Recursive Cumulative Meta-analysis: A Diagnostic for the Evolution of Total Randomized Evidence from Group and Individual Patient Data

John P. A. Ioannidis,1,* Despina G. Contopoulos-Ioannidis,2 and Joseph Lau3

1Therapeutics Research Program, Division of AIDS, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland; 2Department of Pediatrics, George Washington University School of Medicine, Washington, DC; and 3Division of Clinical Care Research, New England Medical Center Hospitals, Tufts University School of Medicine, Boston, Massachusetts

ABSTRACT. Meta-analyses of randomized evidence may include published, unpublished, and updated data in an ongoing estimation process that continuously accommodates more data. Synthesis may be performed either with group data or with meta-analysis of individual patient data (MIPD). Although MIPD with updated data is considered the gold standard of evidence, there is a need for a careful study of the impact different sources of data have on a meta-analysis and of the change in the treatment effect estimates over sequential information steps. Unpublished data and late-appearing data may be different from early-appearing data. Updated information after the end of the main study follow-up may be affected by cross-overs, missing information, and unblinding. The estimated treatment effect may thus depend on the completeness and updating of the available evidence. To address these issues, we present recursive cumulative meta-analysis (RCM) as an extension of cumulative meta-analysis. Recursive cumulative meta-analysis is based on the principle of recalculating the results of a cumulative meta-analysis with each new or updated piece of information and focuses on the evolution of the treatment effect as a more complete and updated picture of the evidence becomes available. An examination of the perturbations of the cumulative treatment effect over sequential information steps may signal the presence of bias or heterogeneity in a meta-analysis. Recursive cumulative meta-analysis may suggest whether there is a true underlying treatment effect to which the meta-analysis is converging and how treatment effects are sequentially altered by new or modified evidence. The method is illustrated with an example from the conduct of an MIPD on acyclovir in human immunodeficiency virus infection. The relative strengths and limitations of both meta-analysis of group data and MIPD are discussed through the RCM perspective. J CLIN EPIDEMIOL 52;4:281–291, 1999. © 1999 Elsevier Science Inc.

KEYWORDS. Meta-analysis, randomized trials, bias, study design, publication bias

INTRODUCTION

Meta-analyses of randomized trials try to incorporate all the available evidence on a given topic to effect a research synthesis. Meta-analyses based on published group data may be affected by the selective nonpublication or late publication of negative findings [1,2]. Such publication bias [1] and publication lag [2] may lead to potentially larger treatment effects in meta-analyses synthesizing early published evidence. The magnitude of the problem may be reduced with inclusion of data from abstracts or communications with experts in the field and with involvement of the pharmaceutical sponsors in the meta-analysis to ensure that no studies are left unearthed. The direct involvement of investigators and sponsors may also enable retrieval of individual patient data from the pertinent studies for performing meta-analyses of individual patient data (MIPD) [3]. Besides further harmonizing of the original trial databases, MIPD often allows the inclusion of follow-up data beyond the original follow-up of the original trial publications. This may further diversify the treatment effect estimates compared with published group data. Thus in the process of performing a meticulous, comprehensive meta-analysis, one is being faced with a sequential accumulation of pieces of information, each of which has its strengths and a set of new problems.

While completeness of the evidence is highly desirable, there is a need to quantify as accurately as possible how much missing (or late-appearing) information may affect the results of a meta-analysis. The relative benefits and drawbacks of updated information are more controversial and also need to be studied systematically. Finally, besides...
of the treatment effect in the RCM shows whether the treatment effect is unstable over information steps, is continuously diminishing over information steps, or is converging to a more stable estimate. Recursive cumulative meta-analysis can be implemented both for retrospective and for prospective meta-analyses.

Pooling of the data from individual studies at each information step may be performed with standard methods (fixed [10] and random effects [11] models for group data; and with study-stratified proportional hazards models [12] either unadjusted or adjusted for important predictors of the outcome for time-to-event individual patient data; exact methods may depend on the form of the data [13]). Heterogeneity at each step can similarly be assessed with a traditional $\chi^2$ statistic [13]. If the meta-analysis protocol specifies an a priori emphasis on subgroup analyses and meta-regression analyses (because it is strongly anticipated that the treatment effect may be highly variable across subgroups of patients), then the RCM approach may be implemented separately for each one of the prespecified subgroups.

