How Quickly Do Systematic Reviews Go Out of Date? A Survival Analysis

Article *in* Annals of internal medicine · August 2007

DOI: 10.7326/0003-4819-147-4-200708210-00179

1 author:

Kaveh Shojania
University of Toronto
198 PUBLICATIONS **11,131** CITATIONS

SEE PROFILE

ARTICLE | Annals of Internal Medicine

How Quickly Do Systematic Reviews Go Out of Date? A Survival Analysis

Kaveh G. Shojania, MD; Margaret Sampson, MLIS; Mohammed T. Ansari, MBBS, MMedSc, MPhil; Jun Ji, MD, MHA; Steve Doucette, MSc; and David Moher, PhD

Background: Systematic reviews are often advocated as the best source of evidence to guide clinical decisions and health care policy, yet we know little about the extent to which they require updating.

Objective: To estimate the average time to changes in evidence that are sufficiently important to warrant updating systematic reviews.

Design: Survival analysis of 100 quantitative systematic reviews.

Sample: Systematic reviews published from 1995 to 2005 and indexed in *ACP Journal Club*. Eligible reviews evaluated a specific drug or class of drug, device, or procedure and included only randomized or quasi-randomized, controlled trials.

Measurements: Quantitative signals for updating were changes in statistical significance or relative changes in effect magnitude of at least 50% involving 1 of the primary outcomes of the original systematic review or any mortality outcome. Qualitative signals included substantial differences in characterizations of effectiveness, new information about harm, and caveats about the previously reported findings that would affect clinical decision making.

Results: The cohort of 100 systematic reviews included a median of 13 studies and 2663 participants per review. A qualitative or quantitative signal for updating occurred for 57% of reviews (95% CI, 47% to 67%). Median duration of survival free of a signal for updating was 5.5 years (CI, 4.6 to 7.6 years). However, a signal occurred within 2 years for 23% of reviews and within 1 year for 15%. In 7%, a signal had already occurred at the time of publication. Only 4% of reviews had a signal within 1 year of the end of the reported search period; 11% had a signal within 2 years of the search. Shorter survival was associated with cardiovascular topics (hazard ratio, 2.70 [CI, 1.36 to 5.34]) and heterogeneity in the original review (hazard ratio, 2.15 [CI, 1.12 to 4.11]).

Limitation: Judgments of the need for updating were made without involving content experts.

Conclusions: In a cohort of high-quality systematic reviews directly relevant to clinical practice, signals for updating occurred frequently and within a relatively short time.

Ann Intern Med. 2007;147:224-233. www.annals.org

For author affiliations, see end of text.

Systematic reviews have become increasingly common in recent years (1) and are recommended by many as the best sources of evidence to guide both clinical decisions (2) and health care policy (3). For systematic reviews to fulfill these roles, their findings must remain relatively stable for at least several years or effective mechanisms must exist for alerting end users to important changes in evidence. Yet, surprisingly little research has assessed the extent to which systematic reviews become out of date or the rate at which this occurs (4–7). Some organizations, such as the Cochrane Collaboration, recommend updating systematic reviews every 2 years, but few empirical data guide this or other recommendations about updating.

We sought to determine how quickly systematic reviews meet explicitly defined criteria for changes in evidence of sufficient importance to warrant updating. We also sought to identify predictors of "survival time," the time to such important changes in evidence. Survival time might vary depending on many factors, including the type of question posed by the original review (for example, therapeutic or diagnostic), the types of studies included (for example, randomized trials or observational studies), and whether the systematic review provided quantitative synthesis. To limit such variation, we focused on systematic reviews of randomized, controlled trials that evaluated therapeutic benefit or harm by providing quantitative synthesis (meta-analysis) for at least 1 outcome.

See also:

Print
Editors' Notes 225
Editorial comment. 2

Web-Only
Appendix Figures
Conversion of figures and tables into slides
Audio summary

Methods

Study Design and Sample

We used a quasi-random process (alphabetical sort order by author) to select 100 systematic reviews that were indexed in *ACP Journal Club* with an accompanying commentary between January 1995 and December 2005 (with a search date no later than 31 December 2004 to ensure at least 1 full year for new evidence to appear). We chose this sampling frame because *ACP Journal Club* selects systematic reviews that meet explicit quality standards and are deemed directly relevant to clinical practice (8). We regarded the sample size of 100 as sufficiently large to achieve suitably narrow confidence intervals and to permit evaluation of up to 5 potential predictors of survival.

Eligibility Criteria

Eligible reviews evaluated the benefit or harm of a specific drug, class of drug, device, or procedure (invasive

procedure or surgery) and included randomized or quasi-randomized, controlled trials. We excluded evaluations of alternative and complementary medicines because the stability of reviews of such therapies might differ substantially from reviews of conventional therapies.

We required that reviews provide a point estimate and 95% confidence interval for at least 1 outcome in the form of a relative risk, odds ratio, or absolute risk difference for binary outcomes and weighted mean differences for continuous outcomes. We excluded meta-analyses of individual-patient data, meta-regressions, and indirect meta-analyses because of the difficulty of determining whether new data would alter previous quantitative results. Two team members independently assessed eligibility, with disagreements resolved by consensus involving a third reviewer. When more than 1 review on the same topic was identified, only the earliest was included.

Searching

For each review, searches for new trials included identifying new systematic reviews on the same topic, submitting relevant content terms to the Clinical Queries function in Ovid, applying the Related Articles function in PubMed to the 3 largest and the 3 most recent trials in the original review (up to 6 trials in total), and using Scopus (www.scopus.com/scopus/home.url) to identify new randomized trials that cited the original review. When these search strategies yielded no eligible new trials, we conducted more comprehensive electronic searches and reviewed relevant chapters in such sources as *Clinical Evidence* and *UpToDate* to ensure that we had not missed new trials.

Team members who had backgrounds in both medicine and clinical research screened citations retrieved by the preceding methods to identify trials that would have met the inclusion criteria in the original review. Retrieved articles were screened in chronological order to ascertain quantitative or qualitative signals for the need for updating. The review protocol stopped when any criteria for updating were met. Each systematic review was discussed in detail, with the final status—signal for updating was or was not detected—adjudicated by consensus (**Figure 1**).

