least desirable): (1) survived with no event, (2) survived with non-disabling stroke, MI, or PLATO-defined major bleeding (one event), (3) survived with non-disabling stroke, MI, or PLATO-defined major bleeding (>1 event), (4) survived with disabling stroke, and (5) death. The distribution of patients was 92.9%, 2.2%, 0.1%, 3.7%, and 1.0% for ticagrelor, and 92.1%, 2.6%, 0.2%, 4.3%, and 0.9% for aspirin in categories 1–5, respectively (Table 3). The cumulative difference (ticagrelor – aspirin) in proportions can be evaluated by sequentially dichotomizing the DOOR outcome. For example, one can dichotomize ‘survived with no event’ versus other categories resulting in difference of 0.8% (95% CI –0.1, 1.7). At the other extreme, one can dichotomize the last

category (death) versus other categories yielding an analysis of survival resulting in a difference of –0.2% (95% CI –0.5, 0.2).

The DOOR probability is 0.504 (95% CI 0.499, 0.508; p = 0.096) implying that the probability of a more desirable result with ticagrelor is 50.4%. The win ratio = 1.11 (95% CI 0.98, 1.26; p = 0.096), implying that a more desirable result occurs 1.11 more frequently on ticagrelor than aspirin.

Partial credit

To illustrate partial credit analyses, we collapse categories 2 and 3 so that there are four total categories i.e. (1) survived with no event, (2) survived with non-disabling event(s), (3) survived with disabling stroke, and (4) death. Partial credit (between 0 and 100%) is provided to the two middle categories with partial credit for surviving with a non-disabling event (p1) greater than that for surviving with a disabling stroke (p2). Contours of the difference (ticagrelor – aspirin) in mean partial credit scores (Figure 1A) and ratio (ticagrelor/aspirin) of partial credit scores (Figure 1B) can be plotted for varying values of p1 and p2 to evaluate the robustness of the results and to allow for personalized choices regarding the values of p1 and p2.

If partial credit of 100% is given to categories (2) and (3) (Figure 1A, point C), equivalent to analysis of survival, then the difference in means is –0.15 (95% CI –0.49, 0.18). If a partial credit of 0% is given to these two categories (Figure 1A, point A), equivalent to analysis of zero versus any event, then the difference in means is 0.82 (95% CI –0.07, 1.72). Quality-of-life studies can be used to provide patient or clinician perspectives to better inform partial credit scoring. Using partial credit scoring of 68 for a non-disabling event and 38 for a disabling event (scoring derived from Okumura et al, 2015)[14] (Figure 1A, point E), the difference in means is 0.32 (95% CI –0.22, 0.87). Table 4 summarizes the partial credit analyses. Analysis for any other scoring strategy can also be evaluated noting that patients

have their own values and that patients' perspectives can vary from that of clinicians.[15] Results for the ratio of scores are also provided in Figure 1B and Table 4.

Subgroup evaluation for personalized medicine

Since benefits and harms have been synthesized by composing them within patients, subgroup evaluations can be conducted to evaluate how the global effects of therapy may vary by patient characteristics. We examined the ticagrelor – aspirin comparison by subgroups defined by: (1) whether the patient had prior aspirin use, and (2) the time-to-loading dose. Results suggest that the advantages with ticagrelor are for patients with prior aspirin and shorter times to the loading dose (Table 5). Such analyses, more appropriately convey the heterogeneity of effects on patients compared with subgroup analyses based on a single outcome.

Discussion

Dichotomized outcomes provide simplicity at the expense of inefficiencies, thereby possibly resulting in the failure to recognize finer but important gradations of patient responses. Ordinal outcomes may improve efficiencies though require careful construction, pre-specification, and analyses.

