Meta-analysis when only the median survival times are known: A comparison with individual patient data results

Stefan Michiels
Institut Gustave-Roussy

Pascal Piedbois
Henri Mondor Hospital

Sarah Burdett
MRC Clinical Trials Unit

Nathalie Syz
Institut Gustave-Roussy

Lesley Stewart
MRC Clinical Trials Unit

Jean-Pierre Pignon
Institut Gustave-Roussy

Background: The hazard ratio (HR) is the most appropriate measure for time to event outcomes such as survival. In systematic reviews, HRs can be calculated either from the raw trial data obtained as part of an individual patient data (IPD) meta-analysis or from the appropriate trial-level summary statistics. However, the information required for the latter are seldom reported in sufficient detail to allow reviewers to calculate HRs. In contrast, the median survival and survival rates at specific time points are frequently presented. We aimed to evaluate retrospectively the performance of meta-analyses using median survival times and survival rates by comparing them with meta-analyses using IPD to calculate HRs.

Methods: IPD from thirteen published meta-analyses (MAs) in cancers with high mortality rates were used. Median survival and survival rates were calculated from the IPD rather than taken from publications so that the same trials, patients, and extended follow-up are used in each analysis.

Results and Conclusions: We show that using median survival times or survival rates at a particular point in time are not reasonable surrogate measures for meta-analyses of survival outcomes and that, wherever possible, HRs should be calculated. Individual trial publications reporting on time to event outcomes, therefore, should provide more detailed statistical information, preferably logHRs and their variances, or their estimators.

Keywords: Meta-analysis, Summary data, Median survival time, Survival analysis, Time to event outcomes

Individual patient data (IPD) meta-analyses (MAs), which involve the central collection, checking, and re-analysis of updated IPD, have been described as the gold standard of systematic review (7;20). However, this approach is not always practical, often due to economic, resource, or time constraints. Some meta-analyses use summary data, which are supplied by the trialists, but more often than not, meta-analyses are performed by extracting the data from the

published literature. This approach is prone to many biases, such as reporting bias (25), publication bias (8;10), and patient exclusion bias (27).

Furthermore, analyses are often performed at either a single fixed point in time or at a series of time points. This is unlikely to cause major difficulties for binary outcomes. However, for time to event outcomes like survival, it can be problematic. Most commonly, the numbers of individuals who have experienced an event and those who are event-free are used to calculate odds ratios (OR) at fixed points in time. However, such analyses relate only to the point in time at which they are calculated, giving the overall odds of experiencing an event on experimental compared with control treatment, and can easily misrepresent the overall effect if points of maximum or minimum difference between survival curves are chosen (26). Time to event outcomes are most appropriately analyzed by calculating HRs where individual durations of survival are used to calculate the overall instantaneous risk of event on experimental compared with control intervention. Such analyses are most easily done using individual patient data and indeed many IPD meta-analyses are done primarily because time to event analyses are essential to the project.

Methods to make better use of summary statistics to calculate or estimate HRs within meta-analyses that rely on extracting data from published reports have been described (18). However, in practice, the most appropriate statistics, the logHR and the variance, are seldom reported. For example, while carrying out a meta-analysis of the literature that looked at the impact of increasing the number of chemotherapy drugs in metastatic lung cancer, we observed that the logHR was reported in only 3 percent of 131 therapeutic comparisons, it's p value more frequently in 37 percent of cases, but often presented without the number of observed deaths. One-year survival was reported in 28 percent of publications and could be read off the survival curve in an additional 15 percent of cases. The median survival times appeared to be the most frequently reported summary statistic in the retrieved publications (73 percent). These observations are similar to figures reported (3) before the CONSORT (16) statement was published. In view of the paucity of data making it difficult to apply the methods described by Parmar et al. (18), we decided to explore an alternative method using median survival times, which may be more readily available in publications. We aimed to compare this strategy with the standard approach of calculating ORs and with the HR obtained from analyzing individual patient data.

METHODS

Data from several IPD MAs coordinated by the Institute Gustave Roussy (IGR), the Meta-analysis Group in Cancer (MAGIC), and the MRC Clinical Trials Unit (CTU) were used to compare empirically the utility of the OR method and

the median survival method against the actual HRs obtained from analysis of the IPD.

