ELSEVIER

Journal of Clinical Epidemiology 66 (2013) 415–425

Journal of Clinical Epidemiology

Patient-level compared with study-level meta-analyses demonstrate consistency of D-dimer as predictor of venous thromboembolic recurrences

Maura Marcucci[a,*], Catrin Tudur Smith[b], James D. Douketis[c], Alberto Tosetto[d], Trevor Baglin[e], Mary Cushman[f], Sabine Eichinger[g], Gualtiero Palareti[h], Daniela Poli[i], Robert Campbell Tait[j], Alfonso Iorio[a]

[a]*Department of Clinical Epidemiology and Biostatistics and Department of Medicine, McMaster University, 1280 Main Street West, Hamilton, ON, L8S 4K1, Canada*

[b]*Department of Biostatistics, University of Liverpool, Brownlow Street, Liverpool, L69 3GS, UK*

[c]*Department of Medicine, McMaster University, St. Joseph's Healthcare Hamilton, 50 Charlton Ave. East, Hamilton, Ontario, Canada*

[d]*Department of Hematology, S. Bortolo Hospital, via Rodolfi 34, 36100 Vicenza, Italy*

[e]*Department of Haematology, Addenbrookes Hospital, Cambridge University Hospitals NHS Trust, Cambridge CB2 0QQ, UK*

[f]*Department of Medicine, University of Vermont, 208 South Park Dr, Colchester, VT 05446, USA*

[g]*Department of Medicine I, Medical University of Vienna, Vienna, Austria*

[h]*Department of Angiology and Blood Coagulation, S. Orsola-Malpighi University Hospital, Via Albertoni 15, 40138 Bologna, Italy*

[i]*Thrombosis Centre, Departement Heart and Vessels, Azienda Ospedaliero-Universitaria Careggi, v.le Morgagni 85, 50134 Firenze, Italy*

[j]*Department of Haematology, Royal Infirmary, Castle Street, Glasgow, UK*

Accepted 20 August 2012; Published online 8 February 2013

Abstract

Objective: We compared the performance of aggregate data (AD)–based and individual patient data (IPD)–based meta-analyses to synthesize evidence on the ability of D-dimer to distinguish recurrence risk in patients with unprovoked venous thromboembolism (VTE) who stopped anticoagulation.

Study Design and Setting: We compared the results of the published AD-based rate ratio of VTE recurrence for positive vs. negative D-dimer, estimated by a mixed-effect Poisson model, with those of the IPD-based hazard ratio obtained by a Cox regression stratified by trial. We performed three additional analyses to investigate the methodological reasons for differences between the two approaches, comparing the IPD Cox regression with AD generated from IPD Poisson regression (to control for differences in population on study), AD time-to-event meta-analysis, and AD generated from IPD meta-regression.

Results: Published analyses agreed in direction and statistical significance when estimating the prognostic value of D-dimer even if IPD estimates suggested a stronger effect. The additional analyses suggested that differences in study populations might explain this slight difference. Poor reporting in published studies precluded a true comparison of AD- and IPD-based assessments of heterogeneity sources.

Conclusion: AD and IPD meta-analyses yielded similar estimates of D-dimer effect to distinguish risk for recurrent VTE. The IPD approach was justified by the need to investigate sources of heterogeneity.

Keywords: D-dimer; Meta-analysis; Individual patient data; Time to event; Prognosis; Sources of heterogeneity

1. Introduction

Meta-analysis is a recognized statistical tool that pools findings from studies addressing the same clinical question [1]. Classic meta-analysis synthesizes aggregate data (AD) reported in published studies to assess effect sizes or other variables of interest. This method, referred to as AD meta-analysis, is easy to perform and suitable for assessing several clinical questions even if it can become quite complicated when addressing, for example, diagnostic questions.

An alternative approach to pool data consists of collecting original patient data from relevant studies and is referred to as individual patient data or individual participant data (IPD) meta-analysis [2]. This method has several

Conflict of interest/financial disclosure: We state that the submitted material has not been published and is not under consideration for publication elsewhere. We have no conflicts of interest or sources of funding to disclose.

* Corresponding author. Tel.: +1 905 525 9140 ext 22421; fax: +1 905-526-8447.

E-mail address: marcucci.maura@gmail.com (M. Marcucci).

0895-4356/$ - see front matter
http://dx.doi.org/10.1016/j.jclinepi.2012.08.007

What is new?

- When applied to the same set of studies that investigated the utility of postanticoagulation D-dimer to distinguish recurrence risk after a first unprovoked venous thromboembolism, individual patient data (IPD) and aggregate data (AD) meta-analyses provided similar results, but the availability of IPD allowed an assessment of the prognostic utility of D-dimer across age, patient sex, time to D-dimer testing, and assay cut-points.
- Meta-analyses based on AD are often able to provide a reliable summary of the existing evidence around a clinical question but need to address limitations linked to data reporting, such as availability, thoroughness, comparability, and suitability for optimal statistics.
- Even if resource intensive, the IPD approach can be justified by the need to reliably investigate sources of heterogeneity.
- IPD collection and analysis should be encouraged among researchers.

advantages. First, it allows improved data checking and updating and standardization of the study populations (e.g., uniform inclusion/exclusion criteria across studies). Second, it enables standardized analysis of outcomes in patient subgroups according to the meta-analysis protocol [2], which can minimize outcome reporting bias [3]. Third, it is the preferred method to analyze survival data when using the (log) hazard ratio and its variance as it allows for both censoring and time-to-event analyses. The log hazard ratio can be calculated for an individual trial using IPD and has the benefit of allowing model assumptions to be fully explored. On the other hand, the log hazard ratio and its variance may be presented directly in a study publication or may be indirectly derived if the required summary data are published [4,5]. Finally, IPD meta-analysis can summarize information while investigating and accounting for potential across-study heterogeneity of results [6].

