Published in final edited form as:

J Thromb Haemost. 2013 August ; 11(8): 1443–1448. doi:10.1111/jth.12324.

Balancing Risk and Benefit in Venous Thromboembolism Trials:

Concept for a Bivariate Endpoint Trial Design and Analytic Approach

John M. Kittelson¹, Alex C. Spyropoulos², Jonathan L. Halperin³, Craig M. Kessler⁴, Sam Schulman⁵, Gabriel Steg⁶, Alexander G. G. Turpie⁵, Neal R. Cutler⁷, William R. Hiatt⁸, Neil A. Goldenberg⁹, and for the Antithrombotic Trials Leadership and Steering (ATLAS) Group

Neil A. Goldenberg: Neil.Goldenberg@allkids.org

¹Department of Biostatistics and Informatics, University of Colorado, Aurora, CO, USA

²Department of Medicine, Division of Hematology, Hofstra, North Shore/LIJ School of Medicine, Lenox Hill Hospital, NY, NY, USA

³Department of Medicine, Division of Cardiology, Mount Sinai School of Medicine, New York, NY, USA

⁴Department of Medicine and Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Washington, D.C., USA

⁵Department of Medicine, Division of Hematology and Thromboembolism, McMaster University, Hamilton, ON, Canada

⁶Université Paris VII–Denis Diderot, Paris, France

⁷Worldwide Clinical Trials, Beverly Hills, CA, USA

⁸Department of Medicine, Division of Cardiology, University of Colorado School of Medicine—Anschutz Medical Campus and CPC Clinical Research, Aurora, CO

⁹Departments of Pediatrics and Medicine, Divisions of Hematology, Johns Hopkins University School of Medicine and All Children's Hospital-Johns Hopkins Medicine, Baltimore, MD and St. Petersburg, FL, USA

Abstract

Antithrombotic trials in venous thromboembolism treatment and prevention, including those evaluating the new oral anticoagulants, have typically evaluated thromboembolism risk as an efficacy endpoint and bleeding risk as a separate safety endpoint. Findings often occur in opposition (i.e., decreased thromboembolism accompanied by increased bleeding, or vice-versa), leading to variable interpretation of the results, which may ultimately be judged as equivocal. In this paper, we offer an alternative to traditional designs based on the concept of a bivariate primary endpoint that accounts for simultaneous effects on antithrombotic efficacy and bleeding harm. We suggest a bivariate endpoint as a general approach to the assessment of “net clinical benefit” in recently published trials and to the design of future trials. Lastly, we illustrate the bivariate endpoint design using two examples: a recently published superiority trial of rivaroxaban

(RECORD1), and an ongoing non-inferiority trial of the duration of anticoagulant therapy in children with venous thrombosis (Kids-DOTT).

Keywords

Anticoagulants; Venous thromboembolism; Treatment Efficacy; Safety

Rationale for Venous Thromboembolism Trials of New Oral Anticoagulants

Low molecular weight heparin and warfarin are standard approaches to anticoagulation therapy, although the former requires parenteral administration and the latter requires regular coagulation monitoring and dose adjustment. These challenges may lead to suboptimal dosing with attendant risks of bleeding or thromboembolism (TE). New oral anticoagulants that specifically target thrombin or factor Xa offer promise as alternatives to standard anticoagulants for treatment of venous TE (VTE), as well as primary prophylaxis in patients at risk of incident TE, such as adults who have atrial fibrillation or are undergoing joint replacement surgery. Randomized controlled clinical trials (RCTs) of these new anticoagulants have been designed to establish either non-inferior or superior efficacy relative to standard care, based on the *expectation* of an improved risk-to-benefit ratio, superior bleeding profiles and more predictable pharmacodynamics that obviate the need for routine coagulation monitoring. Whether “net clinical benefit” has been convincingly demonstrated is subject to debate, in part due to limitations of the traditional univariate endpoint design.