Various statistical methods can be used to obtain a single "effect" estimate from a particular summary statistic at the trial level (22;30;31). We compared two aggregated data methods: one based on the median survival time and the other on the OR of survival rates. Each of these findings was then compared with the method considered as the "gold standard" for individual patient data with time to event outcomes, the pooled HR method. For each of the three measures, a fixed effect model is applied (30). We compared like with like, that is, the same trials, patients, and extended follow-up are used in each analysis. The main end point used was death. The median survival times and 1-year survival rates were calculated from the IPD database rather than taken from publications. We have not explored how the method affects estimates of heterogeneity.

Hazard Ratio

Assume that we have k trials comparing an experimental arm with a control arm. IPD meta-analysis of survival type data is generally based on the stratified log rank test and the overall pooled HR (31). For each trial, the HR and its variance are derived from the log rank statistic, calculated with time to event for individual patients. The pooled logHR is a weighted average of the logHRs with the weights inversely proportional to the variances of the studies. Asymptotically, the logHR follows a standard normal distribution under the null hypothesis of no treatment effect. The overall pooled HR represents the overall risk of dying on the experimental treatment compared with the control treatment.

Odds Ratio

The OR at a particular point in time is the ratio of the odds of dying in the experimental arm compared with the control arm up to that time point. For each trial, we estimate the logOR and its variance using the Yusuf et al. method (31). The pooled logOR is a weighted average of the logORs with the weights inversely proportional to the variances of the k individual studies. The logOR is asymptotically normal under the null hypothesis of no treatment effect. The OR of survival at 1-year was considered to maximize statistical power (6;26). It has been suggested (29) that, when estimating from published survival curves, the OR method should be corrected for censoring. Because we have IPD data with excellent follow-up in the first year, no adjusting was required. To compare HRs and ORs on the same scale, transformations into absolute survival differences are applied (9;26).

Median Survival Ratio

The median survival time is commonly estimated as the first observed event (death) at which the Kaplan–Meier survival function is less than or equal to 0.5. For combining median survival times across studies a pooled ratio of median

Table 1. Characteristics of the 13 Meta-analyses

Meta-analysis (reference)	No. of trials	No. of patients	Death rate	Control treatment	Experimental treatment	Average trial size	Heterogeneity p value
Lung 1 (5)	7	987	0.86	CT(+/−RT)	CT(+/−RT) + PCI	141	.95
Lung 2 (17)	11	1190	0.96	SC	SC + CT	108	<.01
Lung 3 (19)	13	2536	0.91	CT	CT + RT	195	.10
Lung 4 (17)	25	3033	0.93	RT	RT + CT	96	.56
Lung 5 (21)	9	2128	0.64	SURG	SURG + RT	236	.11
Colorectal 1 (15)	7	1216	0.91	5FU(+/−LV) CI	5FU(+/−LV) bolus	174	.91
Colorectal 2 (2)	9	1380	0.87	5FU	5FU + LV	153	.59
Colorectal 3 (13)	10	1544	0.85	5FU(+/−LV)	5FU(+/−LV) + IFN	154	.24
Colorectal 4 (13)	6	1031	0.79	5FU + LV	5FU + IFN	172	.50
Colorectal 5 (1)	8	1168	0.91	5FU	5FU + MTX	146	.57
Colorectal 6 (14)	5	494	0.92	HAI	No HAI	99	.76
Esophageal (4)	6	1147	0.85	(+/−CT) + SURG	(+/−CT) + SURG + RT	191	.07
Glioma (11)	12	3004	0.89	RT	RT + CT	250	.27

CI, continuous infusion; CT, chemotherapy; RT, radiotherapy; PCI, prophylactic cranial irradiation; SC, supportive care; SURG, surgery; 5FU, 5-fluorouracil; LV, folinic acid; IFN, interferon; MTX, methotrexate; HAI, hepatic arterial infusion.

survival times has been proposed (22). The pooled log ratio of median survival times (log(MR)) is a weighted average of the log ratio of median survival times for each trial, with the weights inversely proportional to the variances of the individual studies.

