Methodological issues in trials assessing primary prophylaxis of venous thrombo-embolism

Gabriel Thabut1,2*, Candice Estellat1, Isabelle Boutron1,2, Charles Marc Samama3, and Philippe Ravaud1,2

1 Département d’Épidémiologie, Biostatistique et Recherche Clinique, Groupe hospitalier Bichat Claude Bernard (Assistance Publique–Hôpitaux de Paris), Faculté Xavier Bichat, Université Paris VII, Cedex 18, 46 rue Henri Huchard, 75877 Paris, France; 2 INSERM U738, Paris, France; and 3 Département d’Anesthésie-Réanimation, INSERM U348, Hôpital Avicenne-125, route de Stalingrad, 93009 Bobigny Cedex, France

Received 14 June 2005; revised 31 August 2005; accepted 22 September 2005; online publish-ahead-of-print 13 October 2005

Aims Many trials have been conducted to assess the efficacy of various strategies in the prevention of venous thrombo-embolism (VTE). Some of these trials have been subject to methodological criticisms. We aimed to assess the methodological issues raised by VTE trials.

Methods and results We searched MEDLINE and the Cochrane Central Register of Controlled Trials for articles assessing primary thromboprophylaxis published between 1994 and 2003 in 60 general medical and specialty journals. A total of 77 articles were analysed by two independent reviewers using a list of items. No primary endpoint was defined in 20% of trials. Although the primary endpoint was collected before day 15 in 75% of trials, there were ≥20% missing data in 56% of articles and ≥30% in 24.2% of articles. The rate of missing data was 23.7 ± 9.7% in studies using venography-detected deep-vein thrombosis as an endpoint compared with 5.6 ± 6.0% in studies using other endpoints. Among the 47 superiority trials, 27 (57.4%) reported an intention-to-treat (ITT) analysis, but only 10 (21.3%) reported an analysis that complied with this principle. These results were consistent when limiting the analysis to articles published in high-impact journal (impact factor more than 5).

Conclusion Recent randomized controlled trials assessing prophylactic regimens in VTE have important methodological limitations in terms of primary endpoints, missing data, and compliance with the ITT principle. These methodological shortcomings should be addressed when planning future trials.

KEYWORDS Thrombosis; Prevention; Systematic review

Introduction

Venous thrombo-embolism (VTE) is a common disorder with significant morbidity and mortality.1,2 This complex vascular disease results in two major clinical manifestations. The first and more common is deep-vein thrombosis (DVT), which usually develops in the deep veins of the calf and spreads upward. Pulmonary embolism, either symptomatic or not, occurs as a complication of proximal DVT, in ~40–50% of patients.3 Pulmonary embolism ranges from incidental, clinically unimportant thrombo-embolism to massive embolism with sudden death. Hospital-based studies conducted in the United States suggest that the incidence of VTE is approximately 1.5 per 1000 people annually, whereas pulmonary embolism is thought to cause 50 000 deaths every year.4

There is considerable interest in preventing VTE in at-risk patients, such as those undergoing surgery or requiring prolonged hospitalization. Prevention strategies include orally or subcutaneously administered anticoagulants and mechanical prophylactic devices.1,2 Many trials involving thousands of patients have been conducted to assess the efficacy of these various prophylactic strategies. Consensus guidelines based on these trials support the implementation of prophylactic measures in at-risk patients.1,2 However, although randomized controlled trials (RCTs) are widely accepted to be the most reliable method of determining the effectiveness of treatments, methodological issues may affect the reliability of such studies.

The goal of this study was to assess the methodological issues in published reports of RCTs that aimed to assess the efficacy of primary thromboprophylaxis regimens.

Methods

Search strategy

Journal selection

All reports of RCTs assessing primary prophylaxis regimens in VTE published between 1994 and 2003 were selected from the 10 journals with the highest impact factor (Journal Citation Reports, 2001) from each of the following categories: general and internal
medicine, respiratory medicine, cardiovascular medicine, orthopaedic medicine, surgery and haematology. A list of all journals screened is contained in Appendix 1.

We chose high-impact factor because it is a good predictor of high methodological quality of journal articles and because our goal was not to be exhaustive but to point out the methodological soundness in reports of VTE trials.