Univariate Endpoint Design and Limitations

The design of antithrombotic trials is challenging given that typical efficacy and safety endpoints (TE and bleeding, respectively) have a clear mechanistic relationship modulated by treatment and an obvious potential for trade-offs relative to conventional treatment: i.e., an investigational strategy might reduce TE but increase bleeding or, conversely, reduce bleeding at the expense of increased TE. Although one would desire that a new strategy reduces both TE and bleeding risk, this is rarely observed to be the case. Nevertheless, a new strategy may still be deemed an improvement upon the standard of care when it meaningfully reduces *either* TE or bleeding risk, as long as a reduction in the one is not outweighed by an unacceptable increase in the other. In practice, all antithrombotic trial designs and treatment decisions must consider simultaneous effects on TE and bleeding risks. Typically, these endpoints have been considered separately, even though clinically and scientifically the simultaneous (bivariate) balance of risks is most relevant.

Consider three recent VTE trials involving dabigatran (RENOVATE [1]) and rivaroxaban (EINSTEIN-PE [2] and RECORD1 [3]) using a traditional univariate design. RECORD 1 evaluated the efficacy of rivaroxaban and RENOVATE evaluated the efficacy of two different doses of dabigatran for thromboprophylaxis in patients undergoing hip arthroplasty. EINSTEIN-PE evaluated the efficacy of rivaroxaban for pulmonary embolism treatment. Although bleeding and VTE risks were analyzed separately in each trial, the tradeoff between risks and benefits of the investigational treatments were only qualitatively

discussed. Specifically, the univariate endpoint analysis comparing dabigatran 150 mg vs. warfarin in RENOVATE did not pre-specify the criteria by which to judge whether the trial was “positive” or “negative”, given the opposing impact on efficacy and safety endpoints (decreased TE but increased bleeding with dabigatran 150 mg relative to warfarin).

Bivariate Endpoint Design for a Simultaneous Safety-Efficacy Decision

In a bivariate design, the primary efficacy and safety endpoints are formally addressed simultaneously as a two-dimensional outcome (Figure 1a). As reflected in the lower left quadrant, the most desirable result in a trial of antithrombotic therapy achieves reductions in both bleeding and TE. A negative result (concluding inferiority of the investigational strategy) occurs when both bleeding and TE risks increase (upper right quadrant). The upper left and lower right quadrants represent equivocal results, in which the investigational strategy is beneficial with regard to one endpoint but harmful for the other; decision-making then requires assessment of the tradeoff between the risks of TE and bleeding. Such tradeoffs must consider the nature and severity of the events (e.g., whether subclinical or clinically overt). The tradeoff can be formally expressed as net clinical benefit by weighting the relative clinical importance of TE and bleeding events and comparing the average impacts of the alternative treatment strategies.

In three recent evaluations of warfarin for TE prevention in patients with atrial fibrillation [10–12], intracranial hemorrhage (ICH) was judged to have 1.5 times the impact of TE (ischemic stroke or systemic embolism); net clinical benefit was based on a TE-to-ICH tradeoff in which 2 bleeding events were “equal to” 3 TE events. In the bivariate endpoint design, a weighted-average composite endpoint is constructed in two dimensions by dividing the plane into halves delineating regions of benefit and harm (Figure 1a), which shows TE:bleed weighting scenarios of 1:1.5 [dashed line] and 1:1 [solid line]). Since the relative weights must reflect the specific endpoints of a particular trial, net clinical benefit assessments will vary across trials, and should also strongly consider patient preferences regarding the relative importance of the types of primary outcome events captured in the study. While these preferences differ among individual patients, it must be emphasized that the goal of the bivariate approach is not to better foster individualization of findings when compared to the traditional approach, but rather to promote more objective interpretation of trial findings of net clinical benefit in an overall study population sample, particularly in antithrombotic trials. Furthermore, traditional design principles of appropriate patient selection and stratification for known/hypothesized effect modifiers must still be considered (and are indeed accommodated) in the bivariate endpoint design.