We considered three different estimators for the variance of the log ratio of median survival times for each trial: (i) Using the number of deaths in the control arm and in the experimental arm of each trial (22). This estimator is the maximum likelihood estimator for the variance of log(MR) if the survival distributions are exponential (23). (ii) Using the total number of deaths in each trial (18). We suppose that the allocation ratio in the two groups is 1:1 and that the expected treatment effect will be rather small. (iii) Using the total number of patients in each trial. We suppose that the event rate is very high and again that the allocation ratio in the two groups is 1:1 and that the expected treatment effect will be rather small.

Normality of the pooled log(MR) can be assumed without a normality assumption of the individual log(MR)s based on the central limit theorem, but Simes (22) argues that they will be approximately so for large studies with sufficient follow-up. We also considered the logarithm of the ratio of Kaplan–Meier median survival times as normally distributed, although it is essentially a ratio of two nonparametric estimators. When converting the MR into absolute survival differences, we have applied the HR's transformation formula and, hence, implicitly assumed that MR is an estimate for HR.

RESULTS

Data were available from 13 meta-analyses consisting of 128 randomized controlled trials, 20,858 patients and 18,047 deaths in non-small cell lung cancer (5;17;19;21), colorectal cancer (1;2;13–15), esophageal cancer (4), and high-grade glioma (11). The number of trials included in the meta-analyses ranges from 5 to 25. The average mortality across the trials was 88 percent. The highest mortality rate of 96 percent was observed in Lung 2 (17), and the lowest of 64 percent is in Lung 5 (21). Kaplan–Meier median survival times in the control arms range from 3 to 55 months, with an average of 11 months, whereas median survival times in the experimental arms range from 4 to 40 months, also with an average of 11 months. The basic characteristics of the meta-analyses are summarized in Table 1.

Detailed Example Using Lung 1

Prophylactic cranial irradiation (PCI) is known to reduce the incidence of brain metastases in patients with small-cell lung cancer (SCLC). The IPD MA (5), which analyzed 987 patients with SCLC in complete remission, showed that PCI improved overall survival with an HR of 0.84 (95 percent confidence interval [CI], 0.73–0.97; $p = .01$) or a reduction in the risk of death of 16 percent compared with the control group. The results of the meta-analyses using the three different types of summary statistics when applied to this data are shown in Figure 1.