**Article selection**

We identified reports from both the Cochrane Central Register of Controlled Trials (2004, issue 1) and MEDLINE databases using a search strategy with the terms ‘venous thrombosis’ or ‘pulmonary embolism’, a limit to ‘clinical trials’, and publication date of 1994–2003. In addition, we hand-searched four journals: New England Journal of Medicine, Lancet, Annals of Internal Medicine, and JAMA. One of us (G.T.) assessed the titles and abstracts of electronically retrieved reports to identify relevant studies. Reports were included only if the study design was identified as an RCT, published as a full text article, and assessed primary prophylaxis regimens in VTE. Case series, uncontrolled studies, and articles published as abstract only, editorials, news, and correspondence sections were excluded. For duplicate publications, only the most complete report was selected.

**Data abstraction form**

We developed a checklist of items to target the methodological issues in reports of trials of VTE and abstracted general information, including year of publication, number of centres involved, and type of treatment assessed. The reviewers evaluated whether the data of the Consolidated Standards of Reporting Trials (CONSORT) diagram (i.e., flow of participants through each stage of the trial) were reported, either in a flow chart or in the text.

For each article, the reviewers assessed whether a primary endpoint was defined and whether it was a composite. In the case of a composite endpoint, the components of the endpoints were recorded.

We studied whether patients, care providers, and outcome assessors were blinded. For treatments requiring dosage adjustments, reviewers noted whether a sham-adapted dosage was used to blind patients and physicians. They also determined whether radiographic assessment included a centralized reading.

For each trial, according to the extractable data, we evaluated the analysis set performed. We noted whether the trial reported an intention-to-treat (ITT) analysis for the primary endpoint. An analysis was considered to be ITT if (i) all randomized patients were included in the analysis and (ii) all patients were kept in the treatment group to which they were assigned. An analysis was considered to be modified ITT if data from all randomized patients were analysed according to random allocation, except for patients who failed to take at least one dose of the trial medication or failed to satisfy major entry criteria (eligibility violations).

With either ITT or modified ITT analyses, the reviewers considered how missing data were handled, and they recorded the rate of missing data. Each study was thoroughly assessed for missing outcomes.

For each article, the reviewers assessed whether the sample size calculation was reported. For articles that reported a calculation, the \( \alpha \) and the \( \beta \)-risks were recorded, as was the hypothesized difference between groups.

Quality was determined by the use of the Jadad scale, ranging from 0 to 5, and the Delphi list, ranging from 0 to 9.\(^6^,\(^7\) For both scales, the higher the score, the better the quality.

**Abstraction of data**

All trials were assessed by two reviewers (G.T. and C.E.), who independently examined the selected articles in a computer-generated random sequence. As a rehearsal before the study, the reviewers each evaluated 10 articles, and then they discussed the interpretation of the different items. The reviewers were not blinded to the journal name or authors. Discrepancies between the reviewers in the assessment of the selected articles were resolved by consensus: for each discrepancy, reviewers re-read the article and came to an agreement. The data presented here result from this consensus.

**Statistical analysis**

Continuous variables are described by their mean and standard deviation values and compared with use of the Student’s \( t \)-test. Categorical variables are described by frequencies and percentages and compared with use of the \( \chi^2 \) test. The degree of agreement between the two reviewers involved use of the kappa coefficient for categorical variables. Inter-rater reliability was assessed by use of the intraclass correlation coefficient (ICC) for continuous variables. The Spearman’s \( \rho \) correlation coefficient was used to assess the correlation between continuous variables. All tests were two-sided with \( P < 0.05 \) indicating statistical significance. Tests were not corrected for multiple testing. All data analyses involved use of SAS version 8.2 (SAS Institute Inc., Cary, NC, USA).