For this later aim, meta-regression is used as an AD meta-analytic approach to investigate study characteristics as sources of heterogeneity. When patient-level characteristics are of interest, meta-regression can be applied; however, in the absence of IPD, only aggregate summary values of the variable of interest can be used, leading to aggregation or ecological bias [7]. Indeed, depending on how patients are distributed across pooled studies, the relationship between the outcome and a variable observed by pooling study-level values might differ from the relationship observed by pooling data for every patient [8]. In addition, when we perform a meta-regression, the studies represent the units of analyses forming the sample size: it is usually a small sample size, which limits the number of covariates that can be included in the model without significantly reducing the power of the analysis. Thus, the capacity of meta-regression to represent the relationships among data existing at the individual patient level is not only limited but also may be misleading and potentially leading to incorrect conclusions. Otherwise, an IPD approach has greater statistical power than meta-regression to identify clinically moderate interactions [6].

The approach that is most suited to pooled analyses of studies evaluating prognostic markers is debatable. The accessibility to published data makes AD meta-analysis attractive, and methodological approaches are available to derive and summarize time-to-event AD [4,5]. On the other hand, methodological improvements cannot address limitations because of the quality of primary studies, in particular the inadequate reporting of statistics and results, which can occur in prognosis research [9], and variability in methods, which makes AD meta-analysis unable to adjust for potential confounders. Patient categorization according to different cut-points of the same quantitative prognostic marker represents a relevant limitation for a comparison of AD estimates. Overall, IPD meta-analysis can overcome many of these potential problems [10].

An AD meta-analysis assessed the prognostic value of postanticoagulation D-dimer to distinguish disease recurrence risk after stopping anticoagulant therapy in patients with unprovoked venous thromboembolism (VTE) [11]. D-dimer is a fibrin-specific degradation product that is reliably detected by specific immunoassays [12]. An increased D-dimer level indicates the activation of coagulation and in patients with a prior VTE who have stopped anticoagulation; it may reflect an ongoing prothrombotic state and increased risk for VTE recurrence. This meta-analysis reported that an elevated postanticoagulation D-dimer could distinguish patients at low and high risks for recurrent VTE. However, several issues remained were not addressed, including the prognostic utility of D-dimer according to the patient age, timing of postanticoagulation D-dimer testing, and effect of the D-dimer cut-point and assay used. Another groups of researchers aimed to address these unresolved issues by pooling the same studies in an IPD meta-analysis [13].

The authors of the IPD meta-analysis, in collaboration with a statistician with expertise in this field, aimed to compare the performance of IPD and AD meta-analyses when applied to the prognostic utility of postanticoagulation D-dimer. There were three objectives for this analysis: (1) to qualitatively compare the methods used and results shown in the two articles [11,13] (i.e., compare IPD and AD meta-analyses based on published data), (2) to investigate the methodological reasons for differences for the two meta-analytic approaches, and (3) to discuss similarities and differences with these approaches, focusing on the advantages of IPD meta-analysis in this setting.

2. Methods

2.1. Comparison of the two articles

We qualitatively compared the clinical utility of the estimates obtained in the two meta-analyses, identifying those available in both approaches and those unique to the IPD method. We did not plan formal statistical tests for the comparison of these different approaches.

2.1.1. Meta-analyses to be compared

Verhovsek et al. [11] performed a literature search for their AD meta-analysis until June 2008 and pooled randomized trials or prospective cohort studies that involved patients with symptomatic VTE who received standard anticoagulant therapy, had D-dimer testing after anticoagulation was stopped, and had clinical follow-up to document recurrent VTE. Seven studies satisfied these criteria [14−20]. Relevant patient and D-dimer assay characteristics, duration of anticoagulation, and timing of D-dimer testing were extracted from the published reports; when data were missing or not clearly reported, the primary authors were contacted for clarification [11]. The same studies were then used to perform an IPD meta-analysis [13]. The literature search for potentially eligible studies was extended until July 2010, but no additional studies were found. Details about data search methods, extraction of AD and IPD, and development of study-level and patient-level databases are provided in both articles [11,13].

2.1.2. Characteristics of studies and patients included in the AD and IPD meta-analyses

All the seven source studies, although slightly differing in patient eligibility criteria, included at least some patients with a first unprovoked VTE. The authors of both the AD and IPD analyses defined inclusion criteria to select a homogeneous population of patients.

(1) *Inclusion criteria and definitions common to AD and IPD meta-analyses.*

- Exclusive inclusion of patients with a first unprovoked VTE, defined as that occurring in the absence of an antecedent major risk factor such as surgery, trauma, or cancer. For the AD meta-analysis, data clarification was requested from two studies [18,20] to ensure that data extraction from the published articles was limited to patients with unprovoked VTE.
- First, recurrent VTE after stopping anticoagulation as outcome of interest. For the AD meta-analysis, the authors of one study [19] enrolling patients with more than one recurrence provided AD recalculated considering only the first recurrence. No patient received anticoagulant therapy during follow-up; patients from two randomized trials [16,19] who were allocated to resume anticoagulation after D-dimer testing were excluded from both AD and IPD meta-analyses.
- Definition for D-dimer status (positive vs. negative) as it originally was in the source studies, in which D-dimer was detected by a quantitative assay in six of the seven studies [14,15,17−20] (two studies used the same type of quantitative test and the others four different assays; two studies used a cut-point for a positive test of $\geq$250 ng/mL and the others of $\geq$500 ng/mL); in one study, D-dimer status was defined by a qualitative assay [16]. For the IPD meta-analysis, the source authors provided raw quantitative D-dimer data (in one study [16], raw data were available for most patients and according to four different quantitative assays). The Appendix at www.jclinepi.com provides details of D-dimer assays and cut-points used in the primary studies.

(2) *Inclusion criteria and definitions that differ in AD and IPD meta-analyses.*

- VTE related to estrogens exposure (oral contraceptive and hormone replacement therapy). Only estrogens-related VTE classified as unprovoked VTE in the original articles was included in the AD meta-analysis. In the IPD meta-analysis, the study authors agreed to consider estrogens-related VTE as unprovoked (with the exception of the women exposed to estrogen therapy plus another major clinical risk factor); then, using IPD analysis, each index VTE was reclassified according to this definition.
- Isolated distal deep vein thromboses. These patients were excluded only from the IPD meta-analysis.
- Timing for D-dimer testing. Whereas in the AD meta-analysis only studies with D-dimer testing done 3−8 weeks postanticoagulation were included (or the study authors [19] provided data only for patients in whom D-dimer was measured within this time interval), for the IPD meta-analysis, there were no limitations as to the timing of postanticoagulation D-dimer testing.