For this particular example, the estimates for the overall treatment effect do not differ considerably from the overall HR. The overall pooled OR is 0.83 (95 percent CI, 0.65–1.08; $p = .16$), and the overall pooled MR is 0.88 (95 percent CI, 0.77–1.01; $p = .07$). The HR corresponds to an absolute survival difference of 5.9 percent at 1-year in favor of PCI, the OR also to a survival difference of 4.5 percent, and the MR to 4.3 percent. Neither the OR or ratio of medians are statistically significant at the 0.05 level. In contrast, the HR is statistically significant at the 0.01 level. For the second trial, in Figure 1, the treatment effect as estimated by the OR (1b) is in the opposite direction to the HR (1a) and MR (1c) estimates, owing to a higher percentage of deaths in the experimental arm in the beginning of the trial, whereas at the
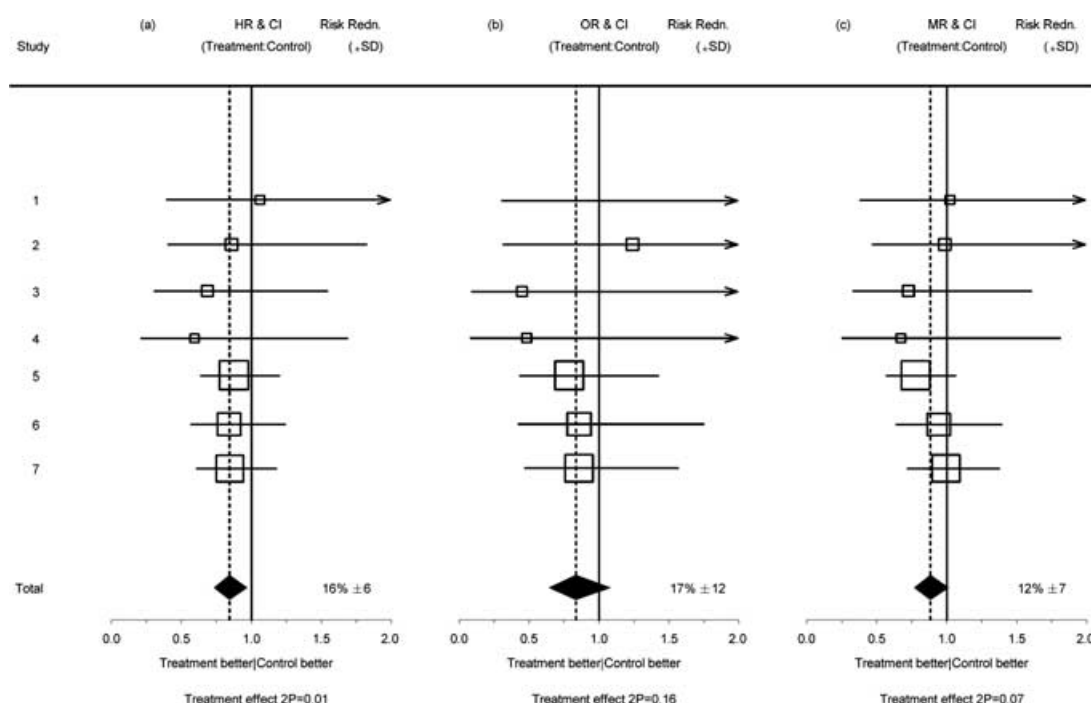
Figure 1. Forest plots for the prophylactic cranial irradiation meta-analysis obtained with (a) the hazard ratio (HR) method, (b) the odds ratio (OR) method, (c) the median survival ratio (MR) method using method i for estimating variance.

end, the percentages are very similar. The median survival times of the two arms of the 7th study are equal, leading to an estimated MR of 1.0.

Trial Level Results

Figure 2 plots the log(OR) versus the log(HR), and the log(MR) versus log(HR), divided by their respective standard errors (denoted as "standardized"). The plots reveal a correlated pattern but indicate that some under- and over-estimations of the treatment effect may occur for both the OR and the MR method. We also evaluate the performance of the three methods by counting the number of significant results with the 5 percent significance level. The HR yields 12/128 trials with a significant treatment effect. The MR method (with variance approximation (i) detects seven of these twelve significant trials, and falsely indicates a further 10 trials as significant. Applying variance approximation (ii) or (iii) gives more false significant trial results, 16 and 18, respectively. The OR method detects fewer true significant trials (5 of 12) but also fewer false significant trials (8 of 12).

Meta-Analysis Level Results

The overall HR method indicates ten treatment effects significant at the $p = .05$ level in thirteen meta-analyses. All three approximations of the MR variance yield the same results for the pooled MR, identifying 6 of 10 correct significant results and 1 false significant result. The pooled OR shows 5 of 10 true significant results but has no false significant results.

The estimations of the overall treatment effects calculated by each method together with 95 percent confidence limits, p values, and transformations into 1-year survival differences are given for individual meta-analyses in Table 2.

The confidence intervals for the MR are similar in width to the confidence intervals for the overall pooled HR, whereas the confidence intervals for the overall OR are much wider. There was no clear tendency that the OR performed much better than the MR method when considering the absolute survival differences. The largest difference between the survival benefit estimated with the OR and HR method can be found in the meta-analysis with the largest treatment effect [Colorectal 6 (14)]. It is well known that Peto's one-step OR estimator (31) tends to be biased for pooling results from trials involving large treatment effects (12).