Theoretically, treatment effects even from highly effective treatments could be largely dissipated if data continued to be updated after patients had crossed over (with loss of the advantage of the treatment over the original control arm) or (the extreme) until all patients had died. Therefore, the inclusion of unpublished information also needs to be studied separately from information updates. We propose that two different diagnostic approaches be used, one including all information steps in the RCM, and the other excluding the steps that reflect the addition of updated information beyond the main (original) follow-up of each study. The latter diagnostic is strictly more appropriate for assessing the effect of publication lag and bias and heterogeneity.

We recommend that for practical purposes, RCM should be used as a graphical test. For modeling potential trends, several regression models for the treatment effect may be fit to the data, including linear, logarithmic, and inverse functions of the information step number. However, the highly correlated nature of the discrete information step estimates means that drawing inferences from the statistical significance of the coefficients obtained from such regression approaches may not be valid. This problem may be partly bypassed by time series modeling, but this is unlikely to be useful, unless the number of information steps is very large. A different way to analyze with precision the trajectory of the treatment effects over information steps is to calculate the ratio of the cumulative treatment effect (e.g., odds ratio) at each information step compared with the treatment effect at the previous step. In the absence of bias and genuine heterogeneity, the plot of this cumulative treatment effect ratio should show fairly symmetrical fluctuations around 1, as in some steps, the cumulative treatment effect would increase and at others, it would decrease; also with more evidence, this ratio should tend to stabilize toward 1.
Recursive Cumulative Meta-analysis

graph would show fluctuations of decreasing height, and the rapidity of the decrease in the height of the fluctuations would depend on the amount of accumulated evidence and its consistency over information steps. In the presence of bias or heterogeneity, values substantially different than 1 would be observed and the fluctuations will not decrease over information steps or may even increase.

ILLUSTRATIVE EXAMPLE

We performed a meta-analysis of updated individual patient data on the clinical efficacy of high dose acyclovir in patients infected with human immunodeficiency virus (HIV). The primary end point of interest was survival, and all randomized controlled trials with any death events were included with a total of 1792 patients and 2974 patient-years of total follow-up. The detailed results of the meta-analysis have been published elsewhere [14]. Here, we present the process of retrieving and updating the information used in the MIPD.

In 1994, we performed a preliminary meta-analysis of all the available published pertinent data [15–19] that suggested that acyclovir provided a survival benefit. The finding was robust when data from meeting presentations were also considered [20,21]. This meta-analysis was updated in February of 1995, with the inclusion of the abstract results of a trial that found no survival benefit from acyclovir [22]. After this presentation, which seemed to contradict previous beliefs among specialists in the field (even though the data were not significantly heterogeneous among trials), we decided to extend the meta-analysis protocol to include data directly from the trialists and sponsors in a meta-analysis of individual patient data. The meta-analysis team contacted not only the industry sponsors but also a large number of experts in the field and asked them regarding the potential for any unpublished studies. A total of almost 200 people were approached. The pharmaceutical sponsors were then asked to retrieve the pertinent databases that had been identified. Several experts from the United States, Europe, and Australia were contacted. The investigators of each trial verified the validity of these unpublished data and reviewed and verified all of the analyses of their pertinent databases. Moreover, all unpublished data were contributed as raw databases of individual patient data with extensive covariate information. These were analyzed by the meta-analysis team, and the end points as well as the mode of analysis had been prespecified in the protocol. All of

