ELSEVIER

Journal of Clinical Epidemiology 62 (2009) 13–21

Journal of Clinical Epidemiology

# Bayesian statistical inference enhances the interpretation of contemporary randomized controlled trials

Duminda N. Wijeysundera[a,b,c,*], Peter C. Austin[a,b,d], Janet E. Hux[a,b,e], W. Scott Beattie[c], Andreas Laupacis[a,b,f,g]

[a]*Department of Health Policy Management and Evaluation, University of Toronto, Toronto, Ontario, Canada*
[b]*Institute for Clinical Evaluative Sciences, Toronto, Ontario, Canada*
[c]*Department of Anesthesia, Toronto General Hospital and University of Toronto, Toronto, Ontario, Canada*
[d]*Department of Public Health Sciences, University of Toronto, Toronto, Ontario, Canada*
[e]*Department of Medicine, Sunnybrook Health Sciences Centre and University of Toronto, Toronto, Ontario, Canada*
[f]*Keenan Research Centre, Li Ka Shing Knowledge Institute of St Michael's Hospital, Toronto, Ontario, Canada*
[g]*Department of Medicine, St. Michael's Hospital and University of Toronto, Toronto, Ontario, Canada*

Accepted 26 July 2008

**Abstract**

**Objective:** Randomized trials generally use ''frequentist'' statistics based on *P*-values and 95% confidence intervals. Frequentist methods have limitations that might be overcome, in part, by Bayesian inference. To illustrate these advantages, we re-analyzed randomized trials published in four general medical journals during 2004.

**Study Design and Setting:** We used Medline to identify randomized superiority trials with two parallel arms, individual-level randomization and dichotomous or time-to-event primary outcomes. Studies with $P < 0.05$ in favor of the intervention were deemed ''positive''; otherwise, they were ''negative.'' We used several prior distributions and exact conjugate analyses to calculate Bayesian posterior probabilities for clinically relevant effects.

**Results:** Of 88 included studies, 39 were positive using a frequentist analysis. Although the Bayesian posterior probabilities of any benefit (relative risk or hazard ratio $< 1$) were high in positive studies, these probabilities were lower and variable for larger benefits. The positive studies had only moderate probabilities for exceeding the effects that were assumed for calculating the sample size. By comparison, there were moderate probabilities of any benefit in negative studies.

**Conclusion:** Bayesian and frequentist analyses complement each other when interpreting the results of randomized trials. Future reports of randomized trials should include both. 

*Keywords:* Randomized controlled trial; Bayesian inference; Frequentist statistics; Probability; Systematic review; Evidence-based medicine

## 1. Introduction

The randomized controlled trial (RCT) is a major research tool in medicine, with approximately 248,000 RCTs listed in PubMed by March 2008. Consequently, consensus-based guidelines have emerged, with the goal to improve RCT reporting [1]. These guidelines have, in turn, improved the quality of the literature [2]. Nonetheless, an important aspect of RCT methodology remains largely unchanged, namely the methods for statistical inference. In this article, we describe the limitations of ''frequentist'' statistical inference that rely on *P*-values and confidence intervals (CIs) to test whether interventions are efficacious. We contrast these limitations with the advantages of Bayesian statistical inference for interpreting RCTs. We provide a concrete example by using Bayesian methods to re-analyze RCTs published recently in high-impact general medical journals.

### 1.1. Limitations of the P-value

Most RCT reports use ''frequentist'' statistical inference, where an intervention is deemed efficacious based on $P < 0.05$, or a 95% CI excluding a null effect [3–7].

* Corresponding author. Department of Anesthesia, Toronto General Hospital and University of Toronto, EN 3-450 200 Elizabeth Street, Toronto, Ontario M5G 2C4, Canada. Tel.: +416-340-4800 ext. 8981; fax: +416-340-3698.

*E-mail address*: d.wijeysundera@utoronto.ca (D.N. Wijeysundera).