The discrepancies between the overall HR and the pooled MR are likely to stem from differences at the trial level, in particular when the differences between the survival curves are not constant over time. This finding can be illustrated by application of a two-sample test sensitive to crossing hazards alternatives, developed by Stablein and Koutrouvelis (24). The two meta-analyses (17;19) with the highest percentage of trials for which the two-sample test rejects the null hypothesis of no crossing hazards at $p = .05$ (5 of 11 and 8 of 13 trials, respectively), correspond to large differences between the estimated overall HR and MR. It is also noteworthy that one of these meta-analyses (17) was the only meta-analysis with strong evidence for heterogeneity using the HR method. On the whole, the number of trials in
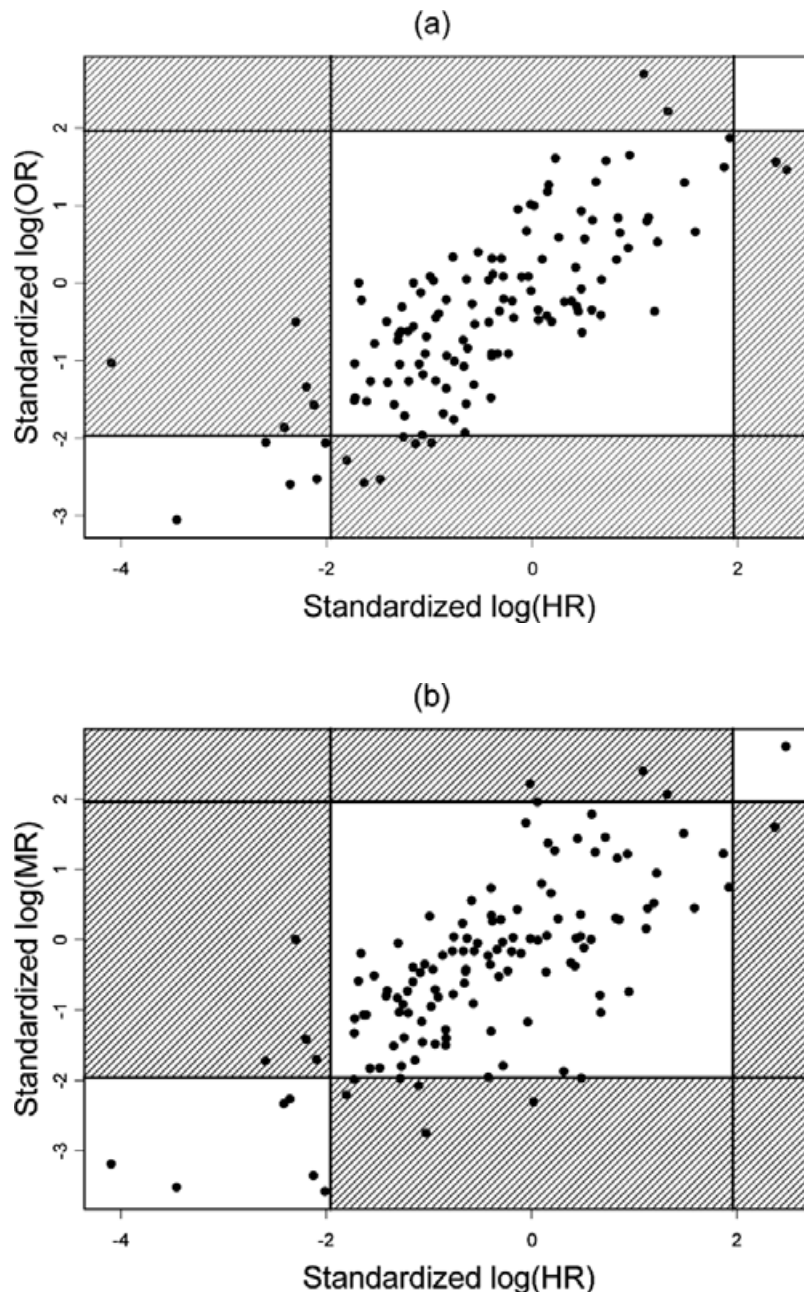
Figure 2. (a) Standardized log(OR) vs. standardized log(HR) and (b) standardized log(MR) vs. standardized log(HR) for all 128 trials from the individual patient data database with 95 percent confidence limits (bold lines). Each trial is represented by one point. Shaded areas correspond to discordant results of significance test for a treatment effect (the 95 percent confidence limits are based on normal approximations).

the meta-analysis or number of patients in the trial, did not seem to influence the occurrence of the shifts in the estimation of the overall effect or the statistical significance for the MR.