2.1.3. Comparison between published AD and IPD meta-analyses

The two meta-analyses summarize available information on the predictive utility of postanticoagulation D-dimer by providing an estimate of the risk ratio for recurrent VTE in patients with a positive vs. negative D-dimer. We compared the published results according to the method of analysis used in the relevant publication: (1) pooled rate ratio [95% confidence interval (CI) and *P*-value] for recurrent VTE for D-dimer positive vs. negative patients obtained in AD meta-analysis by an univariable Poisson regression model and (2) pooled hazard ratio (HR) (95% CI and

P-value) for recurrent VTE for D-dimer positive vs. negative patients obtained with IPD by an univariable Cox regression model. The two approaches retained the definition of positive/negative D-dimer status according to the assays used in the original studies. Table 1 shows the models built in the AD and IPD meta-analyses with the methods used for investigating across-study heterogeneity. In the IPD meta-analysis, a pooled HR adjusted for potential confounders was also provided, and two sensitivity analyses were performed: (1) two different Cox regression analyses were modeled using prespecified cut-points (250 and 500 ng/mL) for D-dimer positive or negative status and recoding IPD according to these cut-points and (2) quantitative D-dimer as a continuous variable (nanograms per milliliter) was modeled in a Cox regression analysis to obtain a trend of the risk for recurrent VTE according to each 1-U D-dimer increase.

Both AD and IPD meta-analysss present the summary estimates of the risk for recurrent VTE as annualized rates and cumulative hazards. We refer the reader to the original articles for further details [11,13].

2.2. Investigation of differences between the performance of AD and IPD meta-analyses

We performed three additional analyses by remodeling the IPD or published data to thoroughly explore the two meta-analytic approaches.

(1) AD generated from IPD Poisson regression vs. IPD Cox regression: comparison of performance of IPD and AD when analyzed using different statistical models after controlling for differences in population on study.

We compared two different statistical methods, AD Poisson regression and IPD Cox regression, after minimizing the different inclusion criteria of the two meta-analyses. We used IPD to generate AD for each study and then reanalyzed these data using the AD Poisson regression and compared them with the IPD Cox regression. We first used this set of AD and IPD to plot the effect size of each original study, as both rate ratio (obtained by a generalized linear model for Poisson distribution) and hazard ratio (Cox regression), to graphically show the consistency among studies. Measures of between-study heterogeneity were also reported. We then modeled a mixed-effect Poisson regression by the user-written GLLAMM command in STATA software (version 9.2; StataCorp, College Station, TX, USA). As in the AD meta-analysis, the model included a fixed effect for D-dimer status and random effect (assumed to be normally distributed) for study and used person time (obtained by summing the follow-up time of all the patients in each D-dimer group) as offset variable. Finally, we compared this AD from IPD generated Poisson regression with an IPD univariable Cox regression with fixed effect for D-dimer and random effect for study (shared frailty) to allow a comparable analysis.

Table 1. Statistical methods used in the two published meta-analyses to explore the effect of D-dimer on the risk of VTE

Meta-analyses	Treatment effect	Method	Heterogeneity assessment	Sources of heterogeneity exploration
AD	Pooled rate ratio	Poisson model (person time as offset variable)	A mixed-effect model with random effects (assumed to be normally distributed) for study and fixed effects for D-dimer were predefined. Between-study heterogeneity was tested separately for recurrence rates in positive and negative D-dimer subgroups; statistically significant heterogeneity was found only for the pooled rate of the positive D-dimer patients.	Not performed
IPD	Pooled hazard ratio	Cox regression model (proportional hazards assumption verified by the analysis of Schoenfeld residuals)	Both a Cox regression model stratified by study with fixed effects for D-dimer and Cox regression model with random effects for study (shared frailty option) and fixed effects for D-dimer were performed: a nonsignificant variance for the shared frailty model was found suggesting a fixed-effects assumption for the study variable as reasonable. A formal test for heterogeneity (between-study variation in D-dimer effects) was obtained by comparing the −2log(likelihood) of the model stratified by study with fixed effects for D-dimer with the −2log(likelihood) of the model stratified by study including the D-dimer by study interaction term: no statistical heterogeneity was found.	By a Cox regression model stratified by study with fixed effect for D-dimer including potential confounders (age, BMI, sex, with or without use of hormonal therapy, thrombophilia status, and timing of D-dimer postanticoagulation testing) and their by D-dimer interaction terms.

Abbreviations: VTE, venous thromboembolism; AD, aggregate data; IPD, individual patient data.

(2) AD time-to-event meta-analysis vs. IPD Cox regression: comparison of the performance of AD and IPD when pooled to evaluate the same end point.

To compare the performance of AD and IPD when used to evaluate the same end point, we attempted to meta-analyze the published AD in a survival data setting. We extracted from each available source article, the summary statistics needed to derive the natural logarithm of HR (lnHR) and its variance (V) according to the methods described by Parmar et al. [4] and Williamson et al. [5] and exemplified by Tierney et al. [21]. We pooled the lnHR and V estimates from each study using both a fixed (inverse variance method) and random (Der Simonian and Laird method) treatment effects meta-analysis. Only the studies that provided the data needed to derive the lnHR and V estimates for patients with a first unprovoked VTE were included in the AD time-to-event meta-analysis. We planned to compare the results of this AD additional analysis with

(a) published results of the IPD-based stratified Cox regression and
(b) IPD-based stratified Cox regression performed on the same set of studies included in the AD time-to-event meta-analysis.

(3) AD generated from IPD meta-regression vs. published IPD Cox regression: comparison of the performance of AD and IPD in the exploration of the sources of heterogeneity.