The bivariate analyses of VTE and bleeding in RENOVATE, RECORD1, and EINSTEIN-PE [1–3] are plotted in Figure 1b. Among them, only the results of EINSTEIN-PE fall in the unequivocally favorable lower left quadrant (although the 95% confidence intervals cross into other quadrants, as is common in non-inferiority trials). Depending upon the clinical impact of the TE endpoint relative to bleeding, some elevation in bleeding risk may be acceptable if there were strong evidence of reduced TE risk (upper left quadrant), and *vice-versa* (lower right quadrant); the former scenario occurred in RECORD1. In RENOVATE, results with the 220 mg dabigatran arm also fell in the upper left quadrant, a reduction in TE

relative to warfarin was offset by increased bleeding (assuming equal weighting of these events, which should be pre-specified). Outcomes of the 150 mg dabigatran arm in RENOVATE fell in the lower right quadrant (inferiority), in which a minor reduction in bleeding relative to warfarin was accompanied by a larger increase in TE.

The bivariate model recognizes that, irrespective of TE-to-bleed weighting, there are limits to the amount of increase in risk of one endpoint that can be offset by even a substantial decrease in the other. The linear tradeoff in Figure 1a (solid line) allows a 4% increase in risk of bleeding as long as there is a corresponding 4% reduction in TE risk. Such a large increase in bleeding may not be acceptable if bleeding is defined as symptomatic and TE as a composite of symptomatic and asymptomatic events. Thus, a linear division of the bivariate space may not adequately reflect the clinical tradeoffs that determine the acceptability of a given treatment. To better represent these tradeoffs, we suggest a curved decision rule, as shown in Figure 2. Although beyond the scope of this review, it has been shown elsewhere that the bivariate endpoint approach with curved decision rules also offers potential gains in trial efficiency, with respect to power/sample size and interim analyses [13]. We now illustrate, using VTE trials, how such decision rules would be constructed for superiority and non-inferiority designs.

Illustrations of the Bivariate Endpoint Design for VTE Trials

Example 1: Bivariate superiority design based on RECORD1

RECORD1 [3] was an RCT that compared rivaroxaban to enoxaparin for thromboprophylaxis in approximately 3,000 patients undergoing hip arthroplasty. The trial was designed to test for both the non-inferiority and superiority of rivaroxaban, although we focus here on specifying a bivariate curve for assessment of superiority. The primary efficacy endpoint was a composite of deep vein thrombosis, nonfatal pulmonary embolism, or death from any cause 36 days after randomization (observed composite event rate 2.4%/year). The rate of major bleeding was <0.3%/year and of major plus clinically relevant non-major bleeding was approximately 6%/year.

The tradeoff between efficacy and safety in this trial can be formalized using the red decision curve in Figure 2a. Superiority is defined by results to the lower left of the curve, while lack of superiority is defined by results to the upper right. Superiority is established when there is a significant reduction in both VTE and bleeding with rivaroxaban compared to enoxaparin (point B). A conclusion of superiority of rivaroxaban is also warranted when there is a highly significant reduction in the risk of VTE in the circumstance of non-inferiority with respect to bleeding risk (point C), and when there is a highly significant reduction in bleeding in the context of non-inferiority with respect to TE risk (point A).

The decision curve in Figure 2a was selected to pass through the three aforementioned design points (A, B, C), and Table 1a shows the inferences derived at these points. If VTE events were more important than bleeding events, it might be reasonable to shift point A to the left, requiring superiority (as opposed to non-inferiority) in VTE risk reduction for prevention of VTE even when bleeding risk is significantly reduced. The reference points that determine the shape of the curve are defined during trial design, and reviewed by the

Steering Committee; for pivotal studies, these reference points should be considered during regulatory review of the protocol.

Example 2: Bivariate non-inferiority design in Kids-DOTT

The ongoing Kids-DOTT trial (NCT00687882) is an RCT investigating the duration of anticoagulation in pediatric VTE [7]. Children with first-episode, provoked venous thrombosis are randomized to short-duration (6 weeks) vs. conventional-duration (12 weeks) anticoagulation. The primary efficacy endpoint is recurrent VTE at 2 years; the primary safety endpoint is clinically-relevant (major and clinically-relevant non-major) bleeding.