0895-4356/09/$ – see front matter 
doi: 10.1016/j.jclinepi.2008.07.006

**What is new?**

Randomized controlled trials (RCTs) generally use frequentist statistics that rely on *P*-values and 95% confidence intervals.

To determine whether Bayesian inference may help overcome some limitations of frequentist methods, we re-analyzed 88 RCTs that were published in high-impact general medical journals in 2004.

For RCTs deemed "positive" by frequentist methods, the Bayesian posterior probabilities of any benefit (relative risk or hazard ratio $< 1$) were high but were lower and variable for larger benefits.

By comparison, there were still moderate probabilities of any benefit in "negative" RCTs.

Given that Bayesian and frequentist analyses provide complementary interpretations, future reports of RCTs should include both.

P-values and, to a lesser degree, CIs have important limitations. *P*-values do not represent what most readers think they do. For a "positive" study with $P = 0.05$, many readers conclude that there was a 5% chance of no treatment effect [8–10]. For a "negative" study with $P = 0.70$, they conclude that there was a 70% chance of no treatment effect. Neither interpretation is correct. What $P = 0.05$ means is that under the null hypothesis (typically, that the intervention has no effect), there was a 5% chance of observing results at least as extreme as seen in the study. *P*-values describe probabilities for data, based on the assumption that the null hypothesis was true. Consequently, they represent deductive inference, which begins with a hypothesis about the world, and tests whether observations are consistent with that hypothesis [5]. In contrast, clinical practice involves inductive inference, which begins with observations, and then determines which hypothesis most likely explains those observations [5,11]. Most clinicians, therefore, intuitively want to know probabilities for clinically relevant effects, based on the observed data; hence, they incorrectly interpret *P*-values using inductive inference [5,11]. Thus, *P*-values are not the most clinically relevant representation of RCT results [8].

*P*-values also have practical limitations. The definition of statistical significance as $P < 0.05$ is arbitrary [12,13]. Indeed, Fisher promoted flexibility when defining statistical significance: "no scientific worker has a fixed level of significance at which from year to year, he rejects hypotheses" [4]. Categorization of studies as "positive" or "negative" based on this arbitrary criterion ignores biological plausibility and previous evidence [14]. It leads to problems, especially when *P*-values fall close to either side of 0.05. Biologically implausible and spurious associations might be identified, especially when sample sizes are large or multiple nonprespecified hypotheses are tested [15]. Difficulties arise when interpreting studies that do not achieve statistical significance. For example, a $P = 0.15$ might be interpreted as a "trend that approaches statistical significance" or "no benefit" depending, in large part, on previous evidence and readers' pre-existing beliefs. These factors are not explicitly acknowledged in many research articles that use frequentist analyses [5,16].

In addition, the *P*-value does not convey the magnitude of an effect. The decision to use an intervention depends, in part, on whether its effect exceeded the minimum clinically important difference (MCID): the smallest treatment effect that would alter patient management [17]. *P*-values are limited because they are influenced by sample size: $P = 0.01$ might be consistent with a large treatment effect in a small study, or an unimportant effect in a large multicenter trial [18,19]. Indeed, with increasing sample size, just about any effect, regardless of how small it is, can reach statistical significance [20,21].

### 1.2. Confidence intervals: a suboptimal solution

The limitations of *P*-values have led to recommendations to instead emphasize CIs when presenting results [22,23]. The CI is an improvement. Instead of a simple "yes" or "no" answer, it communicates the magnitude and precision of the treatment effect [22]. It might also help determine when an effect is clinically important [24,25]. If the lower limit of the CI excludes the MCID, the effect is likely to be important. Conversely, if the upper limit excludes the MCID, the effect is unlikely to be important.

CIs, nonetheless, have important disadvantages. Because they are still based on frequentist inference, they suffer from theoretical problems [26]. Readers frequently conclude that there is a 95% probability that the true treatment effect lies with the 95% CI [27]. This interpretation is erroneous. The 95% refers to the fact that if the same study were repeated many times and the CI similarly calculated for each case, 95% of such intervals would include the true treatment effect [27]. Advocates of the CI do acknowledge this strict definition; however, many clinicians do not share this insight [22].

