Pooling of Individual Patient Data from Clinical Trials

Improvement of analyses of subgroups

Laura Koopman

Pooling of Individual Patient Data from Clinical Trials
Improvement of analyses of subgroups

Utrecht, Universiteit Utrecht, Faculteit Geneeskunde

Thesis University Utrecht, with a summary in Dutch
Proefschrift Universiteit Utrecht, met een samenvatting in het Nederlands

ISBN	978-90-393-47232
Author	Laura Koopman
Cover	Eveline van den Berg
Lay-out	Gildeprint BV, Enschede, the Netherlands
Print	Gildeprint BV, Enschede, the Netherlands

Pooling of Individual Patient Data from Clinical Trials

Improvement of analyses of subgroups

Poolen van Individuele Patiënten Data van Klinische Studies

Verbetering van analyse van subgroepen

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit van Utrecht op gezag van de rector magnificus, prof. dr. J.C. Stoof, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op

donderdag 31 januari 2008 des ochtends te 10.30 uur

door

Laura Koopman

geboren op 29 augustus 1980 te Rotterdam

Promotor:	Prof.dr. D.E. Grobbee
Co-promotoren:	Dr. M.M. Rovers Dr. G.J.M.G. van der Heijden

The studies presented in this thesis were financially supported by the Netherlands Organisation for Scientific Research (grant number 916.46.090).

Financial support by the Julius Center for Health Sciences and Primary Care and the Dutch ENT-society (KNO-vereniging) for the publication of this thesis is gratefully acknowledged. Additional financial support by Solvay Pharmaceuticals BV is also gratefully acknowledged.

Manuscripts based on the studies presented in this thesis

Chapter 2

Koopman L, van der Heijden GJMG, Glasziou PP, Grobbee DE, Rovers MM. A systematic review of analytical methods used to study subgroups in (individual patient data) meta-analyses. *J. Clin. Epidemiol.* 2007 Oct;60(10):1002-9.

Chapter 3

Koopman L, Rovers MM, Glasziou PP, Grobbee DE, van der Heijden GJMG. Reporting of subgroup effects in individual patient data meta-analyses is driven by the number of included patients, testing for heterogeneity, the use of a fixed effect model, and testing for interaction effects. *In revision J. Clin. Epidemiol.*

Chapter 4

Koopman L, van der Heijden GJMG, Glasziou PP, Grobbee DE, Rovers MM. A systematic review on the difference in subgroup effects between conventional and individual patient data meta-analyses.

Chapter 5

Koopman L, van der Heijden GJMG, Hoes AW, Glasziou PP, Grobbee DE, Rovers MM. An empirical comparison of subgroup effects of conventional and individual patient data meta-analyses.

Chapter 6

Koopman L, Hoes AW, Glasziou PP, Appelman CLM, Burke P, Damoiseaux RA, Little P, Le Saux N, McCormick DP, Rovers MM. Antibiotic therapy to prevent the development of asymptomatic middle ear effusion in children with acute otitis media. A meta-analysis of individual patient data. *Arch. Otolaryngol.* 2008 Feb.

Chapter 7

Koopman L, van der Heijden GJMG, Grobbee DE, Rovers MM. Comparison of methods of handling missing data in individual patient data meta-analyses: an empirical example on antibiotics in children with acute otitis media. *Am. J. Epidemiol.* 2008.

Contents

Chapter 1

General introduction

Evidence-based medicine is the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients [1]. Use and application of the principles of evidence-based medicine intends to improve the effectiveness and efficiency of health care. Systematic reviews and meta-analyses have become essential tools to keep up with accumulating new evidence.

Average treatment effect and the individual patient

Meta-analyses estimate the clinical effectiveness of healthcare interventions by combining results of similar studies, most often randomised controlled trials, quantitatively to produce a single and more precise overall estimate of the average effect [2,3]. The direction and magnitude of these average effects are intended to guide decisions regarding clinical practice for a wide range of patients. The physician is thus being asked to treat his/her patients as though they were all represented by the patients in the clinical trials included in the meta-analysis [3-5]. This runs against the intuitive approach physicians have, that is, most physicians would like to move beyond the average effects and use several factors of the patient (i.e., age, gender, genetic profile, life-style), disease (i.e., stage, severity, co-morbidities), and intervention characteristics (i.e., compliance, adjuvant care), and his/her own experience to decide on a patients' individual treatment [4-7]. Hence, information on effects of treatment in subgroups according to strata of patient and disease characteristics at baseline is required, and optimal application of trial results in clinical practice demands targeting interventions to subgroups of patients most likely to benefit [8-11]. Factors responsible for the occurrence of varying treatment effects in subgroups are called "effect modifiers" [12]. The sample size of a single trial is, however, rarely large enough to allow for reliable subgroup analyses [3], and it is hardly possible to perform a new trial for every potential subgroup.

Subgroup analyses and conventional meta-analyses

In conventional meta-analyses based on published data differences in treatment effects between groups of study participants can be assessed by relating outcome to some characteristic (of treatment or study participant) on a continuous or ordered scale by meta-regression analyses [3,11]. Several studies have shown that meta-regression, often used in conventional meta-analyses, has limitations in studying treatment effects in subgroups [11,13-17]. Meta-regression can be useful in investigating differences according to characteristics of study design. Still, the exploration of

patient-level characteristics is problematic because typically only summary values of the characteristic of interest can be obtained from study reports [14-16]. The use of such "mean covariate values" reduces power but more importantly can lead to ecological bias, that is, average baseline data for populations are handled as if they were individual baseline observations, thereby leading to biased associations and inter-correlations between variables that may limit the control of confounding [13,18]. Furthermore, stratified subgroup analyses can only be performed for those papers that published identical subgrouping information, whereas in most instances covariates are coded or stratified differently.

Subgroup analyses and individual patient data meta-analyses

Meta-analyses that use individual patient data (IPD), that is, the raw data of individual trials, rather than simply the overall results of each trial have been proposed as a major improvement in subgroup analyses over conventional meta-analyses. Since IPD meta-analyses often include more detailed data (i.e., the inclusion of unpublished data, data on effects of treatment according to strata of baseline risk, time-to-event data, and updated follow up data), they usually have greater statistical power to carry out informative subgroup analyses [13,14,16,19]. Furthermore, by using the individual patient data the flexibility of subgroup analyses may be enhanced [14]. For instance, individual patient data permits straightforward categorization of individuals for subgroup analyses by single or multiple factors and meaningful translations of covariate-data between scales of measurement [16]. Consequently, the estimated subgroup effects may be less influenced by misclassification and (ecological) bias [20]. IPD meta-analyses, therefore, allow a more thorough assessment as to whether differences in treatment effects between subgroups are spurious or not [4,10,13,14,16,19,21,22].

The merits of the IPD meta-analytic approach to study subgroup effects have, however, not yet been studied properly. Thus far, there is no consensus on the best methodology of analysing subgroup effects in IPD meta-analyses.

General objective

The general objective of this thesis is to study the accuracy, flexibility, and validity of IPD meta-analyses in analysing subgroup effects.

Outline of this thesis

In **chapter 2, 3 and 4** the results of an extensive literature study are described in which we tried to identify all published IPD meta-analyses and their related conventional meta-analyses. In **chapter 2** the methodology used to study subgroup effects in both conventional and IPD meta-analyses is described. In **chapter 3** we describe which characteristics appear to be associated with reported subgroup effects in published IPD meta-analyses. In **chapter 4** we compare the subgroup effects of both meta-analyses that address the same clinical questions. In **chapter 5** we compare subgroup effects of conventional and IPD meta-analyses, using the data of six trials on the effectiveness of antibiotics in children with acute otitis media. In **chapter 6** we show an empirical example of an IPD meta-analysis, namely on the (subgrouping) effects of antibiotics in preventing the development of asymptomatic middle ear effusion in children. **Chapter 7** focuses on the influence of various methods of handling missing data (complete case analyses, single imputation within and over trials, and multiple imputations within and over trials) on subgroup effects of IPD meta-analyses. In **chapter 8** the findings reported in this thesis are discussed and recommendations on how and when treatment effects in subgroups should best be analysed and reported in IPD meta-analyses are provided.

References

1. Offringa M, Assendelft WJJ, Scholten RJ. Inleiding in evidence-based medicine. Klinisch handelen gebaseeerd op bewijsmateriaal. Second ed. Houten/Antwerpen: Bohn Stafleu Van Loghum, 2003.
2. Egger M, Smith GD, O'Rourske K. Rationale, potentials, and promise of systematic reviews. In: Egger M, Smith GD, Altman DG, eds. Systematic reviews in health care: meta-analysis in context. London: BMJ Publishing group, 2001;3-42.
3. Smith GD, Egger M. Going beyond the grand mean: subgroup analysis in meta-analysis of randomized trials. In: Egger M, Smith GD, Altman DG, eds. Systematic reviews in health care: meta-analysis in context. London: BMJ Publishing group, 2001;143-56.
4. Smith GD, Egger M. Incommunicable knowledge? Interpreting and applying the results of clinical trials and meta-analyses. J Clin Epidemiol 1998;51:289-95.
5. Wittes RE. Problems in the medical interpretation of overviews. Stat Med 1987;6:269-80.
6. Mc Alister FA. Applying the results of systematic reviews at the bedside. In: Egger M, Smith GD, Altman DG, eds. Systematic reviews in health care: meta-analysis in context. London: BMJ Publishing group, 2001;373-85.
7. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". Lancet 2005;365:82-93.
8. Glasziou PP, Irwig LM. An evidence based approach to individualising treatment. BMJ 1995;311:1356-9.
9. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. Ann Intern Med 1992;116:78-84.
10. Oxman AD, Clarke MJ, Stewart LA. From science to practice; meta-analyses using individual patient data are needed. JAMA 1995;274:845-6.
11. Thompson SG, Higgins JP. Treating individuals 4: can meta-analysis help target interventions at individuals most likely to benefit? Lancet 2005;365:341-6.
12. Rothman KJ. Epidemiology: an introduction. Oxford University Press, Inc., 2002.
13. Berlin JA, Santanna J, Schmid CH et al. 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.
14. Lambert PC, Sutton AJ, Abrams KR et al. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. J Clin Epidemiol 2002;55:86-94.

15. 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.
16. Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. Eval Health Prof 2002;25:76-97.
17. Teramukai S, Matsuyama Y, Mizuno S et al. Individual patient-level and study-level meta-analysis for investigating modifiers of treatment effect. Jpn J Clin Oncol 2004;34:717-21.
18. Rothman KJ. Modern Epidemiology. 6th ed. Boston/Toronto: Little, Brown and Company, 1986.
19. Clarke MJ, Stewart LA. Obtaining individual patient data from randomised controlled trials. In: Egger M, Smith GD, Altman DG, eds. Systematic reviews in health care: meta-analysis in context. London: BMJ Publishing group, 2001;109-21.
20. Yusuf S, Wittes J, Probstfield J et al. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. JAMA 1991;266:93-8.
21. Stewart LA, Parmar MKB. Meta-analysis of the literature or of individual data: is there a difference? Lancet 1993;341:418-22.
22. Sutton AJ, Abrams KR, Jones DR et al. Systematic reviews of trials and other studies. Health Technol Assess 1998;2:1-276.

Chapter 2

A systematic review of analytical methods used to study subgroups in (individual patient data) meta-analyses

Abstract

Objective

To determine whether individual patient data (IPD) meta-analyses are used to perform subgroup analyses and to study whether the analytical methods regarding subgroup analyses differ between conventional and IPD meta-analyses.

Methods

IPD meta-analyses were identified with a comprehensive literature search. Subsequently, conventional meta-analyses on similar research questions were traced. Methods for studying subgroups were compared for IPD and conventional meta-analyses that were matched with respect to domain, type of treatment, and outcome measure.

Results

Of all 171 identified IPD meta-analyses and 102 conventional meta-analyses, 80% and 45% presented subgroup analyses, respectively. For 35 IPD meta-analyses and 37 "matched" conventional meta-analyses, subgroup analytic methods could be compared. The number of performed subgroup analyses did not differ between IPD and conventional meta-analyses. Both meta-analyses often do not report adequate information on methods of analyses. Interaction tests were often not performed in IPD meta-analyses (69%) and IPD was often not directly modelled (74%).

Conclusion

Many IPD meta-analyses performed subgroup analyses, but the overall treatment effects were more emphasised than the subgroup effects. To study subgroups, a wide variety of analytical methods was used in both IPD and conventional meta-analyses. In general, the use and reporting of appropriate methods for subgroup analyses should be promoted. Recommendations for improvement of methods of analyses are provided.

Introduction

One of the main aims of a meta-analysis is to pool results of similar studies quantitatively in order to produce a single and more precise overall estimate of the average effect [1,2]. The direction and magnitude of the average effect of meta-analyses are intended to guide decisions regarding clinical practice for a wide range of patients. Most physicians, however, would like to use the specific characteristics of a patient to decide on patients' individual treatment [3-5]. The application of trial results in clinical practice, therefore, requires discrimination between subgroups of patients who do and do not benefit from the intervention [6-8].

Compared to randomised trials, meta-analyses offer a better basis for subgroup analysis because they have a larger sample size [9]. In conventional meta-analyses based on published data differences in treatment effects between groups of study participants can be assessed by relating outcome to some characteristic (of treatment or study participant) on a continuous or ordered scale by meta-regression analyses [1]. When subgroup analyses are repeated in either randomised trials or meta-analyses, they do, however, often not confirm earlier findings [1].

Systematic reviews and meta-analyses that use individual patient (i.e., raw) data (IPD) rather than simply the summary results of each trial have been proposed as a major improvement in subgroup analyses. The advantages of using raw data are that more exact information is available on individual patient level about subgroup status, and it offers the opportunity to recode variables (i.e., making them more comparable between trials); to include all randomised patients; and to improve the overall follow-up. IPD meta-analyses usually have greater statistical power to carry out informative subgroup analyses, especially for patient-level subgroups (e.g., age or diabetic status), allowing a more thorough assessment as to whether differences are spurious or not [4,6,10-13]. They may enhance the flexibility and precision of subgroup analyses and thereby allow more informative subgroup analyses.

Previous studies showed that the main effects of IPD and conventional meta-analyses were comparable [10,14-18]. To date, however, no systematic study has been performed that studied whether IPD meta-analyses are indeed used to perform subgroup analyses and whether the methodology used for subgroup analyses differs between IPD and conventional meta-analyses. We, therefore, performed such a systematic review in which we compared the methods used for studying subgroup effects between IPD and conventional meta-analyses.

Methods

Search

A comprehensive literature search in Pubmed, Embase, the Cochrane library and Web of Science was conducted to identify all IPD meta-analyses of randomised controlled trials. For this search keywords from the systematic catalogue or alphabetic index were used (see Appendix for detailed search strategy). The last search was conducted on April 24, 2006. To identify conventional meta-analyses on the same objective a "related articles" search in Pubmed was conducted for every identified IPD meta-analyses (see Appendix for detailed search strategy).

Selection

In first instance, titles and abstracts were screened to identify eligible IPD meta-analyses. Selection of potential eligible IPD meta-analyses was not restricted to particular treatments or disease outcome. Full text papers were retrieved when meta-analytic techniques for raw data (i.e., individual patient data) of randomised trials were used. Potential eligible titles for conventional meta-analyses were included when the objective was comparable with the IPD meta-analyses (i.e., similar for the research question, for example when the effect of treatment X on outcome Y was studied in population Z), and meta-analytic techniques for published randomised trials were used. If obvious duplicate papers were available, only the most recently published paper was included.

To compare the analytical methods used to study subgroups in IPD and conventional meta-analyses, only those IPD and conventional meta-analyses that both performed subgroup analyses could be included. Moreover, only those conventional meta-analyses were included that could be "matched" to IPD meta-analyses on domain (certain type of patients in certain situations for which the objective is studied [19]), and type of treatment, and outcome measure.

Data extraction and analysis

Data from all included IPD and conventional meta-analyses were extracted with respect to specific characteristics, that is, publication year, number of included trials and patients, duration of follow-up, domain, type of treatment, outcome measured, effect measure, heterogeneity tests, fixed or random effects analysis, number and types of subgroups studied, and methods for subgroup analysis. All subgroups

considered as such in the original meta-analyses were counted as subgroups. Tests that were considered appropriate to study heterogeneity were chi-square, I-square, Q-statistic and Breslow-Day. We also recorded whether an appropriate interaction test was performed, that is, whether an interaction term (= treatment ^{x} subgrouping variable) was included in a regression model.

The methods for studying subgroup effects were counted and described for IPD and conventional meta-analyses. These methods were compared for all IPD meta-analyses and their "matched" conventional meta-analyses. Differences in frequencies (e.g., the difference in number of studies included in conventional and IPD meta-analyses) and the corresponding 95% confidence interval (CI) were calculated [20].

Results

Search

In the search for IPD meta-analyses, 1,808 potential eligible papers were identified. Another 39 potential eligible titles were found, while searching for conventional meta-analyses. They covered a range of medical fields, but the majority of the papers concerned oncology and cardiovascular diseases. After studying the abstracts, full-text was retrieved of 302 papers for detailed evaluation; 171 papers were finally included in the analyses. The remaining 1,676 papers were excluded for one of the following reasons: published instead of raw data; IPD meta-analyses on cohort studies, case-control studies, or case reports instead of randomised controlled trials; only one treatment arm evaluated; methodological review; or duplicate publication (Figure 2.1).

The "related articles" search for conventional meta-analyses identified 11,149 potential eligible papers. After studying the abstracts, full-text was retrieved of 362 papers for detailed evaluation; 102 papers were finally included in the analysis. The remaining 11,047 papers were excluded for one of the following reasons: raw instead of published data; conventional meta-analyses on cohort studies, case-control studies or case reports instead of randomised controlled trials; only one treatment arm evaluated; methodological or tutorial review; research question was not similar between conventional and IPD meta-analyses; or duplicate publication (Figure 2.2).

Figure 2.1. Flowchart of the inclusion of individual patient data meta-analyses
Note: IPDMA = individual patient data meta-analyses; CMA = conventional meta-analyses; RCT = randomised controlled trial.

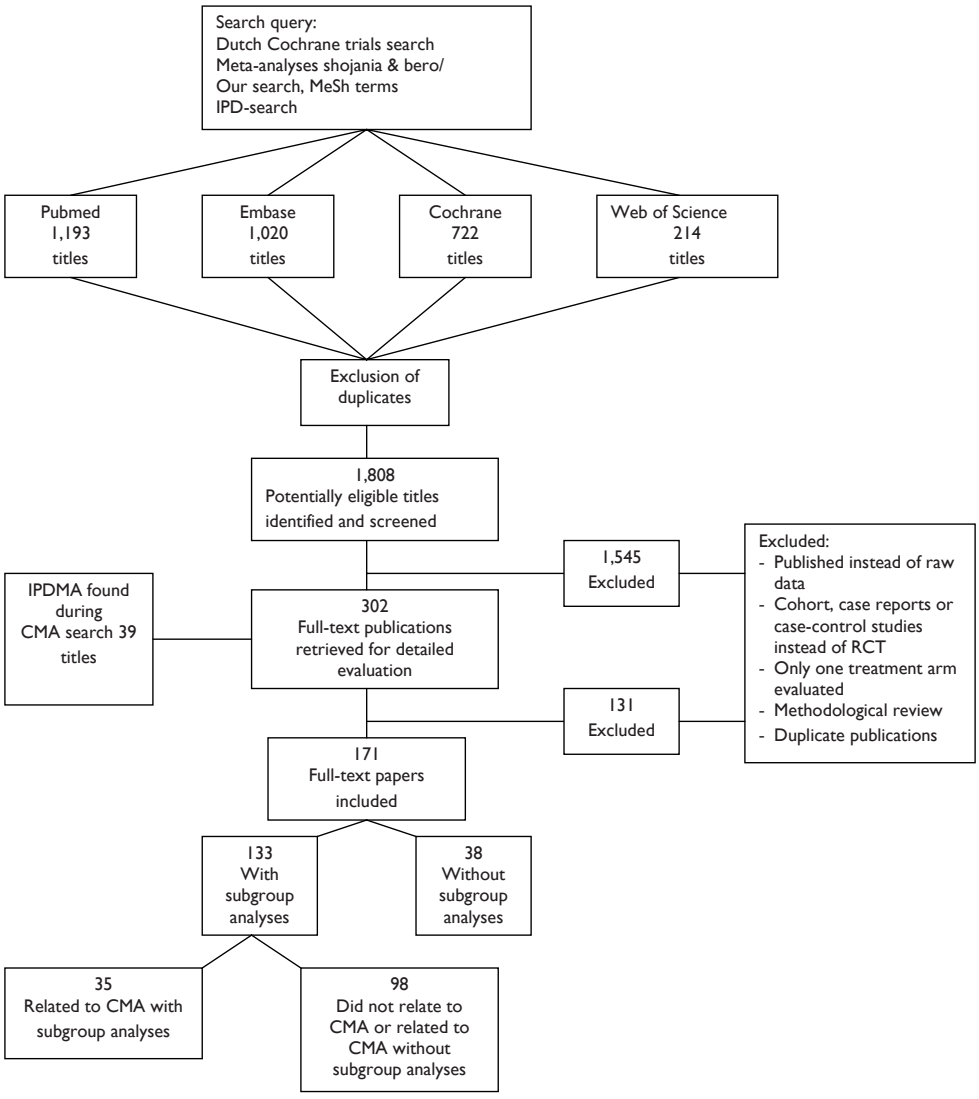
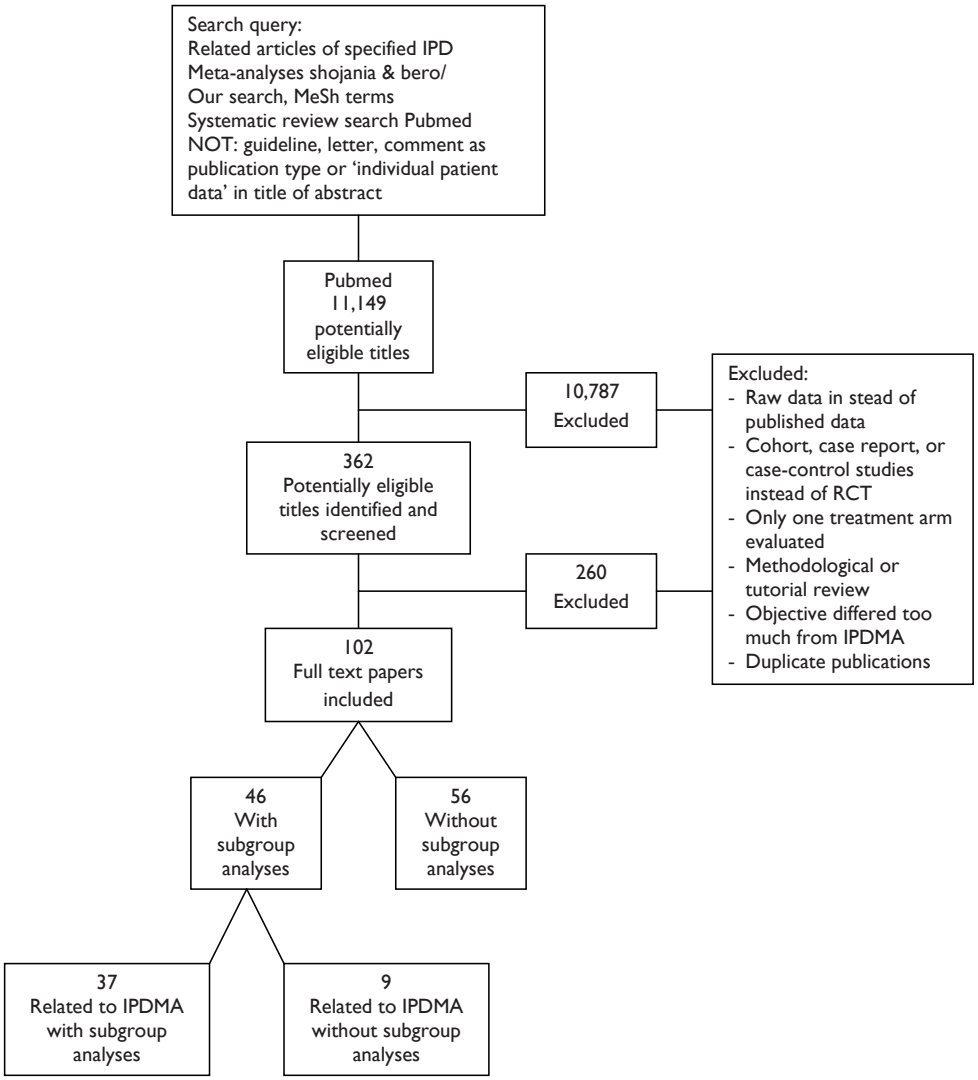


Figure 2.2. Flowchart of the inclusion of conventional meta-analyses
Note: IPDMA = individual patient data meta-analyses; CMA = conventional meta-analyses; RCT = randomised controlled trial.



Summary of all IPD and conventional meta-analyses that studied the same research question

The 171 IPD meta-analyses papers were published between 1993 and 2006. In 136 (80%) IPD meta-analyses subgroup analyses were presented. Of these, 35 could be "matched" to a conventional meta-analysis that also studied subgroups (Figure 2.1). The 102 identified conventional meta-analyses were published between 1990 and 2005. In 46 (45%) conventional meta-analyses subgroup analyses were presented. Of these, 37 could be "matched" to an IPD meta-analysis that also studied subgroups (Figure 2.2). The risk difference (RD) of performing subgroup analyses in IPD and conventional meta-analyses (i.e., the risk of studying subgroups for IPD meta-analyses minus the risk of studying subgroups for conventional meta-analyses) was 34% (95% CI 23; 46%).