The authors of the AD meta-analysis did not investigate the effect of patient- and D-dimer test–related factors on prognostic value of D-dimer. With the aim to complete the comparison between the two meta-analytic approaches by investigating their potential to explore sources of heterogeneity and look for effect modifiers, we performed an AD meta-regression. As described previously, we chose to overcome the variability linked to incompleteness of data reporting, and we regenerated the dependent and independent aggregate variables from IPD (patients with a first unprovoked proximal VTE). We obtained the lnHR by an IPD univariable Cox regression for each study. D-dimer and other patient-level characteristics were included as ''study-level aggregate variables,'' which indicate for each study the proportion of patients with a given characteristic in case of covariates for binary status (e.g., for D-dimer status, the proportion of patients with positive D-dimer; for estrogen therapy, the proportion of users among female patients) or the mean value in case of continuous covariates (e.g., age and timing of D-dimer postanticoagulation testing). Using a similar approach to a previously published comparison of meta-regression with IPD or AD [6], the logHR of jth study ($\log HR_j$) was assumed to be independently normally distributed according to the following model

$$\log HR j \sim N(\alpha + \beta x_j, v_j),$$

where v_j is the variance of the logHR in the jth study, which corresponds to a fixed-effect meta-regression model. The model was extended to incorporate an additive between-study variance component τ^2 (random effect model). To enable a direct comparison between the two approaches, a D-dimer by covariate interaction term (which corresponds to what the meta-regression estimates) was added to each IPD Cox regression model. In addition, for the IPD analysis, sex and estrogen therapy covariates were explored in two separate models to compare the performance of respective AD covariates.

3. Results

3.1. Comparison of the two articles

The study selection process and the characteristics of the individual studies included in AD and IPD meta-analyses are shown in the Appendix at www.jclinepi.com.

3.1.1. Comparison of published AD and IPD meta-analyses

Table 2 compares the characteristics of patients included in the AD and IPD meta-analyses and the results of the summary estimate of the pooled rate/HR. Although the same studies were included, the number of patients and patient-years analyzed in the two articles slightly differs based on different study-level and patient-level selection criteria and definitions. Where a comparison between AD and IPD was possible, both analyses agreed as to the estimate of effect for postanticoagulation D-dimer to distinguish recurrence risk. The IPD unadjusted estimate suggests a somewhat stronger effect of D-dimer. Only the availability of patient-level data allowed adjustment of the pooled HR for possible confounders; in the multivariable model, only patient sex and prior estrogen-associated VTE (coded together as a three-level categorical variable: one level for men, one level for women with estrogen-associated VTE, and one level for women without estrogen-associated VTE) has a significant effect on recurrence risk, without significantly modifying the effect of D-dimer (i.e., no significant interaction). The sensitivity analyses demonstrated that the choice of two different prespecified cut-points for D-dimer status does not affect its prognostic value and that every unit increase in D-dimer as a continuous variable confers a 0.053% increased risk for recurrent VTE.

3.2. Comparison of performance of AD and IPD meta-analyses

Table 3 provides the results of the additional analyses performed by repooling available patient level or published data to further explore differences between the two meta-analytic approaches. Controlling for differences in the patient populations in the two meta-analyses considerably

Table 2. Comparison of published AD and IPD meta-analyses

Meta-analysis characteristics	AD meta-analysis		IPD meta-analysis	
Trials, *n*	7		7	
Patients (positive/negative D-dimer), *n*	907/981[a]		826/992[a]	
Patients-years of follow-up (positive/negative D-dimer)	2,461.6/2,040.2		1,891.3/2,197.6	
Method used (heterogeneity management)	Poisson regression (random effects for study and fixed effects for D-dimer)		Cox regression (stratified by study and fixed effects for D-dimer)	
Pooled effect size[b]	Rate ratio (95% CI)[b]		Hazard ratio (95% CI)[b]	
D-dimer status definition	**Unadjusted**	**Adjusted**	**Unadjusted**	**Adjusted[c]**
D-dimer status according to source authors	2.2 (1.6, 2.9)	—	2.5 (1.9, 3.3)	2.6 (1.9, 3.5)
D-dimer status according to 500 ng/mL cutoff	—	—	2.1 (1.6, 2.8)[d]	2.5 (1.7, 3.8)[d]
D-dimer status according to 250 ng/mL cutoff	—	—	2.4 (1.6, 3.4)[d]	2.4 (1.5, 3.7)[d]
D-dimer as continuous variable (ng/mL)	—	—	1.0005 (1.0003, 1.0006)	1.0005 (1.0003, 1.0007)

Abbreviations: AD, aggregate data; IPD, individual patient data; CI, confidence interval; VTE, venous thromboembolism.

[a] Discrepancy in sample size is because of the slight differences in inclusion criteria definition (e.g., idiopathic vs. provoked index event and extension of index VTE) among source articles.

[b] Pooled effect size (rate or hazard ratio) >1 indicates higher recurrence risk/hazard for positive D-dimer patients.

[c] The models were adjusted for the following supposed confounders: age, sex, hormone therapy use, body mass index, D-dimer test timing, and inherited thrombophilia.

[d] Data not provided in published article in which only graphic results were shown.

minimized the difference between the AD and IPD effect sizes even if obtained by different methods (Poisson and Cox regression, respectively) (Table 3, part A). Fig. 1 plots the effect size of each single study as a rate ratio (Fig. 1A) and hazard ratio (Fig. 1B) and graphically confirms the consistency among studies. When analyzed as time-to-event data (Table 3, part B), the AD results from published studies are consistent with those obtained by IPD using data from all studies and data from the seven studies that reported AD variables needed to estimate (ln)HR and its variance [14–16,18,19]. Even when analyzing the same subset of trials (five trials), the total number of patients included in the AD and IPD analyses differed (we calculated sample size for the AD analysis as the total patients with unprovoked VTE reported in each original study).