The CI also has practical limitations. First, many readers fail to consider the range of values within the interval [7,28]. Given that the conventional 95% CI implicitly uses the 5% cut-off entailed by $P < 0.05$, they simply classify a treatment effect as significant if its 95% CI excludes the null effect [5,7]. Second, the CI does not report the information that clinicians are interested in, namely the probabilities for clinically important benefits. Unless an intervention's 95% CI excludes its MCID, readers cannot easily determine whether it has clinically relevant benefits [25]. Clinical care involves making decisions with less than 95% confidence [12]. Clinicians might be interested in

knowing whether an intervention has an 80% probability of exceeding a meaningful effect. Third, biologically implausible and spurious associations might still be identified because CIs cannot explicitly incorporate external factors (e.g., biological plausibility) when interpreting a study finding.

### 1.3. Bayesian inference

Bayesian inference overcomes several limitations of frequentist statistics. First, it permits inductive inference by reporting the clinically relevant probabilities for specified treatment effects [9,14]. Second, it can determine probabilities for varying magnitudes of therapeutic response. Third, it can explicitly incorporate external information when interpreting the results of a study. Previous evidence, biological plausibility and pre-existing beliefs can all influence the interpretation of *P*-values. Instead of burying this subjectivity within the discussion of a manuscript, Bayesian inference quantitatively incorporates this external information when calculating the probability of a therapeutic response.

Bayesian inference involves specific components: the prior, likelihood, and posterior. The prior is the probability of hypothesized treatment effects, based on information independent of the study (e.g., previous evidence). The likelihood summarizes the data within the study, using a frequentist analysis. The posterior is the end-result: the probability of hypothesized effects, based on data from the study and prior external information. It is calculated using Bayes' theorem, which states that the posterior is directly proportional to the product of the likelihood and prior [29]. Bayes' theorem is not new to clinical practice, where it is used for interpreting diagnostic tests [30].

The prior has been a particularly controversial aspect of Bayesian inference [31,32]. Because the prior describes information independent of the study, Bayesian inference has been criticized as too subjective, in contrast to ''objective'' frequentist analyses [9,18,33]. The ''objectivity'' of frequentist analyses is, however, illusory. Interpretation of *P*-values and CIs are affected by factors that are not explicitly defined in frequentist analyses: external evidence, biological plausibility, and pre-existing beliefs. Readers are unlikely to ignore the biological implausibility of an association between astrological sign and fractures, despite $P = 0.01$ [15]. Similarly, readers' pre-existing opinions influence the interpretation of negative trials, as evidenced by responses to a recent trial of percutaneous coronary intervention [34,35].

Given the subjectivity of the prior, Bayesian analyses use several different priors [36–38]. First, most analyses include a noninformative prior, an essentially flat distribution that permits the posterior to be determined almost entirely by the study data [31,37,38]. A typical noninformative prior is a normal distribution with mean zero and a large SD [8]. Second, external evidence, potentially summarized using meta-analysis, might be used to construct a prior [20,37]. Third, opinions of content experts may be elicited to approximate a prior distribution [37,39]. Fourth, researchers might construct an enthusiastic prior, where the best estimate (median) corresponds to the anticipated treatment effect, with a small probability (e.g., 5%) of no benefit [38]. Finally, a skeptical prior might be constructed. The latter assumes that the best estimate (median) corresponds to no difference between the intervention and control arms, with a small probability (e.g., 5%) that the treatment exceeded an important effect (e.g., MCID) [21,38]. When the study data are sufficiently strong, differing priors have minimal influence on the calculated posteriors [40]. In contrast, if the study data are relatively weak, the posteriors will not agree; nonetheless, this lack of consensus is likely appropriate given the absence of sufficiently compelling data [40].