Signals for the Need to Update Systematic Reviews

In designing criteria for comparing new findings with those in a previous review, we adapted methods used by other investigators to address similar problems with comparing 2 sets of results relating to the same question (9–13), such as randomized and nonrandomized studies of the same intervention. These investigators identified conflicting findings among different publications using a combination of quantitative thresholds for differences in effect magnitude and qualitative judgments about the language used to describe the results. We have similarly conceptualized quantitative and qualitative signals of potential changes in evidence that are sufficiently important to warrant updating previous systematic reviews.

Context

Clinicians rely on systematic reviews for current, evidence-based information.

Contribution

This survival analysis of 100 meta-analyses indexed in *ACP Journal Club* from 1995 to 2005 found new evidence that substantively changed conclusions about the effectiveness or harms of therapies occurred frequently within relatively short time periods. The median survival time without substantive new evidence for the meta-analyses was 5.5 years. Significant new evidence was already available for 7% of the reviews at the time of publication and became available for 23% within 2 years.

Implication

Clinically important evidence that alters conclusions about the effectiveness and harms of treatments can accumulate rapidly.

—The Editors

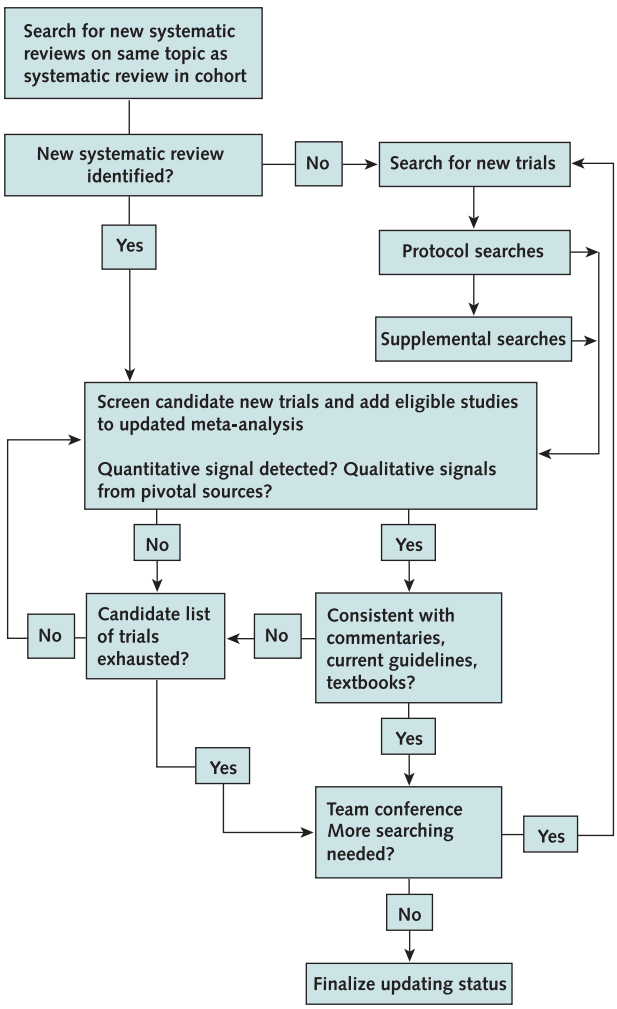
Quantitative Signals

Quantitative signals consisted of a change in statistical significance or relative change in effect magnitude of at least 50%. We restricted these changes to those involving 1 of the primary outcomes of the original review or any mortality outcome. We also ignored trivial changes in statistical significance—when the original and updated meta-analytic results both had *P* values between 0.04 and 0.06—so that quantitative signals of changes in evidence would represent robust indicators of the need to update previous reviews. Quantitative signals were detected by combining data from eligible new trials with the previous results using a fixed-effects approach. Use of fixed-effect models allowed pooling of the new trials with the previous meta-analytic result, as opposed to having to obtain original data from all of the included trials in each of 100 systematic reviews. Although random-effects models are usually preferred to avoid spurious precision in the face of heterogeneity, our goal consisted of detecting potential changes in evidence that would warrant a formal update, not producing exact estimates of the updated results.

Qualitative Signals

Qualitative signals included new information about harm sufficient to affect clinical decision making, important caveats to the original results, emergence of a superior alternate therapy, and important changes in certainty or direction of effect. Qualitative signals were detected by using explicit criteria for comparing the language in the original review with descriptions of findings in new systematic reviews that addressed the same topic, pivotal trials, clinical practice guidelines, or recent editions of major textbooks (for example, *UpToDate*). Pivotal trials were defined as tri-

Figure 1. **Overall process for determining updating status.**



Includes the search protocols to identify candidate new trials, application of criteria from the original review to identify eligible new trials, meta-analytic pooling of new results with previous meta-analytic result, and identification of new systematic reviews on the same topic or "pivotal trials" (published in 1 of the 5 highest-impact general medical journals or more than 3 times the sample size of the previous largest trial) that met any of our criteria for qualitative signals for updating. An individual reviewer reached a tentative conclusion about the presence of quantitative and qualitative signals for updating, but each review was discussed in detail by the project team to reach a final consensus decision. For reviews that did not have any signals for updating, the group also decided whether the searches had been adequate or whether more comprehensive searching for new trials might be required, including more detailed electronic searching and hand-searching of *ACP Journal Club* to look for new trials relevant to the original review.

als that had a sample size at least 3 times larger than that of the previous largest trial or were published in 1 of the 5 highest-impact general medical journals (*The New England Journal of Medicine*, *Lancet*, *Journal of the American Medical Association*, *Annals of Internal Medicine*, and the *British Medical Journal*).