DOOR is a composite ordinal outcome that provides the opportunity to obtain global assessments of the benefits and harms of interventions that more closely reflect the experience of patients. Rank-based methods (the DOOR probability, the proportion in favor of treatment, and the win ratio) and a score-based method, partial credit are available to compare the DOOR outcome between interventions to more optimally inform patient treatment by enhancing the understanding of the totality of intervention effects on patients. Critical components include: (1) using outcomes to analyze patients rather than using patients to analyze outcomes, (2) incorporating patient values, and (3) evaluating personalized effects. A critical factor in this approach requires improved understanding of how to analyze one patient before analyzing many patients.

Construction of DOOR and other composite outcomes is challenging, requiring great care to ensure the composition appropriately recognizes the relative importance of individual outcomes.[16] The importance of recognizing the differential impact of events of interest has been noted. For example, in weighing the risks and benefits of prasugrel for the reduction of thrombotic cardiovascular events in patients with acute coronary syndrome (unstable angina or MI) who undergo percutaneous coronary intervention, the Food and Drug Administration (FDA) review team noted that the components of the primary endpoint (death, MI, stroke) represented irreversible tissue damage. They concluded that the benefit of preventing such events is generally worth the risk of bleeding events that are generally transient and have no irreversible consequences. Bleeding events may have serious consequences, but most of

those that led to irreversible harm (e.g. intracranial hemorrhage) were included in the primary endpoint.[17] Similarly, during the evaluation of dabigatran for the reduction of the risk of stroke and systemic embolism in patients with nonvalvular atrial fibrillation, the FDA described why they approved a higher dose rather than a lower dose. The FDA again noted that nonfatal and extracranial bleeding episodes are less clinically significant than strokes for most patients, as the irreversible effects of strokes and systemic emboli have greater clinical significance than nonfatal bleeding. They concluded that benefit:risk assessment in which strokes and systemic emboli are given more weight than non-fatal bleeding events results in the higher dabigatran dose being more favorable.[18]

The SOCRATES trial was used as an example to illustrate application of the methodologies. DOOR was constructed using death, stroke, MI, and PLATO-defined major bleeding events. The construction attempted to recognize that: (i) fatal events are worse than non-fatal events, (ii) disabling events are worse than non-disabling events, and (iii) multiple events are worse than fewer events. Elements of this construction can be questioned. For example, some data suggests that patients view certain strokes (i.e. mRS=5) as worse than death.[19] Others may question the assumption of equal importance of MIs, non-disabling strokes and major bleeds. Non-disabling stroke is moderately homogeneous in severity and effect on patient function but MIs and major bleeds have more variable sequelae. Alternative constructions of DOOR could be pursued shaped by studies that evaluate the importance of component outcomes.[20,21] Additional categories could be constructed to recognize the heterogeneous nature of the effects of MIs and bleeds. Co-morbidities such as pulmonary embolism and aspiration pneumonia could be integrated. Utilization of levels of the ordinal Modified Rankin scale may provide greater sensitivity.

Application of the new methodologies does not mitigate the necessity of the evaluation of component outcomes. Analyses of component outcomes provide value and should be considered part of the application of the new methodologies. For example, analyses of

safety outcomes may uncover toxicity that may be reduced by modification of the dose or delivery system.

Although we have discussed the methodologies within the context of the SOCRATES trial focused on whether ticagrelor prevents short-term (within 90 days) cardiovascular events, application of the methodologies can be modified and tailored to trials for which the event time is important to consider. For example, a long-term trial evaluating survival time may categorize survival duration with long-term survival being more desirable than short-term survival. The methods are being tailored and employed in other disease settings such as infectious disease[22–24] and oncology.[25] The methods have recently been proposed as a tool for data monitoring committees for conducting benefit:risk assessment during interim data monitoring.[26]

Acknowledgments

Editorial support (only formatting tables and figures, coordinating reviews and preparing the manuscript for submission and re-submission in response to journal comments, i.e., no writing support) was provided by Jackie Phillipson (Zoetic Science, an Ashfield company, part of UDG Healthcare plc, Macclesfield, UK) and funded by AstraZeneca.