Table 2.1 shows the characteristics of the methods used for studying subgroups in IPD meta-analyses (IPDMA) and conventional meta-analyses (CMA). They both used a wide variety of methods to study subgroups, but often do not report adequate information about the use of heterogeneity tests (N_{IPDMA}=69 (51%); N_{CMA}=21 (46%)), fixed or random effects models (N_{IPDMA}=78 (57%); N_{CMA}=15 (33%)), and definition of subgroups prior to data-analyses (N_{IPDMA}=27 (20%); N_{CMA}=12 (26%)). Moreover, direct modelling of IPD and interaction tests were rarely reported (N=29 (21%)), and finally random effects meta-analyses were only occasionally reported (N_{IPDMA}=12 (9%); N_{CMA}=10 (22%)).

Comparison of methods of subgroup analyses in the "matched" conventional and IPD meta-analyses

In total, 37 conventional meta-analyses could be "matched" with 35 IPD meta-analyses. The analytical methods used to study subgroups in both the IPD meta-analyses and their "matched" conventional meta-analyses are shown in table 2.2. Small differences were seen between IPD and conventional meta-analyses in both the (median) number of studies (N_{IPDMA}=8, N_{CMA}=12) and the (median) number of patients included (N_{IPDMA}=2,045, N_{CMA}=4,008). The (median) duration of follow-up in IPD meta-analyses (N=24 months) was twice compared to conventional meta-analyses (N=12 months). Even though on average IPD meta-analyses have a longer follow-up period, the publication dates of the included IPD and conventional meta-analyses do not differ, (median publication date 2000 with a quartile range (1998; 2004) and 2001 (1998; 2003), respectively).

Table 2.1. Characteristics of the methods of studying subgroups in individual patient data meta-analyses (IPDMA) and conventional meta-analyses (CMA)

Characteristics	**IPDMA** (N=136)	**CMA** (N=46)
Number of studies included (median, quartile range)	7 (4; 14)	12 (7; 22)
Number of patients included (median, quartile range)	2,045 (1144; 4953)	3,075 (1397; 5805)
Follow-up in months (median, quartile range)	26 (6; 60)	12 (6; 41)
Number of subgroups studied (median, quartile range)	3 (1; 6)	2 (1; 5)
Time of defining subgroups (N, %)		
A priori	101 (74)	31 (67)
A posteriori	6 (4)	3 (7)
Both	2 (2)	-
No adequate information available	27 (20)	12 (26)
Interaction test (N, % yes)[§]	38 (28)	1 (2)
Stratification per trial (N, % yes)[‖]	107 (79)	42 (91)
Metaregression (N, %)	-	11 (24)
Effect measure (N, %)[*,†]		
Difference/change score	16 (6)	11 (24)
Ratio		
Relative risk	39 (29)	18 (39)
Odds ratio	54 (40)	18 (39)
Risk difference	7 (5)	6 (13)
Hazard ratio	37 (27)	5 (11)
Poisson	2 (1)	-
Heterogeneity (N, %)[*,‡]		
Breslow-Day	2 (1)	1 (2)
Chi square	58 (43)	18 (39)
I square	9 (7)	2 (4)
Q statistic	5 (4)	8 (17)
No adequate information available	69 (51)	21 (46)
Analyses (N, %)[*]		
Random effects	12 (9)	10 (22)
Fixed effect	41 (30)	17 (37)
Both	5 (4)	4 (9)
No adequate information available	78 (57)	15 (33)

[*] In most articles the use of fixed or random effects models, use of effect measures and testing of heterogeneity is not stated specificly for the subgroup analysis
[†] 19 IPDMA and 12 CMA presented two effect measures
[‡] 8 IPDMA and 4 CMA presented two heterogeneity tests
[§] Interaction test i.e., interaction term (= treatment * covariate/factor/subgroup) included in a regression model
[‖] stratification per trial i.e., two-stage method

A small difference in (median) number of subgroups studied (N_{IPDMA}=3, N_{CMA}=2) was seen. There were minor differences in time of defining subgroups between IPD meta-analyses ($N_{a\ priori}$=22 (63%); $N_{a\ posteriori}$=3 (9%)) and conventional meta-analyses ($N_{a\ priori}$=25 (68%)); $N_{a\ posteriori}$=3 (8%)). More IPD meta-analyses (N=11 (31%)) than conventional meta-analyses (N=1 (3%)) performed an interaction test. More conventional meta-analyses (N=34 (92%)) than IPD meta-analyses (N=26 (74%)) stratified their analysis per trial before pooling the results (i.e., two-stage method). The use of meta-regression to study subgroups was reported in 8 conventional meta-analyses (22%). Compared to conventional meta-analyses, IPD meta-analyses expressed their measure of effect more often as a hazard ratio (N_{IPDMA}=11 (31%); N_{CMA}=4 (11%)) and less often as a relative risk (N_{IPDMA}=8 (23%); N_{CMA}=4 (11%)) or a risk difference (N_{IPDMA}=1 (3%); N_{CMA}=5 (14%)). However, conventional meta-analyses bear serious problems extracting hazard ratios from published reports [21]. More IPD meta-analyses reported the use of Breslow-Day (N_{IPDMA}=1 (3%); N_{CMA}=0) as heterogeneity test, while more conventional meta-analyses reported the use of Chi-square (N_{IPDMA}=12 (34%); N_{CMA}=16 (43%)) and Q-statistic (N_{IPDMA}=1 (3%); N_{CMA}=5 (14%)) as heterogeneity test. More conventional meta-analyses reported the use of random effects models (N_{IPDMA}=1 (3%); N_{CMA}=9 (24%)), while more IPD meta-analyses reported the use of fixed effect models (N_{IPDMA}=14 (40%); N_{CMA}=12 (32%)). Finally, both IPD and conventional meta-analyses often do not report adequate information about the use of heterogeneity tests (N_{IPDMA}=21 (60%); N_{CMA}=17 (46%)), fixed or random effects models (N_{IPDMA}=18 (51%); N_{CMA}=12 (32%)) and definition of subgroups prior to data-analyses (N_{IPDMA}=9 (26%); N_{CMA}=9 (24%)).

The type of subgroups studied varied between the compared IPD (N=35) and conventional meta-analyses (N=37): 23 (66%) IPD and 10 (27%) conventional meta-analyses studied patient characteristics (RD 39%, 95% CI 17; 60), for example, age or gender; 24 (69%) IPD and 14 (38%) conventional meta-analyses studied disease characteristics (RD 31%, 95% CI 9; 53), for example, severity or staging; 15 (43%) IPD and 27 (73%) conventional meta-analyses studied treatment related subgroups (RD -30%, 95% CI -52; -8), for example, regimen or dose; 4 (11%) IPD and 13 (35%) conventional meta-analyses studied outcome related subgroups (RD -24%, 95% CI -42; -5), for example, timing; 10 (27%) conventional meta-analyses studied subgroups related to the quality of included trials (RD -27%, 95% CI -41; -13), for example, concealment of allocation, blinding or completeness of follow-up.

Table 2.2. Characteristics of the methods of studying subgroups within the compared individual patient data meta-analyses (IPDMA) and conventional meta-analyses (CMA)

Characteristics	**IPDMA**	**CMA**	**Risk difference**
	(N=35)	(N=37)	% (95% CI)
Number of studies included (median, quartile range)	8 (4; 13)	12 (8; 18)	-
Number of patients included (median, quartile range)	2,045 (1093; 9387)	4,008 (1570; 6116)	-
Follow-up in months (median, quartile range)	24 (5; 63)	12 (5; 48)	
Number of subgroups studied (median, quartile range)	3 (1; 7)	2 (1; 5)	-
Time of defining subgroups (N, %)			
A priori	22 (63)	25 (68)	-5 (-27; 17)
A posteriori	3 (9)	3 (8)	1 (-12; 14)
Both	1 (3)	-	3 (-3; 9)
No adequate information available	9 (26)	9 (24)	2 (-18; 22)
Interaction test (N, % yes) [§]	11 (31)	1 (3)	28 (12; 44)
Stratification per trial (N, % yes) [‖]	26 (74)	34 (92)	-18 (-35; -1)
Metaregression (N, %)	-	8 (22)	-22 (-35; -9)
Effect measure (N, %) [*,†]			
Difference/change score	6 (17)	6 (16)	1 (-16; 18)
Ratio			
Relative risk	7 (20)	18 (49)	-29 (-50; -8)
Odds ratio	15 (43)	14 (38)	5 (-18; 28)
Risk difference	1 (3)	5 (14)	-11 (-24; 2)
Hazard ratio	11 (31)	4 (11)	20 (2; 38)
Heterogeneity (N, %) [*,‡]			
Breslow-Day	1 (3)	-	3 (-3; 9)
Chi square	12 (34)	16 (43)	-9 (-31; 13)
I square	2 (6)	2 (5)	1 (-10; 12)
Q statistic	1 (3)	6 (16)	-13 (-26; 0)
No adequate information available	21 (60)	17 (46)	14 (-9; 37)
Analyses (N, %) [*]			
Random	1 (3)	9 (24)	-21 (-36; -6)
Fixed	14 (40)	12 (32)	8 (-14; 30)
Both	2 (6)	4 (11)	-5 (-18; 8)
No adequate information available	18 (51)	12 (32)	19 (-3; 41)

[*] In most articles the use of fixed or random effects models, use of effect measures and testing of heterogeneity is not stated specificly for the subgroup analysis
[†] 5 IPDMA and 10 CMA used two different effect measures
[‡] 2 IPDMA and 4 CMA used two different heterogeneity tests
[§] Interaction test i.e., interaction term (= treatment * covariate/factor/subgroup) included in a regression model
[‖] stratification per trial i.e., two-stage method

Discussion

In 136 (80%) of the full set of 171 IPD meta-analyses assessed, subgroup analyses were performed to examine whether certain patients benefit more from a specific treatment than others. In total 35 IPD meta-analyses and 37 "matched" conventional meta-analyses could be compared with respect to subgroup analytic methods. A wide variety of methods was used to study subgroups in both IPD and conventional meta-analyses, and the methods to study these subgroups were not reported appropriately. It was often not reported which heterogeneity test was used, whether random or fixed effects models were used, and at what time subgroups were defined.

Some of our findings deserve further discussion. First, we identified 171 IPD meta-analyses, which is much more than expected. However, this still represents only about 2% of the total of over 8,600 systematic reviews published since 1996 (Bastian, Glasziou Cochrane Colloquium, 2005). The majority addressed cardiovascular diseases and oncology, which agrees with the overall available literature in the medical field. A rather small group of IPD meta-analyses could be compared to conventional meta-analyses with respect to subgroup analytic methods. This is due to the small number of conventional meta-analyses that reported subgroup analysis, and the limited number of conventional meta-analyses that could be "matched" to IPD meta-analyses.

Second, although many IPD meta-analyses performed subgroup analyses, the overall treatment effect was usually the main focus of the paper. Only occasionally the subgroup analyses were emphasized. The general lack of differences between subgroups with respect to treatment effects found, may explain this tendency. Other reasons could include the exploratory nature of the subgroup analyses or the absence of formulated hypotheses. It is generally accepted that subgroups should be defined prior to data-analyses since post-hoc analyses are known to be sensitive to spurious associations being found [22,23]. Nevertheless, around 30% of both IPD and conventional meta-analyses did not define their subgroups a priori. In IPD meta-analyses subgroups were often based on patient and disease characteristics, whereas treatment or outcome related subgroups were studied more frequently in conventional meta-analyses. Patient or disease characteristics should be studied using raw data rather than aggregated data since these provide more statistical power to detect patient-level interactions [11,24]. Moreover, IPD meta-analyses offer the opportunity to stratify the subgroup analyses instead of using "mean" covariates,

which may lead to ecological bias [13].

Third, a variety of methods is used to study subgroup effects. Meta-regression was expected to be the main form of subgroup analysis in conventional meta-analyses. However, only 22% of the conventional meta-analyses report the use of meta-regression analyses. Even when original papers reported the use of meta-regression, most papers did not clearly define how the covariates were incorporated in the meta-regression, for example, just one paper reported the use of "mean" covariates, and three papers reported categorized covariates. Most other conventional meta-analyses used stratification methods whereby subgroup data that were available in all trials were pooled, or dummy variables were made for specific variables, for example, average age of participants above or below median age for all trials.

Fourth, in order to reduce the chance of false positive and false negative findings, only subgroup effects for which significant interactions were found, should be studied [25]. However, of those IPD meta-analyses performing subgroup analyses, nearly half (45%) did not report such an interaction test. Furthermore, many of the interaction tests reported in the papers, for example likelihood ratio tests, chi-square tests, or comparisons of hazard ratios are inappropriate. As a result, the number of meta-analyses using an appropriate interaction test (28%) is actually much smaller.

Fifth, most IPD meta-analyses (74%) stratified their analyses by trial, that is, the "two-stage" approach [26], where each trial is analysed separately using its raw data before the summary results from each trial are pooled and analysed using a fixed or random effects meta-analysis. This practice is aimed at adjustment for residual confounding by study, which is generally accepted [27,28]. However, if many trials are included in the meta-analysis, unstable estimates might be produced using fixed trial effects or stratified models [17]. We, therefore, believe that the two-stage approach negates many advantages of using individual patient data, and consider it possible to adjust for possible confounding by direct modelling of IPD and including a dummy for study [17,29-31]. When (IPD) meta-analyses are stratified according to trial, random effects models are rarely used. This was unexpected, because in general heterogeneity is not exclusively explained by random variation.

Sixth, in many IPD and conventional meta-analyses the methods on subgroup analyses were not reported appropriately, notably whether a heterogeneity test was used, whether random or fixed effects models were used, and at what time subgroups were defined. This practice triggered us to provide some recommendations on how analyses should be handled at the end of this paper.

Limitations and strengths

To appreciate our findings certain limitations and strengths should be discussed. Our search for conventional meta-analyses may not have been optimal, as due to practical reasons we used a "related articles" search, instead of a complete conventional meta-analyses search. It is, therefore, possible that we have missed some conventional meta-analyses. It is, however, unlikely that eligible conventional meta-analyses were systematically missed. Moreover, only one person (LK) extracted the data, which might have led to misclassification of the results. We, however, believe this misclassification to be minimal as in case of uncertainty the papers were discussed with two other researchers (GvdH, MMR). Furthermore, the proportion of conventional meta-analyses that included subgroup analysis is probably an overestimation. It is more likely to find a related conventional meta-analysis that reports subgroup analyses because IPD meta-analyses are often performed when subgroup effects, for instance based on results of a meta-regression performed in conventional meta-analyses, are expected. Finally, the small number of IPD and conventional meta-analyses that could be included in the direct comparison on subgroup analytic methods might not be representative of all identified IPD meta-analyses. We, therefore, compared the characteristics of the small groups of "matched" conventional and IPD meta-analyses with the characteristics of all identified IPD and conventional meta-analyses that presented subgroup analyses. This comparison revealed no differences with respect to time of defining subgroups, the use of interaction test, stratification, and meta-regression (data not shown).

One of the major strengths of this study is that, with the developed search strategy, we identified many more IPD meta-analyses than expected, which confirms the effectiveness of our search strategy. Furthermore, this is the first paper that compared the analytical methods to study subgroups between IPD and conventional meta-analyses systematically and directly. Our findings are in agreement with Simmonds *et al.* [26]. They described analytical methods used in 44 IPD meta-analyses published in 1999-2001, and also found evidence of poor reporting, rare use of direct pooling of IPD, and rare reporting of random effects meta-analyses. They finally concluded that the statistical methodology varied substantially.

Recommendations

As shown in this paper a variety of methods were used to study subgroups. To improve the analyses and results of future IPD meta-analyses certain standards should be developed. We would like to give some recommendations on how analyses should be handled.

First, besides using appropriate methods to study main and subgroup effects of specific treatments, it is also important to report them appropriately.

Second, to decide whether pooling of different studies is justified, heterogeneity should be tested using the I-square, which describes the percentage of total variation across studies that can be attributed to heterogeneity rather than chance [32].

Third, if the raw data are available for all studies, we recommend direct pooling of this raw data instead of the two-stage method. A dummy variable for study should be included to adjust for possible confounding [17,29-31].

Fourth, if possible the performance of subgroup analyses should be specified a priori in the study protocol.

Fifth, prognostic modelling techniques should be used to select subgrouping variables since this method has shown to maintain statistical power, while preventing multiple testing [33].

Sixth, before stratified analyses can be performed, an interaction term (treatment ^{x} subgrouping variable) should be included in a regression model, and should reach statistical significance [8,23,25,33,34].

Conclusion

We showed that many IPD meta-analyses performed subgroup analyses, but the overall treatment effects were more frequently emphasised than the subgroup effects. To study subgroups, a wide variety of analytical methods was used in both IPD and conventional meta-analyses. In general, the use and reporting of appropriate methods for subgroup analyses should be promoted. So far, it has been shown that, when possible, subgroups should be defined prior to data-analyses, and appropriate interaction tests should be used to identify relevant subgroups. Nevertheless, this study shows that the principles and methods of studying subgroups in IPD meta-analyses need further study.

References

1. Smith GD, Egger M. Going beyond the grand mean: subgroup analysis in meta-analysis of randomized trials. In: Egger M, Smith GD, Altman DG, eds. Systematic reviews in health care: meta-analysis in context. London: BMJ Publishing group, 2001;143-56.
2. Egger M, Smith GD, O'Rourske K. Rationale, potentials, and promise of systematic reviews. In: Egger M, Smith GD, Altman DG, eds. Systematic reviews in health care: meta-analysis in context. London: BMJ Publishing group, 2001;3-42.
3. Mc Alister FA. Applying the results of systematic reviews at the bedside. In: Egger M, Smith GD, Altman DG, eds. Systematic reviews in health care: meta-analysis in context. London: BMJ Publishing group, 2001;373-85.
4. Smith GD, Egger M. Incommunicable knowledge? Interpreting and applying the results of clinical trials and meta-analyses. J Clin Epidemiol 1998;51:289-95.
5. Wittes RE. Problems in the medical interpretation of overviews. Stat Med 1987;6:269-80.
6. Oxman AD, Clarke MJ, Stewart LA. From science to practice; meta-analyses using individual patient data are needed. JAMA 1995;274:845-6.
7. Glasziou PP, Irwig LM. An evidence based approach to individualising treatment. BMJ 1995;311:1356-9.
8. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. Ann Intern Med 1992;116:78-84.
9. Hahn S, Williamson PR, Hutton JL et al. Assessing the potential for bias in meta-analysis due to selective reporting of subgroup analyses within studies. Stat Med 2000;19:3325-36.
10. Stewart LA, Parmar MKB. Meta-analysis of the literature or of individual data: is there a difference? Lancet 1993;341:418-22.
11. Lambert PC, Sutton AJ, Abrams KR et al. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. J Clin Epidemiol 2002;55:86-94.
12. Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. Eval Health Prof 2002;25:76-97.
13. Berlin JA, Santanna J, Schmid CH et al. 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.

14. Duchateau L, Pignon JP, Bijnens L et al. 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.
15. 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.
16. Olkin I, Sampson A. Comparison of meta-analysis versus analysis of variance of individual patient data. Biometrics 1998;54:317-22.
17. 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.
18. Steinberg KK, Smith SJ, Stroup DF et al. Comparison of effect estimates from a meta-analysis of summary data from published studies and from a meta-analysis using individual patient data for ovarian cancer studies. Am J Epidemiol 1997;145:917-25.
19. Grobbee DE, Miettinen OS. Clinical epidemiology. Introduction to the discipline. Neth J Med 1995;47:2-5.
20. Rothman KJ. Epidemiology: an introduction. Oxford University Press, Inc., 2002.
21. Barzi F, Woodward M. Imputations of missing values in practice: results from imputations of serum cholesterol in 28 cohort studies. Am J Epidemiol 2004;160:34-45.
22. Halpern SD, Karlawish JH, Berlin JA. The continuing unethical conduct of underpowered clinical trials. JAMA 2002;288:358-62.
23. Yusuf S, Wittes J, Probstfield J et al. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. JAMA 1991;266:93-8.
24. Schmid CH, Stark PC, Berlin JA et al. Meta-regression detected associations between heterogeneous treatment effects and study-level, but not patient-level, factors. J Clin Epidemiol 2004;57:683-97.
25. Brookes ST, Whitley E, Peters TJ et al. Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives. Health Technol Assess 2001;5:1-56.
26. Simmonds MC, Higgins JP, Stewart LA et al. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. Clin Trials 2005;2:209-17.
27. Deeks JJ. Systematic reviews of published evidence: miracles or minefields? Ann Oncol 1998;9:703-9.

28. Deeks JJ, Altman DG, Bradburn M.J. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger M, Smith GD, Altman DG, eds. Systematic reviews in health care: meta-analysis in context. London: BMJ Publishing group, 2001;285-312.
29. Higgins JP, Whitehead A, Turner RM et al. Meta-analysis of continuous outcome data from individual patients. Stat Med 2001;20:2219-41.
30. Turner RM, Omar RZ, Yang M et al. A multilevel model framework for meta-analysis of clinical trials with binary outcomes. Stat Med 2000;19:3417-32.
31. Whitehead A, Omar RZ, Higgins JP et al. Meta-analysis of ordinal outcomes using individual patient data. Stat Med 2001;20:2243-60.
32. Higgins JP, Thompson SG, Deeks JJ et al. Measuring inconsistency in meta-analyses. BMJ 2003;327:557-60.
33. Hayward RA, Kent DM, Vijan S et al. Multivariable risk prediction can greatly enhance the statistical power of clinical trial subgroup analysis. BMC Med Res Methodol 2006;6:18.
34. Brookes ST, Whitely E, Egger M et al. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. J Clin Epidemiol 2004;57:229-36.

Chapter 3

Reporting of subgroup effects in individual patient data meta-analyses is driven by the number of included patients, testing for heterogeneity, the use of a fixed effect model, and testing for interaction effects

Abstract

Objective

To identify characteristics that drives reporting of subgroup effects in individual patient data (IPD) meta-analyses.

Methods

Multivariate regression analyses were used to identify characteristics that were associated with reporting subgroup effects in 171 IPD meta-analyses retrieved by a comprehensive literature search.

Results

Of the 171 IPD meta-analyses retrieved, 136 (80%) reported subgroup effects. The number of included patients (odds ratio (OR) 1.68; 95% confidence interval (CI) 1.13; 2.52), testing for heterogeneity (OR 3.48; 95% CI 1.34; 9.05), the use of fixed effect analyses (OR 7.22; 95% CI 1.98; 26.31), and testing for interaction effects (OR 10.35; 95% CI 2.24; 47.74) were associated with reporting subgroup effects. For 31 IPD meta-analyses (18%), these characteristics could not accurately predict whether subgroup effects were reported or not.

Conclusion

Reporting of subgroup effects appears to be driven by the number of included patients, testing for heterogeneity, the use of a fixed effect model, and testing for interaction effects. Despite existing recommendations regarding subgroup analytic methods in meta-analyses our data show that different approaches were used. From this it can be concluded that existing recommendations are not followed.

Introduction

Individual patient data (IPD) meta-analyses, that is, meta-analyses that use the raw data of individual trials, are considered to provide the best opportunity to perform subgroup analyses [1-5]. IPD meta-analyses include more detailed data, and they usually have greater statistical power to carry out informative subgroup analyses than conventional meta-analyses on published data [1-8]. In a previous study [9], we found that IPD meta-analyses indeed reported subgroup effects more frequently compared to conventional meta-analyses. The main effects, however, remained most prominent in either approach. In addition, we found that a wide variety of subgroup analytic methods were used. These methods were not always reported appropriately, in particular, testing for heterogeneity and whether a random or a fixed effects model was used was frequently not reported.

We wondered whether the decision to study and report treatment effects in subgroups was the result of specific choices in the design and analyses phase of IPD meta-analyses. The choice for subgroup analyses may first of all depend on curiosity, clinical insight, experience, and prior proof of varying treatment effects across subgroups. Secondly, the choice for certain data-analytic methods in IPD meta-analyses may drive decisions to study subgroup effects.

Two approaches to IPD meta-analyses can be used. First, IPD can be modelled directly, including a dummy for study, as if all data belong to a single trial, also known as a "one-stage" approach [10]. Second, a "two-stage" approach can be used, that is, each trial is analysed separately using its raw data before the summary results from each trial are pooled and analysed using conventional meta-analyses techniques [10]. The two-stage approach is the more conservative method and is reported to be used more often [9-12].

To identify whether certain characteristics may drive reporting of treatment effects of subgroups in IPD meta-analyses, we derived a prediction model based on the two-stage approach of IPD meta-analyses. In this model we included the following characteristics: the number of included patients, testing for heterogeneity, the use of a random effects model, and testing for interaction effects. Subsequently, to improve the accuracy of the model, additional characteristics of IPD meta-analyses such as publication year, the number of included trials, domain, outcome, effect measure, and the use of a fixed effect model were added.

Methods

Search and selection

We searched Pubmed, Embase, the Cochrane library and Web of Science to identify IPD meta-analyses of randomised controlled trials. For this search keywords from the systematic or alphabetic index were used (see Appendix for detailed search strategy). Full text papers were retrieved when meta-analytic techniques were used to analyse the original data of randomised trials. If obviously duplicate papers were selected, only the most recently published paper was included.

Data extraction and analysis

The following characteristics were extracted from the selected IPD meta-analyses: publication year, number of included trials, number of included patients (quartiles), domain (coronary heart disease, cancer, infectious disease, kidney disease, psychiatry, rheumatology, reproduction, and other), outcome (survival, events, event free time, disease score, relapse/relief, efficiency/efficacy, and disease progression/recovery), effect measure (change score, relative risk (RR), odds ratio (OR), risk difference (RD), hazard ratio (HR), and Poisson), two-stage approach (yes vs no), use of a fixed effect model (yes vs no), use of a random effects model (yes vs no), testing for heterogeneity (i.e., χ^2, I^2, Q-statistic, and Breslow-Day tests were considered as appropriate heterogeneity tests; yes vs no), and testing for interaction effects (i.e., the interaction term of the subgrouping variable with treatment was included in the analysis; yes vs no) or significant interaction test (i.e., $p \leq 0.05$; yes vs no). Dummy variables were created for all categorical variables.