### 1.4. Example of Bayesian inference

The prior, likelihood, and posterior are represented by probability distributions. The prior and likelihood are first combined mathematically to produce the posterior [37,38]. The posterior is then used to determine probabilities of specified effects [37,38]. Historically, Bayesian inference was limited by difficulties in calculating the posterior [33]. In very straightforward cases, it could be determined through exact calculations, but was very difficult to calculate in all other situations [18]. Two developments have largely removed this obstacle. First, exact calculations have been adapted for straightforward analyses of dichotomous or time-to-event outcomes [37,38]. Second, computer-intensive Markov Chain Monte Carlo methods can now address more complex problems [18,41].

Consider the example of an RCT comparing paclitaxel-containing chemotherapy against conventional chemotherapy for relapsed ovarian cancer [42]. The researchers anticipated an HR of 0.71 for the primary outcome, all-cause mortality [42]. In the study, the intervention caused a statistically significant benefit (HR 0.81, 95% CI 0.69–0.97, $P = 0.02$). When expressed on the natural logarithm (log) of the HR scale, the observed effect was a normally distributed likelihood with mean −0.20 and SD 0.088 (Fig. 1). For this example, we considered a skeptical prior where the best estimate was no benefit, but there was a 5% probability of exceeding the projected effect (HR 0.71). The prior was expressed as a normal distribution with mean 0 and SD 0.21 on the log-HR scale (Fig. 1). When combined, using methods described in the Appendix, the normally distributed posterior had a mean −0.17 and SD 0.081 (log-HR scale; Fig. 1). This posterior could then be used to determine probabilities for exceeding specific effects, based on corresponding areas-under-the-curve. For example, the probability of no benefit (HR > 1) was very low: 1.9% (Fig. 2). In contrast, if a clinically meaningful

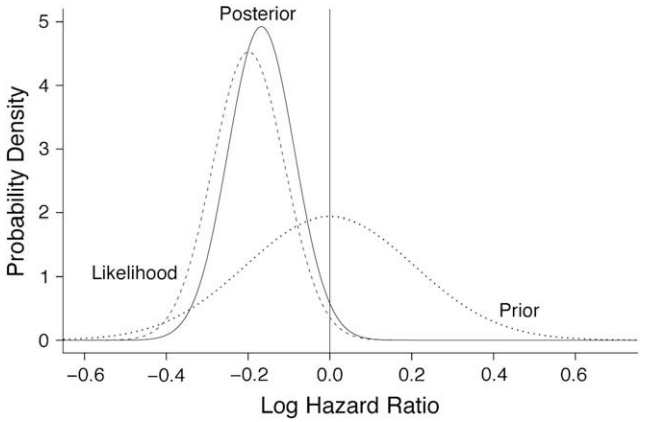


Fig. 1. Derivation of a posterior probability distribution using Bayesian inference. This time-to-event outcome is modeled as the natural logarithm (log) of the hazard ratio, fitted to a normal distribution. The posterior is the final probability of various treatment effects. It is determined by combining the estimated probabilities independent of the study (prior) with evidence from the study (likelihood). Each probability distribution is scaled such that area-under-the-curve is 1.

benefit was defined by HR $< 0.9$, the probability of exceeding a clinically meaningful benefit was 78% (Fig. 2).

In summary, Bayesian inference has important advantages for analyzing RCTs. To demonstrate these advantages, we used Bayesian methods to re-analyze recent RCTs published in high-impact general medical journals.

## 2. Methods

We used Medline to identify RCTs that were published (January 1, 2004 to December 31, 2004) in four general medical journals: New England Journal of Medicine, JAMA, Lancet, and Annals of Internal Medicine. Studies were restricted to those with two parallel arms, individual-level randomization, superiority design, human subjects, and dichotomous or time-to-event primary outcomes. We abstracted data on content area, intervention, control, primary outcome, estimated sample size, treatment effect assumed for sample size estimation, sample size recruited, observed effect size for primary outcome, and need for early termination. Where necessary, authors were contacted to obtain required information. A study was deemed "positive" if it reported $P < 0.05$ in favor of the intervention; otherwise, it was classified as "negative."