DISCUSSION AND CONCLUSIONS

For many diseases, the most important measures of treatment outcome involve time to event analyses. In individual trials, analyses of time to event outcomes, like survival, are common. In systematic reviews and meta-analyses, they are relatively rare. IPD meta-analyses that collect and reanalyze the "raw" trial data are able to use individual survival times *within* trials to calculate pooled HR. However, such projects remain in the minority. Several methods have been developed for calculating/estimating HRs (18;28) that would enable other types of systematic review to calculate or estimate HRs from trial reports. However, the summary statistics

Table 2. Main Results of the Comparison between the Hazard Ratio (HR), Odds Ratio (OR), and Median Survival Ratio Method (MR; using method i for estimating variance) on the Meta-analysis Level

Meta-analysis	HR results			OR results			MR results		
	HR (95% CI)	p value	Survival difference	OR (95% CI)	p value	Survival difference	MR (95% CI)	p value	Survival difference
Lung 1	0.84 (0.73–0.97)	.01	5.9%	0.83 (0.65–1.08)	.16	4.5%	0.88 (0.77–1.01)	.07	4.3%
Lung 2	0.84 (0.74–0.95)	.01	5.3%	0.70 (0.53–0.93)	.01	5.3%	0.73 (0.65–0.82)	.00	9.9%
Lung 3	0.86 (0.79–0.94)	<.01	5.1%	0.84 (0.71–0.99)	.03	4.4%	0.95 (0.87–1.03)	.21	1.8%
Lung 4	0.90 (0.83–0.97)	.01	3.9%	0.87 (0.75–1.00)	.06	3.4%	0.93 (0.86–1.00)	.05	2.7%
Lung 5	1.21 (1.08–1.34)	<.01	−5.3%	1.29 (1.06–1.58)	.01	−4.4%	1.14 (1.02-1.27)	.02	−3.5%
Colorectal 1	0.88 (0.78–0.99)	.04	4.5%	0.84 (0.67–1.06)	.14	4.3%	0.90 (0.80–1.02)	.09	3.6%
Colorectal 2	0.97 (0.86–1.09)	.59	1.1%	0.84 (0.68–1.05)	.13	4.2%	0.97 (0.87–1.09)	.65	1.0%
Colorectal 3	0.98 (0.88–1.09)	.70	0.7%	0.99 (0.81–1.22)	.96	0.1%	0.95 (0.85–1.05)	.32	1.9%
Colorectal 4	1.16 (1.01–1.33)	.04	−5.2%	1.27 (0.99–1.62)	.06	−5.9%	1.15 (1.01–1.32)	.04	−5.0%
Colorectal 5	0.87 (0.77–0.98)	.03	5.1%	0.77 (0.60–0.97)	.03	6.4%	0.87 (0.77–0.98)	.03	5.0%
Colorectal 6	0.74 (0.61–0.89)	<.01	10.4%	0.47 (0.33–0.67)	<.01	18.3%	0.71 (0.59–0.86)	.00	11.7%
Esophagus	0.89 (0.78–1.01)	.06	4.0%	0.86 (0.68–1.10)	.23	3.6%	0.87 (0.77–0.99)	.03	4.8%
Glioma	0.85 (0.78–0.92)	<.01	5.9%	0.76 (0.66–0.88)	<.01	6.5%	0.94 (0.87–1.02)	.12	2.3%

required to make these calculations are not always readily available.