<table>
<thead>
<tr>
<th>Information step (Date)</th>
<th>Pieces of evidence</th>
<th>Total available evidence</th>
<th>Deaths/patients</th>
<th>Total deaths</th>
<th>Meta-analysis P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (August 1991)</td>
<td>Published data on 6 months of follow-up from H56-002 ARC</td>
<td>1/67 3/67</td>
<td>4 3.0</td>
<td>0.33</td>
<td></td>
</tr>
<tr>
<td>2 (Late 1991)</td>
<td>Preliminary 1-year follow-up of H56-002 ARC and AIDS</td>
<td>14/129 35/136</td>
<td>49 18.5</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td>3 (Early 1992)</td>
<td>Preliminary 1-year follow-up of H14-325</td>
<td>44/282 79/285</td>
<td>123 21.7</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>4 (July 1992)</td>
<td>Meeting presentation of preliminary data from P53</td>
<td>45/627 82/633</td>
<td>127 10.1</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>5 (February 1993)</td>
<td>Published data on 1-year follow-up of H56-002 ARC and AIDS</td>
<td>45/627 82/633</td>
<td>127 10.1</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>6 (May 1994)</td>
<td>Published data on 1-year follow-up of H14-325</td>
<td>42/627 81/633</td>
<td>123 9.8</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>7 (October 1994)</td>
<td>Meeting presentation of updated data of H56-002 ARC and AIDS</td>
<td>85/627 127/633</td>
<td>212 16.8</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>8 (January 1995)</td>
<td>Meeting presentation of ACTG063</td>
<td>137/796 181/798</td>
<td>318 19.9</td>
<td>0.003</td>
<td></td>
</tr>
<tr>
<td>9 (May 1996)</td>
<td>Retrieved updated IPD from H14-325</td>
<td>233/796 263/798</td>
<td>496 31.1</td>
<td>0.027</td>
<td></td>
</tr>
<tr>
<td>10 (July 1996)</td>
<td>Retrieved updated IPD from H56-002 ARC and AIDS</td>
<td>233/796 263/798</td>
<td>496 31.1</td>
<td>0.027</td>
<td></td>
</tr>
<tr>
<td>11 (June 1997)</td>
<td>Retrieved updated IPD from ACTG063</td>
<td>233/796 263/798</td>
<td>496 31.1</td>
<td>0.027</td>
<td></td>
</tr>
<tr>
<td>12 (July 1997)</td>
<td>Retrieved IPD from H56-002 KS</td>
<td>239/819 268/821</td>
<td>507 30.9</td>
<td>0.037</td>
<td></td>
</tr>
<tr>
<td>13 (August 1997)</td>
<td>Retrieved updated IPD from P53</td>
<td>245/819 272/821</td>
<td>517 31.5</td>
<td>0.059</td>
<td></td>
</tr>
<tr>
<td>14 (August 1997)</td>
<td>Retrieved updated IPD from ACTG010</td>
<td>246/838 275/843</td>
<td>521 31.0</td>
<td>0.046</td>
<td></td>
</tr>
<tr>
<td>15 (September 1997)</td>
<td>Retrieved IPD from H14-326</td>
<td>247/895 276/897</td>
<td>523 29.2</td>
<td>0.047</td>
<td></td>
</tr>
</tbody>
</table>

Steps 1–5 are reconstructed to show the sequential appearance of the first evidence on acyclovir in HIV infection. The meta-analysis was first performed at step 6 with the publication of data from H14-325. Not included in the calculations are three small studies that had no death events during their limited follow-up [18,26,27]. The meta-analysis P value corresponds to fixed effects calculations with the information available up to each step shown. Random effects P values were very similar because there was no significant heterogeneity between studies at any step.

Abbreviations: ARC = AIDS related complex; AIDS = acquired immunodeficiency syndrome; KS = Kaposi’s sarcoma; IPD = individual patient data; ACTG = AIDS Clinical Trials Group.
these efforts were made in order to avoid the controversy of using summary “data on file” provided by the industry, which previous investigators have suggested may be a biased selection of favorable data or data reported in the most favorable fashion [23–25].