Declaration of conflicting interests

SR Evans is a statistical consultant to AstraZeneca for the SOCRATES and THALES (NCT03354429) trials (significant). **M Knutsson, H Denison, P Ladenvall and J Jonasson** are employees of AstraZeneca. **P Amarenco** receives research grant support (significant) from AstraZeneca, Sanofi, and Bristol-Myers Squibb (BMS) for http://www.TIAregistry.org, the French Government and Pfizer for the Treat Stroke to Target trial, and Boston Scientific for the Warfarin Aspirin Ten(X)-a inhibitors Left Atrial

Appendage Closure in cerebral infarct and Hemorrhage and Atrial Fibrillation registry. He has received modest consultant/advisory board fees from Amgen and BMS, and modest honoraria from Amgen (speaker activities), Pfizer for the SPIRE program Executive Committee, AstraZeneca for the SOCRATES trial Executive Committee, and Kowa for the PROMINENT study (NCT03071692) Executive Committee, and significant honoraria from Bayer for the XANTUS study (NCT01606995) Executive Committee, AstraZeneca for the THALES trial Executive Committee, and Fibrogen as an ALPINE program trials Data and Safety Monitoring Board member. **GW Albers** reports research support (significant) from the NIH/NINDS, equity interest and consulting fees from iSchemaView (significant), and consulting fees from AstraZeneca and Janssen (significant). **P Bath** reports research support (significant) from the NIHR/HTA and British Heart Foundation; is co-chief investigator for the PhEED trial (NCT03358810); and has received honoraria (modest) from DiaMedica, Moleac, Nestle, Phagenesis and Sanofi. **JD Easton** receives research support (significant) from the NIH/the National Institute of Neurological Disorders and Stroke (NINDS) as a co-principal investigator for the POINT trial (U01 NS062835-01A1); POINT received study drug and placebo from Sanofi (NCT00991029). His institution received research grant support from AstraZeneca (significant) for planning and conducting the SOCRATES trial. He also received support (modest) from Boehringer Ingelheim (BI) as a consultant for the planning and conduct of the RE-SPECT ESUS (NCT02239120) trial. **K Minematsu** reports honoraria (all modest for seminar presentations) from Bayer Yakuhin, Otsuka Pharmaceutical, BI, AstraZeneca, Pfizer, Mitsubishi Tanabe Pharma Cooperation, Japan Stryker, Daiichi Sankyo, Astellas Pharma, Nippon Chemiphar, and Fuji Film RI Pharma. He has also received modest fees as a Consultant/Advisory board member from AstraZeneca, CSL Behring, Medico's Hirata, Bayer Yakuhin, EPS Corporation, HEALIOS K.K., and T-PEC Corporation. **CA Molina** reports serving in the Steering Committee (significant) of the CLOTBUST-ER trial (NCT01098981) (Cerevast); SOCRATES (AstraZeneca), ImpACT-24B trial (NCT00826059) (Brainsgate),

REVASCAT trial (NCT01692379) (Fundació Ictus Malaltia Vascular). He has received honoraria for participation in clinical trials, contribution to advisory boards or oral presentations from: AstraZeneca (modest), BI, Daiichi Sankyo, BMS, Covidien, Cerevast, Brainsgate. Dr Molina has no ownership interest and does not own stocks of any pharmaceutical or medical device company. **Y Wang** reports research grant support from AstraZeneca. **KSL Wong** reports honoraria as a member of a steering committee for Johnson & Johnson, AstraZeneca (modest) and Bayer (modest); honoraria for participation in clinical trials, contributions to advisory boards, or oral presentations from Bayer (modest), Sanofi-Aventis (modest), BMS, BI (modest), and Pfizer (modest). **SC Johnston** receives research grant support (significant) from the NIH/NINDS as principal investigator for the POINT trial (NCT00991029); POINT received most study drug and placebo from Sanofi. Also, he was a consultant to AstraZeneca during the planning of the SOCRATES trial and his institution received research support (significant) for the conduct of SOCRATES and THALES.

Funding

SOCRATES and the analyses reported herein were funded by AstraZeneca.

