Thrombosis Research 131 (2013) 204–209

Contents lists available at SciVerse ScienceDirect

Thrombosis Research

journal homepage: www.elsevier.com/locate/thromres

Regular Article

Adjudication of bleeding outcomes in an international thromboprophylaxis trial in critical illness

Donald M. Arnold [a,b], Francois Lauzier [c,d], Christian Rabbat [a], Nicole Zytaruk [e], Bronwyn Barlow Cash [e], France Clarke [e], Diane Heels-Ansdell [d], Gordon Guyatt [a,e], Stephen D. Walter [e], Andrew Davies [f], Deborah J. Cook [a,e,*] and For the PROTECT Investigators, on behalf of the Canadian Critical Care Trials Group and the Australian and New Zealand Intensive Care Society Clinical Trials Group

[a] Department of Medicine, McMaster University, Hamilton, Ontario, Canada
[b] Canadian Blood Services, Hamilton, Ontario, Canada
[c] Division of Critical Care, Medicine and Anesthesiology Departments, Université Laval, Québec, Québec, Canada
[d] Centre de recherche FRQS du Centre hospitalier affilié universitaire de Québec, Axe traumatologie – urgence – Soins Intensifs, Québec, Québec, Canada
[e] Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada
[f] Department of Intensive Care Medicine, The Alfred, Melbourne, Australia

ARTICLE INFO

Article history:
Received 11 October 2012
Received in revised form 26 November 2012
Accepted 7 December 2012
Available online 12 January 2013

Keywords:
Adjudication
Anticoagulation
Bleeding
Clinical trial
Critical care

ABSTRACT

Introduction: Measuring bleeding in critical care trials is challenging. We determined the reliability of adjudicated bleeding assessments in a large thromboprophylaxis trial in the intensive care unit (ICU).
Materials and Methods: PROphylaxis for ThromboEmbolism in Critical Care Trial (PROTECT) was an international randomized controlled trial that compared dalteparin to unfractionated heparin for the prevention of deep vein thrombosis in the ICU. Daily bleeding data were collected prospectively using a validated tool. Bleeds were adjudicated in duplicate by 2 of 4 members comprising a central adjudication committee. Bleeds were stratified by severity and study drug, then randomly assigned to adjudicator pairs. Adjudicators were blinded to treatment allocation, study centre and peer-assessments. We calculated agreement on bleeding severity and examined the effect of adjudication on overall trial results.
Results: In PROTECT, 491 patients had bleeding events including 208 with major bleeding and 283 with minor bleeding only. Of 491 patients, 446 were adjudicated in duplicate: 182 with major, 250 with minor and 14 with no bleeding. After adjudication, 52 of 244 bleeds were downgraded to minor; whereas only 15 of 244 were upgraded to major. Overall agreement among adjudicators was excellent (crude agreement = 86.3%; kappa = 0.76). Hazard ratios for major or any bleeding with dalteparin or unfractionated heparin were similar when analyzed using non-adjudicated events.
Conclusions: Major bleeds were sometimes over-called by research coordinators in a large ICU thromboprohylaxis trial. Adjudicator agreement was excellent. Central adjudication allowed reliable bleeding assessment and enhanced the rigor and validity of this major safety outcome.

Introduction

The measurement of bleeding outcomes in clinical trials is prone to random and systematic error. Even with objective definitions of major and minor events, the rigorous assessment of bleeding is complex because of difficulty quantifying the volume of blood loss, and the need for population-specific criteria for major bleeding. Measuring bleeding in critically ill patients is particularly challenging because bleeding is a common occurrence in the intensive care unit (ICU), is often due to invasive procedures, and frequently occurs at multiple anatomical sites [1].

The PROphylaxis for ThromboEmbolism in Critical Care Trial (PROTECT; NCT00182143) [2] was an international randomized, double-blind and concealed trial that compared the low molecular weight heparin dalteparin to unfractionated heparin (UFH) for thromboprophylaxis in 3,764 medical-surgical ICU patients. The primary outcome was proximal leg deep vein thrombosis (DVT) diagnosed with twice weekly leg ultrasounds. A secondary endpoint was bleeding, which was the primary safety outcome. Other secondary endpoints

Abbreviations: ICU, intensive care unit; PROTECT trial, The PROphylaxis for ThromboEmbolism in Critical Care Trial; UFH, unfractionated heparin; DVT, deep vein thrombosis; APACHE, Acute Physiology and Chronic Health Evaluation.

* Corresponding author at: Room D176 St Joseph's Healthcare, 50 Charlton Avenue East, Hamilton, Ontario, Canada.
E-mail address: debcook@mcmaster.ca (D.J. Cook).

0049-3848/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.thromres.2012.12.005

were DVT at other sites (e.g., upper limb, distal leg DVT), pulmonary embolism, and any venous thromboembolism. The methods [3] and results [2] have been previously described. The main findings of the PROTECT trial were that rates of DVT and bleeding were not different between groups, but use of dalteparin was associated with a significantly lower rate of pulmonary embolism.

To accurately capture both the number and severity of bleeding episodes in PROTECT, every reported bleeding event was independently adjudicated in duplicate by a central adjudication committee. The objective of this study was to describe the methods, results, and outcomes of the adjudication process for bleeding outcomes in PROTECT.