Multivariate regression analyses were started with the following predictors: number of included patients, use of a random effects model (yes vs no), testing heterogeneity (yes vs no), and testing interaction (yes vs no). The model was reduced by excluding predictors with a p-value>0.05. The Hosmer & Lemeshow test [13] was used to estimate the accuracy (goodness-of-fit) of the model. The model's ability to discriminate between IPD meta-analyses that did or did not report subgroup analyses was estimated by the area under the receiver-operating curve (ROC) of the model [14]. Depending on the predictive ability of the basic model, other independent characteristics that were univariately associated with the outcome ($p \leq 0.10$) were added to see whether the model's accuracy could be improved. Finally, discrepancies between the predictions and observations regarding subgroup analyses were examined.

Results

Search and selection

In the search for IPD meta-analyses, 1,847 potentially eligible papers were identified. They covered a range of medical fields but the majority of the papers concerned oncology and cardiovascular diseases. Based on the abstracts, full-text of 302 papers was retrieved for detailed evaluation. Finally, 171 papers were eligible for analyses. The remaining 1,676 papers were excluded for one of the following reasons: the inclusion of published instead of raw data; IPD meta-analyses on cohort studies, on case-control studies, or on case reports instead of randomised controlled trials; methodological review; or duplicate publication.

Of the 171 IPD meta-analyses included in our analyses, 60 (35%) studied a cancer related topic, 136 (80%) reported subgroup analyses, 132 (77%) used the two-stage approach to IPD meta-analyses, 72 (42%) tested for heterogeneity tests, in 21 of these (29%) the heterogeneity test was significant, 51 (30%) used a fixed effect model, 22 (13%) used a random effects model, 41 (24%) tested for interaction effects, and in 15 of these (37%) the interaction effect was significant. The most frequently used outcome was survival (43%), while the most frequently used effect measure was the odds ratio (40%) (Table 3.1).

Analysis

The basic multivariate model showed that reporting of subgroup analyses was determined by the number of included patients (odds ratio (OR) 1.56; 95% confidence interval (CI) 1.08; 2.27) and whether or not heterogeneity was tested (OR 2.43; 95% CI 1.0; 5.89). The accuracy of this model was modest with an area under the ROC of 0.69 (95% CI 0.59; 0.79), and a goodness-of-fit of p=0.50 (Table 3.2). Subsequently, additional characteristics with a statistically significant univariate association (p<0.10) with reporting subgroup effects (i.e., rheumatology (domain), disease progression, number of included patients, number of included trials, testing heterogeneity, use of a fixed effect model, and testing interaction (Table 3.1)) were included to see whether the accuracy of the model improved. The extended model showed that the choice for reporting subgroup analyses was predicted by: the number of included patients (OR 1.70; 95% CI 1.14; 2.55), testing for heterogeneity (OR 3.37; 95% CI 1.30; 8.73), the use of a fixed effect model (OR 7.09; 95% CI 1.95; 25.81), and testing for interaction (OR 9.79; 95% CI 2.12; 45.23). Thus, the

Table 3.1. Characteristics of all 171 individual patient data (IPD) meta-analyses

Variables	171 IPD meta-analyses	136 Subgroup effects reported	Univariate association with reporting subgroup effects
	N (%)	N (%)	OR (90% CI)*
Trial features			
Publication year (median, interquartile range)	2001 (1998; 2004)		0.98 (0.89; 1.08)
Domain			
Cancer	60 (35)	50 (37)	1.45 (0.74; 2.87)
Coronary heart disease	32 (19)	26 (19)	1.14 (0.50; 2.60)
Infectious diseases	21 (12)	17 (13)	1.11 (0.42; 2.93)
Psychiatry	19 (11)	13 (10)	0.51 (0.21; 1.23)
Rheumatology	11 (6)	6 (4)	**0.28 (0.10; 0.79)**
Reproduction	7 (4)	6 (4)	1.57 (0.26; 9.54)
Kidney diseases	6 (4)	6 (4)	0.00 (0.00; -)
Other	15 (9)		
Outcome			
Survival	73 (43)	62 (46)	1.83 (0.94; 3.55)
Events	32 (19)	25 (18)	0.90 (0.41; 1.98)
Disease score	23 (14)	19 (14)	1.26 (0.48; 3.30)
Disease progression/recovery	20 (12)	13 (10)	**0.42; 0.18; 0.98)**
Event free time	11 (6)	10 (7)	2.70 (0.47; 15.59)
Relapse/relief	9 (5)	6 (4)	0.49 (0.15; 1.65)
Efficiency/efficacy	3 (2)	1 (1)	
Number of included patients (median, interquartile range)	1,866 (846; 4012)		**1.69 (1.25; 2.29)**
Number of included trials (median, interquartile range)	7 (4; 12)		**1.08 (1.02; 1.14)**
Analyses features			
Effect measure			
Odds ratio	69 (40)	57 (42)	1.38 (0.72; 2.65)
Hazard ratio	46 (27)	38 (28)	1.31 (0.63; 2.72)
Risk difference	36 (21)	27 (20)	0.72 (0.35; 1.48)
Relative risk	21 (12)	14 (10)	0.46 (0.20; 1.06)
Change score	18 (11)	14 (10)	0.89 (0.33; 2.39)
Poisson	6 (4)	4 (3)	
Heterogeneity tested	72 (42)	64 (47)	**3.00 (1.46; 6.16)**
Heterogeneity significant	21(29)	17 (27)	1.11 (0.42; 2.93)
Fixed effect model	51 (30)	48 (35)	**5.82 (2.06; 16.40)**
Random effects model	22 (13)	18 (13)	1.18 (0.45; 3.11)
Interaction tested	41 (24)	39 (29)	**6.63 (1.92; 22.87)**
Interaction significant	15 (37)	15 (39)	0.00 (0.00; -)
Two-stage approach IPD meta-analyses	132 (77)	107 (79)	1.48 (0.73; 2.99)

* OR = Odds ratio; CI = Confidence interval

probability that subgroup analyses were reported was higher when more patients were included, when heterogeneity was tested, when a fixed effect model was used, and when interaction was tested. The accuracy of the final model was good with an area under the ROC of 0.82 (95% CI 0.74; 0.91), and a goodness-of-fit of p=0.40 (Table 3.3).

Discrepancies

For 31 IPD meta-analyses (18%), the final model could not accurately predict whether subgroup analyses were reported or not. For 18 of these IPD meta-analyses (58%), reporting of subgroup effects was not predicted, whereas they were actually reported. Of these 18 IPD meta-analyses, which were generally of modest size (mean number of patients per study was 1,673), 18 did not test for heterogeneity or interaction, and 14 did not mention whether a fixed or a random effects model was used. Six IPD meta-analyses specified subgrouping variables prior to data analysis.

For the remaining 13 of the 31 IPD meta-analyses (42%), reporting of subgroup effects was predicted, but they were not reported. Of these 13 IPD meta-analyses, which included on average a larger group of patients (mean number of patients per study was 3,909), eight tested for heterogeneity, three used a fixed effect model, and two tested for interaction effects. These interaction tests were not significant, and in agreement with recommendations [15,16] that propose that subgroup analyses are only allowed when a significant interaction effect is found, no subgroup effects were reported.

Comment

Our results show that the choice for reporting subgroup analyses in IPD meta-analyses is driven by 1) a larger sample size, 2) testing for heterogeneity, 3) the use of a fixed effect model, and 4) testing for interaction effects. These factors accurately identified whether subgroup analyses were reported for 82% of the 171 included IPD meta-analyses.

Although some recommendations are available on the analysis and interpretation of subgroups [1,15-19], there is no consensus as to how and when subgroups should best be legitimately identified and analysed. We, therefore, would like to comment upon our findings.

Table 3.2. Multivariate predictors of subgroup analyses in individual patient data meta-analyses - Basic model

Variable	Odds ratio (95% confidence interval)	Regression coefficient
Constant	0.99	0.04
Number of included patients	1.56 (1.07; 2.26)	0.44
Heterogeneity tested	2.41 (1.00; 5.84)	0.88
Goodnes-of-fit	p=0.50	
ROC area	0.69 (0.59; 0.79)	

ROC = Receiver operator curve

Table 3.3. Multivariate predictors of subgroup analyses in individual patient data meta-analyses - Final model

Variable	Odds ratio (95% confidence interval)	Regression coefficient
Constant	0.35	-1.07
Number of included patients	1.68 (1.13; 2.52)	0.52
Heterogeneity tested	3.48 (1.34; 9.05)	1.25
Fixed effect analyses	7.22 (1.98; 26.31)	1.98
Interaction tested	10.35 (2.24; 47.74)	2.34
Goodnes-of-fit	p=0.40	
ROC area	0.82 (0.74; 0.91)	

ROC = Receiver operator curve

First, as in any analysis, a larger sample size enhances the power of IPD meta-analyses and thereby the reliability of subgroup analyses. IPD meta-analyses are performed to increase the statistical power to improve the possibilities for, in particular, subgroup analyses [3-5]. Despite that, some IPD meta-analyses still not include enough patients to perform informative subgroup analyses. Apart from sufficient statistical power of IPD meta-analyses, the possibilities for subgroup analyses may also depend on the case mix, the distribution, and the range of characteristics that may truly modify the effect of treatment.

Second, according to prevailing theory, heterogeneity with respect to the overall and subgroup results needs to be explored first [15,17,19]. It is questionable; however, whether testing for heterogeneity is useful when IPD are modelled directly, because summary statistics of each trial should be derived in order to test for heterogeneity. Adding a covariate for study could be considered to adjust for possible heterogeneity in the direct modelling approach of IPD meta-analyses.

Third, the use of a fixed effect model was associated with the report of subgroup analyses. This is remarkable, since most included IPD meta-analyses used the so-called "two-stage" approach [9,10] for which a random effects model is recommended[18].

Fourth, we showed that testing interaction (i.e., the interaction term of the subgrouping variable with treatment was included in the analysis; yes vs no) was associated with reporting subgroup effects. It has been suggested, however, that performance of subgroup analyses are only justified when this interaction test is significant [15,16]. We found that a significant interaction test was not associated with reporting subgroup effects. Of all 171 IPD meta-analyses, 41 performed interaction tests but only 15 of these (37%) reported that stratified subgroup analyses were justified according to the results of the interaction test. Subgroup effects were presented in 23 IPD meta-analyses (56%), even though the tests for interaction were not significant. For 17 of these 23 IPD meta-analyses (74%) small relevant trends or statistical significant subgroup effects were reported, whereas the other six (26%) did not report (significant) subgroup effects. Consequently, many IPD meta-analyses found relevant subgroup effects irrespective of the results of the interaction tests.

As all meta-analyses, our analyses might be subject to publication or reporting bias, that is, statistical significant treatment effects, in particular of subgroups, are more likely to be published. For some IPD meta-analyses included in our study subgroup analyses may have been performed but not reported, because either the tests for heterogeneity or interaction or subgroup effects may not have reached statistical significance.

In conclusion, our analysis of 171 IPD showed that reporting of subgroup effects appears to be driven by the number of included patients, testing for heterogeneity, the use of a fixed effect model, and testing for interaction effects. So, despite existing recommendations regarding subgroup analytic methods in meta-analyses our data show that different approaches were used. From this it can be concluded that existing recommendations are not followed.

References

1. Berlin JA, Santanna J, Schmid CH et al. 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.
2. Lambert PC, Sutton AJ, Abrams KR et al. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. J Clin Epidemiol 2002;55:86-94.
3. Oxman AD, Clarke MJ, Stewart LA. From science to practice; meta-analyses using individual patient data are needed. JAMA 1995;274:845-6.
4. Smith GD, Egger M. Incommunicable knowledge? Interpreting and applying the results of clinical trials and meta-analyses. J Clin Epidemiol 1998;51:289-95.
5. Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. Eval Health Prof 2002;25:76-97.
6. Clarke MJ, Stewart LA. Obtaining individual patient data from randomised controlled trials. In: Egger M, Smith GD, Altman DG, eds. Systematic reviews in health care: meta-analysis in context. London: BMJ Publishing group, 2001;109-21.
7. Stewart LA, Parmar MKB. Meta-analysis of the literature or of individual data: is there a difference? Lancet 1993;341:418-22.
8. Sutton AJ, Abrams KR, Jones DR et al. Systematic reviews of trials and other studies. Health Technol Assess 1998;2:1-276.
9. Koopman L, Heijden GJMGvd, Glasziou P et al. A systematic review of analytical methods used to study subgroups (individual patient data) meta-analyses. J Clin Epidemiol 2007;in press.
10. Simmonds MC, Higgins JP, Stewart LA et al. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. Clin Trials 2005;2:209-17.
11. Deeks JJ. Systematic reviews of published evidence: miracles or minefields? Ann Oncol 1998;9:703-9.
12. Deeks JJ, Altman DG, Bradburn M.J. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger M, Smith GD, Altman DG, eds. Systematic reviews in health care: meta-analysis in context. London: BMJ Publishing group, 2001;285-312.
13. Hosmer D.W., Lemeshow S. Applied Logistic regression. New York: John Wiley & Sons, Inc., 1989;140-5.

14. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:29-36.
15. Brookes ST, Whitely E, Egger M et al. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. J Clin Epidemiol 2004;57:229-36.
16. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. Ann Intern Med 1992;116:78-84.
17. Higgins JP, Thompson SG, Deeks JJ et al. Measuring inconsistency in meta-analyses. BMJ 2003;327:557-60.
18. Knapp G, Biggerstaff BJ, Hartung J. Assessing the amount of heterogeneity in random-effects meta-analysis. Biom J 2006;48:271-85.
19. 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.

Chapter 4

A systematic review on the difference in subgroup effects between conventional and individual patient data meta-analyses

Abstract

Objective

To compare subgroup effect estimates of published individual patient data and conventional meta-analyses that addresses the same clinical questions.

Methods

Conventional and IPD meta-analyses were identified by a comprehensive literature search. Conventional and IPD meta-analyses with the same domain, objective, and one or more outcome variables were paired. Study characteristics, effect measures, effect estimates and their confidence intervals were extracted.

Results

In total, 35 IPD meta-analyses and 37 conventional meta-analyses were identified as similar. The types of subgroups studied varied between conventional meta-analyses (CMA) and IPD meta-analyses (IPDMA). IPD meta-analyses focused more on analyses of patient characteristics (N_{IPDMA}=23 (66%); N_{CMA}=10 (27%)) and disease characteristics (N_{IPDMA}=24 (69%); N_{CMA}=14 (38%)), whereas subgroup analyses in conventional meta-analyses were often based on treatment characteristics (N_{IPDMA}=15 (43%); N_{CMA}=27 (73%)) or outcome characteristics (N_{IPDMA}=4 (11%); N_{CMA}=13 (35%)). IPD meta-analyses studied 192 subgroups and conventional meta-analyses 149; 23 pairs of subgroup effect estimates could be made. For 22 pairs (96%) these estimates were in the same direction, and similar conclusions were reached for 14 pairs (61%).

Conclusion

Similar subgroup effect estimates were reported for 14 of the 23 subgrouping pairs. These comparisons were related to subgroups based on characteristics of treatment and outcome. IPD meta-analyses appear to provide more relevant clinical information, since they more frequently reported subgroups based on individual patient and disease characteristics.

Introduction

As individual studies are often underpowered, meta-analysis aims to pool results of similar studies quantitatively to produce a more precise overall estimate of the average effect [1,2]. The direction and magnitude of the average effect of meta-analyses are intended to guide decisions regarding clinical practice for a wide range of patients. Most physicians, however, would like to move beyond the average effects and use the specific characteristics of a patient to decide on a patients' individual treatment [3-5]. This requires information on effects of treatment in subgroups according to strata of patient and disease characteristics (i.e., risk profile) at baseline. Optimal application of trial results to individual patients in clinical practice, therefore, demands discrimination between subgroups of patients who may or may not benefit from the intervention [6-8].

Several studies have shown that meta-regression, often used in conventional meta-analyses (i.e., meta-analyses using aggregated data), has limitations in studying treatment effects in subgroups [9,10]. Meta-regression can be useful in investigating differences according to characteristics of study design. Still, the exploration of patient-level characteristics is problematic because only summary values of the characteristic of interest may be obtained from each study [9-11]. The use of those "mean covariate values" reduces power but more importantly can lead to (ecological) bias, that is, average baseline data for populations are handled as if they were individual baseline observations, thereby leading to attenuated associations and inter-correlations between variables that may limit the control of confounding [12,13]. Furthermore, stratified subgroup analyses can only be performed for those papers that published identical subgrouping information, whereas in most instances covariates are coded or stratified differently.

Individual patient data meta-analyses, that is, meta-analyses that use individual patient data rather than the summary results of each trial, have been proposed as a major improvement in meta-analyses, and in particular for subgroup analyses. Since individual patient data meta-analyses often include more detailed data, they usually have greater statistical power to carry out informative subgroup analyses, allowing a more thorough assessment as to whether subgroup differences are spurious or not [4,8,9,11,13-16]. Furthermore, by using individual patient data the flexibility and precision of subgroup analyses may be enhanced [9].

Previous studies [10,15,17-19] have shown that the estimates of main effects from

individual patient data and conventional meta-analyses are in the same direction, especially when comparable methods of analysis are used [20]. So far, no systematic study has been performed that compared the effect estimates of subgroups between published conventional meta-analyses and published individual patient data (IPD) meta-analyses. We, therefore, performed a systematic review on the difference in effect estimates between published conventional and IPD meta-analyses, thereby focussing on the comparability of subgroup effect estimates for pairs of conventional and IPD meta-analyses for which the domain (certain type of patients in certain situations for which the objective is studied [21]), treatment, outcome, and subgrouping variable are similar.

Methods

Search

A comprehensive literature search in Pubmed, Embase, the Cochrane library and Web of Science was conducted to identify all IPD meta-analyses of randomised controlled trials (last search on April 26, 2006). For this search keywords from the systematic catalogue or alphabetic index were used. To identify conventional meta-analyses on the same objective a "related articles" search in Pubmed was conducted for every identified IPD meta-analyses (see Appendix for detailed search strategy).

Selection

For IPD meta-analyses full text papers were retrieved when meta-analytic techniques were used to analyse the original data of randomised trials. If obviously duplicate papers (same study published in multiple journals), or two or more papers including the same analyses with a varying number of studies included were selected, only the most recently published paper was included. Potentially eligible titles for conventional meta-analyses were retrieved when the objective was comparable with the IPD meta-analyses. For further analyses, only conventional and IPD meta-analyses that performed subgroup analyses were included. A conventional and an IPD meta-analysis were compared, when they comprised the same domain, objective, and one or more outcome variables.

Data extraction and analysis

Data from all included conventional and IPD meta-analyses were extracted regarding publication year, number of included trials, time of follow-up, number of included subgroups, type of subgroups, domain, treatment, outcome and endpoints, effect measures, effect estimates and their confidence intervals. To assess whether the subgroup effect estimates of conventional and IPD meta-analyses were in agreement, the direction of the subgroup effects, their effect size, and their precision were compared. Subgroup effect estimates were considered to be in disagreement when different conclusions were reached regarding statistical significance, or when clinically relevant differences were found in the magnitude of the effect estimates (i.e., a difference of 15 percent points or more on a ratio measure, or 4 points or more on the outcome measure [22]), or both.

The following types of subgroups were identified in conventional and IPD meta-analyses: subgroups related to patient characteristics (e.g., age or gender), disease characteristics (e.g., severity or staging), treatment characteristics (e.g., dose or duration of treatment), outcome characteristics (such as timing) or quality of included trials (e.g., concealment of allocation, blinding or completeness of follow-up). We compared the conventional and IPD meta-analyses for the total numbers of reported types of subgroups (e.g., a conventional meta-analysis reported on three subgroups related to treatment characteristics, while an IPD meta-analysis reported on two subgroups related to patient characteristics). Subsequently, we compared the numbers of reported types of subgroups for those subgroups that were similar with respect to subgrouping variable and outcome. For the last two comparisons risk differences (RD) are presented.

Results

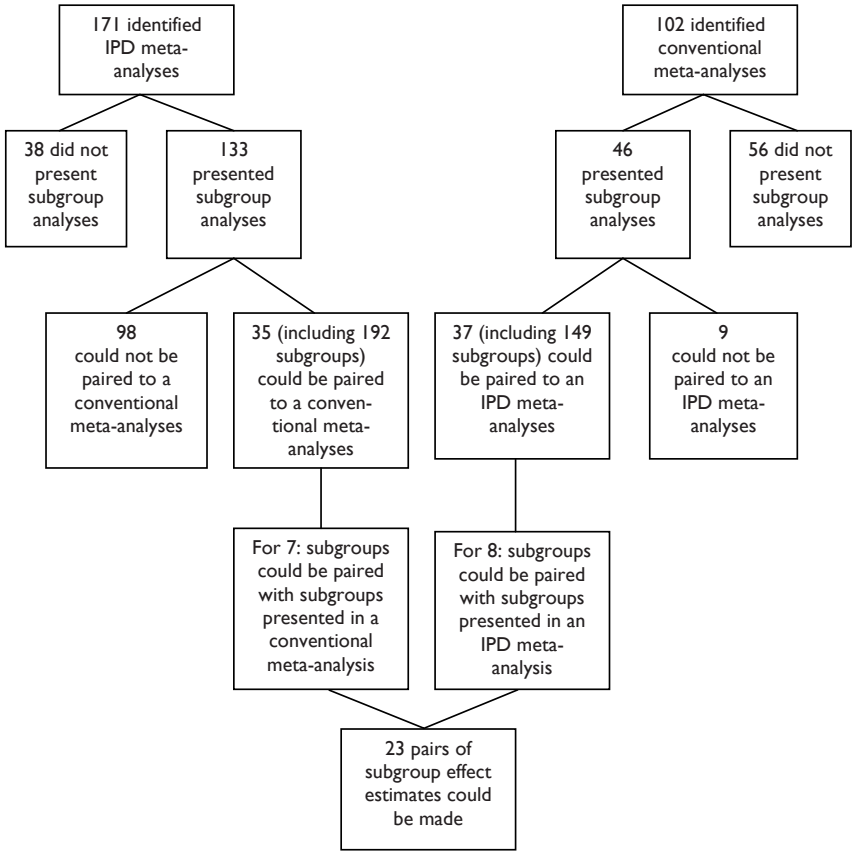
Search and selection

In the search for IPD meta-analyses, 1,808 potentially eligible papers were identified. An additional 39 potentially eligible titles were identified while searching for conventional meta-analyses. Although they covered a range of medical fields, the majority of the papers concerned oncology and cardiovascular diseases. Based on the abstracts, full-text of 302 papers was retrieved for detailed evaluation, and 171 papers were eligible for analyses. The remaining 1,676 papers were excluded for

one of the following reasons: inclusion of published data besides IPD; IPD meta-analysis on cohort studies, on case-control studies, or on case reports instead of randomised controlled trials; methodological review; or duplicate publication.

The "related articles" search for conventional meta-analyses identified 11,149 potentially eligible papers. Based on the abstracts, full-text of 362 papers was retrieved for detailed evaluation, and 102 papers were eligible for analyses. The remaining 11,047 papers were excluded for one of the following reasons: inclusion of published data besides IPD; conventional meta-analyses on cohort studies, on case-control studies, or on case reports instead of randomised controlled trials; methodological or tutorial review; objective differed too much from the IPD meta-analyses; or duplicate publication.

Figure 4.1. Flowchart of included conventional and individual patient data (IPD) meta-analyses and their subgroups.



Of 171 identified IPD meta-analyses, 133 (78%) presented subgroup analyses, and 35 of these could be paired to one or more conventional meta-analyses with respect to domain, objective, and one or more outcome variables. Of 102 identified conventional meta-analyses, 46 (45%) presented subgroup analyses, and 37 of these could be paired to one or more IPD meta-analyses. Thus, for some IPD meta-analyses several similar conventional meta-analyses papers were included, while for some conventional meta-analyses more than one similar IPD meta-analyses were included. Consequently, 35 "matched" IPD meta-analyses including 192 subgroups, and 37 "matched" conventional meta-analyses including 149 subgroups were used for further analyses (Figure 4.1).

Comparison of subgroup effect estimates

For seven published IPD meta-analyses [22-28] and eight published conventional meta-analyses [29-36], 23 subgroup effect estimates could be compared directly. Of these paired subgroup effects, all except one were in the same direction (Figures 4.2A-C). For 4 of the 13 pairs in figure 4.2A a different conclusion was reached regarding the statistical significance of the results. Furthermore, for 6 pairs relevant differences of more than 15 percent points on a ratio scale were found between the effect estimates. Figure 4.2B shows no relevant differences in the results for the 4 pairs, as only differences of ≥4 points on the outcome measure were relevant. For 3 of the 6 pairs in figure 4.2C relevant differences of more than 15 percent points on a proportion measure were found.

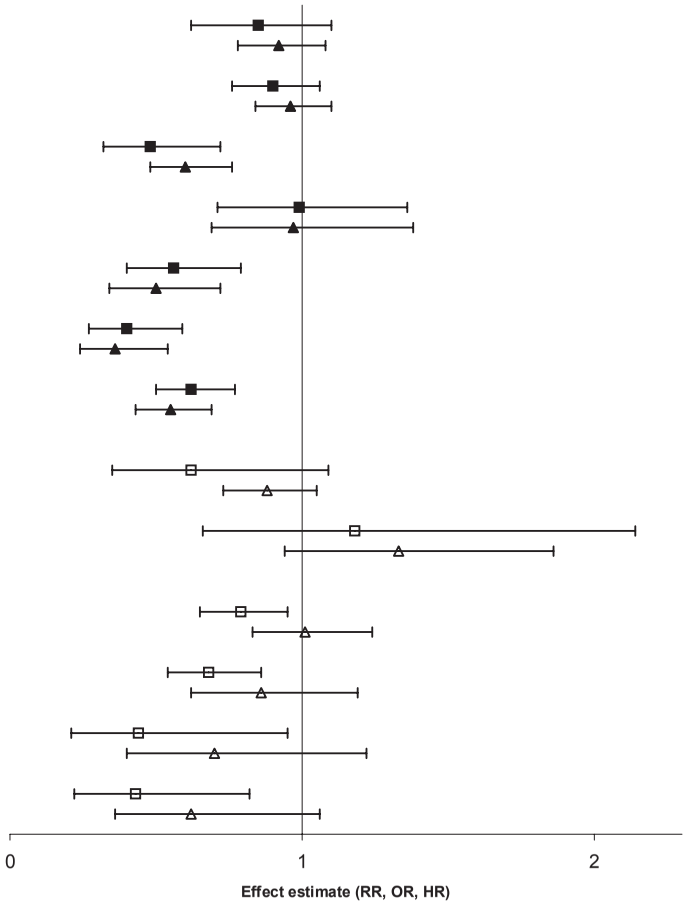
The confidence intervals of the conventional meta-analyses were wider in seven of the 23 pairs of subgroup effect estimates (30%), in another seven pairs (30%) the confidence intervals of the IPD meta-analyses were wider, in five pairs (22%) the widths of the confidence intervals of the conventional and IPD meta-analyses were similar, and for the last four pairs (17%) no confidence intervals were presented (or could be calculated) in the conventional meta-analyses.