**Results**

**Selected articles**

Of 352 articles describing RCTs of primary prophylactic regimens in the prevention of VTE identified between 1994 and 2003, 77 were selected for assessment (Figure 1). A total of 275 articles were rejected, mainly because they did not report a therapeutic intervention or were not randomized. Fifty articles (64.9%) were published in four journals: New England Journal of Medicine (11), Lancet (10), Thrombosis and Haemostasis (15), and Journal of Bone and Joint Surgery—American edition (14). Among the 77 reports, 58 (75.3%) assessed prevention of VTE in orthopaedic surgery [hip surgery (n = 32), hip or knee surgery (n = 8), knee surgery (n = 14), arthroscopic knee surgery (n = 2), and major trauma (n = 2)], whereas 10 concerned non-orthopaedic surgery [surgery for cancer (n = 3), neurosurgery (n = 3), abdominal surgery (n = 2), gynaecological surgery (n = 1), cardiac surgery (n = 1), and nine medicine [acute medical illness (n = 7), immobilization (n = 2)]. Fourteen articles (18.2%) assessed non-pharmacological treatments (e.g. mechanical compression stockings) and 63 (81.8%) pharmacological treatments. The pharmacological treatments tested were subcutaneous low-molecular-weight heparins (38), thrombin inhibitors (nine), factor Xa inhibitors (eight), vitamin K antagonists (three), aspirin (two), and unfractionated heparin (three).

**Reproducibility**

The inter-rater reliability was good for the assessment of blinding \([\kappa = 0.86; 95\% \text{ confidence interval (CI): } 0.76–0.94]\), sample size \([\kappa = 0.90; \text{ CI: } 0.87–0.95]\), primary endpoint \([\kappa = 0.89; \text{ CI: } 0.86–0.96]\), and analysis set \([\kappa = 0.89; \text{ CI: } 0.86–0.96]\). The inter-rater reliability for the assessment of the study quality was fair for the Jadad scale \([\text{ICC} = 0.73; \text{ CI: } 0.58–0.83]\) and the Delphi list \([\text{ICC} = 0.69; \text{ CI: } 0.55–0.83]\).

**Methodological issues**

In 54 reports (70.1%), the trials were multicentre, involving two to 156 centres. The CONSORT diagram was included in the text or in a flow chart in 51 of 65 reports (78.5%)
published since the CONSORT publication in 1996. All studies had a parallel-group design. Trials were defined as superiority trials in 47 (61.0%) reports, equivalence or non-inferiority trials in 11 (14.3%), and not determined in 19 (24.7%).

Primary endpoint
A primary endpoint was defined in 62 reports (80.5%) and was a composite endpoint in 41 (66.1%): two components in 33 reports (e.g. DVT detected by a mandatory venography and symptomatic pulmonary embolism) and three components in eight. In only two reports did the composite endpoint combine efficacy and safety criteria. The manner of VTE diagnosis in the 62 articles with a defined primary endpoint is summarized in Figure 1. In 57 reports (91.9%), the primary endpoint was DVT detected by a mandatory exam, performed in 75% of cases before day 15. The systematic exam was venography (49), Doppler (five), fibrinogen uptake (two), or magnetic resonance imaging (one).

Blinding
In 42 trials (54.5%), patients, care providers, data collectors, and outcome assessors were reported to be blinded to treatment. In seven of these, therapeutic dosage adjustments were required (anticoagulation monitoring). In all cases, a sham-adapted dosage was used to ensure blinding. In 27 trials (35.1%), only the outcome assessor was reported to be blinded, whereas the intervention was not masked in eight trials (10.4%). When radiographic assessment was an endpoint in multicentre studies (48), blinded centralized reading to assess the exams was reported for 29 (60.4%) trials.

Missing data
The rate of missing data was assessed from the 62 articles with a defined primary outcome. A total of 35 articles (56.5%) had >20% missing data and 15 (24.2%) had >30% missing data. The rate of missing data was related to the endpoint, with 23.7 ± 9.7% (49) missing data for trials investigating venography to detect DVT as an endpoint and 5.6 ± 6.0% missing with other primary endpoints. The rate of missing data was also related to the number of centres involved in the study (Figures 2 and 3).