We defined 2 levels of importance for qualitative signals: "potentially invalidating changes in evidence," which would make one no longer want clinicians or policymakers to base decisions on the original findings (such as a pivotal trial that characterized treatment effectiveness in terms opposite of those in the original systematic review), and "major changes in evidence," which would affect clinical decision making in important ways without invalidating the previous results (such as the identification of patient populations for whom treatment is more or less beneficial). Major changes also included differing characterizations of effectiveness that were less extreme than those for potentially invalidating signals but that would still affect clinical decision making (for example, a change from "possibly beneficial" to "definitely beneficial"). Of importance, such characterizations as "possibly effective," "probably effective," and "promising," were all categorized as "possibly effective." Thus, qualitative signals for changes in evidence captured substantive differences in the characterization of treatment effects, not merely semantic differences. Full definitions for each of the specific signals can be found at www.ohri.ca/UpdatingSystRevs.

Data Collection

For each review, we characterized the clinical content area, eligibility criteria for included trials, definitions of reported outcomes, number of included trials and participants, meta-analytic result for each outcome, identification of statistical heterogeneity, and excerpted quotations of the authors' characterizations of the main results. We also abstracted whether a given outcome was explicitly identified as 1 of the "primary" or "main" outcomes. We discounted identification of more than 3 such outcomes as inconsistent with the concept of a primary outcome.

Survival Analysis

For each systematic review, we defined "birth" as publication date and "death" as the occurrence of a qualitative or quantitative signal for updating. Observations were censored on 1 September 2006, the midpoint of the 4-month period during which searches were done for the entire cohort.

We fit nonparametric Kaplan–Meier curves and used multivariable Cox proportional hazards models to examine the association between survival and various features of the systematic reviews, including clinical content area, number of included trials, identification of heterogeneity, and "activity in the field"—defined as present if the review included at least 1 trial published within the last year of its search period or if the review identified ongoing trials eligible for inclusion. We also assessed a potential predictor known only by reviewing the literature published after publication of the original review: the magnitude of the increase in the number of eligible new trials. In addition to the proportional hazards analysis to estimate predictors of survival, we used logistic regression to identify predictors of survival for less than 2 years. All analyses were done with SAS, version 9.0 (SAS Institute, Cary, North Carolina).

Role of the Funding Source

This work was done under contract with the Agency for Healthcare Research and Quality. The funding source did not have a role in the study design; data collection, analysis, or interpretation; or the decision to submit the manuscript for publication.

RESULTS

A search of the Ovid database for *ACP Journal Club* retrieved 651 potential systematic reviews. Achieving the target cohort size of 100 reviews necessitated that we screen the first 325 of these records (**Appendix Figure 1**, available at www.annals.org).

Each of the 100 systematic reviews included a median of 13 studies (interquartile range, 8 to 21) and 2663 participants (interquartile range, 1281 to 8371) (**Table 1**). Most reviews evaluated drug therapies; the most common clinical content area was cardiovascular medicine (**Table 1**). We were able to identify at least 1 new eligible trial for 85 systematic reviews, with a median of 4 new trials (interquartile range, 1 to 7) and 1160 patients (interquartile range, 170 to 3689) per review.

Table 1. **Characteristics of the Cohort of 100 Systematic Reviews**

Characteristic	Composition of Cohort, *n*
Publication type	
Peer-reviewed journal article	72
Cochrane review	27
Health technology assessment	1*
Therapy evaluated	
Medications	85
Medical devices	8
Procedures	7
Clinical topic area	
Cardiovascular	20
Gastroenterology	13
Neurology	11
Other 10 categories	<10 each
Publication period	
January 1995–February 1997	16
March 1997–April 1999	22
May 1999–June 2001	25
July 2001–August 2003	20
September 2003–December 2005	16
Median included trials	13 (interquartile range, 8–21)
Median included participants	2663 (interquartile range, 1281–8371)

* Published by the Canadian Agency for Drugs and Technologies in Health.

Signals for Updating

A quantitative signal for updating occurred for 20 of the 100 systematic reviews. Qualitative signals occurred for 54 reviews, including 8 that met criteria for potentially invalidating changes in evidence. Qualitative signals were derived from new systematic reviews in 23 cases and from pivotal trials in 25 cases. The primary event of interest, a quantitative or qualitative signal for updating, occurred for 57% of reviews (95% CI, 47% to 67%) in the cohort.

Table 2 (9–22) presents examples of signals for updating. The 3 reviews (9, 11, 13) that had a qualitative signal for "opposing findings" are self-explanatory. For example, in the first case, the original review reported that "for every 20 critically ill patients treated with albumin there is one additional death" (9). A subsequent trial (10) with almost 5 times the sample size of previous trials combined showed no such increase. Of the 2 reviews with important differences in characterization short of "opposing findings," 1 of them borders on opposing findings (15). For the prevention of stroke in high-risk patients, the original review (15) mentioned that the addition of dipyridamole to aspirin was associated with a nonsignificant 6% reduction in serious vascular events, but it concluded that the "addition of dipyridamole to aspirin produced no significant further reduction in vascular events compared with aspirin alone." Consistent with our efforts to avoid overcalling changes in evidence, we characterized the original review as consistent with "possible benefit." Thus, the change from this characterization to the definite benefit reported in a subsequent large trial (16) fell short of opposing (and potentially invalidating) the previous findings but still met our criteria for a major change in the characterization of effectiveness. All 3 examples of reviews with qualitative signals for "opposing findings" and the 2 examples of reviews with important differences in characterization short of "opposing findings" also generated at least 1 quantitative signal.

Table 2 also shows an example of a clinically significant caveat (lack of sustained benefit reported from allergen immunotherapy for asthma) (20) and an example of expansion of evidence to a new patient population (secondary prevention for patients with recent stroke) (22). Expansion of benefit for statins from the indications established in the original review (21) to secondary prevention in patients with recent stroke may strike some as not a major change in evidence. However, as emphasized in the new trial itself (22), the editorial that accompanied it (23), and the commentary in *ACP Journal Club* (24), this trial was the first to evaluate the effects of statins on patients who had cerebrovascular disease but not known coronary artery disease. The commentaries also characterized this trial as providing evidence for the increasingly widespread practice of adding statins to the standard treatment for patients with acute stroke, recommendations that had previously been derived from analogies with the treatment for cardiac ischemia. Thus, we regarded this new trial as meeting our criterion of expanding the evidence from the original review in a manner that would be expected to affect practice. More detailed explanations and additional examples of signals for updating can be found at www.ohri.ca/UpdatingSystRevs.