Types of subgroups studied

The types of subgroups studied varied between the 37 conventional meta-analyses (CMA) and 35 IPD meta-analyses (IPDMA) (Table 4.1): IPD meta-analyses generally focused more on analyses of patient characteristics (N_{IPDMA}=23 (66%); N_{CMA}=10 (27%)) and disease characteristics (N_{IPDMA}=24 (69%); N_{CMA}=14 (38%)) rather than on treatment characteristics (N_{IPDMA}=15 (43%); N_{CMA}=27 (73%)) or outcome characteristics (N_{IPDMA}=4 (11%); N_{CMA}=13 (35%)).

Figure 4.2A. Comparison of subgroup effects (ratios) between conventional (square) and individual patient data meta-analyses (triangle), matched for domain, objective and outcome.

ID	Therapy*	Outcome†	Effect measure‡	Effect estimate	95% CI§
35	Amiodarone vs usual care	Mortality, post MI patients	OR	0.85	0.62; 1.10
23			OR	0.92	0.78; 1.08
35	Amiodarone vs usual care	Mortality, placebo patients	OR	0.90	0.76; 1.06
23			OR	0.96	0.84; 1.10
33	Chemotherapy+radiotherapy vs radiotherapy	Survival, concomitant therapy	HR	0.48	0.32; 0.72
25			HR	0.60	0.48; 0.76
33	Chemotherapy+radiotherapy vs radiotherapy	Overal survival, adjuvant therapy	HR	0.99	0.71; 1.36
25			HR	0.97	0.69; 1.38
31	Risedrone vs placebo	Vertebral fracure, 1 year	RR	0.56	0.40; 0.79
24			RR	0.50	0.34; 0.72
31	Risedrone vs placebo	Vertebral fracure, 2 year	RR	0.40	0.27; 0.59
24			RR	0.36	0.24; 0.54
31	Risedrone vs placebo	Vertebral fracure, 2 year	RR	0.62	0.50; 0.77
24			RR	0.55	0.43; 0.69
36	Streptokinase vs placebo	Death/disability, treat ≤3 hours	POR	0.62	0.35; 1.09
26			RR	0.88	0.73; 1.05
36	Streptokinase vs placebo	Death, 3 months, treat ≤3 hours	POR	1.18	0.66; 2.14
26			RR	1.33	0.94; 1.86
34	LMW heparin vs U heparin	DVT, general surgery	RR	0.79	0.65; 0.95
28			OR	1.01	0.83; 1.24
34	LMW heparin vs U heparin	DVT, orthopaedic surgery	RR	0.68	0.54; 0.86
28			OR	0.86	0.62; 1.19
34	LMW heparin vs U heparin	PE, general surgery	RR	0.44	0.21; 0.95
28			OR	0.70	0.40; 1.22
34	LMW heparin vs U heparin	PE, orthopaedic surgery	RR	0.43	0.22; 0.82
28			OR	0.62	0.36; 1.06



* LMW = low molecular weight; U = unfractionated; † MI = myocardial infarction; DVT = deep venous trombosis; PE = pulmonary embolism; ‡ HR = hazard ratio; OR = odds ratio; POR = peto odds ratio; RR = risk ratio; § CI = confidence interval

The six pairs with hollow figures show results that were less in agreement than the pairs with solid figures

Figure 4.2B. Comparison of subgroup effects (mean differences) between conventional (square) and individual patient data (triangle), matched for domain, objective and outcome.

ID	Therapy	Outcome*	Effect measure†	Effect estimate	95% CI‡
29	Donepezil vs placebo	ADAS-score, 5 mg, 12 wks	MD	-2.27	-3.16; -1.39
22		ADAS-score, 5 mg, 12 wks	MD	-2.11	-2.16; -1.61
29	Donepezil vs placebo	ADAS-score, 10 mg, 12 wks	MD	-2.98	-3.98;-1.98
22		ADAS-score, 10 mg, 12 wks	MD	-2.54	-3.09; -1.98
29	Donepezil vs placebo	ADAS-score, 5 mg, 24 wks	MD	-2.01	-2.69; -1.34
22		ADAS-score, 5 mg, 24 wks	MD	-1.98	-2.69; -1.27
29	Donepezil vs placebo	ADAS-score, 10 mg, 24 wks	MD	-2.90	-3.65; -2.16
22		ADAS-score, 10 mg, 24 wks	MD	-3.13	-3.91; -2.36

* ADAS = Alzheimer's disease assessment scale (score 0-70, a mean difference of at least 4 points is considered a relevant change);† MD = mean difference; ‡ CI = confidence interval

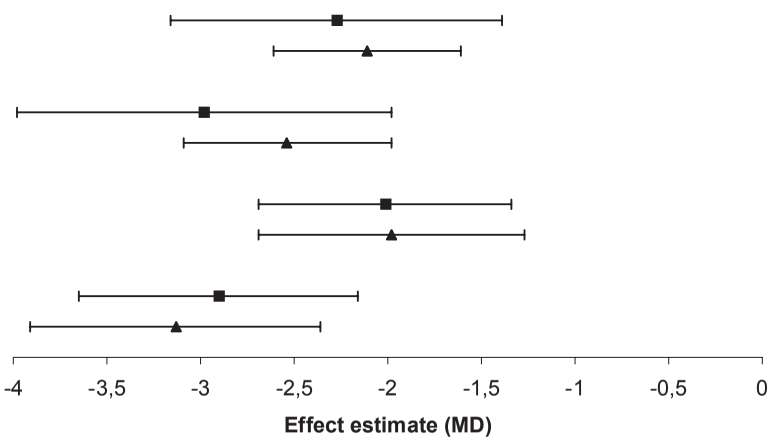
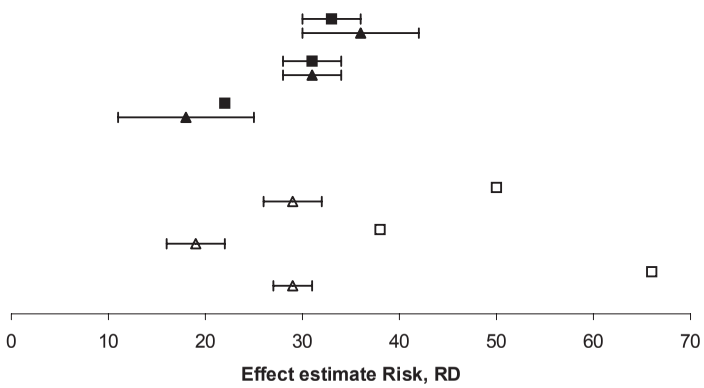


Figure 4.2C. *Comparison of subgroup effects (proportions) between conventional (square) and individual patient data meta-analyses (triangle), matched for domain, objective and outcome.*

ID	Therapy	Outcome*	Effect measure†	Effect estimate	95% CI‡
32	Rizatriptan	Pain relief,	RD	33	30; 36
27	vs placebo	2 hr, 10 mg	RD	36	28; 40
32	Rizatriptan	Pain free,	RD	31	28; 34
27	vs placebo	2hr, 10 mg	RD	31	27; 34
30	Sumatript	Pain free,	RD	22	-
27	an vs	2hr, 50 mg	RD	18	11; 25
30	Sumatript	Pain free,	Risk	50	-
27	an vs	2hr, 50 mg	Risk	29	26; 31
30	Sumatript	Pain free,	RD	38	-
27	an vs	2hr, 100 mg	RD	19	17; 22
30	Sumatript	Pain free,	Risk	66	-
27	an vs	2hr, 100 mg	Risk	29	27; 30

* riz = rizatriptan; sum = sumatriptan; † RD = risk difference; ‡ CI = confidence interval

For three estimates no confidence interval could be calculated. The three pairs with hollow figures show results that were less in agreement than the pairs with solid figures



Less than half of all subgroup effect estimates could be compared directly since both different subgroup characteristics and outcomes were studied in conventional and IPD meta-analyses.

Possible reasons for disagreement

In nine (39%) of the 23 pairs of subgroup effect estimates statistical significant or clinically relevant differences between the effect estimates were found. These nine pairs resulted from three comparable conventional and IPD meta-analyses. The most likely explanation for these disagreements was either or both the varying number of included trials and patients (Table 4.2 & 4.3). In addition, more recent published IPD meta-analyses might have included more recent trials than conventional meta-analyses or the other way around. (Table 4.2, Figure 4.3).

Table 4.1. Types of subgroups studied.

Type of subgroup	IPDMA* N (%)	CMA† N (%)	RD‡ % (95% CI)
All subgroups studied	35	37	
Patient characteristics§	23 (66)	10 (27)	39 (17; 60)
Disease characteristics‖	24 (69)	14 (38)	31 (9; 53)
Treatment characteristics¶	15 (43)	27 (73)	-30 (-52; -8)
Outcome characteristics**	4 (11)	13 (35)	-24 (-42; -5)
Trial quality††	-	10 (27)	-27 (-41; -13)
Comparable subgroups	6	7	
Patient characteristics	-	-	-
Disease characteristics	2 (29)	2 (25)	4 (-41; 49)
Treatment characteristics	6 (86)	7 (88)	-2 (-36; 33)
Outcome characteristics	1 (14)	1 (13)	2 (-33; 36)
Trial quality	-	-	-

* IPDMA = individual patient data meta-analyses; †CMA = conventional meta-analyses; ‡ RD = risk difference; CI = confidence interval; § Examples: age, gender, BMI, smoker, genotype; ‖ Examples: nodal status, arrhytmia, serum creatinine, disease site (oncology), prior events; ¶ Examples: dose, time of treatment, duration, technique; ** Example: time of outcome; †† Examples: allocation concealment, blinding, methodologic quality
The type of subgroup studied was scored in each meta-analyses

Table 4.2. Characteristics that may explain discrepancies of subgroup effects between conventional meta-analyses (CMA) and individual patient data meta-analyses (IPDMA).

ID	Type of meta-analyses*	Number of discrepant subgroup effects	Therapy†	Publication date	Included trials (same)‡	Included patients	Other reasons§
36	CMA	2	Streptokinase vs placebo	2003	18 (3)	1,292	OR
26	IPDMA	2	Streptokinase vs placebo	2000	4 (3)	5,675	RR
34	CMA	4	LMW heparin vs U heparin	1992	23 (4)	8,172	RR
28	IPDMA	4	LMW heparin vs U heparin	2001	23 (4)	12,919	OR
30	CMA	3	Sumatriptan vs placebo	2000	3 (3)	226	Some SE not reported (only graphical outcomes)
27	IPDMA	3	Sumatriptan vs placebo	2001	53 (3)	24,089	(only graphical outcomes)

* LMW = low molecular weight; U = unfractionated; † same trials: the trials that are included in both the CMA and the IPDMA; ‡ OR = odds ratio, RR = risk ratio, SE = standard error

Figure 4.3. Comparison of publication dates of the included pairs of individual patient data meta-analyses (IPDMA) and conventional meta-analyses (CMA).

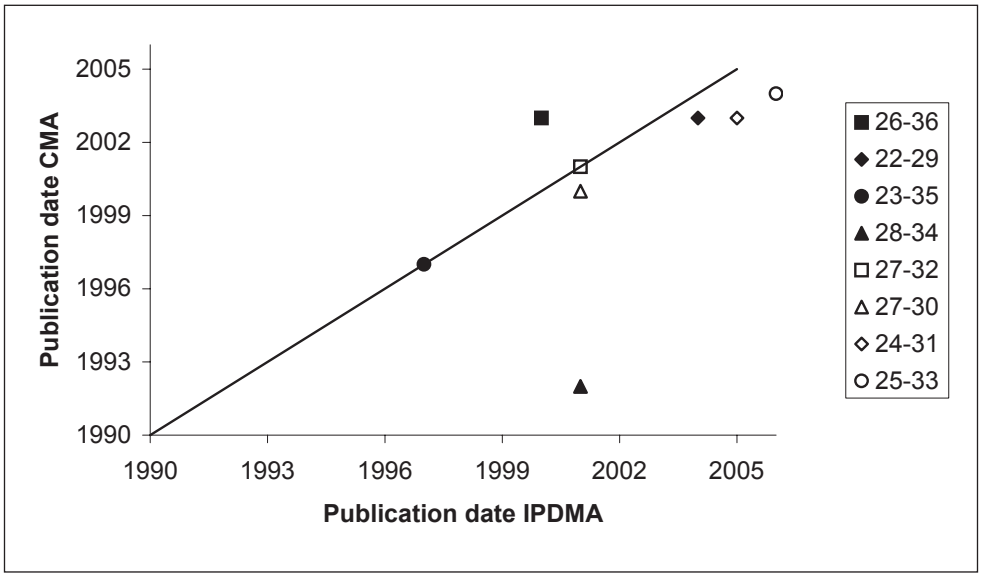


Table 4.3. Number of trials included in both meta-analyses that are comparable

Comparison* IPDMA-CMA	Number of trials in IPDMA	Number of trials in CMA	Number of same trials
23-35	13	15	9
25-33	8	10	7
24-31	5	5	5
26-36†	**4**	**18**	**3**
28-34†	**23**	**23**	**4**
22-29	10	16	5
27-32	53	7	5
27-30	**53**	**3**	**3**

The numbers in the comparison column refer to the reference ID
* CMA = conventional meta-analysis, IPDMA = individual patient data meta-analysis; † bold numbers refer to pairs with disagreeing conclusions

Discussion

For seven published IPD meta-analyses and eight published conventional meta-analyses, subgroup effect estimates could be compared. These comparisons related to subgroups based on characteristics of treatment and outcome. Of the 23 subgroup effect comparisons that could be made, 22 (96%) were in the same direction, and for 14 (61%) a similar conclusion was reached.

This is the first study that compared subgroup effect estimates between published conventional and published IPD meta-analyses. So far, four previous studies compared the main effect estimates between conventional and IPD meta-analyses [15,17-19]. They showed that the main effects are in the same direction, which we would also have concluded if we had studied the main effects (data not shown). We could not find any previous studies that compared subgroup effect estimates from published conventional and IPD meta-analyses.

We identified 171 IPD meta-analyses, but a rather small number of conventional and IPD meta-analyses remained for our comparison. This was because the domain, objective, outcome and subgrouping variable of the identified IPD meta-analyses were often not in agreement with the conventional meta-analyses. Moreover, a rather small number of conventional meta-analyses reported subgroup analyses. The inclusion of different studies in both the conventional and IPD meta-analyses due to

either the limited availability of IPD or the time period in which they are performed might be another reason for the identified discrepancies. Still, we believe that our results are informative.

Our data show that most of the subgroup effects of the compared conventional and IPD meta-analyses are in agreement. Few comparisons, however, concerned clinically relevant effect estimates for subgroups, that is, analyses of subgroups based on patient and disease characteristics. Comparison of subgroup effect estimates from conventional and IPD meta-analyses were limited by the relatively few conventional meta-analyses that studied effect estimates for subgroups based on patient and disease characteristics as subgrouping variables.

Individual patient data included in IPD meta-analyses allow for subgroup analyses according to strata of either patient and disease characteristics, or characteristics of treatment and outcome, and trial quality [2]. In conventional meta-analyses, only data published in the original papers will be available [14]. Meta-regression, which is often used in conventional meta-analyses to perform subgroup analyses, will rarely allow subgroup analyses according to the strata of patient or disease characteristics [9,10,14,37,38], but(10) can be useful in investigating differences according to characteristics of study design.

Overall, we showed that subgroup analyses of IPD meta-analyses were focused on patient and disease characteristics as compared to conventional meta-analyses that more often studied treatment or outcome characteristics. Therefore, the available data do not allow comparisons of subgroup effect estimates between conventional and IPD meta-analyses for clinically relevant subgroups.

Conclusion

In conventional and IPD meta-analyses similar effect estimates were reported for 14 of the 23 (61%) paired subgrouping variables. These comparisons were related to subgroups based on characteristics of treatment and outcome. IPD meta-analyses appear to provide more relevant clinical information, since they more frequently reported subgroups based on individual patient and disease characteristics.

References

1. Egger M, Smith GD, O'Rourske K. Rationale, potentials, and promise of systematic reviews. In: Egger M, Smith GD, Altman DG, eds. Systematic reviews in health care: meta-analysis in context. London: BMJ Publishing group, 2001;3-42.
2. Smith GD, Egger M. Going beyond the grand mean: subgroup analysis in meta-analysis of randomized trials. In: Egger M, Smith GD, Altman DG, eds. Systematic reviews in health care: meta-analysis in context. London: BMJ Publishing group, 2001;143-56.
3. Mc Alister FA. Applying the results of systematic reviews at the bedside. In: Egger M, Smith GD, Altman DG, eds. Systematic reviews in health care: meta-analysis in context. London: BMJ Publishing group, 2001;373-85.
4. Smith GD, Egger M. Incommunicable knowledge? Interpreting and applying the results of clinical trials and meta-analyses. J Clin Epidemiol 1998;51:289-95.
5. Wittes RE. Problems in the medical interpretation of overviews. Stat Med 1987;6:269-80.
6. Glasziou PP, Irwig LM. An evidence based approach to individualising treatment. BMJ 1995;311:1356-9.
7. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. Ann Intern Med 1992;116:78-84.
8. Oxman AD, Clarke MJ, Stewart LA. From science to practice; meta-analyses using individual patient data are needed. JAMA 1995;274:845-6.
9. Lambert PC, Sutton AJ, Abrams KR et al. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. J Clin Epidemiol 2002;55:86-94.
10. 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.
11. Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. Eval Health Prof 2002;25:76-97.
12. Rothman KJ. Modern Epidemiology. 6th ed. Boston/Toronto: Little, Brown and Company, 1986.
13. Berlin JA, Santanna J, Schmid CH et al. 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.

14. Clarke MJ, Stewart LA. Obtaining individual patient data from randomised controlled trials. In: Egger M, Smith GD, Altman DG, eds. Systematic reviews in health care: meta-analysis in context. London: BMJ Publishing group, 2001;109-21.
15. Stewart LA, Parmar MKB. Meta-analysis of the literature or of individual data: is there a difference? Lancet 1993;341:418-22.
16. Sutton AJ, Abrams KR, Jones DR et al. Systematic reviews of trials and other studies. Health Technol Assess 1998;2:1-276.
17. Duchateau L, Pignon JP, Bijnens L et al. 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.
18. 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.
19. Steinberg KK, Smith SJ, Stroup DF et al. Comparison of effect estimates from a meta-analysis of summary data from published studies and from a meta-analysis using individual patient data for ovarian cancer studies. Am J Epidemiol 1997;145:917-25.
20. Olkin I, Sampson A. Comparison of meta-analysis versus analysis of variance of individual patient data. Biometrics 1998;54:317-22.
21. Grobbee DE, Miettinen OS. Clinical epidemiology. Introduction to the discipline. Neth J Med 1995;47:2-5.
22. Whitehead A, Perdomo C, Pratt RD et al. Donepezil for the symptomatic treatment of patients with mild to moderate Alzheimer's disease: a meta-analysis of individual patient data from randomised controlled trials. Int J Geriatr Psychiatry 2004;19:624-33.
23. Effect of prophylactic amiodarone on mortality after acute myocardial infarction and in congestive heart failure: meta-analysis of individual data from 6500 patients in randomised trials. Amiodarone Trials Meta-Analysis Investigators. Lancet 1997;350:1417-24.
24. Adachi JD, Rizzoli R, Boonen S et al. Vertebral fracture risk reduction with risedronate in post-menopausal women with osteoporosis: a meta-analysis of individual patient data. Aging Clin Exp Res 2005;17:150-6.
25. Baujat B, Audry H, Bourhis J et al. Chemotherapy in locally advanced nasopharyngeal carcinoma: an individual patient data meta-analysis of eight randomized trials and 1753 patients. Int J Radiat Oncol Biol Phys 2006;64:47-56.

26. Cornu C, Boutitie F, Candelise L et al. Streptokinase in acute ischemic stroke: an individual patient data meta-analysis : The Thrombolysis in Acute Stroke Pooling Project. Stroke 2000;31:1555-60.
27. Ferrari MD, Roon KI, Lipton RB et al. Oral triptans (serotonin 5-HT(1B/1D) agonists) in acute migraine treatment: a meta-analysis of 53 trials. Lancet 2001;358:1668-75.
28. Koch A, Ziegler S, Breitschwerdt H et al. Low molecular weight heparin and unfractionated heparin in thrombosis prophylaxis: meta-analysis based on original patient data. Thromb Res 2001;102:295-309.
29. Birks JS, Harvey R. Donepezil for dementia due to Alzheimer's disease. Cochrane Database Syst Rev 2003;CD001190.
30. Cady RK, Sheftell F, Lipton RB et al. Effect of early intervention with sumatriptan on migraine pain: retrospective analyses of data from three clinical trials. Clin Ther 2000;22:1035-48.
31. Cranney A, Waldegger L, Zytaruk N et al. Risedronate for the prevention and treatment of postmenopausal osteoporosis. Cochrane Database Syst Rev 2003;CD004523.
32. Ferrari MD, Loder E, McCarroll KA et al. Meta-analysis of rizatriptan efficacy in randomized controlled clinical trials. Cephalalgia 2001;21:129-36.
33. Langendijk JA, Leemans CR, Buter J et al. The additional value of chemotherapy to radiotherapy in locally advanced nasopharyngeal carcinoma: a meta-analysis of the published literature. J Clin Oncol 2004;22:4604-12.
34. Nurmohamed MT, Rosendaal FR, Buller HR et al. Low-molecular-weight heparin versus standard heparin in general and orthopaedic surgery: a meta-analysis. Lancet 1992;340:152-6.
35. Sim I, McDonald KM, Lavori PW et al. Quantitative overview of randomized trials of amiodarone to prevent sudden cardiac death. Circulation 1997;96:2823-9.
36. Wardlaw JM, Zoppo G, Yamaguchi T et al. Thrombolysis for acute ischaemic stroke. Cochrane Database Syst Rev 2003;CD000213.
37. Schmid CH, Stark PC, Berlin JA et al. Meta-regression detected associations between heterogeneous treatment effects and study-level, but not patient-level, factors. J Clin Epidemiol 2004;57:683-97.
38. Teramukai S, Matsuyama Y, Mizuno S et al. Individual patient-level and study-level meta-analysis for investigating modifiers of treatment effect. Jpn J Clin Oncol 2004;34:717-21.

Chapter 5

An empirical comparison of subgroup effects of conventional and individual patient data meta-analyses

Abstract

Background

Previous studies have shown that the main effects of interventions estimated by conventional or individual patient data (IPD) meta-analyses are in the same direction. So far, effects in clinically relevant patient subgroups obtained from conventional or IPD meta-analyses using identical data have not been compared. We compared subgroup effects of conventional and individual patient data (IPD) meta-analyses using the data of six trials (n=1,643) on the effectiveness of antibiotics in children with acute otitis media (AOM).

Methods

Effects (relative risks (RR), risk differences (RD) and their confidence intervals (CI)) of antibiotics in subgroups of children with acute otitis media resulting from 1) conventional meta-analysis using summary statistics derived from published data (CMA), 2) two-stage approach to IPD meta-analysis where summary statistics derived from IPD are used (IPDMA-2), and 3) one-stage approach to IPD meta-analysis where the IPD is pooled into a single data set (IPDMA-1) were compared.

Results

In the conventional meta-analysis only two of the six studies were included because only these reported on relevant subgroup effects. The conventional meta-analysis showed larger (age <2 years) or smaller (age ≥2 years) subgroup effect estimates and wider CIs than both IPD meta-analyses (Age <2 years: RD_{CMA} -21%, $RD_{IPDMA-1}$ -16%, $RD_{IPDMA-2}$ -15%; Age ≥2 years: RD_{CMA} -5%, $RD_{IPDMA-1}$ -11%, $RD_{IPDMA-2}$ -11%). The most important reason for these discrepant results is that the two studies included in the conventional meta-analysis reported outcomes that were different both from each other and from the IPD meta-analyses.

Conclusion

This empirical example shows that conventional meta-analyses do not allow proper subgroup analyses, whereas both one- and two-stage approach IPD meta-analyses allow for accurate and precise subgroup effect analyses. We found no differences between the one- and two-stage meta-analytic approaches.

Background

Individual patient data (IPD) meta-analyses, that is, meta-analyses that use individual patient data rather than simply the overall results of each trial, have been proposed as a major improvement in meta-analytic methods. As IPD meta-analyses are typically based on more detailed data, they usually have greater statistical power to carry out informative subgroup analyses. Moreover, IPD meta-analyses allow more accurate classification of patients based on individual characteristics, and may, therefore, allow a more thorough assessment as to whether differences in effect estimates between clinically relevant patient subgroups are spurious or not [1,2].

Previous studies [3-5] have shown that the overall effects of an intervention obtained from individual patient data or conventional meta-analyses are in the same direction, especially when similar methods of data-analysis are applied [6]. Thus far, no study has been performed that compared effects in patient subgroups estimated by conventional or individual patient data meta-analyses using identical data. We compared the subgroup effects between three meta-analytic approaches: 1) conventional meta-analysis, 2) two-stage approach IPD meta-analysis, and 3) one-stage approach IPD meta-analysis.