Updated databases as well as data on unpublished studies were gradually retrieved with coordinated efforts at several data archives. The retrieval, cleaning, and harmonization of different trial databases proceeded at various speeds. Eventually, updated individual patient data on five studies were fully retrieved, cleaned, and analyzed by May 1996 (protocol H14-325), July 1996 (protocols H56-002 ARC and H56-002 AIDS), June 1997 (AIDS Clinical Trials Group protocol 063), and early August 1997 (protocol P53). On further investigation, it was identified that updated individual patient data from one additional unpublished trial were available, as well as on one previously published pilot trial for which no mortality data had been accumulated in the original publication. These data were also searched in the archives of the sponsor and fully retrieved, cleaned, and analyzed by July 1997 (protocol H56-002 KS) and mid-August 1997 (AIDS Clinical Trials Group protocol 010). Meanwhile, the meta-analysis results were continuously being updated as new data were becoming available. A first draft of the meta-analysis manuscript accompanied with detailed tables for the main calculations was distributed to all the investigators in late August 1997, when it was believed that all the known trial databases had been retrieved with individual patient data. Still, we made further exhortations to the investigators and sponsors to point to and retrieve potentially missing pieces of information. This resulted in the recollection of one more unpublished trial whose database was retrieved in September 1997 (protocol H14-326). We continued to exhort investigators even at this stage while revised versions of the manuscript were being distributed for feedback, and we continued to check with the pharmaceutical sponsor for other unpublished or forgotten data. Even at this final stage, in October 1997, two small ($n = 12$ [26] and $n = 21$ [27]) published trials were brought to our attention that had been missed in the initial search because they had a different primary objective (pathogenesis and evaluation of the efficacy of a different drug). Neither of these studies had any death end points.

Table 1 shows the sequential information steps in the retrieval and updating of evidence for the meta-analysis. For

![FIGURE 1. Graphic presentation of the recursive cumulative meta-analysis, showing the pooled odds ratio and confidence intervals (95%) at each information step. The information steps correspond to the incorporation of pieces of evidence listed in Table 1. The estimate at each information step is the result of a cumulative meta-analysis of all the data available up to this specific information step. The pooled estimate is calculated with the Mantel-Haenszel fixed effects model. DerSimonian and Laird random effects estimates were practically identical, as there was no significant heterogeneity at any step of the cumulative retrieval meta-analysis ($\chi^2$ for heterogeneity $P > 0.1$ at all steps).]
some studies, data are completely new at some point, whereas for other studies, previous data are being replaced over time with more precise (published in contrast to preliminary) or updated follow-up information. In some cases, the retrieved individual patient data have exactly the same counts of events as the previous summary data in the database (see, for example, steps 10 and 11 in Table 1).

As of December 1997, only four of the eight studies included in the meta-analysis had been published. Consistent with our previous investigation [2] describing the extent of publication lag for efficacy trials, all three trials showing a survival benefit had been published, whereas four of the five trials that did not show a significant survival benefit were still unpublished 5 to 9 years after starting enrollment. One of the four unpublished studies was eventually published in mid-1998, almost a decade after starting enrollment [28].

Figure 1 shows a graphical representation of the RCM as more complete, corrected, and updated information becomes available. As seen, early estimates of the treatment effect based on early published trials tend to be larger than later cumulative effects that give a more complete picture of the whole evidence. An inspection of the RCM trajectory suggests that a meta-analysis limited to information from the first few steps may be misleading, as there is strong evidence that the accumulating information is continuously attenuating the magnitude of the treatment effect, and there is no suggestion that the treatment effect is yet converging to a more stable estimate. Thus, the RCM suggests that a meta-analysis of published data may be misleading, as the evidence has not converged yet and the conclusion of large treatment benefit that is highly significant (e.g., \( P = 0.0003 \) at step 6) would be premature. Moreover, the RCM suggests that one may be misled in judging the results as conclusive simply on the basis of very extreme levels of statistical significance achieved in early steps (\( P = 0.002 \) by step 2, \( P < 0.001 \) consistently in steps 3–6). By contrast, the late additions of evidence do not seem to affect the magnitude of the treatment effect in any substantive way. The magnitude and confidence intervals of the treatment effect seem to converge after the ninth information step. Evidence is probably more reliable at this stage and suggests a cautious interpretation of the meta-analysis results, which suggest the presence of a modest treatment effect (pooled odds ratio 0.75), which is more likely to be real, although the 95% confidence intervals extend almost to no benefit.