References

1. DeMets DL and Califf RM. A historical perspective on clinical trials and leadership: where have the academics gone? *JAMA* 2011; 305: 713–714.

2. Johnston SC, Amarenco P, Albers GW, al. Ticagrelor versus aspirin in acute stroke or transient ischemic attack. *N Engl J Med* 2016; 375: 35–43.

3. Pocock SJ, Ariti CA, Collier TJ, et al. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *Eur Heart J* 2012; 33: 176–182.

4. Evans SR, Rubin D, Follmann D, et al. Desirability of outcome ranking (DOOR) and response adjusted for duration of antibiotic risk (RADAR). *Clin Infect Dis* 2015; 61: 800–806.

5. Buyse M. Generalized pairwise comparisons of prioritized outcomes in the two-sample problem. *Statist Med* 2010; 29: 3245–3257.

6. Follmann DA. Regression analysis based on pairwise ordering of patients' clinical histories. *Stat Med* 2002; 21: 335–367.

7. Follmann D, Fay MP, Hamasaki T, Evans S. Analysis of ordered composite endpoints. *Stat Med* 2020; 39: 602–616.

8. Evans SR and Follmann D. Using outcomes to analyze patients rather than patients to analyze outcomes: a step toward pragmatism in benefit:risk evaluation. *Stat Biopharm Res* 2016; 8: 386–393.

9. Johnston SC, Amarenco P, Albers GW, et al. Acute stroke or transient ischemic attack treated with aspirin or ticagrelor and patient outcomes (SOCRATES) trial: rationale and design. *Int J Stroke* 2015; 10: 1304–1308.

10. Easton JD, Aunes M, Albers GW, et al. Risk for major bleeding in patients receiving ticagrelor compared with aspirin after TIA or acute ischemic stroke in the SOCRATES Study. *Circulation* 2017; 136: 907–916.

11. He W, Bloomfield D, Mai Y, Evans SR, et al. Practical considerations for benefit risk assessment and implementation: vorapaxar TRA-2P TIMI 50 case study. Benefit-risk assessment methods in medicinal product development: bridging qualitative and quantitative assessments. CRC Press Taylor & Francis Group 2016. Chapter 11, pp 235–249.

12. Bath PMW, Geeganage C, Gray LJ, et al. Use of ordinal outcomes in vascular prevention trials: comparison with binary outcomes in published trials. *Stroke* 2008; 39; 2817–2823.

13. Bath PM, Woodhouse LJ, Appleton JP, et al. Antiplatelet therapy with aspirin, clopidogrel, and dipyridamole versus clopidogrel alone or aspirin and dipyridamole in patients with acute cerebral ischaemia (TARDIS): a randomised, open-label, phase 3 superiority trial. *Lancet* 2018; 391: 850–859.

14. Okumura K, Inoue H, Yasaka M, et al. Comparing patient and physician risk tolerance for bleeding events associated with anticoagulants in atrial fibrillation — evidence from the United States and Japan. *Value Health Reg Issues* 2015; 6: 65–72.

15. Yuan Z, Levitan B, Burton P, Poulos C, Brett Hauber A, Berlin JA. Relative importance of benefits and risks associated with antithrombotic therapies for acute coronary syndrome: patient and physician perspectives. Curr Med Res Opin 2014; 30: 1733–1741.

16. Neaton JD, Gray G, Zuckerman BD, et al. Key issues in end point selection for heart failure trials: composite end points. *J Cardiac Fail*ure 2005; 11: 567–575.

17. Unger, EF. Weighing benefits and risks — the FDA's review of prasugrel. *N Engl J Med* 2009; 361: 942–955.

18. Beasley, B. Nhi, Unger EF and Temple R. Anticoagulant options — why the FDA approved a higher but not a lower dose of dabigatran. *N Engl J Med* 2011; 364: 1788–1790.

19. Sprigg N, Selby J, Fox L, et al. Very low quality of life after acute stroke: data from the Efficacy of Nitric Oxide in Stroke trial. *Stroke* 2013; 44:3458–3462.