Methods

For this empirical study the data of six trials were available, which were previously used for an IPD meta-analysis assessing the effectiveness of antibiotics in children with acute otitis media (AOM). Its methods and main results are extensively described elsewhere [7]. In brief, individual patient data on 1,643 children aged 6 months to 12 years were included. The primary outcome measure was pain or fever or both at 3 to 7 days, and age (<2 and ≥2 years) and uni / bilaterality appeared to be the clinically relevant subgrouping variables.

For the present study, the summary statistics for the conventional meta-analysis based on published data were extracted from the published reports of the six included trials. The summary statistics for each trial included in the two-stage approach to IPD meta-analysis were extracted from the available individual patient data, subsequently, the summary statistics were pooled and analysed using conventional meta-analyses techniques [8]. For the one-stage approach to IPD meta-analysis the individual patient

data are modelled directly as if all data belong to a single trial, including a covariate for "study" to adjust for potential study differences [8].

Stratified analyses were performed to quantify the effect in the subgroups studied. Relative risks (RR), risk differences (RD), and their 95% confidence intervals (CI) of the subgrouping effects were calculated and compared between each meta-analytic approach.

Results

Only two of the six included studies reported subgroup effects according to age in the trial publications; none of these reported on subgroup effects for uni- and bilateral AOM. Thus, only the results of the subgroup age could be compared between the conventional and the one- and two-stage approach to IPD meta-analyses, whereas for both IPD meta-analyses the subgroup effects of age and bilaterality could be studied. However, the two studies that could be included in the conventional meta-analyses reported endpoints that were different both from each other and from the outcome of the IPD meta-analyses. Appelman *et al.* [9] presented subgroup effects for age on an irregular course at three days, while McCormick *et al.* [10] presented subgroup effects for age regarding treatment failure between 0 and 12 days. For the IPD meta-analyses the original primary and/or secondary outcome variables could be recoded into one similar outcome variable, notably, having ear pain, fever or both at 3-7 days.

Table 5.1 displays the effect estimates (RR and RD) with their 95% confidence intervals of the three types of meta-analyses. The conventional meta-analysis (CMA) showed larger treatment effects for the subgroup age <2 years than both IPD meta-analyses (IPDMA-1 and IPDMA-2), which showed similar treatment effects (RD_{CMA} -21%, $RD_{IPDMA-1}$ -16%, $RD_{IPDMA-2}$ -15%; RR_{CMA} 0.33, $RR_{IPDMA-1}$ 0.67, $RR_{IPDMA-2}$ 0.68; Table 5.1). For the subgroup age ≥2 years the RD of the conventional meta-analysis showed a smaller effect as compared to the one- and two-stage approach IPD meta-analyses, while both IPD meta-analyses showed similar treatment effects (RD_{CMA} -5%, $RD_{IPDMA-1}$ -11%, $RD_{IPDMA-2}$ -11%; Table 5.1). The RRs for the subgroup age ≥2 years were comparable between the three meta-analytic approaches (RR_{CMA} 0.62, $RR_{IPDMA-1}$ 0.63, $RR_{IPDMA-2}$ 0.64; Table 5.1). The confidence intervals of the conventional meta-analyses were, however, wider (Table 5.1).

The treatment effects for the subgroups, in which age and bilaterality were combined, were similar for the one- and two-stage approach to IPD meta-analyses (Table 5.1).

Table 5.1. Subgroup results (relative risk (RR), risk difference (RD), and their 95% confidence intervals (CI)) of the three meta-analytic approaches

Subgroup*	**Studies† (n)**	**Events/Totals**		**RR (95% CI)**	**RD (95% CI)**
		antibiotic	***control***		
Age <2 years					
CMA	2 [1,6]	8/80	19/62	0.33 (0.16; 0.68)	-21 (-33; -8)
IPDMA-2	5 [1,3,4,5,6]	86/268	130/273	0.68 (0.55; 0.84)	-15 (-23; -7)
IPDMA-1	5 [1,3,4,5,6]	86/268	130/273	0.67 (0.54; 0.84)	-16 (-24; -8)
Age ≥2 years					
CMA	2 [1,6]	8/96	12/92	0.62 (0.25; 1.55)	-5 (-13; 4)
IPDMA-2	6	103/528	160/519	0.64 (0.52; 0.79)	-11 (-16; -6)
IPDMA-1	6	103/528	160/519	0.63 (0.51; 0.79)	-11 (-16; -6)
Unilateral AOM, age <2 yrs					
CMA	0				
IPDMA-2	4 [1,3,4,6]	43/122	50/127	0.88 (0.64; 1.21)	-5 (-17; 7)
IPDMA-1	4 [1,3,4,6]	43/122	50/127	0.90 (0.65; 1.24)	-4 (-16; 8)
Unilateral AOM, age ≥2 yrs					
CMA	0				
IPDMA-2	5 [1,2,3,4,6]	56/296	76/295	0.75 (0.56; 1.02)	-6 (-13; 0)
IPDMA-1	5 [1,2,3,4,6]	56/296	76/295	0.73 (0.54; 1.00)	-7 (-14; 0)
Bilateral AOM, age <2 yrs					
CMA	0				
IPDMA-2	4 [1,3,4,6]	39/134	70/125	0.53 (0.39; 0.72)	-26 (-38; -15)
IPDMA-1	4 [1,3,4,6]	39/134	70/125	0.52 (0.38; 0.71)	-27 (-39; -15)
Bilateral AOM, age ≥2 yrs					
CMA	0				
IPDMA-2	5 [1,2,3,4,6]	21/94	29/83	0.64 (0.40; 1.02)	-13 (-26; 0)
IPDMA-1	5 [1,2,3,4,6]	21/94	29/83	0.64 (0.40; 1.03)	-13 (-26; 0)

CMA = conventional meta-analyses; IPDMA-1 = individual patient data meta-analyses were the IPD is modelled directly; IPDMA-2 = individual patient data meta-analyses were summary statistics are extracted from the IPD; AOM = acute otitis media

† 1 = Appelman et al., 2 = Burke et al., 3 = Damoiseaux et al., 4 = Le Saux et al., 5 = Little et al., 6 = McCormick et al. The two studies included in the CMA reported different endpoints, while for the IPD meta-analyses the original outcome variables were recoded into one similar outcome variable.

Discussion

To our knowledge, this is the first study comparing intervention effects in patient subgroups resulting from conventional meta-analysis and the one- and two-stage approach IPD meta-analyses using identical data. Our data confirm earlier studies showing that the performance of subgroup analyses in conventional meta-analyses is hampered because most papers do not report on subgroup effects [2,11]. Furthermore, the conventional meta-analysis showed larger (age <2 years) or smaller (age ≥2 years) subgroup effect estimates and wider confidence intervals than both IPD meta-analyses. Differences in the results of conventional and IPD meta-analyses may be due to the use of other data analytical techniques, discrepancies in outcome scales, limited availability of outcome data in subgrouping variable strata, or missing data in the subgrouping variables. More likely, however, these discrepancies can be explained by the fact that the two studies that could be included in the conventional meta-analysis reported outcomes that were different both from each other and from the outcome used in the IPD meta-analyses. For the purpose of the comparison we pooled the different outcomes of the studies included in the conventional meta-analysis anyway, but of course in other circumstances we would probably have decided that the outcomes were too heterogeneous to pool.

In addition, fewer studies reported on effects in clinically relevant patient subgroups, reducing the precision of subgroup effect estimates from the conventional meta-analysis. It should, however, be noted that lack of reporting subgroup effects does not mean that this subgroup information is not available. The IPD actually showed that almost all subgrouping variables were measured in all trials.

We did not find differences in subgroup effects between the one- and two-stage approach IPD meta-analyses, in our example. Most likely this is a result of the rather straightforward analytical techniques that were used. The two-stage approach, that is, analysing each trial separately using its IPD before the summary results from each trial are pooled and analysed using conventional meta-analyses techniques, has been recommended to prevent confounding [2,8]. The one-stage approach in analysing IPD meta-analyses, that is, direct pooling of the IPD, however, allows more flexibility in more complex situations without loss of power due to stratification by trial [2]. Furthermore, by adding a variable "study" in fixed effect regression analyses it is also possible to adjust for potential confounding.

To overcome the problem of non-reporting of subgroup effects in trial reports,

investigators might ask original trialists for stratified analyses for the subgroups of interest, which is comparable with the two-stage approach in analysing IPD meta-analyses. However, in that case, one could probably better ask for the IPD, since IPD offer the opportunity to recode variables, update follow-up data, impute the missing data, and allows flexible analyses, and more advanced modelling techniques [2].

In conclusion, our data confirm that conventional meta-analyses do not allow proper subgroup analyses, whereas both one- and two-stage approach IPD meta-analyses allow for accurate and precise subgroup effect analyses. We found no differences between the one- and two-stage approaches to IPD meta-analyses.

References

1. Clarke MJ, Stewart LA. Obtaining individual patient data from randomised controlled trials. In: Egger M, Smith GD, Altman DG, eds. Systematic reviews in health care: meta-analysis in context. London: BMJ Publishing group, 2001;109-21.
2. Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. Eval Health Prof 2002;25:76-97.
3. 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.
4. Steinberg KK, Smith SJ, Stroup DF et al. Comparison of effect estimates from a meta-analysis of summary data from published studies and from a meta-analysis using individual patient data for ovarian cancer studies. Am J Epidemiol 1997;145:917-25.
5. Stewart LA, Parmar MKB. Meta-analysis of the literature or of individual data: is there a difference? Lancet 1993;341:418-22.
6. Olkin I, Sampson A. Comparison of meta-analysis versus analysis of variance of individual patient data. Biometrics 1998;54:317-22.
7. Rovers MM, Glasziou P, Appelman CL et al. Antibiotics for acute otitis media: a meta-analysis with individual patient data. Lancet 2006;368:1429-35.
8. Simmonds MC, Higgins JP, Stewart LA et al. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. Clin Trials 2005;2:209-17.
9. Appelman CL, Claessen JQ, Touw-Otten FW et al. Co-amoxiclav in recurrent acute otitis media: placebo controlled study. BMJ 1991;303:1450-2.
10. McCormick DP, Chonmaitree T, Pittman C et al. Nonsevere acute otitis media: a clinical trial comparing outcomes of watchful waiting versus immediate antibiotic treatment. Pediatrics 2005;115:1455-65.
11. 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.

Chapter 6

Antibiotics to prevent the development of asymptomatic middle ear effusion in children with acute otitis media: an individual patient data meta-analysis

Abstract

Objectives

To determine the predictors of developing asymptomatic middle ear effusion (MEE) in children with acute otitis media (AOM), and to examine the effect of antibiotics in preventing the development of MEE in these children.

Methods

The data of five randomised controlled trials were included in the individual patient data meta-analysis (total 1,328 children aged 6 months to 12 years). We identified independent predictors of developing asymptomatic MEE at one month and studied whether these children benefited more from antibiotics. The primary outcome was MEE (defined as a type B tympanogram) at one month.

Results

The overall relative risk (RR) of antibiotics in preventing the development of asymptomatic MEE after one month was 0.9 (95% CI 0.8; 1.0; p=0.19). Independent predictors of developing asymptomatic MEE were age less than two years and recurrent AOM. No significant interaction effects with treatment were found.

Conclusion

Due to the marginal effect of antibiotics on developing asymptomatic MEE and the known negative effects of prescribing antibiotics, such as the development of antibiotic resistance and side effects, we do not recommend antibiotics to prevent MEE.

Introduction

Otitis media (OM) is one of the most common diseases in infants and children [1,2]. Acute otitis media (AOM) and otitis media with effusion (OME) are different stages of the OM continuum [3]. Children with OME suffer up to five times more episodes of AOM than those without OME, whereas 50% of the children with AOM will develop asymptomatic middle ear effusion (MEE) after an episode of acute infection [4]. The effusion might lead to a conductive hearing loss of 15-40 dB, and this hearing loss could have an adverse effect on language development, cognitive development, behaviour, and quality of life [5]. However, past research has not been conclusive as more recent studies demonstrate little or no effect of MEE on language and cognitive development [6-8].

Findings from a recent study showed that antibiotics are mainly beneficial in children younger than 2 years of age with bilateral AOM, and in children with AOM and otorrhea with respect to pain and/or fever at 3-7 days. For most other children with AOM an observational policy seems justified [9]. Since antibiotics may also influence the development of asymptomatic MEE [10-12], the question is whether treatment might be more beneficial in preventing the development of this MEE in specific subgroups of children. Our individual patient data (IPD) meta-analysis, that is, a meta-analysis on the individual original data of performed trials, offers the unique opportunity to identify subgroups that are more or less likely to benefit.

In this IPD meta-analysis we therefore aimed to 1) determine the predictors of developing asymptomatic MEE in children with AOM, and 2) examine the effect of antibiotics in preventing the development of asymptomatic MEE in these children.

Methods

Selection of the trials

A systematic literature search was performed from Pubmed, Embase, the Cochrane library, and the proceedings of international symposia on recent advances in otitis media. To be selected for the IPD meta-analysis, trials had to be randomised, had to include children aged 0 to 12 years with AOM, the comparison had to be between antibiotics and placebo or no treatment, and MEE at one month had to be measured.

Data collection and endpoints

The primary investigators of all selected trials were asked to provide the raw data of their trials. The obtained data were thoroughly checked for consistency, plausibility, and integrity of randomisation and follow-up. Any queries were resolved by the responsible trial investigator or statistician. Based on a literature search and the availability of information in routine clinical practice, the following baseline candidate predictors for MEE were selected: age (<2 years vs ≥2 years), gender (boys vs girls), season (autumn/winter vs spring/summer), having been breast fed (yes vs no), smoking in the household (yes vs no), siblings (yes vs no), recurrent AOM (yes vs no), fever (yes vs no), pain (yes vs no), bilateral AOM (yes vs no), otorrhea (yes vs no), common cold (yes vs no), crying (yes vs no), coughing (yes vs no), red tympanic membrane (yes vs no), bulging tympanic membrane (yes vs no), and perforation of the tympanic membrane (yes vs no). The primary outcome was MEE at one month diagnosed with tympanometry (type B tympanograms were indicative of the presence of MEE).

Data analyses

Data were available for 85% of the predictor variables (range 18 to 100%), and for 91% of the outcome variable (range 76 to 99%). To decrease bias and to increase statistical efficiency, we imputed the missing data per trial using the linear regression method (Missing Value Analysis function) available in SPSS software (SPSS for Windows, version 12.0, SPSS Inc.) [13,14]. Such imputation is based on the correlation between each variable with missing values and all other variables, as estimated from the complete set of subjects from that particular trial.

To decide whether pooling of the data was justified, heterogeneity between studies was assessed using the I^2 [15]. Since this I^2 was lower than 25%, pooling was indeed performed.

First, independent predictors of developing asymptomatic MEE at one month were identified [16]. To get rid of possible influence of antibiotic therapy on the findings, we only included the children from the control groups in this prognostic analysis. Predictors with a univariate association with the outcome ($p \leq 0.10$) were included in multivariate logistic regression analyses. The model was reduced by excluding predictors from the model with a p-value>0.05. The predictive accuracy of the model was estimated by the reliability (goodness-of-fit) using Hosmer & Lemeshow tests [17]. The prognostic model's ability to discriminate between children with and without

a poor outcome was estimated by the area under the receiver-operating curve (ROC) of the model [18]. In addition, we calculated the absolute risks of developing asymptomatic MEE across combinations of independent predictors.

Subsequently, the individual predictors were used to study whether the children at risk of asymptomatic MEE at one month benefited more from antibiotics than those with a lower risk. Fixed effect logistic regression analysis were performed to study whether the interaction effect between the subgrouping variable and treatment was statistically significant [19]. This approach tests and estimates the difference between treatment effects across subgroups directly, that is, it involves one statistical test irrespective of the number of subgroups. Stratified analyses were performed to quantify the effect in the subgroups studied. Relative risks (RR), rate differences (RD), and their 95% confidence intervals (CI) were calculated. Finally, sensitivity analyses were performed, including either trials that included a placebo treatment, or trials that included the same dose regimen. All analyses were performed according to the intention-to-treat principle.

Results

Nineteen trials were identified that studied the effectiveness of antibiotics in AOM children. Thirteen trials were excluded due to: inadequate randomisation, control arm receiving another antibiotic treatment, or non-availability of information on the outcomes included in our meta-analysis. Five research groups provided their data [20-24], the data of the other trial were not available [25]. The mean number of children in the trials ranged from 121 to 512. In total, 44% of the total number of 1,328 children were younger than two years of age, 50% were male, 52% had recurrent AOM, 34% had bilateral AOM, and 94% had a red tympanic membrane (Table 6.1 & 6.2).

Prognostic model

In total, 660 children in the control arms of the trials were included in the prognostic analyses. Of these, 334 (51%) developed asymptomatic MEE. The independent predictors associated with the development of asymptomatic MEE at one month were age less than two years (odds ratio (OR) 2.2; 95% CI 1.6; 3.0), and recurrent AOM (OR 1.5; 95% CI 1.1; 2.1). The prognostic model showed a good fit (goodness-of-fit test p=0.88), and the area under the ROC was 0.61 (95% CI 0.57; 0.66).

Table 6.1. Characteristics of the five included trials

Trial, year	Number of patients	Participants	Interventions	Duration intervention	Outcomes
Appelman, 1991	121	Children aged 6 months – 12 years visiting a general practitioner (GP) with recurrent acute otitis media (AOM)	Amoxycillin/ cluvulanate vs. placebo	7 days	- fever after 3 days
					- pain after 3 days
					- otorrhea
					- otoscopy and tympanometry after 1 month
Burke, 1991	232	Children aged 3 to 10 years with AOM	Amoxycillin vs. placebo	7 days	- symptom kept diary by the parents (including fever + earpain)
					- home visits by researcher after 24 hours and 5-7 days
					- otoscopy and tympanometry after 1 and 3 months
Damoiseaux, 2000	240	Children aged 6 months to 2 years visiting a GP with AOM	Amoxycillin vs. placebo	10 days	- symptoms at day 4 assessed by a GP (incl. fever + earpain)
					- otoscopy and tympanometry after 6 weeks and 3 months
Le Saux, 2005	512	Children aged 6 months to 5 years presented to clinics or the emergency department with AOM	Amoxycillin vs. placebo	10 days	- Telephone follow-up at day 1,2, 3 and between 10 and 14 days (incl. fever)
					- Tympanometry at 1 and 3 months
McCormick, 2005	223	Children aged 6 months to 12 years with AOM	Immediately antibiotics (Amoxycillin) vs. delayed treatment	10 days	- Symptom diary kept by the parents (incl. fever + earpain)
					- Analgesics consumption
					- Nasophayngeal carriage
					- Adverse events
					- Absence from school
					- Tympanometry after 12 and 30 days

Table 6.2. Baseline characteristics of the 1328 included children.

Characteristics	N(%)		
	Placebo	***Antibiotics***	***Total***
Age <2 years	290 (44)	294 (44)	584 (44)
Male gender	330 (50)	337 (51)	667 (50)
Recurrent AOM*	355 (54)	333 (50)	688 (52)
Siblings	472 (78)	456 (76)	928 (77)
Winter season	500 (76)	505 (76)	1,005 (76)
Being ever breastfed	137 (59)	138 (60)	275 (59)
Passive smoke exposure	187 (38)	192 (39)	379 (39)
Crying	415 (84)	404 (83)	819 (83)
Cough	381 (77)	365 (75)	746 (76)
Common cold	427 (78)	430 (77)	857 (78)
Earpain	567 (86)	579 (87)	1,146 (86)
Fever	272 (50)	271 (49)	543 (50)
Bilateral AOM*	220 (33)	236 (35)	456 (34)
Otorrhea	19 (15)	16 (14)	114 (21)
Perforation	8 (7)	7 (6)	15 (6)
Red tympanic membrane	621 (94)	633 (95)	1,254 (94)
Bulging tympanic membrane	264 (40)	271 (41)	535 (40)

* AOM = Acute otitis media

Table 6.3 shows the absolute risks of developing asymptomatic MEE at one month in children with certain combinations of independent prognostic factors. The absolute risk of developing asymptomatic MEE was highest in children aged less than two years with recurrent AOM, that is, 64% (95% CI 58; 70). The absolute risk was lowest in children aged two years or older without recurrent AOM, that is, 30% (94% CI 25; 35).

Table 6.3. Absolute risks of developing middle ear effusion at one month for the overall effect and each of the subgroups of children with acute otitis media (AOM).

Predicting variable	N (%) of all children	Absolute risk (95% CI)
Overall effect	648 (100)	49% (46; 52)
<2 years, no recurrent AOM	174 (27)	55% (50; 60)
<2 years, recurrent AOM	171 (26)	64% (58; 70)
≥2 years, no recurrent AOM	96 (15)	30% (25; 35)
≥2 years, recurrent AOM	207 (32)	49% (44; 54)

Overall effect of antibiotics on OME development

There was no significant overall effect of antibiotics in preventing the development of asymptomatic MEE at one month (p=0.19). The relative risk (RR) on developing MEE was 0.9 (95% CI 0.8; 1.0) and the risk difference (RD) showed a small benefit for antibiotic treatment (4%, 95% CI -2; 9).

Effect of antibiotics on OME in subgroups of children

The largest treatment effect was found in children aged two years and older without recurrent AOM: 36% in the placebo group and 24% in the antibiotic group developed asymptomatic MEE at one month (RD -11%; 95%CI -21; -2%). The smallest effect was found in children aged younger than two years with recurrent AOM: 65% in the placebo group and 63% in the antibiotics group developed asymptomatic MEE after one month (RD -1%; 95% CI -13; 10%). However, none of the (combined) subgrouping variables showed a statistically significant interaction effect with treatment (Table 6.4).

Sensitivity analyses including either trials that were placebo controlled, or trials including the same dose of antibiotics, produced similar results.

Table 6.4. Overall effect and stratified subgroup results presented as risk differences (RD), relative risks (RR), and their confidence intervals (95% CI).

Subgroup	N	Placebo (%)	Antibiotics (%)	RD (95% CI)	RR (95% CI)	p-value for interaction
Overall effect	1,328	51	47	-4 (-9; 2)	0.93 (0.82; 1.04)	
Age						
<2 years	584	61	57	-3 (-11; 5)	0.95 (0.81; 1.08)	
≥2 years	744	43	39	-4 (-11; 3)	0.91 (0.73; 1.08)	0.89
Recurrent AOM						
No	640	46	39	-8 (-15; 0)	0.83 (0.65; 1.01)	
Yes	688	54	56	1 (-6; 9)	1.02 (0.89; 1.16)	0.10
Age and recurrent AOM*						
<2 years, no recurrent AOM	317	57	53	-4 (-15; 7)	0.93 (0.73; 1.13)	
<2 years, recurrent AOM	267	65	63	-1 (-13; 10)	0.98 (0.80; 1.16)	
≥2 years, no recurrent AOM	323	36	24	-11 (-21; -2)	0.68 (0.34; 1.02)	
≥2 years, recurrent AOM	421	48	51	3 (-7; 13)	1.06 (0.87; 1.26)	0.44

*AOM = Acute otitis media

Discussion

Combining data from the control groups of five randomised trials, we found that age less than two years and recurrent AOM were independent predictors of developing asymptomatic MEE at one month. Only a small beneficial effect of antibiotics was found in preventing the development of asymptomatic MEE in children aged two years and older without recurrent AOM. However, none of the (combined) subgrouping variables showed a statistically significant interaction effect with treatment.

The results are in agreement with the findings of two previous studies regarding the efficacy of antimicrobial prophylaxis for middle ear effusion, which also reported small beneficial effects [11,26]. The children in these previous studies, however, were not suffering from AOM when entering the trial. The results are also in agreement with current guidelines [27,28], which do not recommend prescribing antibiotics with the aim to prevent the development of MEE.

The major strength of our study is the large number of children that could be analysed. By re-analysing the data of five trials, we were able to include 1,328 children (of whom 660 were not initially treated with antibiotics), which gave us the power to predict the absolute risks of developing asymptomatic MEE in children with AOM and the opportunity to study whether these subgroups benefited more from treatment with antibiotics.

To appreciate our results, some possible limitations should also be discussed. First, we were only able to study asymptomatic middle ear effusion at one month, since this was the only point in time at which tympanograms were made in all available trials. However, we expect that after a longer period (e.g., three or six months) the spontaneous resolution rate will be higher resulting in an even smaller antibiotic effect.

Second, since myringotomy is considered to be unethical in the treatment of AOM in most western countries, MEE was defined as a type B tympanogram. This might have resulted in some misclassification. However, as both the sensitivity and specificity of a type B tympanogram are known to be quite high, that is, 81 and 74% respectively [29], we believe that the results accurately reflect the treatment effect.

Third, we did not study all possible subgroups, but we selected only those at higher risk of developing MEE at one month. The strength of this approach is that our prognostic analyses revealed only a few relevant subgroups, limiting the number of subgroup analyses performed and subsequent false positive findings (type I error)

due to multiple testing. Furthermore, other subgroups that might benefit more from treatment with antibiotics, for example, those with Down syndrome or cleft palate, could not be studied in this IPD meta-analysis as these subgroups were excluded from the individual trials. The experience of many clinicians that these subgroups of children might benefit more from antimicrobial prophylaxis has not yet been evidenced in randomised controlled trials. As the question whether to treat these children with antimicrobial prophylaxis is very relevant for clinical practice, future trials studying these specific subgroups seems justified.

Fourth, the children in the included trials were prescribed antibiotics for either seven or ten days. Prescribing antibiotics for a longer period might be more effective. However, in the current era of increasing antibiotic resistance, we should study the effectiveness regarding the recommended duration, and a period of five to ten days is recommended in all international AOM guidelines [30-33].

In conclusion, only a small beneficial effect of antibiotics was found in preventing the development of asymptomatic MEE in children aged two years and older without recurrent AOM. Due to this marginal effect and the known negative effects of prescribing antibiotics, such as the development of antibiotic resistance and side effects, we do not recommend prescribing antibiotics to prevent MEE. More research is, however, needed to identify relevant subgroups of children that have middle ear effusion that might benefit from other treatments.