When the updated follow-up steps were excluded to separate the effect of potential retrieval bias from the effect of
updated information, the RCM looked similar (Fig. 2). Again, there was evidence that the estimate of the treatment effect converged after the eighth step. The magnitude of the converging treatment effect seemed slightly larger when updating steps were excluded, but a conservative interpretation would still be appropriate given the confidence intervals.

Table 2 shows for comparison the results of typical meta-analyses based on published versus published and presented versus all published and unpublished data. Each analysis is reported with the inclusion or not of updated follow-up data. The treatment effect seems to shrink across these categories as missing information is included and the updated follow-up data tend to be even more conservative. Nevertheless, with the exception of the results of a meta-analysis limited only to published data, the difference across the other categories is rather small as far as the magnitude of the treatment effect is concerned. Also shown are results of MIPD analyses including all data either censored at the specified follow-up of the primary trial analysis for each study or including all updated follow-up information. There is hardly any meaningful difference between summary and individual level estimates once similarly complete data are considered in the calculations.

Figure 3 shows the evolution of the ratio of the cumulative treatment effect (odds ratio) at each information step divided by the treatment effect at the previous step. Large initial fluctuations eventually lead to a more stabilized treatment effect with little relative change between different steps. To put this plot into perspective, two more examples are given using data from two very well-known published cumulative meta-analyses: intravenous streptokinase in acute myocardial infarction, a typical example in which there is widespread consensus on the validity of the meta-analysis results and the intervention unquestionably works [7] (Fig. 4); and magnesium salts in acute myocardial infarction [29], a typical example in which a mega-trial (ISIS-4 [30]) arrived at opposite conclusions compared with a previous meta-analysis, causing overt debate on the efficacy of this intervention (Fig. 5) [31]. The fluctuations of the treatment effect of streptokinase expectedly diminish rapidly over information steps. On the other hand, in the case of magnesium, the treatment effect continues to fluctuate as much or even more as more data are accumulated. Increasing fluctuations are noted even before the appearance of ISIS-4 (added at information step 11 in the figure), the mega-trial that found no benefit from magnesium in contrast with previous evidence. Persistence or increase in the treatment effect fluctuations may be a signal of genuine heterogeneity or bias, and this signal was apparent for magnesium very early on in the course of the accumulated evidence.

TABLE 2. Comparison of meta-analysis estimates

<table>
<thead>
<tr>
<th>Meta-analysis of summary data</th>
<th>Acyclovir</th>
<th>Control</th>
<th>Fixed effects (95% CI)</th>
<th>Random effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without updated follow-up information</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Published studies</td>
<td>41/282</td>
<td>78/285</td>
<td>0.43 (0.28–0.66)</td>
<td>0.43 (0.28–0.67)</td>
</tr>
<tr>
<td>Published and presented studies</td>
<td>95/796</td>
<td>136/798</td>
<td>0.61 (0.45–0.83)</td>
<td>0.56 (0.36–0.88)</td>
</tr>
<tr>
<td>All studies</td>
<td>99/876</td>
<td>138/875</td>
<td>0.64 (0.47–0.86)</td>
<td>0.60 (0.39–0.92)</td>
</tr>
<tr>
<td>With updated follow-up information</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Published studies</td>
<td>41/282</td>
<td>78/285</td>
<td>0.43 (0.28–0.66)</td>
<td>0.43 (0.28–0.67)</td>
</tr>
<tr>
<td>Published and presented studies</td>
<td>137/796</td>
<td>181/798</td>
<td>0.64 (0.48–0.86)</td>
<td>0.64 (0.48–0.86)</td>
</tr>
<tr>
<td>All studies</td>
<td>247/895</td>
<td>276/897</td>
<td>0.75 (0.57–1.00)</td>
<td>0.76 (0.57–1.00)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Meta-analysis of individual patient data†</th>
<th>Acyclovir</th>
<th>Placebo</th>
<th>Unadjusted (95% CI)</th>
<th>Adjusted (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without updated follow-up information</td>
<td>109/876</td>
<td>149/897</td>
<td>0.68 (0.53–0.87)</td>
<td>0.67 (0.52–0.85)</td>
</tr>
<tr>
<td>With updated follow-up information</td>
<td>247/895</td>
<td>276/897</td>
<td>0.79 (0.66–0.93)</td>
<td>0.78 (0.65–0.93)</td>
</tr>
</tbody>
</table>