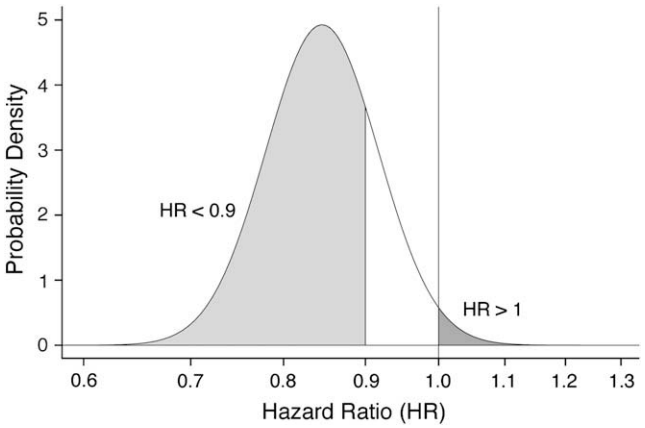


Fig. 2. Estimation of the probabilities for specified treatment effects using a Bayesian posterior distribution. The posterior distribution is scaled such that the area-under-the-curve is 1. Probabilities for specified treatment effects are determined using the area-under-the-curve. For example, the probability of hazard ratio $< 0.90$ is the area-under-the-curve to the left of 0.90. Similarly, the probability of no benefit, namely a hazard ratio $> 1$, is the area-under-the-curve to the right of 1.

### 2.1. Analyses

For each study, the likelihood was expressed as a normal distribution on the log-OR or log-HR scale [37,38,43]. We considered three normally distributed priors: noninformative, skeptical, and enthusiastic (Table 1). For each study, exact conjugate analyses were used to calculate three normally distributed posteriors [37,38]. These methods are described in the Appendix. The posteriors were then used to estimate probabilities for a range of treatment effects, from any benefit (RR or HR $< 1$) to a large effect (RR or HR $< 0.5$; Fig. 3). We also estimated probabilities for exceeding the effects that were used for sample size calculation. All analyses were performed using SAS Version 8.20 (SAS Institute, Cary, NC) and R 2.4.1 [44].

## 3. Results

We included 88 studies (Table 2; Appendix). Thirty-nine studies reported $P < 0.05$ in favor of the intervention. In these positive studies, the median treatment effects for dichotomous and time-to-event outcomes were RR 0.46 (interquartile range 0.22–0.72) and HR 0.62 (interquartile range 0.36–0.73), respectively. In the negative studies, the median effects were RR 0.99 (interquartile range 0.90–1.09) and HR 0.93 (interquartile range 0.84–1.03).

Table 1
Definitions of prior distributions, all of which follow a normal distribution

| Prior distribution | Mean | Boundaries |
|---|---|---|
| Uninformative | 0 | Standard deviation = 10 (log odds ratio or log hazard ratio scale) |
| Skeptical | 0 | 5% probability of exceeding the treatment effect that was assumed by the study investigators for the estimation of sample size |
| Enthusiastic | Treatment effect that was assumed by the study investigators for the estimation of sample size | 5% probability of no benefit (i.e., odds ratio $> 1$ or hazard ratio $> 1$) |