References

1. Freid VM, Makuc DM, Rooks RN. Ambulatory health care visits by children: principal diagnosis and place of visit. Vital Health Stat 13 1998;1-23.
2. Schilder AG, Lok W, Rovers MM. International perspectives on management of acute otitis media: a qualitative review. Int J Pediatr Otorhinolaryngol 2004;68:29-36.
3. Rovers MM, Schilder AG, Zielhuis GA et al. Otitis media. Lancet 2004;363:465-73.
4. Alho OP, Oja H, Koivu M et al. Chronic otitis media with effusion in infancy. How frequent is it? How does it develop? Arch Otolaryngol Head Neck Surg 1995;121:432-6.
5. Rovers MM, Zielhuis GA, Bennett K et al. Generalisability of clinical trials in otitis media with effusion. Int J Pediatr Otorhinolaryngol 2001;60:29-40.
6. McCormick DP, Johnson DL, Baldwin CD. Early middle ear effusion and school achievement at age seven years. Ambul Pediatr 2006;6:280-7.
7. Paradise JL, Campbell TF, Dollaghan CA et al. Developmental outcomes after early or delayed insertion of tympanostomy tubes. N Engl J Med 2005;353:576-86.
8. Rovers MM, Straatman H, Ingels K et al. The effect of ventilation tubes on language development in infants with otitis media with effusion: A randomized trial. Pediatrics 2000;106:E42.
9. Rovers MM, Glasziou P, Appelman CL et al. Antibiotics for acute otitis media: a meta-analysis with individual patient data. Lancet 2006;368:1429-35.
10. Mandel EM, Casselbrant ML. Antibiotics for otitis media with effusion. Minerva Pediatr 2004;56:481-95.
11. Principi N, Marchisio P, Massironi E et al. Prophylaxis of recurrent acute otitis media and middle-ear effusion. Comparison of amoxicillin with sulfamethoxazole and trimethoprim. Am J Dis Child 1989;143:1414-8.
12. Mandel EM, Casselbrant ML, Rockette HE et al. Efficacy of antimicrobial prophylaxis for recurrent middle ear effusion. Pediatr Infect Dis J 1996;15:1074-82.
13. Greenland S, Finkle WD. A critical look at methods for handling missing covariates in epidemiologic regression analyses. Am J Epidemiol 1995;142:1255-64.
14. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15:361-87.
15. Higgins JP, Thompson SG, Deeks JJ et al. Measuring inconsistency in meta-analyses. BMJ 2003;327:557-60.

16. Rovers MM, Glasziou P, Appelman CL et al. Predictors of pain and/or fever at 3 to 7 days for children with acute otitis media not treated initially with antibiotics: a meta-analysis of individual patient data. Pediatrics 2007;119:579-85.
17. Hosmer D.W., Lemeshow S. Applied Logistic regression. New York: John Wiley & Sons, Inc., 1989;140-5.
18. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:29-36.
19. Brookes ST, Whitely E, Egger M et al. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. J Clin Epidemiol 2004;57:229-36.
20. Appelman CL, Claessen JQ, Touw-Otten FW et al. Co-amoxiclav in recurrent acute otitis media: placebo controlled study. BMJ 1991;303:1450-2.
21. Burke P, Bain J, Robinson D et al. Acute red ear in children: controlled trial of non-antibiotic treatment in general practice. BMJ 1991;303:558-62.
22. Damoiseaux RA, Van Balen FA, Hoes AW et al. Primary care based randomised, double blind trial of amoxicillin versus placebo for acute otitis media in children aged under 2 years. BMJ 2000;320:350-4.
23. Le Saux N, Gaboury I, Baird M et al. A randomized, double-blind, placebo-controlled noninferiority trial of amoxicillin for clinically diagnosed acute otitis media in children 6 months to 5 years of age. CMAJ 2005;172:335-41.
24. McCormick DP, Chonmaitree T, Pittman C et al. Nonsevere acute otitis media: a clinical trial comparing outcomes of watchful waiting versus immediate antibiotic treatment. Pediatrics 2005;115:1455-65.
25. Kaleida PH, Casselbrant ML, Rockette HE et al. Amoxicillin or myringotomy or both for acute otitis media: results of a randomized clinical trial. Pediatrics 1991;87:466-74.
26. Casselbrant ML, Kaleida PH, Rockette HE et al. Efficacy of antimicrobial prophylaxis and of tympanostomy tube insertion for prevention of recurrent acute otitis media: results of a randomized clinical trial. Pediatr Infect Dis J 1992;11:278-86.
27. Otitis media with effusion. Pediatrics 2004;113:1412-29.
28. NHG standaard otitis media met effusie. http://nhg.artsennet.nl/upload/104/standaarden/M18/start.htm.
29. Shekelle P, Takata G, Chan LS et al. Diagnosis, natural history, and late effects of otitis media with effusion. Evid Rep Technol Assess (Summ) 2002;1-5.

30. PRODIGY guidance: Otitis media acute. http://www.cks.library.nhs.uk/otitis_media_acute.
31. Diagnosis and management of acute otitis media. Pediatrics 2004;113:1451-65.
32. Appelman CL, Van Balen FA, van de Lisdonk EH et al. NHG-Standaard Otitis Media Acuta. Huisarts en wetenschap 2006;49:615-21.
33. Scottish Intercollegiate Guidelines Network. Diagnosis and management in childhood otitis media in primary care. A national clinical guideline. http://www.sign.ac.uk.

Chapter 7

Comparison of methods of handling missing data in individual patient data meta-analyses: an empirical example on antibiotics in children with acute otitis media

Abstract

Objective

To explore the influence of various methods of handling missing data (complete case analyses, single imputation within and over trials, and multiple imputations within and over trials) on the subgroup effects of individual patient data meta-analyses.

Methods

An empirical data set was used to compare these five methods regarding the subgroup results. Logistic regression analyses were used to determine interaction effects (regression-coefficients, standard errors, and p-values) between subgrouping variables and treatment. Stratified analyses were performed to determine the effects in subgroups (rate ratios, rate differences, and their 95% confidence intervals).

Results

Imputation over trials resulted in different regression-coefficients and standard errors of the interaction term as compared with imputation within trials and complete case analyses. Significant interaction effects were found for complete case analyses, and imputation within trials, whereas imputation over trials often showed no significant interaction effect.

Conclusion

In our empirical example, imputation within trials appears to be the most appropriate approach of handling missing data in individual patient data meta-analyses. Since association of covariates might differ across the included studies, imputation of missing data over trials might lead to bias. Therefore, despite the gain in statistical power, imputation over trials is not recommended.

Introduction

Individual patient data (IPD) meta-analyses, that is, meta-analyses that use IPD rather than simply the overall results of each trial, have been proposed as a major improvement in meta-analyses [1-3]. As IPD meta-analyses generally include more detailed data, they usually have greater statistical power to carry out informative subgroup analyses. Moreover, IPD meta-analyses allow accurate classification of patients based on individual characteristics and may, therefore, allow a more thorough assessment as to whether subgroup differences are spurious or not [1-3]. The assessment of subgroup effects is relevant for clinical practice, as most physicians would like to use the specific characteristics of a patient to decide on a patient's individual treatment [4-6].

Missing data complicate the analyses of IPD meta-analyses, as in any study. For IPD meta-analyses, the same approaches for handling missing data might be used as in a single study. However, not only the frequency of missing data but also the missingness process may vary across studies from which individual data are pooled. Different methods of handling missing data may, therefore, have a different impact on the results of IPD meta-analyses. Moreover, due to pooling in IPD meta-analyses, another type of missing data may occur in the pooled data set; namely, some variables might not be measured at all in a specific included trial. This type of missingness may further complicate the handling of missing data in IPD meta-analyses.

Conventional complete case analyses, that is, removing subjects with a missing value from the analyses, may reduce precision or produce biased results because only part of the data is used [7-9]. But the results may improve when missing data are imputed [7,10]. Common methods of imputation are single and multiple imputations. With single imputation the available data of subjects without missing values in the study are used in a regression model to estimate the distribution of the variables for which values are missing [11]. A random value of the estimated distribution will replace the missing values for the particular variable. With multiple imputations, regression techniques are used to estimate multiple distributions of the variable for which values are missing. Bootstrap techniques are used to draw a value from the estimated distributions to replace the missing value. Each missing value is, thus, imputed several times; consequently multiple data sets are created [11].

An essential difference between imputing data in a single study and imputation in IPD meta-analyses is that imputation in IPD meta-analyses can be performed

within the data set of each trial before these data are pooled into one data set or for the final data set after pooling (over trials). In particular, when IPD are handled as if they belong to one trial, it might seem logical to impute missing data over trials. However, most published IPD meta-analyses used the so-called two-stage approach, where each trial is analysed separately, using its raw data before the summary results from each trial are pooled and analysed with conventional meta-analyses techniques [12]. In this two-stage approach, imputation within trials might be most suitable. With imputation within trials the variables not measured in specific trials are not imputed. With imputation over trials, all data are imputed; that is, variables that were not measured in a specific trial are imputed on the basis of estimates from other trials. Furthermore, imputation of missing data over trials will result in a gain in statistical power. However, the imputation of missing variables over trials might be biased because some variables might be associated with each other in one of the included studies, whereas this association may differ for the other studies. This might result in biased effect estimates [13].

To determine the best strategy to handle missing data in IPD meta-analyses, we explored the impact of various methods of handling missing data on the subgroup effects of IPD meta-analyses. Using empirical data we compared complete case analyses, single imputation within trials, single imputation over trials, multiple imputations within trials, and multiple imputations over trials. Conventionally, significance ($p<0.05$) of the interaction term between treatment and subgrouping variables is considered conditional for studying treatment effects stratified for these subgroups [14]. Therefore, we assessed the impact of the five methods of handling missing data on the results of the interaction tests and the treatment effects in the stratified subgroup analyses.

Methods

For this study the data of an IPD meta-analysis were used, which evaluated the effect of antibiotics versus placebo or no treatment in children with acute otitis media (AOM) as described elsewhere [15]. In our empirical example, the primary outcome measure was pain, fever, or both at 3-7 days, and age, bilateral AOM, and otorrhea were the subgrouping variables.

Imputation techniques

Single (conditional mean) imputation was performed by use of the Missing Value Analysis function available in SPSS (SPSS for Windows, version 14.0; SPSS, Inc.). This method fits a prediction model for each variable with a missing value, the variable with a missing value is the outcome, and all other variables (i.e., all measured covariates, a variable for study, and the outcome variable) [15] are included as predictors. Missing values are replaced by estimates resulting from the prediction model [10,11].

Multiple imputation was done by use of the aregImpute algorithm [16] in S-plus (S-Plus® for windows, version 6.2, Lucent technologies Inc.). AregImpute is a technique that uses additive regression, bootstrapping and predictive mean matching for multiple imputation. Bootstrap techniques are used to impute missing data by drawing predicted values from a full Bayesian predictive distribution. Different bootstrap resamples are used for each of the multiple imputations, in which a flexible additive regression model is fitted on a sample with replacement from the original data. This model takes the uncertainty in the imputations into account that are caused by having to fit imputation models and is used to predict all of the original missing and nonmissing values for the target variable. Thereby, aregImpute uses predictive mean matching with optional weighted probability sampling [17,18]. The same variables, used as predictors in the single imputation process, were used for the multiple imputation process. The imputation process was repeated five times. Consequently five data sets were created.

Since the dichotomous variables were coded as 0 or 1, the imputed values of these variables were rounded to 0 or 1, and the imputed values of continuous variables were rounded to the nearest observed integer.

Although it is likely that a different process gives rise to missing data for each study, we assume similarity of the missingness process across studies.

Subgroup analyses

Fixed effect logistic regression analyses, including a dummy variable for study, were used to determine the interaction effect, that is, the regression-coefficient (β), standard error (SE), and the p-value of the interaction term: subgrouping variable $\times$ treatment. The interaction effects investigated were age $\times$ treatment, bilateral AOM $\times$ treatment, otorrhea $\times$ treatment, and age $\times$ bilateral AOM $\times$ treatment. Stratified subgroup analyses were performed to determine the treatment effects in the subgroups, that is, risk difference (RD), relative risk (RR), and their 95% confidence

intervals (CI)).

The multiple imputation process both within and over trials resulted in five data sets. We analysed these data sets separately. To combine the results of the interaction tests (i.e., β and SE of the interaction terms) and the stratified subgroup analyses (i.e., RR and RD) for the five data sets, we used the formulas of Rubin as described by Schafer [9] to combine point estimates taking into account the variance of the estimates within and between studies.

To examine the influence of the various methods to handle missing data, we compared the interaction effects and the subgroup effects among complete case analyses, single imputation within and over trials, and multiple imputations within and over trials.

Results

In table 7.1 the distribution of missing data of subgrouping variables and outcomes is presented. Of those variables that were available in all trials, only 3% of the values were missing. These missings could both be imputed within and over trials. The variables bilateral AOM and otorrhea were not measured in one (n=316) and four (n=1,118) trials, respectively. That is, for 19% and 66% of all the cases, bilaterality and otorrhea were missing. These missings could only be imputed over trials.

For the complete case analyses, information on otorrhea, bilateral AOM, age, and pain or fever or both at 3-7 days was available for 32%, 81%, 100%, and 97%, respectively. After single and multiple imputations within trials, information on otorrhea, bilateral AOM, age, and pain or fever or both at 3-7 days was available for 34%, 81%, 100%, and 100%, respectively. Expectedly, after single and multiple imputations over trials, the percentage of available information increased to 100 for all variables.

Table 7.1. Distribution of variables and missings in the datasets after five methods of handling missing data.

Variable	Complete cases			Imputation within trials						Imputation over trials					
				Single imputation			Multiple imputations			Single imputation			Multiple imputations		
	Placebo	*Antibiotics*	*Missing*	*Placebo*	*Antibiotics*	*Missing*	*Placebo*	*Antibiotics*	*Missing*	*Placebo*	*Antibiotics*	*Missing*	*Placebo*	*Antibiotics*	*Missing*
	N (%)	*N (%)*	*N (%)*	*N (%)*	*N (%)*	*N (%)*	*N (%)*	*N (%)*	*N (%)*	*N (%)*	*N (%)*	*N (%)*	*N (%)*	*N (%)*	*N (%)*
Subgroup															
Age <2 years	286 (35)	280 (34)	5 (0)	287 (35)	280 (34)	0 (0)	287 (35)	281 (34)	0 (0)	287 (35)	280 (34)	0 (0)	287 (35)	280 (34)	0 (0)
Bilateral AOM	220 (33)	236 (35)	316 (19)	220 (33)	236 (35)	315 (19)	220 (33)	236 (35)	315 (19)	260 (32)	267 (33)	0 (0)	247 (30)	261 (32)	0 (0)
Otorrhea	61 (22)	46 (18)	1,118 (68)	65 (23)	51 (19)	1,088 (66)	63 (22)	50 (19)	1,088 (66)	180 (22)	175 (21)	0 (0)	146 (18)	129 (16)	0 (0)
Outcome															
Pain and/ or fever at 3-7 days	293 (37)	189 (24)	50 (3)	303 (37)	198 (24)	0 (0)	300 (36)	194 (24)	0 (0)	308 (37)	197 (24)	0 (0)	306 (37)	195 (24)	0 (0)

The co-variates included in the imputation model were: age, gender, treatment, season, smoking in the household, having been breastfed, family history of acute otitis media (AOM), having siblings, recurrent AOM, ear-, nose-, throat-surgery, fever at baseline, bilateral AOM at baseline, pain at baseline, runny nose at baseline, normal tympanic membrane at baseline, red tympanic membrane at baseline, bulging tympanic membrane at baseline, crying at baseline, cough at baseline, vomiting at baseline, perforation at baseline, otorrhea at baseline, and fever or ear pain or both at 3-7 days.

Results of interaction tests

The results of the interaction test (i.e., regression-coefficient (β), standard error (SE) and p-value of the interaction terms) differed among the five methods of handling missing data (see Table 7.2). After single imputation within trials, the β and SE of the interaction terms were comparable to those of the complete case method. After multiple imputation within trials the SE of the interaction terms of both "bilateral AOM × treatment" and "otorrhea × treatment" were slightly larger than those of the complete case analyses, whereas the β and SE for the combined subgrouping variable age and bilateral AOM and the SE for age were smaller.

After single imputation over trials the β of the interaction term "age × treatment" was larger than the β's of the complete case analyses and imputation within trials. Both the β's of the other interaction terms and the SE of otorrhea were smaller after single imputation over trials as compared with complete case analyses and imputation within trials. After multiple imputation over trials, the SE of the interaction term "bilateral AOM × treatment" was larger than the SE of the complete case analyses and imputation within trials, whereas the β's of all interaction terms and the SE of age and bilateral AOM combined were considerably smaller.

The interaction effects of "bilateral AOM × treatment", "age × bilateral AOM × treatment", and "otorrhea × treatment" were significant ($p<0.05$) for both the complete case analyses and single imputation within trials. After multiple imputation within trials, the interaction effects reached borderline significance ($0.05<p<0.1$). After imputation over trials, only a significant effect ($p=0.05$) was found for bilateral AOM and age combined when single imputation over trials was used. For multiple imputations over trials none of the interaction terms were significant.

Results of stratified subgroup analyses

Only small differences were found between the five methods of handling missing data regarding the rate ratio (RR), rate difference (RD) and their 95% CI of the stratified subgroup analyses (Table 7.3); that is, the conclusions with respect to the clinically relevant subgroups remained similar for all methods. The stratified subgroup analyses showed relevant treatment effects for the subgroups bilateral AOM, age and bilateral AOM combined, and otorrhea for all methods. For example, the RD for children aged less than 2 years with bilateral AOM, were -27%, -25%, -26%, -25%, and -25% for complete case analyses, single imputation within, single imputation over, multiple imputations within, and multiple imputations over trials, respectively.

Table 7.2. Interaction results presented as beta-coefficient (β), standard error (SE), and p-values for the five methods of handling missing data.

Subgrouping variable	Complete Cases		Imputation within trials				Imputation over trials			
			Single imputation		Multiple imputations		Single imputation		Multiple imputations	
	β (SE)	p-value	β (SE)	p-value	β (SE)	p-value	β (SE)	p-value	β (SE)	p-value
Age (β4)*	-0.05 (0.23)	0.84	-0.05 (0.23)	0.83	-0.05 (0.05)	0.38	0.13 (0.22)	0.55	-0.01 (0.23)	0.98
Bilateral AOM (β4)*	-0.63 (0.26)	0.02	-0.58 (0.25)	0.02	-0.67 (0.33)	0.04	-0.27 (0.23)	0.25	-0.37 (0.37)	0.32
Bilateral AOM and Age (β5)†	-0.64 (0.26)	0.02	-0.58 (0.26)	0.02	-0.14 (0.08)	0.10	-0.48 (0.24)	0.05	-0.12 (0.15)	0.40
Otorrhea (β4)*	-0.94 (0.48)	0.05	-0.95 (0.46)	0.04	-0.98 (0.54)	0.07	-0.18 (0.26)	0.50	-0.42 (0.46)	0.36

* P (pain and/or fever at 3-7 days) = β0 + β1 * study + β2 * treatment + β3 * subgrouping variable + β4 * treatment * subgrouping variable

† P (pain and/or fever at 3-7 days) = β0 + β1 * study + β2 * treatment + β3 * subgrouping variable a + β4 * subgrouping variable b + β5 * treatment * subgrouping variable a * subgrouping variable b

The number of patients included in the logistic regression analyses for the five methods of handling missing data is equal to the number of patients included in the stratified subgroup analyses shown in table 3.

AOM = acute otitis media

Table 7.3. Results of the stratified subgroup analyses presented as risk differences (RD), relative risks (RR) and their 95%-confidence intervals for the five methods of handling missing data.

Subgrouping variable	Complete Cases			Imputation within trials						Imputation over trials					
				Single imputation			Multiple imputations			Single imputation			Multiple imputations		
	N	RD	RR	N	RD	RR	N	RD	RR	N	RD	RR	N	RD	RR
Age															
<2 years	541	-16 (-24; -7)	0.67 (0.46; 0.89)	567	-15 (-23; -7)	0.68 (0.47; 0.89)	567	-15 (-23; -7)	0.67 (0.40; 0.90)	569	-15 (-23; -7)	0.68 (0.47; 0.89)	567	-15 (-23; -7)	0.68 (0.47; 0.88)
≥2 years	1,047	-11 (-17; -6)	0.63 (0.42; 0.85)	1,076	-11 (-16; -6)	0.64 (0.43; 0.85)	1,076	-11 (-16; -6)	0.64 (0.43; 0.85)	1,074	-12 (-17; -7)	0.62 (0.41; 0.83)	1,076	-12 (-17; -7)	0.62 (0.41; 0.83)
Bilateral AOM															
No	840	-6 (-12; 0)	0.79 (0.57; 1.02)	872	-6 (-12; 0)	0.80 (0.58; 1.02)	872	-5 (-11; 0)	0.82 (0.53; 1.10)	1,116	-11 (-16; -6)	0.67 (0.47; 0.87)	1,135	-10 (-16; -5)	0.68 (0.36; 1.01)
Yes	437	-22 (-30; -13)	0.55 (0.29; 0.81)	456	-20 (-29; -11)	0.57 (0.32; 0.83)	456	-21 (-30; -13)	0.55 (0.29; 0.80)	527	-19 (-27; -11)	0.60 (0.37; 0.83)	508	-20 (-28; -12)	0.56 (0.15; 0.98)
Bilateral AOM and Age															
Unilateral, <2 years	249	-4 (-16; 8)	0.90 (0.57; 1.22)	261	-5 (-17; 6)	0.87 (0.55; 1.18)	261	-3 (-15; 8)	0.91 (0.48; 1.33)	283	-5 (-17; 6)	0.87 (0.57; 1.17)	284	-6 (-17; 6)	0.86 (0.46; 1.25)
Unilateral, ≥2 years	591	-7 (-14; 0)	0.73 (0.43; 1.04)	611	-6 (-13; 0)	0.76 (0.46; 1.06)	611	-6 (-13; 0)	0.76 (0.46; 1.06)	833	-13 (-18; -7)	0.59 (0.33; 0.84)	850	-12 (-18; -6)	0.61 (0.25; 0.97)
Bilateral, <2 years	259	-27 (-39; -15)	0.52 (0.21; 0.83)	273	-25 (-36; -14)	0.55 (0.25; 0.84)	273	-26 (-38; -15)	0.52 (0.21; 0.82)	286	-25 (-36; -14)	0.54 (0.25; 0.83)	283	-25 (-37; -14)	0.54 (0.19; 0.88)
Bilateral, ≥2 years	177	-13 (-26; 0)	0.64 (0.16; 1.12)	183	-12 (-25; 1)	0.65 (0.18; 1.12)	183	-14 (-27; 0)	0.63 (0.14; 1.11)	241	-11 (-23; 0)	0.70 (0.32; 1.08)	225	-13 (-25; -1)	0.63 (0.00; 1.03)
Otorrhea															
No	412	-15 (-24; -6)	0.66 (0.40; 0.92)	439	-14 (-23; -5)	0.66 (0.40; 0.93)	441	-14 (-23; -5)	0.66 (0.37; 0.95)	1,288	-12 (-17; -7)	0.66 (0.48; 0.83)	1,368	-12 (-16; -7)	0.67 (0.40; 0.94)
Yes	106	-37 (-55; -20)	0.41 (0.00; 0.94)	116	-36 (-53; -20)	0.39 (0.00; 0.93)	113	-37 (-54; -21)	0.39 (0.00; 0.94)	355	-17 (-27; -7)	0.60 (0.30; 0.90)	275	-21 (-32; -10)	0.53 (0.00; 1.12)

AOM = acute otitis media, RD = risk difference, RR = relative risk

Discussion

We examined the effects of different methods of handling missing data on the subgroup results in an IPD meta-analysis. The results of the five methods of handling missing data showed some remarkable differences with respect to the regression-coefficients, standard error and p-values of the interaction terms. However, no clinically relevant differences were found among the five methods regarding the subgroup effects.

For appreciation of our findings, some aspects deserve further discussion. First, it is remarkable that the interaction results differ among the various approaches. Despite the increased power, the interaction terms "bilateral AOM × treatment" and "otorrhea × treatment" were not significant after single and multiple imputation over trials, whereas the interaction term "age × bilateral AOM × treatment" was not significant after multiple imputation over trials. According to current recommendations [14], stratified subgroup analyses would not have been permitted for these subgroups. Subsequently, the clinically relevant subgroup effects found in the stratified analyses would have been missed.

Second, since some covariates might be associated with each other in one of the included studies, whereas this association may differ for the other studies, imputation over trials might lead to bias. Our results showed that the distribution of bilateral AOM was associated with age, that is, younger children were more frequently diagnosed with bilateral AOM. We could not, however, find any changes in this association after imputation. We did not, therefore, correct for this bias. However, other factors that were not measured in the trials might be associated with variables in the trials and might have introduced bias.

Third, it may seem illogical that single imputation performed better than multiple imputations in our analyses. Single imputation commonly results in an overestimation of the precision of the study associations because too low estimates of the standard error are obtained, while correct estimates of the standard error are obtained with multiple imputations [7,11]. However, certain methodological problems became apparent during the multiple imputation process. The variable "study" could not be included in the multiple imputation process. Because of the complete missingness of variables in some studies, the imputation algorithm could not determine what the distribution of the missing variable should have been in these studies when the variable study was included. Subsequently, it was not possible to take into account

differences in distributions of variables according to study. Variables strongly skewed toward 0 or 1 could not be included in the imputation process for the same reason. To explore the influence of study in the single imputation process, we performed sensitivity analyses both with and without study. Since the results of imputation with or without study gave similar results, we presented only the results without study. Furthermore, combining the results of the analyses of five data sets after multiple imputations appeared to be laborious and time-consuming. These methods should therefore be simplified and improved before they can be applied easily in IPD meta-analyses.

Fourth, to study the influence of two different adjustment methods we have performed some additional analyses. First, we have added the covariate "study" to the regression analyses as this is the best method to adjust for residual confounding. The results of these analyses were, however, in agreement with the earlier results. Second, we have analysed a fully saturated model. The results of these analyses were also in agreement with the earlier findings; notably, there was virtually no difference in the regression-coefficients of treatment and the interaction term, whereas the standard errors increased. It should be noted, however, that the odds ratios of the regression model are not the most relevant effect estimates in clinical practice. We, therefore, presented the stratified analyses without adjustment.