The status of the studies (published, presented, unpublished) is as of September 1997 (time of completion of the meta-analysis of individual patient data). Fixed effects calculations are performed according to the Mantel-Haenszel model, and random effects calculations are performed according to the DerSimonian and Laird model. The two estimates are very similar because there was no significant heterogeneity in any case (χ² for heterogeneity P > 0.1). Unadjusted proportional hazard model estimates are stratified per study; adjusted estimates are also adjusted for baseline CD4 count (square root thereof), baseline hemoglobin, age, gender, race, and AIDS diagnosis at baseline. The count of events when no updating is considered is slightly different in the group data meta-analysis versus the MIPD, as in the MIPD, the results include all events that occurred in the originally specified follow-up, whereas a few such events had not been included in the published analyses of the main follow-up of these studies. One study (ACTG010) is included only in meta-analyses using updated information, as no deaths were recorded in the initial follow-up report.

†Includes all studies (both published and unpublished).  
Abbreviations: AIDS = acquired immunodeficiency syndrome; CI = confidence interval; MIPD = meta-analysis of individual patient data.
DISCUSSION

We have presented an extension of cumulative meta-analysis that can be used to evaluate the composite evidence on a clinical topic as more information is being obtained, retrieved, and updated. Recursive cumulative meta-analysis offers a diagnostic approach that can visualize the accumulating evidence as a continuum and can be helpful in assessing whether early estimates of the treatment effect change over time; whether they are moderated or completely dissipated by overt or hidden publication bias [1] and publication lag [2] for negative results; and what effect updated follow-up has on the data. It deals in a systematic fashion with issues that have been thought to be obstacles to objective effect estimation in meta-analysis.

Clinical trials are typically designed with a frequentist hypothesis in mind (the treatment is not better than control), while physicians usually want to know the magnitude of the treatment effect for estimation purposes (how much it works). The interchange of frequentist and estimation approaches is not always without problems [32–34]. Moreover, there is a growing understanding that clinical evidence is a dynamic process, not a static estimation of a single treatment effect at a single time-point [7,35,36]. Recursive cumulative meta-analysis attempts to bridge this transition from frequentist trials to unbiased estimation.

Until now, meta-analyses have been reported in the literature as final products with little or no information on how information was gathered over time, as in most cases, all of this information has been obtained from literature searches performed at a specific time point. However, initiatives such as the Cochrane Collaboration [37] have introduced and will probably soon establish the concept that systematic reviews should be continuously renewed as new evidence appears, and they are never final. In this process, which is likely to be the future of meta-analysis, RCM is focusing on the evolution of treatment effects over time. Our approach suggests that simply reporting a weighted aggregate may not be adequate, and details on the order in which data appeared can offer insight on whether the treatment effect is changing as a more complete picture evolves. The implications are not limited to practitioners of meta-analysis, but to medicine at large, as the concept can be applied to any meta-analysis regardless of subject matter. Fluctuations in the magnitude of the treatment effect may point to the presence of either bias or genuine heterogeneity that need to be scrutinized. As a general rule, meta-analyses with excessive fluctuations, asymmetrical fluctuations, and