Table 2. **Examples of Quantitative and Qualitative Signals**

Original Systematic Review	Quantitative Signal		Qualitative Signal	
	Change in Statistical Significance (95% CI)	Change in Effect Size ≥50%	Specific Signal	Explanation
Alderson et al., 2002 (9)	Relative risk for death became nonsignificant: 1.68 (1.26–2.23) → 1.04 (0.95–1.13)	Relative increase in risk for death of 0.68 → only 0.04	*Opposing findings:* probable/possible increase in mortality → definite conclusion of no such increase	Original review reported a "strong suggestion" that albumin increases the risk for death, stating that "for every 20 critically ill patients treated with albumin there is one additional death." Trial published in high-impact journal and having almost 5 times the sample size of previous trials combined showed no such increase and concluded: "use of either 4 percent albumin or normal saline for fluid resuscitation results in similar outcomes at 28 days (10)."
Bucher et al., 1996 (11)		Relative reduction in odds of preeclampsia (0.62 → 0.21) and odds of developing hypertension (0.70 → 0.25)	*Opposing findings:* definite benefit in 2 primary outcomes → definite conclusion of no such benefit	Original review reported a definite reduction in preeclampsia and development of hypertension. Trial published in high-impact journal and having almost double the sample size of previous trials combined found no statistically significant improvement in either outcome and concluded that "calcium supplementation during pregnancy did not prevent preeclampsia, pregnancy-associated hypertension, or adverse perinatal outcomes (12)."
Lord et al., 2003 (13)	Based on 3 new trials, increase in ovulation with metformin plus clomifene vs. clomifene alone became nonsignificant: odds ratio of 4.41 (2.37–8.22) → 1.42 (0.98–2.05)	For metformin and clomifene vs. clomifene alone, reduction in increased odds of pregnancy (3.4 → 1.1) and ovulation rate (3.4 → 0.4)	*Opposing findings:* definite improvement in key outcomes and recommendation as first-line agent → no such benefit and recommendation against use as first-line agent	Original review found that metformin statistically significantly increased achievement of ovulation in the comparison of metformin with placebo and metformin plus clomifene vs. clomifene alone, as well as statistically significantly increased pregnancy rates. The authors concluded that "its choice as a first line agent seems justified." Trial published in high-impact journal with sample size more than twice that of previous largest trial showed no such benefits and concluded that "metformin is not an effective addition to clomifene citrate as the primary method of inducing ovulation in women with polycystic ovary syndrome (14)."
Antithrombotic Trialists' Collaboration, 2002 (15)	Odds ratio of 0.94 (0.84–1.06) → 0.82 (0.74–0.91)		*Important differences in characterization short of "opposing findings":* probably/possibly beneficial → definitely beneficial	Original review stated that "the addition of dipyridamole to aspirin was associated with only a nonsignificant further 6% reduction in serious vascular events." It highlighted that "the apparent reduction in non-fatal stroke was derived mainly from one large study" and "this result was not supported by the findings for non-fatal stroke in the other studies." It concluded that the "addition of dipyridamole to aspirin produced no significant further reduction in vascular events compared with aspirin alone." A trial published in a high-impact journal and larger than the previous largest study reported an absolute risk reduction in the primary outcome (a composite of death from all vascular causes, nonfatal stroke, nonfatal myocardial infarction, or major bleeding complication, whichever happened first) of 1.0% per year (CI, 0.1%–1.8%). The authors concluded that "[these] results, combined with the results of previous trials, provide sufficient evidence to prefer the combination regimen of aspirin plus dipyridamole over aspirin alone as antithrombotic therapy after cerebral ischaemia of arterial origin (16)."

Table 2—**Continued**

Original Systematic Review	Quantitative Signal		Qualitative Signal	
	Change in Statistical Significance (95% CI)	Change in Effect Size ≥50%	Specific Signal	Explanation
Birck et al., 2003 (17)	Relative risk for contrast nephropathy with acetylcysteine lost its statistical significance: 0.44 (0.22, 0.88) → 0.81 (0.58, 1.13)	Relative risk reduction for contrast nephropathy decreased: 0.56 → 0.19	*Important differences in characterization short of "opposing findings":* definitely beneficial → probably/possibly beneficial	The original review included 7 trials and found that "compared with periprocedural hydration alone, administration of acetylcysteine and hydration significantly reduced the relative risk of contrast nephropathy by 56% (0.435 [95% CI 0.215–0.879], p=0.02) in patients with chronic renal insufficiency." The authors acknowledged that it remained unclear to what extent acetylcysteine improved harder clinical end points, but the impact on measures of renal function was regarded as robust. They concluded that acetylcysteine "significantly reduces the risk of contrast nephropathy in patients with chronic renal insufficiency." A subsequent meta-analysis that included 20 trials reported a decreased effect magnitude that lost its statistical significance. The authors also emphasized that the trials exhibited significant, unexplained heterogeneity. They concluded that "acetylcysteine may reduce the incidence of contrast-related nephropathy, but this finding is reported inconsistently across currently available trials. High-quality, large clinical trials are needed before acetylcysteine use in this indication can be recommended universally" (18).
Abramson et al., 1995 (19)			*Clinically important caveat:* benefit reported in original review is not sustained	Original review reported statistically significant improvements in symptomatic improvement from immunotherapy with any allergen, statistically significant reduction in ongoing medication requirements after mite immunotherapy, and statistically significant improvements in bronchial hyperreactivity and other intermediate outcomes. The authors thus concluded that "allergen immunotherapy is a treatment option in highly selected patients with extrinsic ('allergic') asthma." A trial published in a high-impact journal (20) reported improvements in various outcomes during the first year of treatment but loss of statistically significant benefits in the second year. The authors also pointed out that reduced medication costs were counterbalanced by the costs of immunotherapy.
Bucher et al., 1998 (21)			*Expansion of treatment:* benefit demonstrated for a new patient population	Original review concluded that "randomized, controlled trials suggests that in hyperlipidemic patients who have not previously had stroke, HMGcoA reductase inhibitors reduce the incidence of stroke." This review included patients with and without previous coronary artery disease but excluded those with previous stroke because of heterogeneity. Thus, for patients who had had stroke, the benefit of statins remained unclear. A large trial published in a high-impact journal showed that "in patients with recent stroke or TIA and without known coronary heart disease, 80 mg of atorvastatin per day reduced the overall incidence of strokes and of cardiovascular events, despite a small increase in the incidence of hemorrhagic stroke (22)."