Intention to treat
In 36 reports (46.8%), the analysis was explicitly described as ITT. This rate was 57.4% for superiority trials (47) and 45.5% for equivalence or non-inferiority trials (11). On the basis of extractable data, only six of the 47 superiority trials (12.8%) complied with the ITT principle (all randomized participants were included in the analysis and kept in their original group) and four (8.5%) adopted a modified ITT scheme, whereby patients who failed to satisfy major entry criteria (eligibility violations) and take at least one dose of the trial medication were excluded (Figure 4). Of the 10 trials (21.3%) belonging to the ITT and modified ITT sets, six had missing data for the primary outcome measure: in these trials, an implicit imputation was used by including patients with missing responses in the denominator but not in the numerator. Among the 47 superiority trials, six were dose-ranging trials (phase II) and 41 were phase III trials. ITT or modified ITT was reported in one of six phase II trials (16.7%) and in nine of 41 phase III trials (22.0%). For equivalence or non-inferiority trials, on the basis of extractable data, one of the 11 trials (9.1%) complied with the ITT principle.
Sample size justification

A sample size justification was reported for 57 trials (74.0%). In only 13 reports did the authors justify the hypothesis on which sample size determination was based. Patients lost to follow-up were taken into account in the sample size calculation in 32 reports (56.1%) only.

Quality of trials and journal effect

The mean (+SD) scores on the Jadad scale and the Delphi list were 3.2 ± 1.8 and 6.1 ± 1.8, respectively. We found a weak but statistically significant relationship between the date of publication and the quality of trials assessed by the Jadad and the Delphi score (Pearson’s correlation coefficient: \( r = 0.26 \) (\( P = 0.02 \)) and \( r = 0.32 \) (\( P = 0.005 \)), respectively). Quality of trials was slightly better for those reporting on thrombin inhibitors and anti-Xa inhibitors than for LMWH: 4 ± 1.1 vs. 3.4 ± 1.4 and 7.1 ± 1.3 vs. 6.3 ± 1.6 for the Jadad and the Delphi scores, respectively. However, these differences were not statistically significant (\( P = 0.12 \) and \( P = 0.08 \)). We analysed the ‘journal effect’ by comparing the characteristics of trials according to the impact factor of journals in which they were published. An impact factor of more than 5 (median impact factor of the 77 articles in this survey) was considered high and one below 5 was low. The results are summarized in Table 1. Articles published in high-impact journals had better-quality scores on the Jadad scale and the Delphi list (Figure 5), were more frequently multicentre, and included more patients than those published in low-impact journals. In addition, they more frequently included a double-blind design. However, we failed to detect a difference in primary endpoint, analysis set, and number of missing data between articles published in high- and low-impact journals.

Discussion

Our survey of articles reporting RCTs that investigated the primary prevention of VTE raised methodological issues. The main concerns relate to the definition of the primary endpoint and the handling of missing data.

The choice of primary endpoint is a critical issue in trial design because it underlies the determination of treatment effectiveness. Because regulatory authorities are often consulted by investigators at the time of protocol setting, variation in the choice of primary endpoint from study to study may reflect the position of regulatory authorities.

Two-thirds of reports used a composite outcome combining symptomatic events (symptomatic DVT, pulmonary embolism, and death) and venographically detected DVT. The use of composite outcome in RCTs is controversial. This is especially the case in VTE trials, because a
considerable imbalance exists between asymptomatic events, accounting for the vast majority of events, and symptomatic events. However, as symptomatic VTE is relatively rare and difficult to detect clinically in a reliable manner, the American College of Chest Physicians (ACCP)\(^2\) consensus statement and the European Health Authorities (EMEA)\(^9\) recommend the use of a composite outcome combining clinical events with asymptomatic DVT. In addition, in most cases, the components of composite outcome were reported alongside the results of the primary analysis as proposed by Freemantle \(^8\).

Few reports \((n = 5)\) considered symptomatic events only, i.e. symptomatic DVT, pulmonary embolism, and death. Although theoretically ideal, the use of symptomatic events only as a primary endpoint has some drawbacks. As the incidence of such events is low, the sample size required to prove efficacy needs to be large: an average of 6109 ± 2652 were enrolled in the five trials with...
symptomatic events as endpoints compared with 760 ± 674 in the 57 trials with asymptomatic events as endpoints. Moreover, the assessment of symptomatic DVT may prove difficult, especially after hip or knee surgery in which lower limb symptoms may not be very specific.