![Graph of the cumulative treatment effect ratio over sequential information steps for the acyclovir recursive cumulative meta-analysis. For each information step, the cumulative treatment effect ratio is defined as the cumulative odds ratio at that step divided by the cumulative odds ratio at the previous step. Values >1 suggest an overestimation of the treatment effect in the previous step compared with the current one. Values <1 suggest an underestimation of the treatment effect at the previous step compared with the current one.](image)
those that do not reach values close to 1 for the cumulative treatment effect ratio should be scrutinized for bias and heterogeneity. It would be imprudent to put specific values that would define an “abnormal” signal in these properties of the trajectory because we do not know what is the absolute unbiased truth in clinical research in order to correlate it with specific trajectories. Although large simple trials have been proposed as the gold standard, several investigations that have compared them to meta-analyses [38–40] have not reached consensus on whether large trials can indeed be gold standards [41]. The RCM approach offers a framework wherein current and future meta-analyses may be presented and scrutinized and empirical evidence of the properties of the trajectories of meta-analyses across different fields can be collected. It is thus a first step toward a more robust interpretation of clinical evidence and how such evidence is obtained.

Recursive cumulative meta-analysis may complement and improve over sequential monitoring approaches that have been proposed for meta-analyses based on the use of alpha-spending functions [36]. Such approaches aim to avoid premature claims on statistical significance when only small portions of the total evidence are available and statistical significance at an early look at the data may be due to chance. However, alpha-spending function approaches are not appropriate when there is bias in the order in which the evidence is accumulated. In contrast to a single large

---

**FIGURE 4.** Graph of the cumulative treatment effect ratio over sequential information steps for a cumulative meta-analysis of intravenous streptokinase in myocardial infarction. This example uses a database that has been published and described in detail elsewhere [7]. For each information step, the cumulative treatment effect ratio is defined as the cumulative odds ratio at that step divided by the cumulative odds ratio at the previous step. Values >1 suggest an overestimation of the treatment effect in the previous step compared with the current one. Values <1 suggest an underestimation of the treatment effect at the previous step compared with the current one.
clinical trial in which there is no reason to believe that the most favorable results will be systematically accumulated first, in a meta-analysis, it is possible that this may happen. Pogue and Yusuf [36] have suggested that alpha-spending functions would not have claimed statistical significance for a controversial meta-analysis (magnesium in acute myocardial infarction), whereas this would have been the case in an undisputed effective medication (streptokinase in myocardial infarction). Still Egger and colleagues [43] have recently shown that the inclusion of abstracts and unpublished data would have resulted in reaching robust significance for magnesium as well, even if an alpha-spending function were to be used. Alpha-spending functions are problematic in meta-analysis when there is bias or genuine heterogeneity of the treatment effect. In contrast, the RCM aims specifically at capturing this potential irregularity in the evolution of the treatment effect over time.

Recursive cumulative meta-analysis may also offer a means for the dissection of sources of disagreement between MIPD and MPL in a consistent and standardized manner as data are accumulated for an MIPD. There has been a long debate on the relative merits of MIPD over MPL, but actual data are scarce. Early evidence suggested that MIPD gives significantly more conservative estimates than MPL [7], but other investigators [8,9] did not observe any discrepancies. In particular, summary-level or individual-level information seems to lead to the same estimates when based on the same databases, as seen also in the example of the acyclovir meta-analysis. Discrepant results probably arise either by publication bias in MPL or retrieval bias in MIPD or by the inclusion of updated information that differentiate the databases used by the two methods.

The collection of updated data beyond the original randomized follow-up presents several problems of implementation and interpretation. Such data should be carefully examined in each meta-analysis and the RCM graphs with versus without the updating should be compared. The updated follow-up analyses are not appropriate when simply the effects of intervention versus placebo are to be studied, as many patients may cross-over once the main trial is over and information is released to the participants. Cross-overs may actually be influenced by the released trial results. Still, the updated follow-up has its advantages, such as the ability of assessing late outcomes of patients, and the com-