Survival Analysis

Median survival free of a signal for updating was 5.5 years (CI, 4.6 to 7.6 years) (**Figure 2**). For the 57 reviews with signals for updating, median time to event was 3.0 years (interquartile range, 0.9 to 5.1 years). However, a signal for updating occurred within 2 years for 23% of reviews (CI, 15% to 33%) and within 1 year for 15% (CI, 9% to 24%). For 7% of reviews (CI, 3% to 14%), a signal had already occurred at the time of publication. Even with restriction only to quantitative signals or "potentially inval-

idated changes in evidence," signals for updating occurred within 2 years for 15% of reviews. Restricting the analysis solely to quantitative signals, 12% of reviews had signals for updating within 2 years and 7% within 1 year, including 4 reviews for which the quantitative signal had already occurred at publication.

In univariate analyses (**Table 3**), shorter survival was associated with a clinical content area of cardiovascular medicine (hazard ratio, 2.58 [CI, 1.39 to 4.78]) (**Appendix Figure 2**, available at www.annals.org) and an increase in the total number of patients by a factor of 2 or more (hazard ratio, 1.79 [CI, 1.03 to 3.10]) (**Appendix Figure 3**, available at www.annals.org). Multivariate analysis produced 3 noteworthy changes to these results: heterogeneity in the original review became a statistically significantly predictor for a signal for updating (hazard ratio, 2.15 [CI, 1.12 to 4.11]), an increase in the total number of patients by a factor of 2 or more lost statistical significance as a predictor, and including more than the median of 13 trials became a borderline statistically significant predictor of increased survival (hazard ratio, 0.56 [CI, 0.30 to 1.03]; $P = 0.06$).

The 5 variables shown in **Table 3** represent those we had considered a priori as the most plausible potential predictors. Other potential predictors that were tested in secondary analyses included the source of the systematic review (Cochrane vs. non-Cochrane), included number of participants greater than the median of 2663, detection or suspicion of publication bias in the original review, and several variables related to increases in the number of trials or participants in the literature since the original review. None of these features statistically significantly predicted signals for updating.

Prediction of Signals for Updating within 2 Years of Publication

No variable statistically significantly predicted a signal for updating within 2 years. However, cardiovascular topics showed a nonsignificant increase in the odds of a signal for updating within 2 years (odds ratio, 2.67 [CI, 0.88 to 8.10] $P = 0.08$), as did an increase in the total number of patients by a factor of 2 or more (odds ratio, 2.29 [CI, 0.84 to 6.25]; $P = 0.11$). Sensitivity analyses involving different time frames, such as occurrence of a signal within 3 years, yielded similar results.

Effect of Production and Publication Times on Survival

The median time between the end of the search period and the publication date for a systematic review was 1.1 years (interquartile range, 0.8 to 1.7 years). Time to publication did not differ substantially between Cochrane and journal reviews, nor did it decrease statistically significantly from 1995 to 2005.

When survival analyses were repeated by using the end of the search period as "birth," median survival was 6.9 years (CI, 6.1 to 9.0 years). A signal for updating occurred within 3 years of the search for 20% of reviews (CI, 13% to 29%), within 2 years for 11% (CI, 6% to 19%), and within 1 year for 4% (CI, 1% to 11%). Predictors of survival did not differ from those identified in the analysis that used publication date as "birth."

DISCUSSION

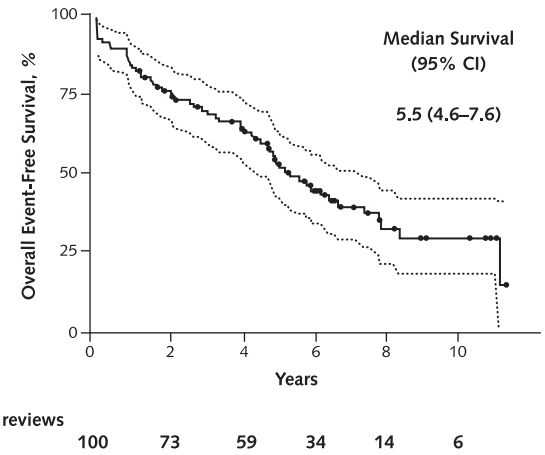
In a cohort of high-quality systematic reviews directly relevant to clinical practice, median survival was 5.5 years. However, signals for updating occurred within 2 years for 23% of reviews, within 1 year for 15%, and before publication for 7%. We found several statistically significant predictors of signals for updating, but no features predicted which reviews would require attention within 2 years.

Our results indicated a far greater need for updating than the only other such evaluation, a comparison of Cochrane reviews from 1998 with their updates in 2002 that reported important changes in conclusions in just 9% of reviews (4). Of importance, that study relied exclusively on interpretations of new evidence by authors of the original review, who might be disinclined to find new evidence or report important changes. Also, only 70% of Cochrane reviews had updates. It is possible that reviewers were less likely to update large increases in the number of new trials or major changes in conclusions given the greater work involved. Finally, Cochrane reviews differ in important respects, such as clinical topic coverage, from other peer-reviewed systematic reviews (25).

We restricted our cohort to systematic reviews of randomized trials of conventional drugs, devices, or procedures that reported meta-analytic results for at least 1 dichotomous outcome. Our exclusion of qualitative reviews, reviews of nontherapeutic topics, meta-analyses of individual-patient data, and meta-regressions reflected our concern that rates of change in evidence might differ across these different types of reviews. Thus, we acknowledge that our results may not generalize to all reviews. That said, as shown in **Appendix Figure 1** (available at www.annals.org), excluding the records retrieved by our initial electronic search that were not systematic reviews, 139 of the first 287 systematic reviews (48%) were eligible for inclusion. Thus, although our cohort may seem highly selected, approximately half of the reviews indexed in *ACP Journal Club* were eligible for inclusion in our cohort. Granted, *ACP Journal Club* itself represents a nonrandom sample of all systematic reviews insofar as it selects reviews that meet certain quality standards and have high potential to affect clinical practice. However, these biases strengthen our results because such reviews represent those one would hope had the greatest stability.