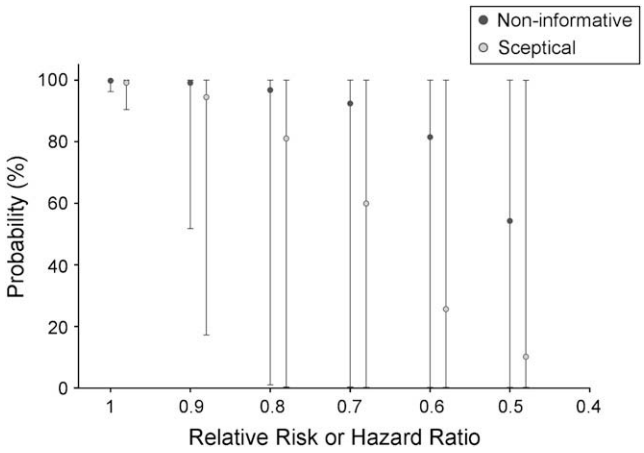


Fig. 3. Probability of exceeding specified risk reductions in positive studies, based on noninformative and skeptical priors. The treatment effects are expressed as relative risks or hazard ratios in these positive studies ($P < .05$ in favor of the intervention). The posterior estimates were calculated using both noninformative and skeptical priors. These treatment effects range from values < 1 (any benefit) to values < 0.5 (>50% reduction in risk). Circles represent median values and error bars denote 5th and 95th percentiles.

### 3.1. Positive studies

When using a noninformative prior, the probability of any benefit (RR or HR < 1) was very high (median 100%, 5th–95th percentiles: 96%–100%) in the positive studies. However, for effects of greater magnitude, this median probability decreased, whereas variation between studies increased (Fig. 3). In nine of the 39 studies, the probabilities of RR or HR < 0.8 were below 70%. The probabilities of RR or HR < 0.6 were below 70% in 17 studies.

Because a strongly positive study should convince a reasonable skeptic, we re-calculated posterior probabilities using a skeptical prior. The skeptical prior decreased median probabilities for specified treatment effects, and increased overall variability (Fig. 3).

### 3.2. Negative studies

Within the 49 negative studies, there was a moderate probability of any benefit (median 67%, 5th–95th percentiles: 2%–97%), based on a noninformative prior (Fig. 4). Fifteen studies were consistent with a greater than 80% probability of any benefit (RR or HR < 1).

In contrast to positive studies, a strongly negative study should convince a reasonable enthusiast; hence, we re-calculated posterior probabilities using an enthusiastic prior. The enthusiastic prior was associated with increased strength of evidence for the intervention (Fig. 4).

### 3.3. Agreement between frequentist results and Bayesian posterior probabilities

For positive studies, Bayesian and frequentist analyses had good agreement for the presence of any benefit (RR or HR < 1). However, in the case of larger effects, such as that assumed for sample size estimation, the analyses yielded differing results (Table 3). For negative studies, the different analyses tended to agree for the presence of larger effects. In contrast, Bayesian analyses suggested that, for many negative studies, there was still a reasonable posterior probability of any benefit (Table 3). These differences were generally increased when informative priors (skeptical or enthusiastic) were used.

## 4. Discussion

Researchers should use statistical methods that maximize the knowledge gained from a study. Our results

Table 2
Characteristics of included randomized controlled trials[a]

| Characteristic | Trials (N = 88) | Positive trials (n = 39) | Negative trials (n = 49) |
|---|---|---|---|
| *Journal* | | | |
| New England Journal of Medicine | 36 (41) | 20 (51) | 16 (33) |
| Journal of the American Medical Association | 16 (18) | 6 (15) | 10 (20) |
| The Lancet | 29 (33) | 10 (26) | 19 (39) |
| Annals of Internal Medicine | 7 (8) | 3 (8) | 4 (8) |
| *Area of investigation* | | | |
| Cardiovascular medicine | 22 (25) | 6 (15) | 16 (33) |
| Oncology | 19 (22) | 13 (33) | 6 (12) |
| Infectious disease | 17 (19) | 10 (26) | 7 (14) |
| Critical care medicine | 6 (7) | 0 (0) | 6 (12) |
| Other | 24 (27) | 10 (26) | 14 (29) |
| *Study characteristics* | | | |
| Participants recruited, median (interquartile range) | 505 (288–1550) | 500 (206–1113) | 518 (333–2159) |
| Outcome type | | | |
| Binary outcome | 43 (49) | 19 (49) | 24 (49) |
| Time-to-event | 45 (51) | 20 (51) | 25 (51) |
| Early termination | 19 (22) | 8 (21) | 10 (20) |
| Statistical significance in favor of intervention ($P < 0.05$) | 39 (44) | 39 (100) | 0 (0) |

[a] Values are expressed as numbers (percentage) unless otherwise indicated.