20. Doernberg SB, Tran TT, Tong SYC, et al. Good studies evaluate the disease while great studies evaluate the patient: Development and application of a DOOR endpoint for *Staphylococcus aureus* bloodstream infection. *Clin Infect Dis* 2018 Oct 12. doi: 10.1093/cid/ciy766. [Epub ahead of print].

21. Miyahara S, Ramchandani R, Kim S, et al. Applying risk-benefit analysis to outcomes in TB clinical trials. *Clin Infect Dis* Accepted for publication.

22. Van Duin D, Lok J, Earley M, et al. SR. Colistin vs. ceftazidime-avibactam in the treatment of infections due to carbapenem-resistant enterobacteriaceae. *Clin Infect Dis* 2018; 66: 163–171.

23. Montepiedra G, Yuen CM, Rich ML, et al. Totality of outcomes: A different paradigm in assessing interventions for treatment of tuberculosis. *J Clin Tuber Other Mycobact Dis* 2016; 4: 9–13.

24. Lodise Jr TP, Rosenkranz SL, Finnemeyer M, et al. The Emperor's new clothes: prospective observational evaluation of the association between initial vancomycin exposure and failure rates among adult hospitalized patients with MRSA bloodstream infections (PROVIDE). *Clin Infect Dis* 2019; Jun 3. pii: ciz460. doi: 10.1093/cid/ciz460. [Epub ahead of print]

25. Peron J, Roy P, Ding K, et al. Assessing the benefit–risk of new treatments using generalised pairwise comparisons: the case of erlotinib in pancreatic cancer. *Br J Cancer* 2015; 112: 971–976.

26. Evans SR, Bigelow R, Chuang-Stein C, et al. Presenting risks and benefits: helping the Data Monitoring Committee to do its job. *Ann Intern Med* 2019; Nov 19. Doi: 10.7326/M19-1491. [Epub ahead of print].

Table 1. Example of Desirability-of-Outcome Rankings (DOOR).

DOOR Category	Partial Credit Scoring
Survival without adverse event (most desirable)	100%
Survival with adverse event	Partial credit
Death (least desirable)	0%

Table 2. Ordinal outcome analyses for efficacy and safety in the SOCRATES trial.

Analysis	Category	Ticagrelor (N = 6589) n (%)	Aspirin (N = 6610) n (%)	OR (95% CI)[a]	p-value
Efficacy Stroke[b,c]				0.862 (0.747–0.995)	0.0419[d]
	Stroke mRS 6[e]	33 (0.5)	26 (0.4)		
	Stroke mRS 4–5	65 (1.0)	86 (1.3)		
	Stroke mRS 2–3	144 (2.2)	152 (2.3)		
	Stroke mRS 0–1	131 (2.0)	167 (2.5)		
	No event	6199 (94.3)	6160 (93.5)		
Bleeding (safety)[f]	PLATO major Fatal	9 (0.1)	4 (0.1)		<0.0001[g]
	PLATO major Life-threatening	13 (0.2)	23 (0.3)		
	PLATO major Other	9 (0.1)	11 (0.2)		
	PLATO minor	75 (1.1)	44 (0.7)		
	PLATO minimal	499 (7.6)	234 (3.6)		
	No event	5944 (90.8)	6265 (95.2)		

[a]Odds ratio analyses were not performed for safety given a violation of the proportional odds assumption; [b]Full analysis set; [c]Patients with stroke with a missing mRS were excluded; [d]Wald test; [e]mRS = 6 includes deaths due to reasons unrelated to stroke; [f]Safety analysis set including events on or after the date of first dose up to 7 days after the date of last dose; [g]Wilcoxon test.

CI: confidence interval; mRS: modified Rankin Scale; OR: odds ratio; PLATO: Platelet Inhibition and Patient Outcomes study.

Table 3. DOOR and component outcomes by treatment.