The main limitation of our findings is that the assessments of the need to update previous reviews did not involve input from experts in the relevant content areas. However, our approach of having investigators apply explicit qualitative and quantitative criteria to compare 2 sets of results addressing the same question of interest repre-

Figure 2. **Overall survival time (95% CI) free of signals for updating.**



The immediate decrease in survival at time zero reflects the 7 systematic reviews for which signals for updating had already occurred at the time of publication. The low number of reviews at risk after 10 years reflects the fact that the sample spanned 1995 to 2005 and censoring occurred on 1 September 2006. Thus, only reviews published before September 1996 and having no signals for updating could have more than 10 years of observation.

sents the norm in methodological research of this type (26–32). The notable exception was an evaluation of the average shelf life of clinical practice guidelines (7). By choosing a few guidelines (17 in total) produced by a single agency, the investigators were able to ask the authors of the original guidelines to assess changes in evidence. Using such an approach was not feasible for our analysis of a much larger sample of 100 systematic reviews. However, we chose quantitative signals of changes in evidence that few would question as important and used explicit criteria for comparing the language of new findings with those of the original review. Moreover, we used expert sources, such as editorials and textbooks, to confirm our assessments wherever possible.

It is also important to note that our judgments concerned signals of the need to update previous systematic reviews, not definitive judgments about actual changes in evidence. If a previous review concluded that a treatment was effective and a trial in a high-impact journal concluded that the treatment had no benefit, we would count the new result as a signal for updating. We regard such a signal as reasonable for 2 reasons. First, a formal update that incorporated the new evidence might in fact yield conclusions that differ substantially from those of the previous review. Second, even if a formal update would not change the conclusions, the publication of a new trial in a high-impact journal would raise important questions for clinicians about the previous review. In fact, they might preferentially act on the trial's conclusions precisely because it appeared in a high-impact journal. Thus, it would be important to reassert the findings of the original review in an update that explicitly addressed the new evidence.

Ideally, readily discernible features of systematic reviews would indicate whether major changes in evidence were likely to appear within short time frames. Although several features statistically significantly predicted survival, no features adequately distinguished reviews that would require updating within 2 years from those that would not. Our modest sample size of 100 reviews limited our ability to test predictors of survival. However, it is unlikely that we would miss associations of the magnitude required to identify reviews that will probably require updating within short time frames with useful positive and negative predictive values.

Our results have important implications for those who produce, publish, and use systematic reviews. Publishers probably cannot reduce the time for the peer review and publication processes for systematic reviews beyond the benchmarks already attempted for submissions of all types.

Table 3. **Predictors of Signals for Updating in the Cohort of 100 Systematic Reviews**

Features	Univariate Hazard Ratio (95% CI)	*P* Value	Multivariate Hazard Ratio (95% CI)	*P* Value
Original review or new evidence				
Clinical topic area				
Cardiovascular	2.58 (1.39–4.78)	0.003	2.70 (1.36–5.34)	0.004
Neurology	1.37 (0.59–3.16)	0.47	1.08 (0.44–2.68)	0.86
Gastroenterology	1.35 (0.58–3.13)	0.48	1.12 (0.47–2.67)	0.80
Other	Reference	–	Reference	–
Heterogeneity present or suspected	1.64 (0.94–2.86)	0.08	2.15 (1.12–4.11)	0.02
Activity in field*	1.36 (0.76–2.44)	0.30	1.31 (0.68–2.52)	0.41
Included > 13 trials (median)	0.79 (0.46–1.33)	0.37	0.56 (0.30–1.03)	0.06
New evidence				
Doubling of total included participants	1.79 (1.03–3.10)	0.04	1.14 (0.65–2.01)	0.65

* Recent activity defined as present if the original systematic review included at least 1 trial published within the final 12 months of the search period or identified ongoing trials eligible for inclusion.

However, authors might consider submitting their work to the journals that are most likely to accept a given review to avoid delays because of multiple iterations of the peer review process. When the process of submission and rejection from other journals has resulted in the passage of more than 1 year from the date of the previous search, authors should update the search before resubmission, as we found that only 4% of reviews had signals for updating within 1 year of the previous search date. In fact, journals might consider requiring that authors update searches more than 1 year old before submitting systematic reviews. Finally, users of systematic reviews need to recognize that changes in evidence relevant to clinical decision making can occur within relatively short time frames. Once the search date is older than even 1 year, users should check for more recent trials on the same topic to see whether new evidence has altered the findings of a given systematic review. In some cases, such changes will already have occurred at the time of publication.

From the Ottawa Health Research Institute, University of Ottawa, Chalmers Research Group, and Children's Hospital of Eastern Ontario Research Institute, Ottawa, Ontario, Canada.

Disclaimer: The authors of this report are responsible for its content. Statements in the report should not be construed as endorsement by the Agency for Healthcare Research and Quality or the U.S. Department of Health and Human Services.

Acknowledgments: The authors thank Keith O'Rourke for statistical advice, Jessie McGowan and Tamara Rader for assistance with searches, and Alison Jennings for assistance with development of the meta-analytic worksheet. They also gratefully acknowledge Dr. David Atkins and the members of the technical advisory panel for the project funded by the Agency for Healthcare Research and Quality from which this work derives: Drs. Paul Shekelle, Evelyn Whitlock, Cynthia Mulrow, Doug Altman, Martin Eccles, and P.J. Devereaux.

Grant Support: By the Agency for Healthcare Research and Quality, U.S. Department of Health and Human Services (contract no. 290-02-0021). Dr. Shojania received additional salary support from the Government of Canada Research Chairs program. Dr. Moher is the recipient of a University of Ottawa Research Chair.

Potential Financial Conflicts of Interest: None disclosed.