Table 3, part C, shows the results of IPD-generated AD meta-regression and IPD Cox regression used to investigate the effect of prespecified covariates on the predictive value of D-dimer. Both for the AD meta-regression and IPD Cox regression, Table 3 shows the effect of covariates obtained by univariable fixed-effect models (for Cox regression, as an example, the models included the effect of D-dimer and the

Table 3. Comparison of performance of AD and IPD meta-analyses

Objectives	AD meta-analysis		IPD meta-analysis	
(A) Same data (seven trials and 1,818 patients)	Poisson regression (random effects for study and fixed effects for D-dimer), RR (95% CI)		Cox regression (random effects for study and fixed effects for D-dimer), HR (95% CI)	
	2.5 (1.9, 3.3)		2.4 (1.8, 3.1)	
(B) Same outcome (time to event): comparison of published AD to IPD meta-analysis	HR (95% CI)		HR (95% CI)[a]	
(i) Different set of trials	2.1 (1.6, 2.8)[b]		2.4 (1.8, 3.2)[b]	
(ii) Same set of trials (five trials and 1,847 AD patients and 1,553 IPD patients)	2.1 (1.6, 2.8)[b]		2.4 (1.7, 3.2)[b]	
(C) Source of heterogeneity investigation	Meta-regression (univariable and fixed effects)		Cox regression (univariable and stratified by trial)	
Covariate (*n* of trials, *n* of patients)	β for covariate (SE)	*P*-value	β for D-dimer × covariate (SE)	*P*-value
Age, yr (7, 1,818)	−0.0005 (0.0247)	0.984	−0.0062 (0.0081)	0.446
Sex, male (7, 1,818)	−0.0329 (0.0361)	0.405	−0.1342 (0.2986)	0.653
Prior hormone therapy use (7, 1,785)	0.0065 (0.0097)	0.528	−0.0555 (0.4990)	0.911
Body mass index (4, 832)	0.0135 (0.1122)	0.350	0.0305 (0.0306)	0.319
D-dimer test timing (6, 1,649)	0.0015 (0.0043)	0.750	0.0010 (0.0033)	0.749
Inherited thrombophilia (7, 1,626)	0.0037 (0.817)	0.817	0.0686 (0.3448)	0.842

Abbreviations: AD, aggregate data; IPD, individual patient data; RR, rate ratio; CI, confidence interval; β, regression coefficient; HR, hazard ratio; SE, standard error; VTE, venous thromboembolism.

[a] HR was calculated on patients with a first unprovoked VTE, without excluding distal VTE, to increase the comparability with the AD-derived outcome.

[b] No statistically significant between-study heterogeneity was found: reported results were obtained by fixed-effect models.

covariate of interest). As shown in the table, none of the IPD interaction terms or AD covariates were identified as statistically significant; AD and IPD estimates for β coefficients differed, even if very slightly in some cases (e.g., β for timing of D-dimer testing), but no systematic patterns could be identified. When meta-regression models were repeated after adding a parameter for between-study variance (τ^2) for each univariable model, τ^2 was estimated to be zero. Despite questionable reliability because of the small number of studies included, there was no unexplained heterogeneity even without any covariates. Multivariable meta-regression models including more than one covariate were not examined because of the lack of statistical power with only seven studies providing data (four in case of body mass index).

4. Discussion

IPD meta-analysis is considered the gold standard method for meta-analysis. However, the IPD approach is often more complex and requires additional resources compared with AD meta-analysis such that there is ongoing debate about the merits and drawbacks of each approach. We believe that we have enriched this debate by comparing AD and IPD meta-analysis for the same set of studies that investigated the utility of postanticoagulation D-dimer to stratify recurrence risk after a first unprovoked VTE. We first compared the published results of the two meta-analyses and then we rearranged the AD and IPD to allow a better comparison of the two approaches. Our objective was to explore how well these meta-analytic methods performed when applied to nonrandomized studies assessing prognosis. In fact, at variance with the several IPD vs. AD comparisons performed on data from randomized controlled trials, our work adds to the relatively few comparisons of meta-analyses of observational studies.

We now discuss which advantages of IPD meta-analysis in this example could have justified this methodological approach.

4.1. Estimate of overall prognostic effect

The performance of AD and IPD meta-analyses for questions of treatment and prognosis has been addressed, with a variable degree of agreement reported using the two approaches. One of the main reason for different