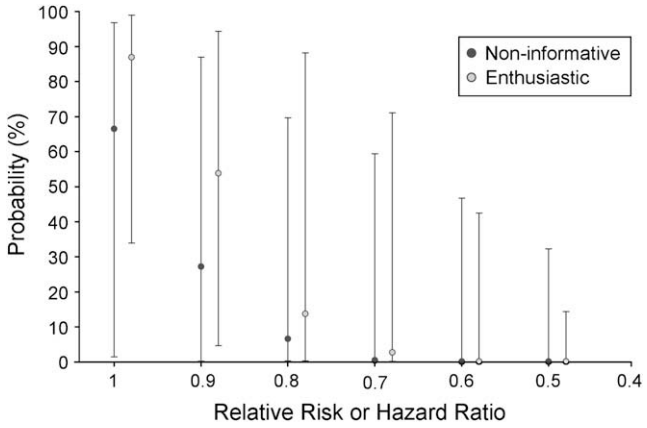


Fig. 4. Probability of exceeding specified risk reductions in negative studies, based on noninformative and enthusiastic priors. The treatment effects are expressed as relative risks or hazard ratios. The posterior estimates were calculated using both noninformative and enthusiastic priors. These treatment effects range from values $< 1$ (any benefit) to values $< 0.5$ ($> 50\%$ reduction in risk). Circles represent median values and error bars denote 5th and 95th percentiles.

suggest that Bayesian inference improves the interpretation of RCTs. We propose, therefore, that RCT reports should include Bayesian posterior probabilities for a range of clinically relevant effects. These posterior probabilities should be based on several explicitly defined priors, including noninformative, skeptical and enthusiastic distributions.

Our recommended approach necessitates strategies for including Bayesian inference in RCT reports. A strong case can certainly be made for replacing frequentist analyses with Bayesian posterior probabilities. Nonetheless, most readers have minimal exposure to Bayesian methods, whereas they are at least familiar with *P*-values and CIs. Thus, a rapid shift toward exclusively Bayesian reporting is not prudent. The transition from frequentist to Bayesian reporting must allow sufficient time to familiarize readers. In the interim, a reasonable compromise would involve Bayesian inference complementing frequentist methods. Spiegelhalter et al. have suggested that an ''Interpretation'' section supplement the ''Results'' section of research reports [45]. The Results section would summarize observed results, using P-values and CIs. The Interpretation section would use several sensible priors to report Bayesian posterior probabilities for clinically relevant effects. Thus, whereas the Results section would report the study data, the Interpretation section would quantitatively place the ''data into context by taking into account other sources of evidence, meta-analyses, the opinions of skeptical observers, and so on'' [45]. This formal integration of results with pre-existing knowledge stands in contrast to many current RCT reports [16].

A widespread use of Bayesian methods also entails consensus-based standards [21,46], which should encompass probability distributions for expressing outcomes, types of priors, and computational methods. Prior experience suggests that such standards are feasible and effective [1,2].

Our recommended approach offers important advantages. Bayesian inference is clinically relevant. Clinicians are interested in probabilities that interventions are beneficial, or Bayesian posterior probabilities. These probabilities can be calculated for a range of clinically relevant effects, thereby informing the decision to implement a new intervention. Consider two examples that demonstrate how posterior probabilities might inform clinical decision making. In the first example, a new drug has a statistically significant benefit, a 70% posterior probability of exceeding a meaningful effect, and important side effects. Given this posterior probability, some clinicians might not use the new drug, despite its ''statistically significant'' benefit. In contrast, if this drug is instead inexpensive and safe, clinicians might still use it, despite having less than 95% confidence that it causes a meaningful effect. Posterior probabilities can similarly inform decision making after a ''negative'' trial. In the second example, a new intervention has an HR of 0.70 (95% CI 0.41−1.20; $P = 0.2$). Using a noninformative prior, the posterior probabilities for this intervention causing HR $< 1$, HR $< 0.9$, and HR $< 0.7$ are 90%, 82%, and 50%, respectively. In specific contexts, such as a serious disease with few therapies, these posterior probabilities might convince clinicians to use the intervention, despite a nonsignificant *P*-value [47].