Reproducible Research Statement: The data set is available to interested readers by contacting Dr. Shojania (e-mail, kshojania@ohri.ca); statistical code can be obtained from Mr. Doucette (e-mail, sdoucette@ohri.ca).

Requests for Single Reprints: Kaveh G. Shojania, MD, The Ottawa Hospital–Civic Campus, 1053 Carling Avenue, Room C403, Box 693, Ottawa, Ontario K1Y 4E9, Canada; e-mail, kshojania@ohri.ca.

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. [PMID: 11525102]

2. **Mulrow CD, Cook DJ, Davidoff F.** Systematic reviews: critical links in the great chain of evidence [Editorial]. Ann Intern Med. 1997;126:389-91. [PMID: 9054284]

3. **Bero LA, Jadad AR.** How consumers and policymakers can use systematic reviews for decision making. Ann Intern Med. 1997;127:37-42. [PMID: 9214251]

4. **French SD, McDonald S, McKenzie JE, Green SE.** Investing in updating: how do conclusions change when Cochrane systematic reviews are updated? BMC Med Res Methodol. 2005;5:33. [PMID: 16225692]

5. **Moher D, Tsertsvadze A.** Systematic reviews: when is an update an update? Lancet. 2006;367:881-3. [PMID: 16546523]

6. **Moher D, Tsertsvadze A, Tricco A, Eccles M, Grimshaw J, Sampson M, Barrowman N.** Systematic review identified few methods and strategies describing when and how to update systematic reviews. J Clin Epidemiology; [Forthcoming].

7. **Shekelle PG, Ortiz E, Rhodes S, Morton SC, Eccles MP, Grimshaw JM, et al.** Validity of the Agency for Healthcare Research and Quality clinical practice guidelines: how quickly do guidelines become outdated? JAMA. 2001; 286:1461-7. [PMID: 11572738]

8. **Shea B, Boers M, Grimshaw JM, Hamel C, Bouter LM.** Does updating improve the methodological and reporting quality of systematic reviews? BMC Med Res Methodol. 2006;6:27. [PMID: 16772030]

9. **Alderson P, Bunn F, Lefebvre C, Li WP, Li L, Roberts I, et al.** Human albumin solution for resuscitation and volume expansion in critically ill patients. Cochrane Database Syst Rev. 2002:CD001208. [PMID: 11869596]

10. **Finfer S, Bellomo R, Boyce N, French J, Myburgh J, Norton R; SAFE Study Investigators.** A comparison of albumin and saline for fluid resuscitation in the intensive care unit. N Engl J Med. 2004;350:2247-56. [PMID: 15163774]

11. **Bucher HC, Guyatt GH, Cook RJ, Hatala R, Cook DJ, Lang JD, et al.** Effect of calcium supplementation on pregnancy-induced hypertension and preeclampsia: a meta-analysis of randomized controlled trials. JAMA. 1996;275: 1113-7. [PMID: 8601931]

12. **Levine RJ, Hauth JC, Curet LB, Sibai BM, Catalano PM, Morris CD, et al.** Trial of calcium to prevent preeclampsia. N Engl J Med. 1997;337:69-76. [PMID: 9211675]

13. **Lord JM, Flight IH, Norman RJ.** Metformin in polycystic ovary syndrome: systematic review and meta-analysis. BMJ. 2003;327:951-3. [PMID: 14576245]

14. **Moll E, Bossuyt PM, Korevaar JC, Lambalk CB, van der Veen F.** Effect of clomifene citrate plus metformin and clomifene citrate plus placebo on induction of ovulation in women with newly diagnosed polycystic ovary syndrome: randomised double blind clinical trial. BMJ. 2006;332:1485. [PMID: 16769748]

15. **Antithrombotic Trialists' Collaboration.** Collaborative meta-analysis of randomised trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. BMJ. 2002;324:71-86. [PMID: 11786451]

16. **Halkes PH, van Gijn J, Kappelle LJ, Koudstaal PJ, Algra A; ESPRIT Study Group.** Aspirin plus dipyridamole versus aspirin alone after cerebral ischaemia of arterial origin (ESPRIT): randomised controlled trial. Lancet. 2006;367:1665-73. [PMID: 16714187]

17. **Birck R, Krzossok S, Markowetz F, Schnülle P, van der Woude FJ, Braun C.** Acetylcysteine for prevention of contrast nephropathy: meta-analysis. Lancet. 2003;362:598-603. [PMID: 12944058]

18. **Nallamothu BK, Shojania KG, Saint S, Hofer TP, Humes HD, Moscucci M, et al.** Is acetylcysteine effective in preventing contrast-related nephropathy? A meta-analysis. Am J Med. 2004;117:938-47. [PMID: 15629733]

19. **Abramson MJ, Puy RM, Weiner JM.** Is allergen immunotherapy effective in asthma? A meta-analysis of randomized controlled trials. Am J Respir Crit Care Med. 1995;151:969-74. [PMID: 7697274]

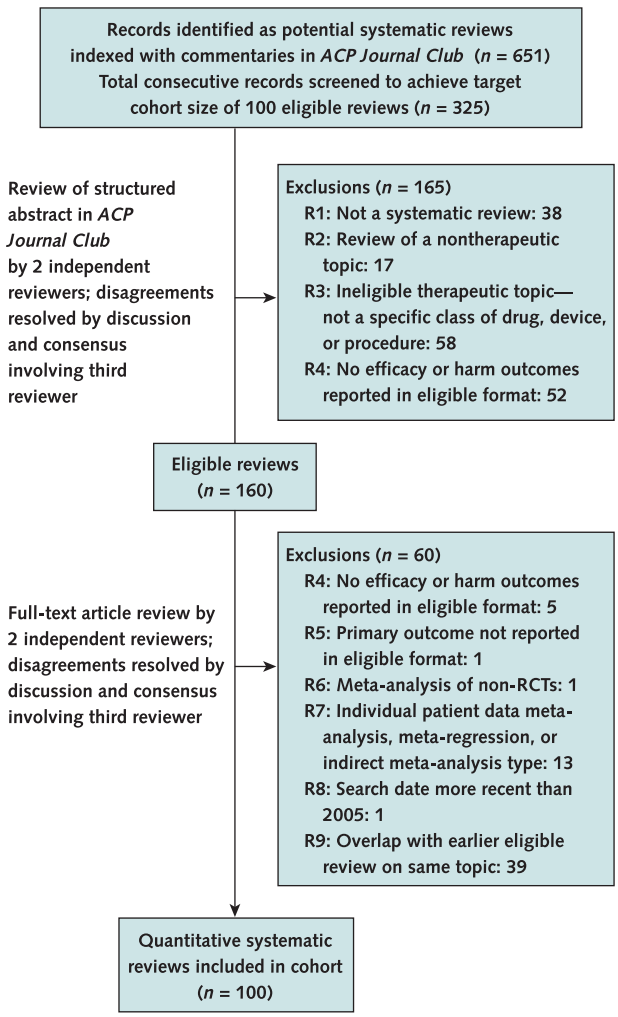
20. **Creticos PS, Reed CE, Norman PS, Khoury J, Adkinson NF Jr, Buncher CR, et al.** Ragweed immunotherapy in adult asthma. N Engl J Med. 1996;334: 501-6. [PMID: 8559203]