Materials and methods

Measurement of bleeding in PROTECT

We previously developed HEME (HEmorrhage MEasurement in the ICU), a comprehensive validated bleeding measurement tool that is highly reliable and specific for medical-surgical patients who are admitted to the ICU [1]. Using the HEME tool, all bleeding events in PROTECT were prospectively measured by dedicated trained research coordinators at each participating site. Bleeding severity was established by physiologic and anatomic parameters and the need for therapeutic interventions. Bleeding was defined as *major* if, in the absence of another cause, it fulfilled any of the 5 following criteria: 1) life threatening bleeding with hypovolemic shock; 2) life threatening bleeding into a critical site (e.g. intracranial or pericardial); 3) other critical bleeding (e.g., epidural, intraocular or intra-articular); 4) bleeding requiring an invasive procedure (e.g., surgery, embolization); or 5) clinically important bleeding requiring transfusion of 2 or more units of packed red blood cells, or associated with a decrease in systolic blood pressure of at least 20 mm Hg or an increase in heart rate of at least 20 beats/minute. Bleeding that did not satisfy these criteria was defined as *minor*. All bleeds were graded locally by research coordinators. Bleeding severity was reported per patient based on the most severe episode.

Calibration exercise

Bleeding events were assessed centrally by members of the bleed adjudication committee, which consisted of 2 intensivists, a hematologist, and a nephrologist who were all experienced in ICU bleeding adjudication research [1,4,5]. All adjudicators initially participated in a calibration exercise to reduce the variability in assessments among raters. Independently and blinded to study drug and to each others' assessment, all 4 adjudicators examined the charts of 20 patients with major bleeding and 20 patients with minor bleeding as determined by local research coordinators. After half of the charts were reviewed, adjudicators discussed their assessments as a group, identified reasons for disagreements and clarified criteria for bleeding categories. Thereafter, the remaining 20 charts were reviewed independently and a second round of discussion was held. We determined *a priori* that a threshold level of excellent agreement among all 4 raters, which we defined as kappa = 0.8 or higher, would be required in the calibration exercise before proceeding with duplicate adjudication for the remainder of the trial.

Bleeding adjudication process

After the initial calibration exercise, the severity of each bleeding event was adjudicated independently by the principal investigator and by one of the 3 other adjudicators as randomly selected. The randomization procedure was concealed and stratified by bleed severity (major or minor) and study drug (dalteparin or UFH) (Fig. 1). All adjudicators were blinded to treatment allocation, participating center and each other's assessment. Disagreements between pairs of raters were resolved by discussion and consensus.

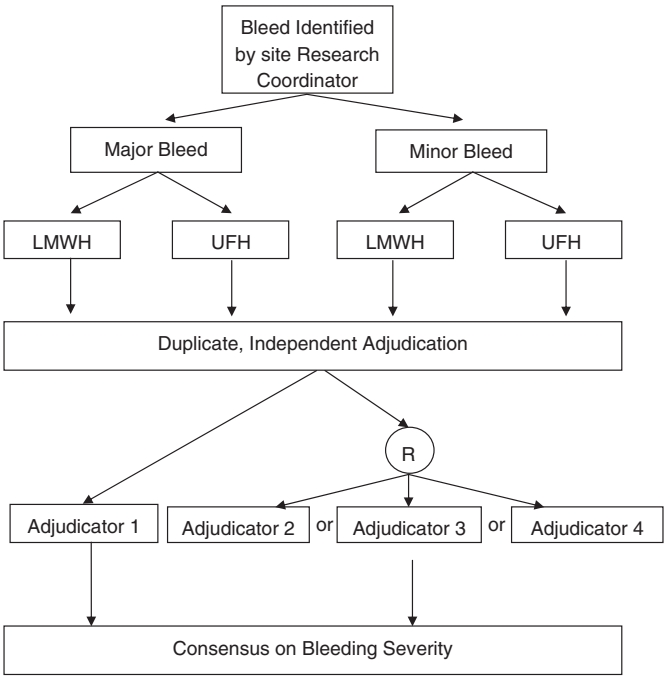

Fig. 1. Algorithm used to adjudicate bleeding event in the PROTECT trial using a 4-member bleeding adjudication committee. Each bleeding event was independently adjudicated by the principal investigator (Adjudicator 1) and 1 of 3 other adjudicators (Adjudicators 2–4) as randomly assigned, stratified by study drug and by bleed severity determined by local research coordinators. LMWH = low molecular weight heparin, UFH = unfractionated heparin.

Source data

Accessing information on a password protected secure website, adjudicators made independent assessments about bleeding severity (Appendix 1) based on data collected by the site research coordinators, including: Acute Physiology and Chronic Health Evaluation (APACHE) II score [6], admitting diagnosis, baseline demographics, daily hematology and coagulation laboratory data and medication use. The anatomical site, severity, start and end date and time, the need for interventions (including blood product transfusions) and vital status was evaluated for each bleed. Additional source documentation was reviewed by adjudicators including physicians' and nurses' notes, results of relevant diagnostic tests (e.g., endoscopy) and procedures (e.g., surgery).

Statistical analysis

Crude agreement and chance-corrected agreement (kappa) was calculated among all 4 adjudicators during the calibration exercise (n = 40 patients). For the remainder of the trial, adjudicators assessed each bleed as major, minor or no bleed; and overall bleeding severity was reported per patient based on the most severe bleed. We report crude agreement and chance-corrected agreement using weighted kappa with quadratic weights for 3-category agreement (major, minor or no bleed) and kappa for 2-category (major or minor) agreement.

To estimate the effect of adjudication, we re-analyzed the hazard ratio and 95% confidence intervals (CI) of the principal safety outcomes in the PROTECT trial: time to major bleeding and time to total bleeding using un-adjudicated events reported by site research coordinators. In the initial pilot phase of PROTECT [7], there were 21 patients whose bleeding severity was not assessed by research coordinators and thus could not be included in this analysis. We performed a best and worst-case scenario sensitivity analysis to evaluate whether the use of