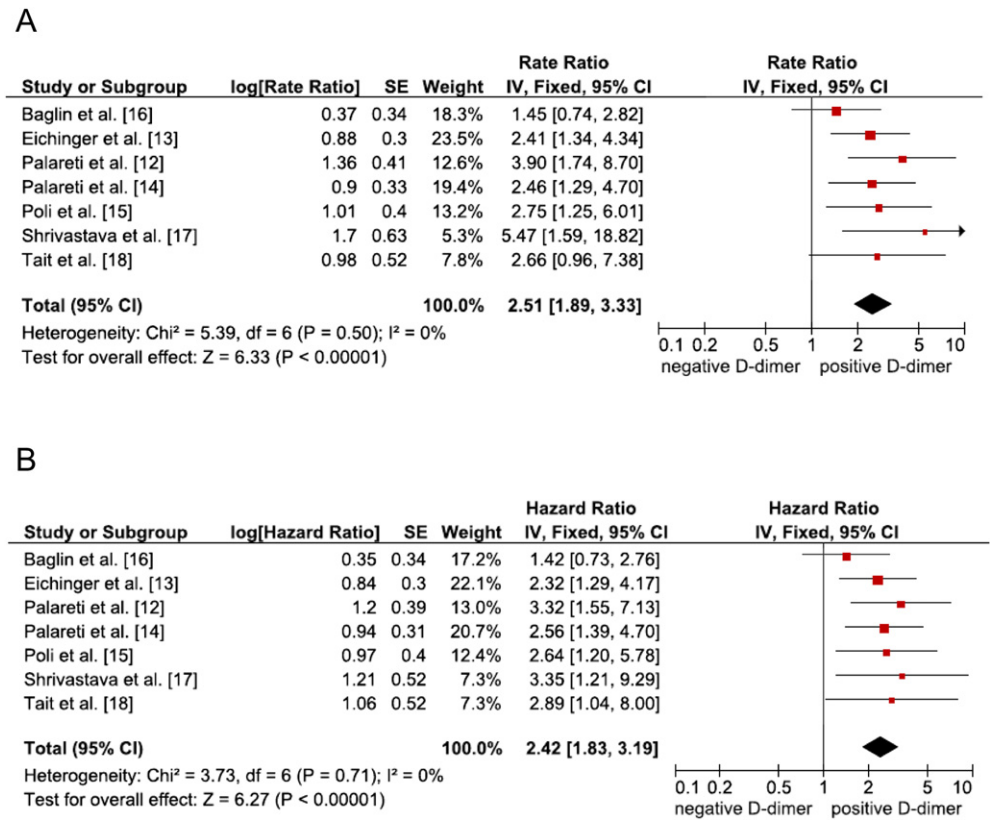

Fig. 1. (A) The figure shows the forest plot of effect size for each source study expressed as rate ratio and 95% CIs for recurrent VTE for positive vs. negative D-dimer. Estimates were obtained for each study by modeling IPD data for patients with a first unprovoked proximal VTE in a Poisson regression. (B) The figure shows the forest plot of effect size for each source study expressed as hazard ratio and 95% CIs for recurrent VTE for positive vs. negative D-dimer. Estimates were obtained for each study by modeling IPD data for patients with a first unprovoked proximal VTE in a Cox regression. Total (95% CI) effect size, for both rate ratio hazard ratio, was obtained by IV. CI, confidence interval; VTE, venous thromboembolism; IPD, individual patient data; SE, standard error; IV, inverse of variance method.

conclusions with either approach was the diversity in data sources (different studies and different patients) [22]. Some authors found that the AD and IPD analyses provided the same results when the same source studies were included, at least when considering the overall estimates of treatment or exposure effects [23]. In other studies, many of which focused on time-to-event data, the authors tried to reduce differences in populations meta-analyzed by an IPD or AD approach, but the results remained significantly different [24,25]. One such study [24] provided a possible explanation in that their meta-analysis was based on a heterogeneous set of trials showing different treatment effects, especially when assessed over time.

In our example, we found that the AD- and IPD-derived HRs for recurrent VTE with a positive vs. negative postanticoagulation D-dimer are consistent in their statistical significance, with a slightly stronger effect observed with the IPD analysis. The similarity in results may be because of the lack of heterogeneity among the source data and the similarity of patients and studies included in both analyses. Beyond the similarity in the patients studied, the low between-study heterogeneity also may be explained by the robustness of D-dimer to predict disease recurrence. When we tried to control for the differences in the source data, the AD-based effect estimate approached the IPD-based effect estimate. Creating identical data sets allowed us to compare the performance of the two statistical approaches used in the AD and IPD analyses. Cox regression is a standard approach for IPD-based survival analysis. Nevertheless, Poisson regression, suited to deal with count data (which follow a Poisson distribution), is widely used to employ a person-time approach, especially when the outcome of interest is rare. This is because it can incorporate follow-up duration in the model and allows for different lengths of patient follow-up. The first empiric demonstrations of good fitting of Poisson regression compared with Cox regression in a single study on time-to-event data dates back many years [26,27]. Successive Poisson models, including fixed or random effects, were proposed as alternative meta-analytic approaches to pool rates and ratios calculated in follow-up studies and were often applied for studies with recurrent events [28,29]. When applied to recurrent event scenarios, Poisson and Cox analyses theoretically would answer two different clinical questions: the first looks at number of recurrences over a period, whereas the latter looks at time to (first) recurrence. In our example, both AD and IPD meta-analyses pool source studies which had a first recurrent VTE event as the outcome. This is another reason for not expecting significant differences between the two approaches.

Results obtained by the AD Poisson regression were more similar to those obtained by the IPD Cox regression than those obtained by the additional AD time-to-event analysis. The difference between AD and IPD time-to-event analyses might be, again, partially explained by incomplete overlap in source data. Indeed, even if differences between the two populations meta-analyzed were minimized by recalculating the IPD Cox regression on the same set of studies for which an AD time-to-event analysis was possible and including IPD for patients with a distal index VTE event, source data could not be considered the same because of incomplete homogeneity in the definition of some inclusion criteria in the pooled studies (e.g., definition of unprovoked VTE and classification of estrogen-associated VTE). We conclude that, in this example, AD-based estimates are reliable compared with the gold standard IPD approach; otherwise, the differences among the study populations on which the published data were obtained, even if slight, make the population on which the results of the AD meta-analysis not clearly definable. Moreover, asking the source study authors for data not provided in published reports may be as demanding as requesting IPD.

4.2. *Heterogeneity assessment and management—sources of heterogeneity investigation*

The authors of the AD meta-analysis tested the between-study heterogeneity separately for the pooled recurrence

Table 4. General advantages and disadvantages of AD and IPD meta-analysis

Properties	AD meta-analysis	IPD meta-analysis
Advantages	• Easier to include data from all or nearly all inherent published and unpublished studies • Well-developed meta-analytic methodology	• Data checking and homogeneity of variable definitions • Consistency of the analyses across studies, especially for time-to-event data • Reliable study of the sources of heterogeneity • Reliable study of the covariate treatment interactions also for continuous and multiple covariates simultaneously • Data updating • Direct collaboration between trialists and reviewers • Scopes of using advanced modeling approaches
Disadvantages	• Limited data reporting about methods and results • Limited availability of suited statistics to pool • Need for additional or new data from primary authors or for compiling tables with detailed AD • Ecological or aggregate bias (meta-regression)	• Time and resource consuming • Difficulties to collect all individualized data from relevant published and unpublished studies • Need for strategies to deal with studies not providing IPD • Limited knowledge and skills with available modeling approaches

Abbreviations: AD, aggregate data; IPD, individual patient data.

rates in positive and negative D-dimer groups; as far as the rate ratio, they chose a priori to perform a Poisson mixed-effect model using random effects for study and fixed effects for D-dimer without a formal assessment of heterogeneity. Although for the first additional analysis (in the second part of this article) we performed Poisson and Cox models with a fixed effect for D-dimer and random effects for study, to allow the comparison with the published AD results, this is not the preferred approach: the inclusion of random effects for study is equivalent to assuming that the studies included are a random sample from a larger population of studies, and this is not consistent in the context of meta-analysis [30,31]. The authors of the AD meta-analysis did not investigate sources of heterogeneity. When we planned to compare AD and IPD in terms of ability to explore sources of heterogeneity, we realized that a true comparison was not possible because the original studies did not sufficiently contain the necessary covariate data. Such insufficient reporting is a common limitation for identifying sources of heterogeneity in AD meta-analyses [6]. Therefore, when we compared the IPD Cox regression with the IPD-generated AD meta-regression, the AD results were considered the ''best possible'' performance obtainable with the available data. With this in mind, in our example, IPD and AD meta-regressions agreed in terms of indicating that none of the covariates significantly modified the prognostic effect of D-dimer. However, there was variability in estimates of effect and standard errors between the two approaches. We cannot draw general conclusions from our example because we were in an uncommon meta-analytic context, in which source data were homogeneous; results elsewhere obtained from such a comparison [6] suggest that ''AD meta-regression can be accurate if there is evidence for a within-study treatment by covariate interaction and sufficient across-study variation for the aggregate value of the covariate. Departures from this condition could mean that meta-regression results using AD are unreliable.''