	Ticagrelor (N = 6589) n (%)	Aspirin (N = 6610) n (%)	Cumulative difference % (95% CI)
DOOR[a,b]			
Survived with no event	6124 (92.9)	6089 (92.1)	0.8 (–0.1, 1.7)
Survived with non-disabling stroke, MI or PLATO major bleeding, 1 event	147 (2.2)	171 (2.6)	0.5 (–0.3, 1.2)
Survived with non-disabling stroke, MI or PLATO major bleeding, >1 event	6 (0.1)	11 (0.2)	0.4 (–0.3, 1.1)
Survived with disabling stroke	244 (3.7)	281 (4.3)	–0.2 (–0.5, 0.2)
Death	68 (1.0)	58 (0.9)	
Component Outcomes			
MI	25 (0.4)	21 (0.3)	
PLATO major bleeding	45 (0.7)	44 (0.7)	
Non-disabling stroke	113 (1.7)	143 (2.2)	

Disabling stroke	277 (4.2)	307 (4.6)

[a]Patients with stroke but missing mRS were characterized as disabling strokes; [b]Intention-to-treat analysis set.

CI: confidence interval; DOOR = desirability of outcome ranking; mRS: modified Rankin Scale; PLATO: Platelet Inhibition and Patient Outcomes study.

Table 4. Partial credit analyses summary.

Partial Credit Non-disabling Events	Partial Credit Disabling Events	Figure 1A &1B Depiction	Difference in means and 95% CI (Ticagrelor – aspirin)	Ratio and 95% CI (Ticagrelor/aspirin)
0	0	Point A	0.8248 (-0.0723, 1.7219)	1.0090 (0.9992,1.0188)
100	0	Point B	0.3934 (-0.3455, 1.1324)	1.0041 (0.9963,1.0120)
100	100	Point C	-0.1546 (-0.4864, 0.1773)	0.9984 (0.9951,1.0018)
50	25	Point D	0.4721 (-0.1713, 1.1155)	1.0050 (0.9982,1.0118)
68	38	Point E	0.3232 (-0.2232, 0.8697)	1.0034 (0.9977,1.0091)

CI: confidence interval

Table 5. Subgroup analyses from the SOCRATES trial.

	Subgroup	Win ratio 95% CI	DOOR probability[a] 95% CI	Partial credit non-disabling events (n = 68) and disabling stroke (n = 38) Ticagrelor:aspirin ratio and 95% CI
Prior aspirin[b]	No	1.061 (0.906,1.245)	0.502 (0.497,0.507)	0.9998 (0.9929, 1.0066)
	Yes	1.221 (0.989,1.510)	0.507 (0.500,0.515)	1.011 (1.0008, 1.0216)
Time to loading dose[c]	<6h	1.295 (0.835,2.038)	0.508 (0.494,0.522)	1.0107 (0.9931, 1.0287)
	6–<12h	1.296 (1.013,1.671)	0.510 (0.501,0.520)	1.0109 (0.9988, 1.0231)
	12–<18h	1.049 (0.795,1.399)	0.502 (0.491,0.512)	0.9988 (0.9854, 1.0124)
	18h+	0.996 (0.822,1.217)	0.500 (0.494,0.506)	0.9993 (0.9913, 1.0074)

[a]A win ratio greater than one and a DOOR probability greater than 0.5 favors ticagrelor

[b]Treatment x prior aspirin interaction test p-value = 0.337

[c]Treatment x time to loading dose interaction test p-value = 0.096

CI: confidence interval; DOOR = desirability of outcome ranking

Figure 1. A) Contours of the difference (ticagrelor – aspirin) in mean partial credit scores as a function of the partial credit provided for survival with a non-disabling event (y-axis) and survival with a disabling event (x-axis). A difference of zero indicates equivalence.

B) Contours of the ratio (ticagrelor/aspirin) in partial credit scores as a function of the partial credit provided for survival with a non-disabling event (y-axis) and survival with a disabling event (x-axis). A ratio of 1 represents equivalence.