FIGURE 5. Graph of the cumulative treatment effect ratio over sequential information steps for a cumulative meta-analysis of intravenous magnesium in acute myocardial infarction. This example uses a database that has been published and described in detail elsewhere [29]. For each information step, the cumulative treatment effect ratio is defined as the cumulative odds ratio at that step divided by the cumulative odds ratio at the previous step. Values >1 suggest an overestimation of the treatment effect in the previous step compared with the current one. Values <1 suggest an underestimation of the treatment effect at the previous step compared with the current one.
parisons are still protected by the initial randomization if one is interested in making an assessment of simply whether early versus deferred implementation of the tested intervention makes a difference in the long run. Updated data are likely to offer more evidence and may even correct biased large estimates of the treatment effects based on interim analyses. They may give better evidence on the durability of a treatment effect, and for clinical purposes, it is important to know whether the efficacy of early treatment is time-limited and, if so, to determine the duration of its benefit [44]. However, besides the effect of cross-overs, updated data may be of inferior quality, may suffer from more extensive missing data [45], end points may not have been verified as completely and as objectively as those during the main follow-up (particularly for masked trials), and the decision to attempt some updating may be influenced by what the early results have shown (e.g., updating may be performed preferentially on studies that show the larger early benefits). The net effect of these forces on the treatment effect is difficult to decipher. Generally, updated survival data are probably more trustworthy than updated data on more subjective end points such as disease progression or change in health status and surrogate markers. This is the reason that the acyclovir MIPD protocol was designed with survival only as the primary end point. It is reassuring when the results of the RCM are consistent regardless of whether updated data are included or not, as in the case of the presented MIPD. When the contrary is observed, the robustness of the updated data needs to be carefully scrutinized or the durability of the treatment benefit may need to be questioned.

In the absence of extensive empirical evidence in the relative validity of MPL and MIPD, strong statements about their relative importance may be premature. Both can offer very useful information, provided their strengths and limitations are understood. Meta-analysis of individual patient data offers undoubtedly the strong advantages of performing risk modeling and more detailed time-to-event analyses, but it is more costly and resource consuming and requires the bona fide collaboration of a large number of investigators who are all enlightened to recognize the importance of such endeavors [3]. Failure to gain approval for access to all individual patient data may add another layer of bias peculiar to MIPD, and full sharing of all data, as was the case in the presented MIPD, is often not attained. Publication bias, retrieval bias, and the effects of updated information need to be evaluated as a whole. This may be performed in the RCM framework.

Finally, for some treatments, different patient populations may experience different treatment benefits from a treatment [46–48]. If this is strongly suspected, we recommend that RCM is modified to adjust also for predictors that may determine the magnitude of the treatment effect. A separate RCM may then be performed for each subgroup of patients that has been hypothesized in the meta-analysis protocol to potentially experience a different treatment effect. In the acyclovir MIPD, we had overall observed no evidence that the magnitude of the survival benefit was different in patients at different levels of risk defined by predictors such as baseline CD4 count, age, acquired immunodeficiency syndrome diagnosis at entry, and baseline hemoglobin level [14]. Of course, subgroup effects are harder to discern [49], but for some treatments, it may be more imperative to adjust for subgroup effects. In particular, new treatments are often likely to be tested first in patients at high risk of the disease outcome and, if promising, then tested in patients of lower risk. Trials with high-risk patients are also likely to accumulate the necessary number of events earlier than trials of low risk patients. Empirical evidence shows that slow accumulation of events is a common reason for the protracted conduct and late publication of randomized trials [2]. The diversity of the studied patient populations needs to be examined carefully to evaluate whether the inclusion criteria of a meta-analysis may be too broad and whether the RCM diagnostics should be applied to all the studied populations or subgroups thereof.

We are grateful to several investigators who contributed data for the meta-analysis of acyclovir and in particular to the principal investigators Ann C. Collier, David A. Cooper, Brian G. Gazzard, Paul D. Griffin, A. Paul Fiddian, Andrew T. Pavia, Michael S. Saag, Spotwood L. Sprunke, and Michael S. Youle. Christopher H. Schmid contributed valuable comments to the manuscript.

References
2. Ioannidis JP. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. JAMA 1998; 279: 281–286.
10. Mantel N, Haenszel W. Statistical aspects of the analysis of