Our study has limitations. We were unable to compare the two approaches based on statistical test and measures. Additional studies are needed to better understand when the added value of an IPD analysis over AD analysis is justified. A related systematic review is in progress as a Cochrane Methodology Review [32]. In addition, in our case, the project of an IPD meta-analysis rose from

Table 5. Overview of general considerations to be made when planning to perform an individual patient data (IPD) meta-analysis of prognostic cohort data with binary or time-to-event outcome

General approaches to pool data	**One-stage analysis (i.e., combining IPD from all studies to perform a single analysis)**	**Two-stage analysis (i.e., conventional pooling of summary statistics obtained for each study from IPD)**
Advantage/suitable circumstances	• Widest choice of methods available • Can address any question in any circumstance	• Standard summary data meta-analysis techniques • Easier to perform and understand for most practitioners
Disadvantages/limits	• Cannot be simply analyzed as a ''mega-trial'' and requires strategies to account for pooling different studies, which can be complex • It strongly depends on statistician's background and skills	• Suboptimal usage of IPD data • Does not take full advantage of exploring variability in patient characteristics • Does not allow continuous covariates

Approaches to take into account heterogeneity across studies	**Fixed-effect model**	**By-study stratified model**	**Random effect model**
Advantage/suitable circumstances	• Straightforward in the two-stage approach when there is not heterogeneity across studies • Statistics easier to be performed and results easier to be interpreted	• The simplest way to take into account IPD origin from different studies in the one-stage approach when there is no heterogeneity	• Recommended for when there is heterogeneity across studies • Models exist for both the one- and two-stages approaches
Disadvantages/limits	• Absence of heterogeneity across studies is an uncommon (for someone an ''unreal'') circumstance which limits applicability	• It's a fixed-effect model (see limits) • Feasibility depends on the software used	• Feasibility depends on the software used and statistician's skills • Results not always easy to be interpreted

Regression modeling	**Logistic**	**Cox**	**Poisson**
Advantage/suitable circumstances	• Suitable when there is no interest for time-dependent events • Usually easier to be performed • Simpler assumptions	• Considered the gold standard for time-to-event analysis • Its use is ''per se,'' a good reason to pool IPD data in prognosis research	• Proper for count data and person-time approach • Can model time-to-event data with different lengths of patient follow-up • Best fitting with sparse data
Disadvantages/limits	• Time-to-event and time-varying covariates cannot be inherently accounted for	• Proportional hazard assumption to be tested • Proper data type and setting not always available	• Model parameter can be difficult to set • Incorporation of random effect (mixed models) might be complex • Overdispersion (consider using a ''special'' Poisson distribution—negative binomial)

the need to answer to clinical question remained unmet by the AD meta-analysis; even if this was a way to seek the best evidence, it may be considered a limit for performing an ''unbiased'' comparison between the two meta-analytic works.

5. Conclusions

In summary, the evidence we provided can be used to answer the following two questions. First, which advantages of IPD meta-analysis could have justified this methodological approach in this specific case? We confirmed that AD-based meta-analyses provide a reliable summary of the existing evidence, more so perhaps when the observed effect is consistent across studies, as in our present example. We believe that in the example provided herein, the IPD approach was primarily justified by the need to investigate consistency of the measured effect across the full range of possible values of potential modifiers not managed and not at the best manageable under the AD approach. Clinically, useful results were demonstrated: the prognostic utility of D-dimer across age, patient sex, time to test, and assay cut-point.

Second, which meta-analytic approach should be recommended to pool prognostic cohort data? In Table 4, we provided a usable brief summary of the possible IPD and AD meta-analyses' advantages and disadvantages. We recognize the difficulties of obtaining all the existing raw data on the topic of interest, but it might be easier for the primary authors to share them than to provide supplemental AD or compile tables with detailed AD [33]. A more interactive synergism between trialists and reviewers, a much deeper data checking and updating, the opportunity to explore the sources of variability among studies, and patients subgroups using continuous and time-dependent covariates and multiple covariates simultaneously [34,35]: these are the principle benefits that IPD methods can offer to the research on prognosis. Furthermore, pooling IPD for a meta-analysis will be the first step in making available those extended data sets that makes it possible to derive and validate prognostic clinical prediction guides that stratify patients according to individual risk for adverse outcomes. Moreover, we recognize that statistical skills required by IPD are not yet widely available [34], but heightened awareness of this technique will encourage further research in this field. We conclude by providing a summary Table 5 intended as a ''beginner's guide'' for planning an IPD meta-analysis, in particular when dealing with prognostic cohorts. Only some of the available models are shown, and many of the specific methodological topics the statistician should consider, such as the choice of classical or Bayesian approach to build hierarchical models or the opportunity to include a random effect for ''treatment,'' are not discussed. Overall, its aim is to encourage clinical researchers to undertake more collaborative studies assessing patient prognosis.

Acknowledgments

The authors thank Qilong Yi for having provided precious details on the statistical methods used in the aggregate data meta-analysis.

Appendix

Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jclinepi.2012.08.007.