We, therefore, explored the utility of using the summary statistics that are most commonly reported in publications of randomized controlled trials in cancer: the median survival time and the survival rates. We have compared the results obtained using a simple method, based on the median survival time (MR), and the commonly used OR of survival rates with the actual HRs calculated from individual patient data. The HR incorporates changes over time, whereas the two other methods only take one point on the survival curve into account. This was done using data obtained for thirteen IPD meta-analyses involving cancers with high mortality rates. As the median survivals and survival rates were calculated directly from the IPD databases (as opposed to using information extracted from publications) we are comparing like with like and other sources of bias have been minimized. We found that both the MR and OR method may result in serious under- or overestimation of the treatment effect and major loss of statistical power. It can be expected that the MR method will be even more biased for meta-analyses involving trials with lower mortality rates than those used for this study, as the estimate for the median survival time becomes more unstable for trials with lower event rates. Furthermore, in 20 percent (25/128) of trials included in this study the log(MR) had an opposite sign to the log(HR), that is, whereas one method suggested that treatment was beneficial, the other suggested that it was detrimental. The most frequently used method in literature-based meta-analyses, the OR method, did not perform much better than the MR ratio method when translated into absolute survival differences to compare them with HRs. More detailed comparisons between the OR and the HR on individual patient data of a large meta-analysis in chemotherapy in head and neck cancer have yielded similar results (9). Conse-

quently, neither the median ratio nor the OR can be recommended as a surrogate method for analyzing time to event outcomes.

It is extremely important, certainly in oncology, to have precise estimates of the difference in survival between treatments. This requires calculating an HR either directly from the individual patient data or from relevant summary statistics. Neither the MR or OR method can provide a reasonable alternative to this. Trial reports of time to event outcomes, therefore, should provide sufficient statistical information in the form of the results of log rank test, the degrees of freedom, direction, and the exact p value or the HR, and its confidence interval to enable readers to calculate HRs directly from these statistics and for those conducting systematic reviews to combine these summary statistics to obtain pooled estimates of HRs from similar trials. However, it should be borne in mind that even if full summary statistics are reported, other biases (unpublished trials, patient exclusion, duration of follow-up) (20;26) can potentially influence the outcome of meta-analyses based on the published literature and that these will also need to be addressed in any high-quality systematic review.

CONTACT INFORMATION

Stefan Michiels, MSc, Biostatistician (michiels@igr.fr), Department of Public Health, Institut Gustave-Roussy, 39 rue Camille Desmoulins, 94805 Villejuif cedex, France
Pascal Piedbois, MD, PhD, Full Professor and Head (pascal.piedbois@hmn.ap-hop-paris.fr), Department of Medical Oncology, Assistance Publique—Hopitaux de Paris, Henri Mondor Hospital, 94000 Creteil, France
Sarah Burdett, MSc, Meta-analyst (sb@ctu.mrc.ac.uk), Meta-analysis Group, MRC Clinical Trials Unit, 222 Euston Road, London NW1 2DA, UK

Nathalie Syz, MSc, Biostatistician (nsyz@igr.fr), Department of Public Health, Institut Gustave-Roussey, 39 rue Camille Desmoulins, 94805 Villejuif, France

Lesley Stewart, PhD, Head (ls@ctu.mrc.ac.uk), Meta-analysis Group, MRC Clinical Trials Unit, 222 Euston Road, London NW1 2DA, UK

Jean-Pierre Pignon, MD, PhD, Head (jppignon@igr.fr), Meta-analysis Team, Department of Public Health, Institut Gustave-Roussey, 39 rue Camille Desmoulins, 94805 Villejuif, France

REFERENCES

1. Advanced Colorectal Cancer Meta-Analysis Project. Meta-analysis of randomized trials testing the biochemical modulation of fluorouracil by methotrexate in metastatic colorectal cancer. *J Clin Oncol*. 1994;12:960-969.

2. Advanced Colorectal Cancer Meta-Analysis Project. Modulation of fluorouracil by leucovorin in patients with advanced colorectal cancer: Evidence in terms of response rate. *J Clin Oncol*. 1992;10:896-903.

3. Altman DG, De Stavola BL, Love SB, Stepniewska KA. Review of survival analyses published in cancer journals. *Br J Cancer*. 1995;72:511-518.

4. Arnott SJ, Duncan W, Gignoux M, et al. Preoperative radiotherapy in esophageal carcinoma: A meta-analysis using individual patient data (Oesophageal Cancer Collaborative Group). *Int J Radiat Oncol Biol Phys*. 1998;41:579-583.