Finally, in the context of this empirical study, there is no "gold" standard available. Simulation studies are needed to further refine the relation between number and nature of missing values and to compare the results with a gold standard.

In conclusion, in our empirical example, imputation within trials appears to be the most appropriate approach of handling missing data in IPD meta-analyses. Despite a gain in statistical power, imputation over trials is not recommended, because it might lead to bias when associations between covariates differ across the included studies.

References

1. Clarke MJ, Stewart LA. Obtaining individual patient data from randomised controlled trials. In: Egger M, Smith GD, Altman DG, eds. Systematic reviews in health care: meta-analysis in context. London: BMJ Publishing group, 2001;109-21.
2. Oxman AD, Clarke MJ, Stewart LA. From science to practice; meta-analyses using individual patient data are needed. JAMA 1995;274:845-6.
3. Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. Eval Health Prof 2002;25:76-97.
4. Mc Alister FA. Applying the results of systematic reviews at the bedside. In: Egger M, Smith GD, Altman DG, eds. Systematic reviews in health care: meta-analysis in context. London: BMJ Publishing group, 2001;373-85.
5. Smith GD, Egger M. Incommunicable knowledge? Interpreting and applying the results of clinical trials and meta-analyses. J Clin Epidemiol 1998;51:289-95.
6. Wittes RE. Problems in the medical interpretation of overviews. Stat Med 1987;6:269-80.
7. Greenland S, Finkle WD. A critical look at methods for handling missing covariates in epidemiologic regression analyses. Am J Epidemiol 1995;142:1255-64.
8. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15:361-87.
9. Schafer JL. Analysis of Incomplete Multivariate Data. Chapman & Hall, 1997.
10. Donders AR, van der Heijden GJ, Stijnen T et al. Review: a gentle introduction to imputation of missing values. J Clin Epidemiol 2006;59:1087-91.
11. Moons KG, Donders RA, Stijnen T et al. Using the outcome for imputation of missing predictor values was preferred. J Clin Epidemiol 2006;59:1092-101.
12. Simmonds MC, Higgins JP, Stewart LA et al. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. Clin Trials 2005;2:209-17.
13. Thompson SG, Higgins JP. Treating individuals 4: can meta-analysis help target interventions at individuals most likely to benefit? Lancet 2005;365:341-6.
14. Brookes ST, Whitely E, Egger M et al. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. J Clin Epidemiol 2004;57:229-36.

15. Rovers MM, Glasziou P, Appelman CL et al. Antibiotics for acute otitis media: a meta-analysis with individual patient data. Lancet 2006;368:1429-35.
16. Harrell FE. The Hmisc library, 2002. http://hesweb1.med.virginia.edu/biostat/s/Hmisc.html.
17. Harrell, F. E. Multiple imputation using additive regression, bootstrapping and predictive mean matching. http://lib.stat.cmu.edu/S/Harrell/help/Hmisc/html/areglmpute.html.
18. Harrell, F. E. Multiple imputation using additive regression, bootstrapping and predictive mean matching. http://hosho.ees.hokudai.ac.jp/~kubo/Rdoc/library/Hmisc/html/areglmpute.html.

Chapter 8

General discussion

Individual patient data (IPD) meta-analyses, that is, meta-analyses that use the raw data of individual trials have been proposed as a major improvement in subgroup analyses. Our findings show that many IPD meta-analyses include results for subgroups, but the overall effects were still reported as the main outcome. Typically, however, IPD meta-analyses produce more accurate subgroup effects than conventional meta-analyses. We also found that, up to now, a wide variety of methods are used to study subgroups in published conventional and IPD meta-analyses. Furthermore, these methods are often not reported appropriately. It appears that there is no consensus about which methods should be used to study subgroup effects in IPD meta-analyses. Inconsistent and inappropriate use of subgroup analytic methods may result in invalid estimates of treatment effects in subgroups. This will consequently lead to incorrect and unjustified conclusions. For example, that treatment is not beneficial in a particular patient subgroup, with the consequence that treatment is withheld from those likely to benefit from it. When we are able to identify clinically relevant subgroups of patients that benefit most from a certain intervention, clinicians can make more adequate individualized decisions regarding the treatment. In this last chapter we, therefore, would like to discuss the results described in this thesis in further detail and suggest appropriate methods to study and report subgroup effects in IPD meta-analyses. Our discussion will focus on 1) two approaches in analysing IPD meta-analyses that may be used; 2) handling missing data in IPD meta-analyses, since due to pooling of individual patient data two types of missing data may occur in the pooled data set; 3) specification of subgrouping variables a priori, which may be difficult since often limited information regarding potential subgroup effects is available; 4) power of subgroup analyses in IPD meta-analyses, since the determination of subgroup effects is an important feature of IPD meta-analyses; 5) stratification of subgroups, which can be used to estimate treatment effects in these subgroups of patients; 6) appropriate subgroup effect estimates; 7) reporting of subgroup methods and effects. In the context of this discussion we propose some recommendations on how to analyse and report subgroup effects in IPD meta-analyses. The proposed methods will lead to more accurately analysed and more valid estimates of subgroup effects, which will subsequently improve individualized health care.

Table 8.1. Recommendations regarding the analyses and report of subgroup effects in individual patient data (IPD) meta-analyses.

Recommendations
1. Use the one-stage approach in analysing IPD meta-analyses
2. Impute missing values within trials to prevent confounding
3. Prognostic modelling techniques should be used to select subgrouping variables.
4. Power IPD meta-analyses according to the considered subgroup analyses
5. Use stratified subgroup analyses to determine relevant treatment effects in subgroups
6. Quantify subgroup effects as risk differences
7. Report the above mentioned methods and the results of the subgroup analyses properly

Use the one-stage approach in analysing IPD meta-analyses

We recommend to use the one-stage approach in analysing IPD meta-analyses, since with this approach 1) power is maintained, 2) it is possible to adjust for confounding, 3) more advanced modelling techniques can be used, 4) it is easier to perform stratified subgroup analyses, and 5) data sets with scarce events can be analysed more conveniently [1-4].

In the one-stage approach [5] to IPD meta-analyses, the individual patient data of each trial included in the meta-analyses are pooled and analysed as if all data belong to a single new trial. Using this approach there is no place to test for heterogeneity. An additional factor "study" can, however, be included in fixed effect regression analyses to adjust for putative differences between the included trials. Another approach in analysing IPD meta-analyses, which is being used most often, is the so called "two-stage approach" [5]. In this approach each trial is analysed separately using the IPD before the summary results from each trial are pooled using conventional meta-analyses techniques. In our empirical example, the one- and two-stage approach IPD meta-analyses gave similar results, probably because we used rather straightforward data analyses techniques. The two-stage approach in analysing IPD meta-analyses has been recommended to prevent direct comparison of patients from one trial with patients from another trial [5-8]. Moreover, some are afraid that randomisation will not be maintained when subgroup data from different trials are pooled, consequently, confounding might be introduced. However, this would already be the case when subgroups within trials are studied. Furthermore, we consider it possible to adjust for confounding in the one-stage approach by adding the variable "study" in the fixed effect regression model. In contrast, the one-stage approach in analysing IPD meta-analyses allows more flexibility in more complex situations. For instance, no power

is lost due to stratification by trial; data sets with scarce events can be analysed more conveniently; and more advanced modelling techniques can be used, such as logistic regression analyses, multivariate regression analyses, or Cox's survival analyses. We, therefore, recommend using the one-stage approach in analysing IPD meta-analyses.

Impute missing values within trials to prevent confounding

We recommend imputation of missing values within trials to prevent confounding. Missing data complicate the analyses of IPD meta-analyses, as in any study. Due to pooling in IPD meta-analyses two types of missing data may occur in the pooled data set. Either for some patients' for some reason specific values of a variable might not be measured, or a variable might not be measured at all in a particular included trial. Subsequently, imputation of missing values in IPD meta-analyses can be performed within the data set of each trial before data is pooled into one data set. Consequently, only the specific values of variables that were not measured for a patient are imputed. Imputation can also be performed for the final data set after pooling (over trials). Consequently, all missing values are imputed, and thus variables that were not measured in a specific trial are imputed based on estimates from other trials. By nature, imputation of missing values over trials will improve statistical power. However, some variables might be associated with other variables in one of the included studies. This association may differ across the included studies. If missing values, in particular of variables that were not measured in certain trials, are imputed over trials, then these missings might be imputed according to associations that exist within other trials. However, this association may not hold for the study in which they were missing. Consequently, the imputation of missing variables over trials might result in biased effect estimates [9,10]. We, therefore, recommend imputing missing values within trials to prevent confounding.

Prognostic modelling techniques should be used to select subgrouping variables

Subgrouping variables can be selected based on prior information that may come from three possible sources: a) results from previous trials, b) therapeutic hypotheses, and c) related biological hypotheses [10-13]. Pre-selection of subgrouping variables may be difficult when limited information regarding potential subgroup effects is available. An innovative approach in selecting subgrouping variables is to

use prognostic modelling techniques. By using this technique, multiple subgrouping variables can be combined in one analysis. That is important, since most health outcomes have multiple independent risk factors. It is, therefore, likely that analysis across combinations of subgroup variables will result in more valid and representative subgroup effects. Furthermore, it prevents multiple testing, and thus reduces the chance of false positive results [14,15]. In contrast, by using single covariates power may be lost, there is often multiple testing, and there is an increased chance of false positive findings.

The use of prognostic modelling techniques partly overcomes the problem of identifying possible subgrouping variables a priori, while it prevents the performance of post hoc subgroup analyses, which are particularly sensitive for false positive findings. We, therefore, believe that prognostic modelling techniques offer a major contribution in selecting subgrouping variables. Furthermore, we suggest to allow adding a maximum of two subgrouping variables based on previous results and hypotheses [10-13], in addition to the variables selected by prognostic modelling. Since some subgrouping variables might not be selected with prognostic modelling analyses, while there is a strong believe that this specific variable may modify the treatment effect.

Power IPD meta-analyses according to the considered subgroup analyses

Since the main purpose of IPD meta-analyses is to estimate subgroup effects, we recommend to power IPD meta-analyses according to the considered subgroup analyses. A larger sample size enhances the accuracy of subgroup analyses [8,11,16-18]. Furthermore, the possibilities to study subgroups in IPD meta-analyses may also depend on the case mix, the distribution, and the range of characteristics that may truly modify the effect of treatment. For instance, baseline risk is often not normally distributed in trials, that is, most trial outcomes occur in a relatively small number of high risk patients, while most patients have a lower than average risk [19,20]. It is thus important to ensure power in strata with high and low risk patients, to prevent false positive and false negative outcomes. We, therefore, recommend to power IPD meta-analyses accordingly.

Use stratified subgroup analyses to determine relevant treatment effects in subgroups

Stratified subgroup analyses should be performed to analyse the treatment effects in subgroups of patients to determine which patients may benefit more from treatment

than others [15]. Using the one-stage approach to IPD meta-analyses implies that it is necessary to perform a fixed effect regression model as recommended before. The subgrouping variables included in these analyses should be selected based on a prognostic model and a maximum of two other variables based on either previous results or biological reasoning. Others have recommended to perform stratified analysis only if a statistically significant treatment-by-subgroup interaction is demonstrated in a regression model [9,11,13,21]. We, however, found that clinically relevant subgroup effects are not always accompanied by significant interaction effects [22-29]. We, therefore, recommend performing stratified subgroup analyses without studying interaction effects using the one-stage approach and pre-specified subgrouping variables.

Risk difference should be used to report subgroup effects

We prefer using risk differences to report subgroup effects. Since a risk difference tells us more about the absolute treatment effect of an individual patient [20,30]. By contrast, a relative risk gives no information about the magnitude of the individual benefit; it only gives us the risk of an event occurring in the treated group relative to the risk of an event occurring in the control group [31]. The application of the odds ratio, which is often generated in logistic regression analyses, to the individual patient is even more complicated. Others have commented that a risk difference estimated in a randomised controlled trial or meta-analysis might be less generalisable to routine clinical practice than the relative risk, partly because patients enrolled in randomised controlled trials might not be representative of those patients in routine clinical practice [32]. However, by presenting risk differences for strata of baseline risk, the results will be more applicable to the individual patient.

Report the above mentioned methods and the results of the subgroup analyses properly

We recommend adequate reporting of 1) the above mentioned methods, and 2) the results of the subgroup analyses. The currently available QUORUM checklist for reporting meta-analyses focuses on main effects of conventional meta-analyses [33,34]. It is, however, important to report that a one-stage approach is used in analysing IPD meta-analyses; missing variables were imputed within trials; subgrouping variables were defined based on prognostic modelling techniques, previous results or hypotheses; which subgrouping variables were defined; analyses were powered with subgroup analyses in mind; regression analyses were used to determine the effect of

study; stratified subgroup analyses were performed; all estimated subgroup effects; and that clinical inferences regarding subgroup effects are discussed.

Further study

The results of this thesis are all based on empirical data, which have a number of obvious advantages in terms of the extent to which the variations in scenarios are realistic. A particular disadvantage is that there is little control over the underlying distributions. Inevitably, multiple parameters will vary simultaneously in an unknown way, making it difficult to draw final conclusions about their separated influences. Simulated data, on the other hand, allow complete control over both the underlying distributions and the nature of the alternations to the parameters. Future simulations studies, using the parameters from the empirical studies described in this thesis, should therefore be performed to further investigate: 1) whether confounding might be introduced when subgroups are studied, and if so, under which circumstances this occurs; 2) the relation between number and nature of missing values and to compare the results of single and multiple imputations within and over trials; 3) the validity and accuracy of prognostic modelling techniques in identifying subgrouping variables; 4) whether risk differences can be used in more complex analytical models, since its mathematical behaviour might be rather complicated. Furthermore, alternative innovative methods with respect to analyzing effect modification, such as a Bayesian approach, may more adequately quantify effect modification than the currently recommended analyses [22,35]. In a first attempt to use Bayesian analyses, we found that direct Bayesian subgroup modeling, that is, including a prior for the size and spread of effect in and between subgroups in WINBUGS, is not yet a straightforward possibility. Therefore, the Bayesian models have to be developed further to make this possible.

Conclusion

In this thesis we showed that IPD meta-analyses are indeed more accurate, flexible, and valid in analysing subgroup effects, and should therefore be performed. Future studies should adhere to our recommendations regarding the analyses and report of subgroup effects in IPD meta-analyses.

References

1. Higgins JP, Whitehead A, Turner RM et al. Meta-analysis of continuous outcome data from individual patients. Stat Med 2001;20:2219-41.
2. 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.
3. Turner RM, Omar RZ, Yang M et al. A multilevel model framework for meta-analysis of clinical trials with binary outcomes. Stat Med 2000;19:3417-32.
4. Whitehead A, Omar RZ, Higgins JP et al. Meta-analysis of ordinal outcomes using individual patient data. Stat Med 2001;20:2243-60.
5. Simmonds MC, Higgins JP, Stewart LA et al. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. Clin Trials 2005;2:209-17.
6. Deeks JJ. Systematic reviews of published evidence: miracles or minefields? Ann Oncol 1998;9:703-9.
7. Deeks JJ, Altman DG, Bradburn M.J. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger M, Smith GD, Altman DG, eds. Systematic reviews in health care: meta-analysis in context. London: BMJ Publishing group, 2001;285-312.
8. Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. Eval Health Prof 2002;25:76-97.
9. Yusuf S, Wittes J, Probstfield J et al. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. JAMA 1991;266:93-8.
10. Thompson SG, Higgins JP. Treating individuals 4: can meta-analysis help target interventions at individuals most likely to benefit? Lancet 2005;365:341-6.
11. Brookes ST, Whitley E, Peters TJ et al. Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives. Health Technol Assess 2001;5:1-56.
12. Higgins J, Thompson S, Deeks J et al. Statistical heterogeneity in systematic reviews of clinical trials: a critical appraisal of guidelines and practice. J Health Serv Res Policy 2002;7:51-61.
13. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. Ann Intern Med 1992;116:78-84.
14. Hayward RA, Kent DM, Vijan S et al. Multivariable risk prediction can greatly enhance the statistical power of clinical trial subgroup analysis. BMC Med Res Methodol 2006;6:18.

15. Kent DM, Hayward RA. Limitations of applying summary results of clinical trials to individual patients: the need for risk stratification. JAMA 2007;298:1209-12.
16. Berlin JA, Santanna J, Schmid CH et al. 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.
17. Lambert PC, Sutton AJ, Abrams KR et al. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. J Clin Epidemiol 2002;55:86-94.
18. Clarke MJ, Stewart LA. Systematic reviews: obtaining data from randomised controlled trials: how much do we need for reliable and informative meta-analyses? British medical journal 1994;309:1007-10.
19. Ioannidis JP, Lau J. The impact of high-risk patients on the results of clinical trials. J Clin Epidemiol 1997;50:1089-98.
20. Rothwell PM. Treating individuals 2. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. Lancet 2005;365:176-86.
21. Assmann SF, Pocock SJ, Enos LE et al. Subgroup analysis and other (mis)uses of baseline data in clinical trials. Lancet 2000;355:1064-9.
22. Simon R. Bayesian subset analysis: application to studying treatment-by-gender interactions. Stat Med 2002;21:2909-16.
23. Yusuf S, Zucker D, Peduzzi P et al. Effect of coronary artery bypass graft surgery on survival: overview of 10-year results from randomised trials by the Coronary Artery Bypass Graft Surgery Trialists Collaboration. Lancet 1994;344:563-70.
24. Loeffler M, Brosteanu O, Hasenclever D et al. Meta-analysis of chemotherapy versus combined modality treatment trials in Hodgkin's disease. International Database on Hodgkin's Disease Overview Study Group. J Clin Oncol 1998;16:818-29.
25. Krogsgaard K, Bindslev N, Christensen E et al. The treatment effect of alpha interferon in chronic hepatitis B is independent of pre-treatment variables. Results based on individual patient data from 10 clinical controlled trials. European Concerted Action on Viral Hepatitis (Eurohep). J Hepatol 1994;21:646-55.
26. Gueyffier F, Boissel JP, Boutitie F et al. Effect of antihypertensive treatment in patients having already suffered from stroke. Gathering the evidence. The INDANA (INdividual Data ANalysis of Antihypertensive intervention trials) Project Collaborators. Stroke 1997;28:2557-62.
27. Levy V, Katsahian S, Fermand JP et al. A meta-analysis on data from 575 patients with multiple myeloma randomly assigned to either high-dose therapy or conventional

therapy. Medicine (Baltimore) 2005;84:250-60.

28. Martinelli BF, Rovaris M, Johnson KP et al. Effects of glatiramer acetate on relapse rate and accumulated disability in multiple sclerosis: meta-analysis of three double-blind, randomized, placebo-controlled clinical trials. Mult Scler 2003;9:349-55.
29. Sin DD, Wu L, Anderson JA et al. Inhaled corticosteroids and mortality in chronic obstructive pulmonary disease. Thorax 2005;60:992-7.
30. Walter SD. Choice of effect measure for epidemiological data. J Clin Epidemiol 2000;53:931-9.
31. Rothman KJ. Epidemiology: an introduction. Oxford University Press, Inc., 2002.
32. Rothwell PM, Mehta Z, Howard SC et al. Treating individuals 3: from subgroups to individuals: general principles and the example of carotid endarterectomy. Lancet 2005;365:256-65.
33. Moher D, Cook DJ, Eastwood S et al. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. Lancet 1999;354:1896-900.
34. Shea B, Dube C, Moher D. Assessing the quality of reports of systematic reviews: the QUOROM statement compared to other tools. In: Egger M, Smith GD, Altman DG, eds. Systematic Reviews in Health Care. Meta-analysis in context. London: BMJ Publishing Group, 2001;122-39.
35. Dixon DO, Simon R. Bayesian subset analysis. Biometrics 1991;47:871-81.

Summary

Meta-analyses that use individual patient data (IPD), that is, the raw data of individual trials, rather than simply the overall results of each trial have been proposed as a major improvement in subgroup analyses. Since IPD meta-analyses often include more detailed data, they usually have greater statistical power to carry out informative subgroup analyses. By using the individual patient data the flexibility of subgroup analyses may be enhanced. Consequently, the estimated subgroup effects may be less influenced by misclassification and (ecological) bias. IPD meta-analyses, therefore, allow a more thorough assessment as to whether differences in treatment effects between subgroups are spurious or not. The merits of the IPD meta-analytic approach to study subgroup effects have, however, not yet been studied properly. Thus far, there is no consensus on the best methodology of analysing subgroup effects in IPD meta-analyses. The general objective of this thesis, therefore, is to study the accuracy, flexibility, and validity of IPD meta-analyses in analysing subgroup effects.

In **chapter 2**, **3** and **4** we described the results of an extensive literature study in which we tried to identify all published IPD meta-analyses and their related conventional meta-analyses. In **chapter 2** we determined whether IPD meta-analyses are used to perform subgroup analyses, and studied whether the analytical methods regarding subgroup analyses differed between conventional and IPD meta-analyses. Of all 171 identified IPD meta-analyses and 102 conventional meta-analyses, 136 (80%) and 46 (45%) presented subgroup analyses. For 35 IPD meta-analyses and 37 "matched" conventional meta-analyses, subgroup analytic methods could be compared. Many IPD meta-analyses performed subgroup analyses, but the overall treatment effects were more frequently emphasised than the subgroup effects. To study subgroups, a wide variety of analytical methods was used in both conventional and IPD meta-analyses. We, therefore, concluded that the use and reporting of appropriate methods for subgroup analyses should be promoted.

In **chapter 3** we described which characteristics appear to be associated with reported subgroup effects. Multivariate regression analyses were used to identify characteristics that were associated with reporting subgroup effects in 171 IPD meta-analyses retrieved by a comprehensive literature search. Of the 171 IPD meta-analyses retrieved, 136 (80%) reported subgroup effects. Our results showed that the choice for reporting subgroup analyses in IPD meta-analyses is driven by a larger sample size (odds ratio (OR) 1.68; 95% confidence interval (CI) 1.13; 2.52), testing for heterogeneity (OR 3.48; 95% CI 1.34; 9.05), the use of a fixed effect

model (OR 7.22; 95% CI 1.98; 26.31), and testing for interaction effects (OR 10.35; 95% CI 2.24; 47.74). These factors accurately identified whether subgroup analyses were reported for 82% of the 171 included IPD meta-analyses. Despite existing recommendations regarding subgroup analytic methods in meta-analyses our data showed that different approaches were used. From this it can be concluded that existing recommendations are not followed.

In **chapter 4** we compared the subgroup effect estimates of published individual patient data and conventional meta-analyses that address the same clinical question. In total, 35 IPD meta-analyses and 37 conventional meta-analyses could be paired with respect to domain, objective, and one or more outcome variables. The types of subgroups studied appeared to vary between conventional meta-analyses (CMA) and IPD meta-analyses (IPDMA). IPD meta-analyses focused more on analyses of patient characteristics (N_{IPDMA}=23 (66%); N_{CMA}=10 (27%)) and disease characteristics (N_{IPDMA}=24 (69%); N_{CMA}=14 (38%)), whereas subgroup analyses in conventional meta-analyses were often based on treatment characteristics (N_{IPDMA}=15 (43%); N_{CMA}=27 (73%)) or outcome characteristics (N_{IPDMA}=4 (11%); N_{CMA}=13 (35%)). IPD meta-analyses studied 192 subgroups and conventional meta-analyses 149; 23 pairs of subgroup effect estimates could be made. Of the subgroup comparisons related to characteristics of treatment and outcome, 22 (96%) were in the same direction, and for 14 (61%) a similar conclusion was reached. We concluded that IPD meta-analyses appear to provide more relevant clinical information, since they more frequently reported subgroups based on individual patient and disease characteristics.

In **chapter 5** we compared subgroup effects of 1) conventional meta-analyses using summary statistics derived from published data, 2) two-stage approach to IPD meta-analyses where summary statistics derived from IPD are used (IPDMA-2), and 3) one-stage approach to IPD meta-analyses where the IPD is pooled into a single data set (IPDMA-1). The data of six trials (n=1,643) on the effectiveness of antibiotics in children with acute otitis media were used. In the conventional meta-analysis only two of the six studies were included because only these reported on relevant subgroup effects. The conventional meta-analysis showed larger (age <2 years) or smaller (age ≥2 years) subgroup effect estimates and wider CIs than both IPD meta-analyses (age <2 years: risk difference $(RD)_{CMA}$ -21%, $RD_{IPDMA-1}$ -16%, $RD_{IPDMA-2}$ -15%; age ≥2 years: RD_{CMA} -5%, $RD_{IPDMA-1}$ -11%, $RD_{IPDMA-2}$ -11%). The most important reason for these discrepant results is that the two studies included in the conventional meta-analyses reported outcomes that were different both from each

other and from the IPD meta-analyses. This empirical example, therefore, shows that conventional meta-analyses do not allow proper subgroup analyses, whereas IPD meta-analyses produce more accurate subgroup effects. We found no differences between the one- and two-stage meta-analytic approaches.

In **chapter 6** we showed an empirical example of an IPD meta-analysis, on the (subgrouping) effects of antibiotics in preventing the development of asymptomatic middle ear effusion (MEE) in subgroups of children. The data of five randomised controlled trials were included in the individual patient data meta-analysis (total 1,328 children aged 6 months to 12 years). The primary outcome was MEE (defined as a type B tympanogram) at one month. The overall relative risk (RR) of antibiotics in preventing the development of asymptomatic MEE after one month was 0.9 (95% CI 0.8; 1.0; p=0.19). Independent predictors of developing asymptomatic MEE were age less than two years and recurrent AOM. No significant interaction effects with treatment were found. Due to the marginal effect of antibiotics on developing asymptomatic MEE and the known negative effects of prescribing antibiotics, such as the development of antibiotic resistance and side effects, we do not recommend antibiotics to prevent MEE.