The principal analysis in Kids-DOTT is based on establishing non-inferiority, since shorter therapy would be preferred if it were at least as effective as conventional-duration anticoagulation. As a univariate VTE design, the non-inferiority bound was chosen as a 5% absolute increase in risk of recurrent VTE (i.e.,; hence, non-inferiority would be decided if the upper limit of the 95% confidence interval excluded a 5% absolute increase in VTE recurrence risk), with an anticipated critical value (i.e., observed absolute difference) in VTE recurrence risk of 0%. This design was selected under the assumption that the shorter course of therapy would reduce the risk of bleeding. If a substantial reduction in bleeding is observed, a minor increase in observed VTE recurrence risk (and an upper bound of the 95% confidence interval of greater than 5% absolute risk increase) would in fact be acceptable.

The decision is best represented by a bivariate curve (Figure 2b) reflecting how the acceptable risk of VTE recurrence depends on the bleeding risk reduction (and vice-versa). As in the RECORD1 example, the Kids-DOTT decision curve was selected to pass through three reference points: point A represents an important reduction in bleeding with a slight increase in recurrent VTE, point B indicates a smaller reduction in bleeding risk with no effect on recurrent VTE, and point C denotes a slight increase in bleeding that is offset by a significant reduction in recurrent VTE. Observations at interim analyses and at the end of the study will be plotted on this figure, and non-inferiority will be assessed based on the decision regions in Figure 2b. Ultimately, the results will be reported as point estimates and 95% confidence intervals (Table 1b). Notably, in Kids-DOTT as in many other trials, a number of secondary and exploratory efficacy and safety endpoints are described. While the bivariate endpoint approach focuses on the net clinical benefit of the primary safety and efficacy outcomes, non-primary endpoints should still be reported among the trial results, in traditional univariate fashion.

Conclusions and Implications

Clinical applications of antithrombotic therapies are inherently bivariate, involving tradeoffs between the risks of bleeding and TE. We therefore advocate for broader use of a bivariate endpoint trial design and decision criteria, as described here. Incorporating a given weighting of TE:bleeding that must reflect the specific primary endpoints of a particular trial, the bivariate design then formalizes the tradeoff decision to recommend for or against a new antithrombotic strategy. In this way, a bivariate endpoint approach to balancing risk and benefit will foster uniformity and clinical transparency in the interpretation of results of future antithrombotic trials.

Acknowledgments

Disclosure of Conflict of Interest

J. M. Kittelson – Receives partial salary support from the University of Colorado through grants from the National Institutes of Health and the Centers for Disease Control. He has received honoraria for participation in Food and Drug Administration activities, and has received consulting fees for work on data and safety monitoring committees from Genentech and BioMarin pharmaceuticals.

A. C. Spyropoulos receives fees from sanofi-aventis, Boehringer Ingelheim, Bristol Myers-Squibb, Johnson & Johnson and Bayer Healthcare for consulting activities, from Astellas Pharma for data and safety monitoring committee activities, from Eisai and Bayer for data and safety monitoring Committee and steering committee activities and receives honoraria from CPC Clinical Research, a non-profit ARO affiliated with the University of Colorado.

J. L. Halperin receives consulting fees for advisory and/or steering committee activities from Bayer Healthcare, Biotronik, Boehringer Ingelheim, Bristol Myers-Squibb, Daiichi Sankyo, Johnson & Johnson, Pfizer and sanofi-aventis, honoraria from AstraZeneca for data and safety monitoring board activities, research support from the National Institutes of Health, National Heart, Lung, and Blood Institute, and honoraria from CPC Clinical Research. He is a member of the Cardiovascular and Renal Drugs Advisory Committee of the U.S. Food and Drug Administration.

C. M. Kessler received consulting fees for advisory, data and safety monitoring board and/or steering committee activities from Baxter Immuno, Bayer Healthcare, CSL Behring, Eisai, NovoNordisk, Octapharma, Pfizer and sanofi-aventis, and honoraria from CPC Clinical Research. Georgetown University receives research support on his behalf from the National Institutes of Health, the Maternal and Child Health Bureau and the Centers for Disease Control and Prevention, as well as from Amgen, Baxter Immuno, Eisai, Genentec, GlaxoSmithKline, Grifols, NovoNordisk, Octapharma and sanofi-aventis.