Similarly, the inclusion of venographically detected DVT in the composite endpoint for thromboprophylaxis trials has drawbacks. Thus, its use leads to an average of 23.7% of missing data as compared with 5.6% with other endpoints. Overall, >50% of the trials had >20% missing data (35 reports) for the primary outcome, and almost 25% of trials had >30% missing data. These findings are especially impressive given the short duration of follow-up (less than 15 days in most trials). Missing data may affect the power of trials and bias the estimation of the treatment effect. In the 35 studies with >20% missing data for the primary endpoint, 16 (45.7%) did not account for it in the sample size determination, which led to a lower than expected statistical power. Moreover, missing data were merely ignored in most of the studies, which used a complete-case approach and therefore did not strictly comply with the ITT principle. The ITT principle was strictly used in only 21.3% of the superiority trials. In addition, this complete-case approach, closer to the per-protocol than the ITT principle, may tend to overemphasize the estimate of efficacy of a given treatment in superiority trials and lead to bias unless missing data are randomly distributed. However, several arguments suggest that this was the case in most studies. First, the distribution of patients not assessed for primary efficacy was well balanced between the two groups in most reports in this survey. Secondly, the few articles (n = 4) comparing the baseline characteristics of patients with and without missing data failed to detect a difference. Thirdly, most missing values were due to technical failures and patient refusal, two situations unlikely to be related to the outcome. Nevertheless, as suggested by the International Conference on Harmonisation Guidelines, the trials should include sensitivity analyses in such situations, in order to fully assess the potential influence of missing responses. Such sensitivity analysis would be particularly useful when the rate of missing data is expected to be high. Only three studies assessed several allocation schemes of missing data (all positive or all negative outcome). No study assessed the potential impact of various allocations of cases with missing outcomes.

Another drawback of including venographically detected DVT in the composite endpoint for thromboprophylaxis trials is that the inter-observer agreement of venograms is known to be low, especially with distal events. Large-scale studies have reported kappa values ranging from 0.40 to 0.75. However, the implementation of guidelines for performing venography and centralized reading may improve inter-observer agreement.

Lastly, venographically detected DVT is only a surrogate endpoint of mortality or morbidity due to venous thrombosis and its clinical relevance is still debated. Several meta-analyses of thromboprophylactic trials have demonstrated a relationship between reduction in asymptomatic DVT and reduction in VTE, regardless of the type of thromboprophylaxis or the thrombotic risk level. This relationship was also found in single studies. However, these findings have been challenged by the results of a recent meta-analysis and a large-scale RCT, where a significantly lower rate of venographically detected DVT in one group contrasted with a significantly higher rate of symptomatic events in the same group.

The quality of blinding reported in the majority of articles must be emphasized. Many efforts have been made to ensure blinding in the complex situations where dosage adjustments are required. In this case, sham-adapted dosages were used. Moreover, when radiographic assessment was an endpoint in multicentre studies, blinded centralized reading was implemented in most cases.

**Limitations of the study**

We restricted our search to articles published in high-impact journals, but trial reports published in less prestigious journals are usually less rigorous in reporting end-points and outcomes.
Methodological issues in trials assessing primary prophylaxis of VTE

233

journals are likely to have the same or lower quality methodology. We assessed only reports of RCTs, not the trials themselves. However, failure to report the methods of a trial does not necessarily mean that the investigators did not carry out these methods. Consequently, such methodological deficiencies may be related to the reporting of the trials rather than to their conduct.

Conclusions

Although many guidelines and meta-analyses have been published on this subject, this study is the first to highlight the methodological issues raised by trials assessing primary prophylaxis of VTE. Most reports had methodological limitations in defined analysis set, primary endpoint, and missing data, whatever the impact factor of the journal in which they were published. Whether these limitations, largely inherent in thromboprophylaxis trials, alter the soundness of their conclusions remains to be determined.

Conflict of interest: none declared.

Appendix 1

List of journals included in the search

- 10 highest impact journals for general and internal medicine: New England Journal of Medicine, JAMA, The Lancet, Annals of Internal Medicine, Annual Review of Medicine, Archives of Internal Medicine, BMJ, American Journal of Medicine, Medicine, Proceedings of the Association of American Physicians.

Appendix 2

List of the articles assessed in this study


Downloaded from https://academic.oup.com/eurheartj/article-abstract/27/2/227/553685 by guest on 09 February 2019
Methodological issues in trials assessing primary prophylaxis of VTE


References