5. Auperin A, Arriagada R, Pignon JP, et al. Prophylactic cranial irradiation for patients with small-cell lung cancer in complete remission (Prophylactic Cranial Irradiation Overview Collaborative Group). *N Engl J Med*. 1999;341:476-484.

6. Buyse M, Ryan LM. Issues of efficiency in combining proportions of deaths from several clinical trials. *Stat Med*. 1987;6:565-576.

7. Chalmers I. The Cochrane Collaboration: Preparing, maintaining and disseminating systematic reviews of the effects of health care. *Ann N Y Acad Sci*. 1993;703:156-165.

8. Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ*. 1994;309:1286-1291.

9. 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;22:538-547.

10. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet*. 1991;337:867-872.

11. Glioma Meta-analysis Trialists (GMT) Group. Chemotherapy in adult high-grade glioma: A systematic review and meta-analysis of individual patient data from 12 randomised trials. *Lancet*. 2002;359:1011-1018.

12. Greenland S, Salvan A. Bias in the one-step method for pooling study results. *Stat Med*. 1990;9:247-252.

13. Meta-Analysis Group in Cancer. Alpha-interferon does not increase the efficacy of 5-fluorouracil in advanced colorectal cancer. *BMJ*. 2001;84:611-620.

14. Meta-Analysis Group in Cancer. Reappraisal of hepatic arterial infusion in the treatment of nonresectable liver metastases from colorectal cancer. *J Natl Cancer Inst*. 1996;88:252-258.

15. Meta-Analysis Group in Cancer. Toxicity of fluorouracil in patients with advanced colorectal cancer: Effect of administration schedule and prognostic factors. *J Clin Oncol*. 1998;163:537-541.

16. Moher D, Schulz KF, Altman DG. The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet*. 2001;357:1191-1194.

17. Non-small Cell Lung Cancer Collaborative Group. Chemotherapy in non-small cell lung cancer a meta-analysis using updated data on individual patients from 52 randomised clinical trials. *BMJ*. 1995;311:899-909.

18. Parmar MKB, Torri W, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Stat Med*. 1998;17:2815-2834.

19. Pignon JP, Arriagada R, Ihde DC, et al. A meta-analysis of thoracic radiotherapy for small-cell lung cancer. *N Engl J Med*. 1992;327:1618-1624.

20. Pignon JP, Hill C. Meta-analyses of randomised clinical trials in oncology. *Lancet Oncol*. 2001;2:475-482.

21. PORT Meta-analysis Trialists Group. Postoperative radiotherapy in non-small cell lung cancer: Systematic review and meta-analysis of individual patient data from nine randomised controlled trials. *Lancet*. 1998;352:257-263.

22. Simes RJ. Confronting publication bias: A cohort design for meta-analysis. *Stat Med*. 1987;6:11-29.

23. Simon R. Confidence intervals for reporting results of clinical trials. *Ann Intern Med*. 1986;105:429-435.

24. Stablein DM, Koutrouvelis IA. A two-sample test sensitive to crossing hazards in uncensored and singly censored data. *Biometrics*. 1985;41:643-652.

25. Stewart LA, Parmar MKB. Bias in the analysis and reporting of randomized controlled trials. *Int J Technol Assess Health Care*. 1996;12:264-275.

26. Stewart LA, Parmar MKB. Meta-analysis of the literature or of individual patient data: Is there a difference? *Lancet*. 1993;341:418-422.

27. Tierney JF, Clarke M, Stewart LA. Is there bias in the publication of individual patient data meta-analyses? *Int J Technol Assess Health Care*. 2000;16:657-667.

28. Tudur C, Williamson PR, Khan S, Best LY. The value of the aggregated data approach in meta-analysis with time-to-event outcomes. *J R Statist Soc A*. 2001;164:357-370.

29. Vale CL, Tierney JF, Stewart LA. Effects of adjusting for censoring on meta-analyses of time-to-event outcomes. *Int J Epidemiol*. 2002;31:107-111.

30. Whitehead A, Whitehead J. A general parametric approach to the meta-analysis of randomized clinical trials. *Stat Med*. 1991;10:1665-1677.

31. Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction: An overview of the randomized trials. *Prog Cardiovasc Dis*. 1985;27:335-371.