S. Schulman receives fees from Merck, Bayer Healthcare and Boehringer Ingelheim for adjudication committee, data and safety monitoring committee and steering committee activities and honoraria from CPC Clinical Research.

Philippe Gabriel Steg receives fees from Amarin, Astrazeneca, Bayer, BristolMyersSquibb, Boehringer-Ingelheim, Daiichi-Sankyo, GlaxoSmithKline, Merck, Novartis, Otsuka, Pfizer, Roche, Sanofi, Servier, The Medicines Company and Vivus for steering committees, data monitoring committees, event committees and consulting activities. He receives honoraria from CPC Clinical Research. His institution receives research grants from Sanofi and Servier;

Neal R. Cutler. is President and CEO of Worldwide Clinical Trials, Inc., a contract research organization.

William R. Hiatt receives support from grants provided by the National Institutes of Health and from the pharmaceutical industry for sponsored research initiatives, partial salary support through research grants provided to CPC Clinical Research and the University of Colorado, as well as fees from the U.S. Food and Drug Administration as a Special Government Employee for several advisory committees. He provides consulting services to the pharmaceutical industry only through CPC Clinical Research. Current relationships include GlaxoSmithKline, TheraVasc, AstraZeneca, and Pluristem.

Neil A. Goldenberg receives partial salary support from grants from the National Institutes of Health and research support from Eisai, Inc., All Children's Hospital Foundation, and CPC Clinical Research.

References

1. Eriksson BI, Dahl OE, Rosencher N, Kurth AA, van Dijk CN, Frostick SP, Prins MH, Hettiarachchi R, Hantel S, Schnee J, Büller HR. RE-NOVATE Study Group. Dabigatran etexilate versus enoxaparin for prevention of venous thrombo-embolism after total hip replacement: a randomised, double-blind, non-inferiority trial. *Lancet*. 2007; 370:949–956. [PubMed: 17869635]
2. Büller HR, Prins MH, Lensin AW, Decousus H, Jacobson BF, Minar E, Chlumsky J, Verhamme P, Wells P, Agnelli G, Cohen A, Berkowitz SD, Bounameaux H, Davidson BL, Misselwitz F, Gallus AS, Raskob GE, Schellong S, Segers A. EINSTEIN-PE Investigators. Oral rivaroxaban for the treatment of symptomatic pulmonary embolism. *N Engl J Med*. 2012; 366:1287–1297. [PubMed: 22449293]