21. **Bucher HC, Griffith LE, Guyatt GH.** Effect of HMGcoA reductase inhibitors on stroke. A meta-analysis of randomized, controlled trials. Ann Intern Med. 1998;128:89-95. [PMID: 9441587]

22. **Amarenco P, Bogousslavsky J, Callahan A, Goldstein LB, Hennerici M, Rudolph AE, et al; Stroke Prevention by Aggressive Reduction in Cholesterol Levels (SPARCL) Investigators.** High-dose atorvastatin after stroke or transient

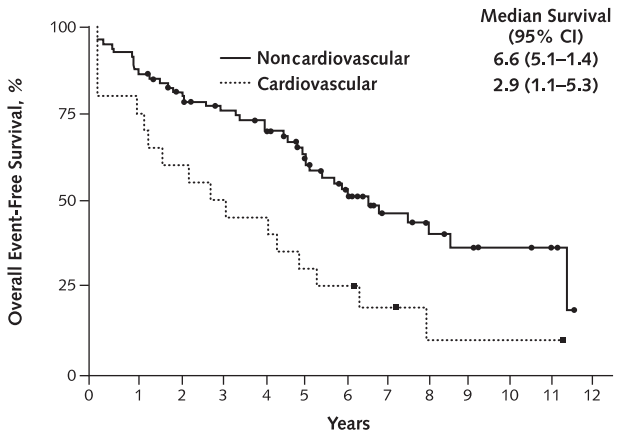
ischemic attack. N Engl J Med. 2006;355:549-59. [PMID: 16899775]
23. **Kent DM.** Stroke—an equal opportunity for the initiation of statin therapy [Editorial]. N Engl J Med. 2006;355:613-5. [PMID: 16899782]
24. **Cheng E.** Atorvastatin reduced stroke and CV events after recent stroke or TIA in patients with no known coronary heart disease. ACP J Club. 2007;146:7. [PMID: 17203927]
25. **Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG.** Epidemiology and Reporting Characteristics of Systematic Reviews. PLoS Med. 2007;4:e78. [PMID: 17388659]
26. **Benson K, Hartz AJ.** A comparison of observational studies and randomized, controlled trials. N Engl J Med. 2000;342:1878-86. [PMID: 10861324]
27. **Cappelleri JC, Ioannidis JP, Schmid CH, de Ferranti SD, Aubert M, Chalmers TC, et al.** Large trials vs meta-analysis of smaller trials: how do their results compare? JAMA. 1996;276:1332-8. [PMID: 8861993]
28. **Concato J, Shah N, Horwitz RI.** Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med. 2000;342:1887-92. [PMID: 10861325]
29. **Ioannidis JP.** Contradicted and initially stronger effects in highly cited clinical research. JAMA. 2005;294:218-28. [PMID: 16014596]
30. **Ioannidis JP, Haidich AB, Pappa M, Pantazis N, Kokori SI, Tektonidou MG, et al.** Comparison of evidence of treatment effects in randomized and nonrandomized studies. JAMA. 2001;286:821-30. [PMID: 11497536]
31. **LeLorier J, Grégoire G, Benhaddad A, Lapierre J, Derderian F.** Discrepancies between meta-analyses and subsequent large randomized, controlled trials. N Engl J Med. 1997;337:536-42. [PMID: 9262498]
32. **Toma M, McAlister FA, Bialy L, Adams D, Vandermeer B, Armstrong PW.** Transition from meeting abstract to full-length journal article for randomized controlled trials. JAMA. 2006;295:1281-7. [PMID: 16537738]

Appendix Figure 1. **Screening of potential systematic reviews for inclusion in cohort.**



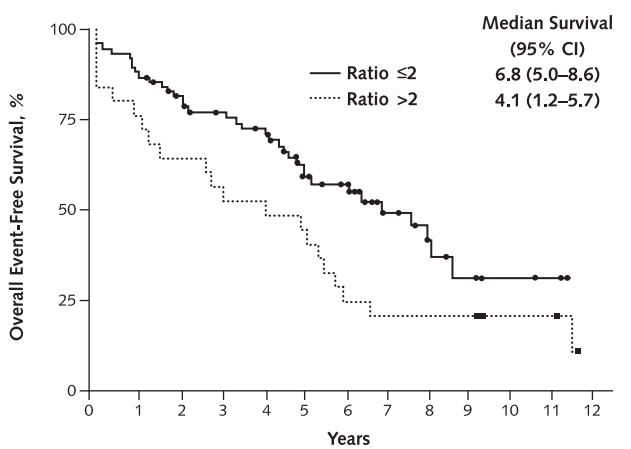
R = reason; RCT = randomized, controlled trial.

Appendix Figure 2. **Survival of the original systematic review by clinical topic area.**



Stratified by cardiovascular reviews (n = 20) versus reviews on all other topics (n = 80).

Appendix Figure 3. **Kaplan–Meier plot showing the effect on survival of increasing the total number of patients by more than a factor of 2.**



Ratio of new total sample size to old total sample size is >2, which occurred for 25% of systematic reviews in the cohort.