In **chapter 7** we studied the influence of various methods of handling missing data (complete case analyses, single imputation within and over trials, and multiple imputations within and over trials) on the subgroup effect estimates of IPD meta-analyses. An empirical data set was used to compare these five methods regarding the subgroup results. Logistic regression analyses were used to determine interaction effects (regression-coefficients, standard errors, and p-values) between subgrouping variables and treatment. Stratified analyses were performed to determine the effects in subgroups (rate ratios, rate differences, and their 95% confidence intervals). Imputation over trials resulted in different regression-coefficients and standard errors of the interaction term as compared to imputation within trials and complete case analyses. Significant interaction effects were found for complete case analyses, and imputation within trials, whereas imputation over trials often showed no significant interaction effect. Imputation of missing data over trials might lead to bias, since association of covariates might differ across the included studies. Therefore, despite the gain in statistical power, imputation over trials is not recommended. In our empirical example, imputation within trials appears to be the most appropriate approach of handling missing data in IPD meta-analyses.

In **chapter 8** we discussed the findings reported in this thesis and provided

recommendations on how and when treatment effects in subgroups should best be analysed and reported in IPD meta-analyses. We showed that IPD meta-analyses are indeed more accurate, flexible, and valid in analysing subgroup effects, and should therefore be performed. Future studies should adhere to our recommendations regarding the analyses and report of subgroup effects in IPD meta-analyses.

Samenvatting

In conventionele meta-analyses worden de gepubliceerde resultaten van vergelijkbare klinische studies gebundeld en opnieuw geanalyseerd. Een dergelijke meta-analyse leidt tot een preciezere gemiddelde schatting van de effectiviteit van een bepaalde interventie (behandeling) voor een gemiddelde groep patiënten. Een clinicus ziet in de dagelijkse praktijk echter geen gemiddelde patiënt maar een individu. Om deze individuen adequaat te kunnen behandelen zouden we eigenlijk rekening willen houden met bepaalde kenmerken van patiënten. Hierbij willen we ook onderscheid maken tussen subgroepen van patiënten die wel en subgroepen van patiënten die geen baat hebben bij de interventie. Meta-analyses die gebruik maken van de individuele (originele) patiënten gegevens, zogenaamde individuele patiënten data (IPD) meta-analyses, bieden de mogelijkheid om de invloed van patiëntkenmerken te bestuderen. Dit komt doordat IPD meta-analyses meestal meer gedetailleerde gegevens van de studies includeren (zoals ongepubliceerde gegevens en gegevens van de behandelingseffecten gestratificeerd naar patiëntkenmerken) waardoor er meer mogelijkheden zijn om informatieve subgroepanalyses uit te voeren. Met behulp van IPD meta-analyses is het mogelijk om meer gedegen te beoordelen of verschillen in behandelingseffect tussen subgroepen vals of echt zijn. Tot nu toe is echter niet goed onderzocht in hoeverre er middels IPD meta-analyses werkelijk nauwkeuriger en meer valide subgroepeffecten worden verkregen. Ook is er nog geen consensus bereikt over de beste methoden om subgroepeffecten te bestuderen. Het doel van dit proefschrift is daarom de nauwkeurigheid, de flexibiliteit en de validiteit te bestuderen van IPD meta-analyses die subgroepeffecten analyseren.

In de **hoofdstukken 2**, **3** en **4** beschrijven we de resultaten van een uitgebreide literatuur-studie waarin we probeerden om alle gepubliceerde IPD meta-analyses en een vergelijkbare conventionele meta-analyse te identificeren. Vervolgens bepaalden we in **hoofdstuk 2** of IPD meta-analyses worden gebruikt om subgroepanalyses uit te voeren. Bovendien bestudeerden we of de methoden die voor het uitvoeren van subgroepanalyses worden gebruikt verschillen tussen IPD meta-analyses en conventionele meta-analyses. Van alle 171 geïdentificeerde IPD meta-analyses en 102 conventionele meta-analyses, waren er respectievelijk 136 (80%) en 46 (45%) die subgroep analyses presenteerden. Desondanks lag de nadruk bij de rapportage van veel IPD meta-analyses op de gemiddelde interventie-effecten. Voor 35 IPD meta-analyses en 37 gepaarde conventionele meta-analyses, konden de methoden van de subgroep analyses worden vergeleken. Het bleek dat de onderzoekers verschillende methoden gebruikten om subgroepen te analyseren. Wij concludeerden daarom

dat het gebruik en het rapporteren van de meest geschikte methoden voor het analyseren van subgroepen in IPD meta-analyses moet worden gestimuleerd.

In **hoofdstuk 3** beschreven we welke kenmerken met het rapporteren van subgroep effecten blijken te worden geassocieerd. Onze resultaten laten zien dat de keuze om subgroep effecten te rapporteren in IPD meta-analyses wordt gedreven door een grotere onderzoekspopulatie, het testen van heterogeniteit (ongelijkheid in bijvoorbeeld de patiëntenpopulaties), het gebruik van een fixed-effect model, en het testen op interactie-effecten (zijn er verschillen in behandelingseffect te verwachten tussen subgroepen van patiënten). Deze factoren werden bij 82% van de 171 IPD meta-analyses die subgroep effecten rapporteren gebruikt. Ondanks bestaande aanbevelingen voor methoden voor het analyseren van subgroepen in meta-analyses laten onze resultaten zien dat afwijkende analysetechnieken werden gebruikt. Hieruit kunnen we concluderen dat bestaande aanbevelingen niet worden gevolgd.

In **hoofdstuk 4** vergeleken we de geschatte subgroepeffecten van gepubliceerde IPD meta-analyses en conventionele meta-analyses met een vergelijkbare onderzoeksvraag. In totaal konden 35 IPD meta-analyses en 37 conventionele meta-analyses worden gepaard met betrekking tot hun domein (patiënten op wie de onderzoeksvraag/doel van toepassing is), determinant (factor die bv. het ontstaan van een ziekte mede bepaald) en een of meer uitkomstvariabelen. De typen subgroepen die in respectievelijk de conventionele meta-analyses (CMA) en individuele patiënten data meta-analyses (IPDMA) werden bestudeerd, bleken te verschillen. Binnen IPD meta-analyses bleken voornamelijk subgroepen op basis van patiëntkenmerken (bv. leeftijd, geslacht) en ziektekenmerken (bv. ernst van de ziekte) bestudeerd te worden. Binnen conventionele meta-analyses bleken voornamelijk subgroepen op basis van kenmerken van een interventie (bv. dosis van medicijn) of van de uitkomst (bv. tijdstip van de uitkomst) bestudeerd te worden. In totaal werden binnen de 35 door ons geselecteerde IPD meta-analyses 192 subgroepen bestudeerd, en binnen de 37 conventionele meta-analyses waren dat er 149. Uiteindelijk konden er 23 subgroep effecten, die voornamelijk betrekking hadden op kenmerken van de interventie en/of de uitkomst, vergeleken worden. Bij 22 (96%) paren waren beide effecten in gelijke richting (dus beiden positief of negatief), en bij 14 (61%) paren was ook de conclusie vergelijkbaar. Wij concludeerden op basis van deze resultaten dat IPD meta-analyses meer relevante klinische informatie lijken te geven, omdat ze vaker subgroepen rapporteren die betrekking hebben op individuele patiënt- en

ziektekenmerken.

In **hoofdstuk 5** vergeleken we subgroep effecten van 1) een conventionele meta-analyse waarbij de resultaten van gepubliceerde studies werden gebruikt, 2) een twee-staps benadering van een IPD meta-analyse waarbij samengevatte resultaten worden afgeleid van individuele patiënten data (IPDMA-2), 3) een één-staps benadering van een IPD meta-analyse waarbij de individuele patiënten data worden samengevoegd in één enkele dataset (IPDMA-1). Voor deze studie werd de data van zes gepubliceerde studies (n=1643 patiënten) naar de effectiviteit van antibiotica bij kinderen met een oorontsteking gebruikt. In de conventionele meta-analyse konden maar twee van de zes studies worden geïncludeerd omdat alleen deze studies relevante subgroepeffecten rapporteerden.Voor de conventionele meta-analyse waren de behandelingseffecten voor de subgroepen met kinderen jonger dan twee jaar groter en voor de subgroepen met kinderen ouder dan twee jaar kleiner dan in beide IPD meta-analyses. De belangrijkste reden voor de verschillende resultaten is dat de gerapporteerde uitkomstvariabelen voor de twee studies die zijn geïncludeerd in de conventionele meta-analyse van elkaar verschilden. Bovendien verschilden ze ook van de uitkomstvariabele die werd gebruikt in beide IPD meta-analyses. Dit (empirische) voorbeeld laat daarom zien dat het bij conventionele meta-analyses moeilijk is om goede subgroep analyses uit te voeren, terwijl middels IPD meta-analyses nauwkeurige subgroepeffecten verkregen kunnen worden. We vonden echter geen verschil tussen de één- en twee-staps benadering van IPD meta-analyses.

In **hoofdstuk 6** is een IPD meta-analyse uitgevoerd naar het effect van antibiotica op het voorkomen van het ontstaan van effusie (vocht achter het trommelvlies) in het middenoor bij kinderen met een oorinfectie. De gegevens van vijf studies werden geïncludeerd in de IPD meta-analyse (in totaal 1328 kinderen in de leeftijd van 6 maanden tot 12 jaar). De primaire uitkomst was effusie in het middenoor na één maand. Het bleek dat de kans op het ontstaan van effusie in het middenoor in de groep die antibiotica kreeg ongeveer 10% kleiner was ten opzichte van de groep die geen antibiotica kreeg. Kinderen jonger dan 2 jaar die al eerder een oorontsteking gehad hadden bleken een grotere kans te hebben om effusie te ontwikkelen dan oudere kinderen die nog nooit een oorontsteking gehad hadden. Het effect van de behandeling was echter niet verschillend voor deze groepen. Gezien het marginale effect van antibiotica in het voorkomen van het ontstaan van effusie in het middenoor en de bekende negatieve effecten van het voorschrijven van antibiotica,

zoals de ontwikkeling van antibiotica resistentie en bijwerkingen, adviseren we geen antibiotica voor te schrijven om effusie te voorkomen.

In **hoofdstuk 7** bestudeerden we de invloed van verschillende methoden om met missende data om te gaan (complete data analyse, enkelvoudige imputatie (binnen en over studies), en meervoudige imputatie (binnen en over studies) op subgroep effecten in een IPD meta-analyse. Een empirische dataset werd gebruikt om de vijf methoden te vergelijken met betrekking tot de subgroep resultaten. De verschillende analyses lieten zien dat het imputeren van missende waarden over studies andere interactie-effecten opleverden dan de complete data analyse en het imputeren van missende waarden binnen studies. Terwijl de gestratificeerde subgroep effecten voor alle methoden vergelijkbaar waren. Het imputeren van missende waarden over studies kan leiden tot een vertekening van het effect, omdat de samenhang tussen twee of meer variabelen kan verschillen over studies. In ons empirisch voorbeeld lijkt imputatie binnen trials de beste methode om met missende data in IPD meta-analyses om te gaan.

In **hoofdstuk 8** bediscussieerden we de bevindingen gerapporteerd in dit proefschrift en presenteren we aanbevelingen over hoe en wanneer behandelings effecten in subgroepen het best kunnen worden geanalyseerd en gerapporteerd in IPD meta-analyses. We kunnen concluderen dat IPD meta-analyses inderdaad nauwkeuriger, flexibeler en meer valide zijn in het analyseren van subgroep effecten. Door meer van deze meta-analyses uit te voeren is het mogelijk om nog meer "zorg op maat" te leveren en om die reden zouden er dus meer van deze analyses uitgevoerd moeten worden.

Dankwoord

In de afgelopen drieëneenhalf jaar heb ik ongelofelijk veel geleerd. In de eerste plaats natuurlijk veel over de methodologie die ik in dit proefschrift heb beschreven en de epidemiologie in het algemeen, maar op de tweede plaats (en niet onbelangrijk) ook zeer veel over mezelf. Ik had het niet willen missen! Ik wil daarom ook alle personen bedanken die hebben bijgedragen aan de tot stand koming van dit proefschrift.

Geachte Co-promotores, beste Maroeska en Geert, wat had ik zonder jullie moeten beginnen? Maroeska, ons allereerste (sollicitatie) gesprek herinner ik me vooral als gezellig. Maar toen ik werkelijk begon was het voor ons beiden toch nog even aftasten wie we nou precies voor ons hadden, al hadden we snel een prettige werksfeer gecreëerd. Prettige en soms intense discussies volgden over het hoe, wie, wat, waar en waarom van de methodologie van subgroep analyses en IPD meta-analyses. Geert, je zou me in eerste instantie alleen begeleiden tijdens het eerste zwangerschapsverlof van Maroeska, maar uiteindelijk vonden we je bijdragen zo waardevol dat je bent gebleven, en bovendien kwam er nog een tweede verlof. Ook de samenwerking met jou heb ik als prettig en zeer waardevol ervaren. Je was een belangrijke factor in de vele discussies. Soms was ik meer een toeschouwer als de discussies tussen Maroeska en jou sneller gingen dan ik kon bijhouden. Maar ik denk dat ik op het eind zo af en toe tussen beiden kon komen. Verder heb ik altijd gedacht dat ik koppig was maar daarin moet ik jullie toch echt als mijn meerdere erkennen. Bedankt voor jullie vertrouwen, steun en motivatie, dat had ik soms wel nodig.

Geachte promotor, beste Rick, onze eerste ontmoeting in mei 2004 verliep niet zo soepel. Het was tijdens mijn tweede sollicitatiegesprek waarin het vuur me zeer na aan de schenen werd gelegd. Ik dacht dit is een verloren zaak, maar u zag genoeg lichtpuntjes om mij aan te nemen. Hiervoor wil ik u bij deze nog hartelijk bedanken. Verder verraste u mij in de afgelopen jaren wel vaker met uw bijdragen in de discussies en de suggesties om mijn artikelen te verbeteren.

Collega-promovendi en met name mijn kamergenoten van 6.119 en 6.103, bedankt voor alle gezellige werk-uurtjes, goede discussies, lunches en uitjes. Zonder jullie was het promoveren niet zo leuk geweest.

Tot slot wil ik ook alle andere niet nader bij naam genoemde collega's in het Julius bedanken voor de gezelligheid.

I would like to thank Cees Appelman, Peter Burke, Roger Damoiseaux, Paul Little, David McCormick, and Nicole Le Saux for providing the data of their trials on the effectiveness of antibiotics in children with acute otitis media and for commenting

on my manuscripts.

I would like to thank Keith Abrams for giving me the opportunity to study Bayesian modelling techniques at the department of Health Sciences at Leicester University. Unfortunately, we could not finish our investigation regarding a Bayesian subgroup model in time to include it in this thesis. But I hope we will finish it some day. Thanks for all your time, while I know you were very busy. Thanks for your invitation to speak together at the Royal Statistical Society in London; I had much fun doing it. My visit was a good experience, and I have learned a lot.

I would also like to thank Paul Glasziou for his careful reading and useful suggestions on my manuscripts. It always amazed me how quick you responded to all my e-mails.

Naast mijn begeleiders en collega's wil ik ook mijn lieve vrienden en familie bedanken. Vaak hadden jullie geen idee waar ik nou mee bezig was. Al mijn uitleg ten spijt. Misschien dat enkelen van jullie wat meer begrijpen als ze dit proefschrift (de nederlands samenvatting dus) lezen. Maar gelukkig hangt onze vriendschap daar niet van af. Bedankt voor jullie gezelligheid, de vele borrels, en de interesse in mijn onderzoek.

Lieve Annemarie, we begonnen als huisgenoten maar werden al snel vriendinnen. We hadden zoveel te bespreken en konden allebei ons ei bij elkaar kwijt. En dan ben je afgestudeerd en ga je promoveren. Ik vond dat een moedige stap maar ook goed bij je passen. Vervolgens, bijna een jaar later begin ik er zelf ook aan. Vanaf dat moment konden we ook onze promotie-frustraties bij elkaar kwijt. Bedankt voor je gezelligheid, je luisterend oor en je adviezen. Hopelijk rond ook jij je promotie binnenkort met goed gevolg af. Ik ben blij dat je mijn paranimf wilt zijn.

Lieve Eveline, ook wij begonnen als huisgenoten en tevens als dispuutgenoten maar inmiddels zijn we vooral goede vriendinnen. Ik ben ontzettend blij dat ik jouw zo goed hebben leren kennen. Onze vele uren samen voor de tv met weer een of andere film en popcorn, onze tutavondjes, onze vele gesprekken en onze gezamenlijke uitjes en vakanties met onze mannen. Ik hoop dat we onze vriendschap nog heel lang in stand kunnen houden. Bedankt voor je samenwerking om een mooie kaft te maken, hopelijk rond je je studie snel af en kun je nog meer mensen helpen met je creatieve ideeën. Ik ben blij dat je mijn paranimf wilt zijn.

Lieve Ineke, Klaas, Bas en Hanneke, bedankt voor de gezelligheid jullie interesse en jullie steun. Het is altijd gezellig als we weer eens een weekendje langs komen, of

als we er een paar dagen tussenuit gaan met zijn allen.

Willem-Jan, lief broertje, bedankt dat je er altijd voor me bent. Ik kan niet zonder je gezelligheid! Ik ben blij met onze goede band, dat we overal over kunnen praten en onze ervaringen met van alles en nog wat kunnen delen. Misschien kun je uit dit boekje ook wat motivatie halen om je eigen studie snel af te ronden?

Lieve paps en mams, bedankt dat jullie er altijd voor me zijn. Zonder jullie had ik dit echt niet voor mekaar gekregen. Vanaf het moment dat ik begon te studeren in Wageningen wilden jullie graag op de hoogte blijven. Of het nou mijn studie of mijn vele activiteiten daaromheen betrof. Wat betreft de studie werd dat steeds moeilijker, en in de afgelopen drie jaar snapten jullie er vaak niets meer van. Jullie steun, vertrouwen, gezelligheid en goede zorgen hebben me gemaakt tot wat ik nu ben. Ik hoop nog lang van jullie steun en gezelligheid te mogen genieten.

Tot slot, lieve Tom, mijn grote steun en toeverlaat in gezellige en moeilijke tijden. Ik heb het je de afgelopen maanden niet makkelijk gemaakt. Eerst vertrek ik naar Leicester en moeten we elkaars gezelschap zoeken via internet en de telefoon. En vervolgens ben ik niet altijd even goedgehumeurd vanwege de stressvolle laatste maanden, waarin ik de gedachte aan de afronding van mijn proefschrift niet vaak los kon laten. Jouw vertrouwen in mij helpt me om door te zetten, dankjewel.

Curriculum Vitae

Laura Koopman was born on August 29th, 1980 in Rotterdam, the Netherlands. In 1998 she graduated at the o.s.g. Willem Blaeu in Alkmaar and received a VWO degree. She started her university training in Nutrition and Health at Wageningen University that same year. As part of this study two research projects were conducted. Her first project at the department of Metabolism & Genomics of Wageningen University concerned the expression and regulation of PPAR-α, -β, and -γ in human blood cells. For the second project, an intervention study into web-based nutrition counselling intervention on the risks of cardiovascular diseases, she undertook an internship at the Family Medicine Centre, Hotel Dieu Hospital in Kingston (Ontario), Canada. When she returned to Wageningen she performed data-analyses on part of the data she collected during her internship, at the department of Nutrition and Epidemiology of Wageningen University. In March 2004 she obtained her Master of Science degree in Nutrion and Health. In July 2004 she started the work described in this thesis at the Julius Center for Health Sciences and Primary Care (supervised by Prof. Dr. DE Grobbee, Dr. MM Rovers and Dr. GJMG van der Heijden). She obtained her Master of Science degree in Clinical Epidemiology at the Netherlands Institute for Health Sciences, Erasmus Medical Center Rotterdam, in June 2006. From April until June 2007 she performed part of the work for her thesis at the department of Health Sciences, at the University of Leicester, under supervision of Prof. Dr. KR Abrams. As of November 2007, she is working as a researcher at NIVEL, the Netherlands institute for research in health care.

Appendix

Search strategy

Table 1. Search strategy for individual patient data (IPD) meta-analyses for varying databases

	Search strategy "IPD meta-analyses"	
Database	Search strategy	Number
PubMed	#1 Shojania & Bero	
	#2 Our search	
	#3 Dutch Cochrane trial search	
	#4 Individual patient data	
	#5 #1 OR #2	
	#6 #5 AND #3 AND #4	1,193
Embase	MeSH terms: trials, meta-analysis, individual patient data	1,02
Web of Science	#1 Meta-analysis (topic)	
	#2 Trial (topic)	
	#3 Individual patient data (topic and/or title)	
	#4 #1 AND #2 AND #3	214
Cochrane library	#1 Individual patient data from 1800-2004 [all products]	
	#2 Meta-analysis	
	#3 Clinical trials	
	#4 Controlled clinical trials	
	#5 Randomized controlled clinical trials	
	#6 #1 AND #2 AND (#3 OR #4 OR #5)	722

Table 2. Search strategy for conventional meta-analyses in PubMed

	Search strategy "conventional meta-analyses"	
Database	Search strategy	Number
PubMed	#1 Shojania & Bero	
	#2 Our search	
	#3 Meta-analysis MeSH	
	#4 Related article search specific IPD-article	
	#5 Systematic review search Pubmed for #4	
	#6 #4 AND #1	
	#7 #4 AND #2	
	#8 #4 AND #3	
	#9 #5 OR #6 OR #7 OR #8 NOT guideline[pt] NOT editorial[pt] NOT comment[pt] NOT letter[pt]	11,149

Shojania & Bero – meta-analysis [1]

(((meta-analysis [pt] OR meta-analysis [tw] OR metanalysis [tw]) OR ((review [pt] OR guideline [pt] OR consensus [ti] OR guideline* [ti] OR literature [ti] OR overview [ti] OR review [ti]) AND ((Cochrane [tw] OR Medline [tw] OR CINAHL [tw] OR (National [tw] AND Library [tw])) OR (handsearch* [tw] OR search* [tw] OR searching [tw]) AND (hand [tw] OR manual [tw] OR electronic [tw] OR bibliographi* [tw] OR database* OR (Cochrane [tw] OR Medline [tw] OR CINAHL [tw] OR (National [tw] AND Library [tw])))))) OR ((synthesis [ti] OR overview [ti] OR review [ti] OR survey [ti]) AND (systematic [ti] OR critical [ti] OR methodologic [ti] OR quantitative [ti] OR qualitative [ti] OR literature [ti] OR evidence [ti] OR evidence-based [ti]))) BUTNOT (case* [ti] OR report [ti] OR editorial [pt] OR comment [pt] OR letter [pt]))

Our search – meta-analysis

("Meta-Analysis"[MH] OR "Review Literature"[MH] OR "meta-analysis"[pt] OR "meta-anal*"[tw] OR "metaanal*"[tw] OR "quantitativ* review*"[tw] OR "quantitative* overview*"[tw] OR "systematic* review*"[tw] OR "systematic* overview*"[tw] OR "methodologic* review*"[tw] OR "methodologic* overview*"[tw] OR (("review"[pt] OR "review"[tw]) AND ("medline"[tw] OR "cinahl"[tw] OR "embase"[tw] OR "excerpta"[tw] OR "odds ratio"[tw] OR "pooled"[tw] OR "pooling"[tw]))) NOT (letter[pt] OR editorial[pt] OR comment[pt] OR in vitro[mh] OR ("animal"[mh] NOT ("human"[mh] AND "animal"[mh])))

Dutch Cochrane trials search – trials [2]

randomized controlled trial[PTYP] OR randomized controlled trials OR controlled clinical trial[PTYP] OR clinical trial[PTYP] OR clinical trials OR (clinical AND trial) OR random allocation OR random* OR double blind method OR single blind method OR (singl* OR doubl* OR trebl* OR tripl*) OR blind* OR mask* OR placebo* OR placebos OR research design OR comparative study OR evaluation studies OR follow up studies OR prospective studies OR control OR controlled OR prospectiv* OR volunteer* ("individual patient data"[All Fields] OR "individual patient data meta"[All Fields] OR "individual patient data meta analysis"[All Fields])

Individual patient data

"individual patient data"[All Fields] OR "individual patient"[All Fields] OR "patient data"[All Fields] OR "individual data"[All Fields] OR "individual patient data meta"[All Fields] OR "individual patient data meta analysis"[All Fields] OR "individual patient's data"[All Fields] OR "original patient data"[All Fields] OR "original data"[All Fields]) OR "individual data analysis"[All Fields] OR "raw data"[All Fields] OR "raw patient data"[All Fields] OR "raw data analyses"[all Fields])

Embase MeSH terms

('randomized controlled trials'/exp OR 'randomized trials' OR 'controlled trials' OR 'clinical trials'/exp OR (clinical AND trial) OR 'random allocation'/exp OR random* OR 'double blind method'/exp OR blind* OR mask* OR placebo* OR 'research design'/exp OR comparative study'/exp OR evaluation studies' OR 'follow-up studies'/exp OR 'prospective studies'/exp OR control* OR prospective*) AND ('meta analyses'/exp OR 'review literature'/exp OR 'meta analyses' OR review/exp OR 'quantitative review' OR 'qualitative review' OR 'systematic review'/exp OR pooled OR pooling OR evidence) AND ('individual patient data' OR 'individual patient' OR 'patient data'/exp OR 'individual data')

Meta-analysis MeSH

MeSH descriptor Meta-analysis in MeSH products AND MeSH descriptor review in MeSH products

Clinical trials

MeSH descriptor Clinical trials in MeSH products

Controlled clinical trials

MeSH descriptor Controlled clinical trials in MeSH products

Randomized controlled trials

MeSH descriptor Randomized controlled trials in MeSH products

References

1. Shojania KG, Bero LA. Taking advantage of the explosion of systematic reviews: an efficient MEDLINE search strategy. Eff Clin Pract 2001;4:157-62.
2. Robinson KA, Dickersin K. Development of a highly sensitive search strategy for the retrieval of reports of controlled trials using PubMed. Int J Epidemiol 2002;31:150-3.