3. Eriksson BI, Borris LC, Friedman RJ, Haas S, Huisman MV, Kakkar AK, Bandel TJ, Beckmann H, Muehlhofer E, Misselwitz F, Geerts W. RECORD1 Study Group. Rivaroxaban versus enoxaparin for thromboprophylaxis after hip arthroplasty. *N Engl J Med*. 2008; 358:2765–2775. [PubMed: 18579811]
4. McClure LA, Brown MB. A likelihood approach to analyzing clinical trial data when treatments favor different outcomes. *Contemp Clin Trials*. 2006; 27:340–352. [PubMed: 16426899]
5. Kosorok MR, Yuanjun S, DeMets DL. Design and analysis of group sequential clinical trials with multiple primary endpoints. *Biometrics*. 2004; 60:134–145. [PubMed: 15032783]
6. Zhao L, Hu XJ, Lagakos SW. Statistical monitoring of clinical trials with multivariate response and/or multiple arms: a flexible approach. *Biostatistics*. 2009; 10:310–323. [PubMed: 19015160]
7. Goldenberg NA, Tripputi M, Crowther M, Abshire TC, DiMichele D, Manco-Johnson MJ, Hiatt WR. The "parallel-cohort RCT": Novel design aspects and application in the Kids-DOTT trial of pediatric venous thromboembolism. *Contemp Clin Trials*. 2010; 31:131–133. [PubMed: 19941974]
8. Hung HMJ, Wang SJ, Tsong Y, Lawrence J, O'Neil RT. Some fundamental issues with non-inferiority testing in active controlled trials. *Stat Med*. 2003; 22:213–225. [PubMed: 12520558]
9. Goodman SN, Berlin JA. The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. *Ann Intern Med*. 1994; 121:200–206. [PubMed: 8017747]
10. Singer DE, Chang Y, Fang MC, Borowsky LH, Pomernacki NK, Udaltsova N, Go AS. The net clinical benefit of warfarin anticoagulation in atrial fibrillation. *Ann Intern Med*. 2009; 151:297–305. [PubMed: 19721017]
11. Friberg L, Rosenqvist M, Lip GY. Net clinical benefit of warfarin in patients with atrial fibrillation: a report from the Swedish atrial fibrillation cohort study. *Circulation*. 2012; 125:2298–2307. [PubMed: 22514252]
12. Connolly SJ, Eikelboom JW, Ng J, Hirsh J, Yusuf S, Pogue J, de Caterina R, Hohnloser S, Hart RG. ACTIVE (Atrial Fibrillation Clopidogrel Trial with Irbesartan for Prevention of Vascular Events) Steering Committee and Investigators. Net clinical benefit of adding clopidogrel to aspirin therapy in patients with atrial fibrillation for whom vitamin K antagonists are unsuitable. *Ann Intern Med*. 2011; 155:579–586. [PubMed: 22041946]
13. Tripputi, M. PhD Dissertation. George Washington University; 2009. Use of Mediation in Designing Clinical Trials with two Primary Endpoints.

Figure 1.

(a) Decision-making with two endpoints: Decision-making in bivariate space is straightforward if a new anticoagulation therapy results in reduction of *both* bleeding and VTE risk or if it increases risk of *both* events (southwest or northeast quadrants). Decision-making is difficult if results are equivocal (northwest or southeast quadrants). **(b) Bivariate results from recent VTE trials of new oral anticoagulants:** Primary VTE and bleeding results from recent oral anticoagulation trials plotted as a bivariate outcome, calculated from data provided in the references. Axes denote the estimated difference in risk for VTE and bleed (experimental therapy minus control; i.e., negative values indicate reduced risk with experimental). Although new anticoagulation treatments are always evaluated on both bleeding and VTE endpoints, the results are usually presented separately and the tradeoff is left to the readership.

Figure 2.

(a) Superiority trial based on RECORD1: Bivariate decision rule maps the possible 2-dimensional outcomes into decisions in favor of superiority (blue shaded) or against superiority (red shaded). Points A,B,C used to define curve: A represents superiority on bleeding and non-inferiority on VTE, B represents significant reductions on both endpoints, and C represents superiority on VTE and non-inferiority on bleeding. **(b) Non-inferiority trial based on Kids-DOTT:** Bivariate decision rule maps the possible 2-dimensional outcomes into decisions in favor of non-inferiority (blue shaded) or against non-inferiority (red shaded). Points A,B,C define the curve: A represents highly significant benefit on bleeding with slight inferiority on VTE, B represents no effect on bleeding with a statistically significant benefit on bleeding, and C highly significant benefits on VTE with slight harm on bleeding.

Table 1

Inference on the decision boundary at 3 reference points.

Hypothetical Result ^I	VTE risk difference		Bleed risk difference	
	Observed	95% CI	Observed	95% CI
(a) Superiority decision criteria (Figure 2a)				
A	-0.008	(-0.019, 0.003)	-0.044	(-0.061, -0.026)
B	-0.011	(-0.022, 0.000)	-0.017	(-0.034, 0.000)
C	-0.028	(-0.039, -0.017)	-0.004	(-0.021, 0.013)
(b) Non-inferiority decision criteria (Figure 2b)				
A	0.010	(-0.039, 0.059)	-0.120	(-0.169, -0.071)
B	0.000	(-0.049, 0.049)	-0.050	(-0.090, -0.001)
C	-0.050	(-0.009, -0.001)	0.040	(-0.009, 0.089)

^ILabel for points in Figure 2.