References

[1] Thacker SB. Meta-analysis. A quantitative approach to research integration. JAMA 1988;259:1685−9.
[2] Sutton AJ, Higgins JPI. Recent developments in meta-analysis. Stat Med 2008;27:625−50.
[3] Hutton JL, Williamson PR. Bias in meta-analysis due to outcome variable selection within studies. Appl Stat 2000;49:359−70.
[4] Parmar MKB, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. Stat Med 1998;17:2815−34.
[5] Williamson PR, Smith CT, Hutton JL, Marson AG. Aggregate data meta-analysis with time-to-event outcomes. Stat Med 2002;21:3337−51.
[6] Smith CT, Williamson PR, Marson AG. An overview of methods and empirical comparison of aggregate data and individual patient data results for investigating heterogeneity in meta-analysis of time-to-event outcomes. J Eval Clin Pract 2005;11:468−78.
[7] Lambert PC, Sutton AJ, Abrams KR, Jones DR. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. J Clin Epidemiol 2002;55:86−94.
[8] Berlin JA, Santanna J, Schmid CH, Szczech LA, Feldman HI. Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. Stat Med 2002;21:371−87.
[9] Altman DG. Systematic reviews in health care—systematic reviews of evaluations of prognostic variables. BMJ 2001;323:224−8.
[10] Riley RD, Abrams KR, Sutton AJ, et al. Reporting of prognostic markers: current problems and development of guidelines for evidence-based practice in the future. Br J Cancer 2003;88:1191−8.
[11] Verhovsek M, Douketis JD, Yi QL, et al. Systematic review: D-dimer to predict recurrent disease after stopping anticoagulant therapy for unprovoked venous thromboembolism. Ann Intern Med 2008;149:481−W94.
[12] Hirsh J, Lee AY. How we diagnose and treat deep vein thrombosis. Blood 2002;99:3102−10.
[13] Douketis JD, Tosetto A, Marcucci M, et al. Patient-level meta-analysis: effect of measurement timing, threshold, and patients age on ability of D-dimer testing to assess recurrence risk after unprovoked venous thromboembolism. Ann Intern Med 2010;153:523−31.
[14] Palareti G, Legnani C, Cosmi B, et al. Predictive value of D-dimer test for recurrent venous thromboembolism after anticoagulation withdrawal in subjects with a previous idiopathic event and in carriers of congenital thrombophilia. Circulation 2003;108:313−8.
[15] Eichinger S, Minar E, Bialonczyk C, et al. D-dimer levels and risk of recurrent venous thromboembolism. JAMA 2003;290:1071−4.
[16] Palareti G, Cosmi B, Legnani C, et al. D-dimer testing to determine the duration of anticoagulation therapy. N Engl J Med 2006;355:1780−9.

[17] Poli D, Antonucci E, Ciuti G, Abbate R, Prisco D. Combination of D-dimer, F1+2 and residual vein obstruction as predictors of VTE recurrence in patients with first VTE episode after OAT withdrawal. J Thromb Haemost 2008;6:708−10.
[18] Baglin T, Palmer CR, Luddington R, Baglin C. Unprovoked recurrent venous thrombosis: prediction by D-dimer and clinical risk factors. J Thromb Haemost 2008;6:577−82.
[19] Shrivastava S, Ridker PM, Glynn RJ, et al. D-dimer, factor VIII coagulant activity, low-intensity warfarin and the risk of recurrent venous thromboembolism. J Thromb Haemost 2006;4:1208−14.
[20] Tait RC, Lowe GD, McColl MD, McMahon AD, Robertson L, King L. Predicting risk of recurrent venous thrombosis using a 5-point scoring system including fibrin D-dimer. J Thromb Haemost 2007;5(suppl 2):O-M-060.
[21] Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR. Practical methods for incorporating summary time-to-event data into meta-analysis. Trials 2007;8:16.
[22] Jeng GT, Scott JR, Burmeister LF. A comparison of meta-analytic results using literature vs individual patient data—paternal cell immunization for recurrent miscarriage. JAMA 1995;274:830−6.
[23] Olkin I, Sampson A. Comparison of meta-analysis versus analysis of variance of individual patient data. Biometrics 1998;54:317−22.
[24] Duchateau L, Pignon JP, Bijnens L, Bertin S, Bourhis J, Sylvester R. Individual patient-versus literature-based meta-analysis of survival data: time to event and event rate at a particular time can make a difference, an example based on head and neck cancer. Control Clin Trials 2001;22:538−47.
[25] Michiels S, Piedbois P, Burdett S, Syz N, Stewart L, Pignon JP. Meta-analysis when only the median survival times are known: a comparison with individual patient data results. Int J Technol Assess Health Care 2005;21:119−25.
[26] Pearce N, Checkoway H, Dement J. Exponential models for analyses of time-related factors, illustrated with asbestos textile worker mortality data. J Occup Med 1988;30:517−22.
[27] Callas PW, Pastides H, Hosmer DW. Empirical comparisons of proportional hazards, poisson, and logistic regression modeling of occupational cohort data. Am J Ind Med 1998;33:33−47.
[28] Glynn RJ, Buring JE. Ways of measuring rates of recurrent events. BMJ 1996;312:364−7.
[29] Sturmer T, Glynn RJ, Kliebsch U, Brenner H. Analytic strategies for recurrent events in epidemiologic studies: background and application to hospitalization risk in the elderly. J Clin Epidemiol 2000;53:57−64.
[30] Smith CT, Williamson PR, Marson AG. Investigating heterogeneity in an individual patient data meta-analysis of time to event outcomes. Stat Med 2005;24:1307−19.
[31] Higgins JP, Whitehead A, Turner RM, Omar RZ, Thompson SG. Meta-analysis of continuous outcome data from individual patients. Stat Med 2001;20:2219−41.
[32] Clarke MJ, Stewart L, Tierney J, Williamson PR. Individual patient data meta-analyses compared with meta-analyses based on aggregate data (Protocol). Cochrane Database Syst Rev 2007;(Issue 2). Art. No.: MR000007.
[33] Clarke M, Godwin J. Systematic reviews using individual patient data: a map for the minefields? Ann Oncol 1998;9:827−33.
[34] Simmonds MC, Higgins JP, Stewart LA, Tierney JF, Clarke MJ, Thompson SG. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. Clin Trials 2005;2:209−17.
[35] Hemingway H, Riley RD, Altman DG. Ten steps towards improving prognosis research. BMJ 2009;339:b4184.