In addition, the subjectivity of Bayesian prior, although considered by some to be a limitation, might actually help better interpret RCTs [9]. If there are sufficient data, varying priors should only minimally influence the posterior probabilities [40]. Thus, major qualitative differences between noninformative, enthusiastic and skeptical posterior probabilities are a sensitivity analysis to assess the strength of evidence from a study [32,48]. After a ''positive'' study, major differences between noninformative and skeptical posteriors might suggest that confirmatory studies are warranted [49]; in addition, they may explain the incomplete uptake of evidence-based therapies into clinical practice [50,51]. After a ''negative'' trial, major differences between noninformative and enthusiastic posteriors might suggest when further study of promising interventions is warranted, and explain the differing responses to negative studies [34,52,53].

The flexibility of Bayesian inference also helps it address some limitations of the MCID. The MCID can vary with perspective (patients, clinicians, payers, society) [54], underlying methodology [54], and individual clinicians [55]. In addition, RCTs are rarely designed to detect the MCID because the entailed sample size might be unfeasible [24]. Bayesian inference is flexible; hence, it reports posterior probabilities for varying definitions of the MCID, regardless of the sample size assumptions of the individual RCT.

Bayesian inference also remains applicable to readers distrustful of subjective priors. These readers can focus on noninformative posterior probabilities. Because these probabilities are essentially determined by the study results

Table 3
Agreement between frequentist results and Bayesian posterior probabilities[a]

| Treatment effect | Bayesian posterior probability of specified treatment effect (%) | Positive trials[b] (n = 39) | Negative trials[b] (n = 49) |
|---|---|---|---|
| Any benefit (relative risk or hazard ratio < 1) | Noninformative prior | | |
| | >95 | 39 (100) | 5 (10) |
| | >90 | 39 (100) | 9 (18) |
| | >50 | 39 (100) | 31 (63) |
| | Skeptical prior | | |
| | >95 | 34 (87) | 0 (0) |
| | >90 | 38 (97) | 7 (14) |
| | >50 | 39 (97) | 31 (63) |
| | Enthusiastic prior | | |
| | >95 | 37 (94) | 15 (31) |
| | >90 | 37 (94) | 21 (43) |
| | >50 | 38 (97) | 44 (90) |
| Effect projected for sample size calculation | Noninformative prior | | |
| | >95 | 9 (23) | 0 (0) |
| | >90 | 11 (28) | 0 (0) |
| | >50 | 21 (54) | 1 (2) |
| | Skeptical prior | | |
| | >95 | 4 (10) | 0 (0) |
| | >90 | 4 (10) | 0 (0) |
| | >50 | 9 (23) | 0 (0) |
| | Enthusiastic prior | | |
| | >95 | 6 (15) | 0 (0) |
| | >90 | 7 (18) | 0 (0) |
| | >50 | 20 (51) | 1 (2) |

[a] Values are expressed as numbers (percentage) unless otherwise indicated.

[b] These values represent the proportion of positive or negative trials (based on a frequentist analysis) that had exceeded various Bayesian posterior probabilities for specific benefits (either any benefit or the effect assumed when determining the sample size). For example, under a noninformative prior, 100% of positive trials had a Bayesian posterior probability > 95% for achieving any benefit, whereas 10% of negative trials had a posterior probability > 95% for achieving any benefit.

alone, they usually result in similar point estimates of treatment effect as frequentist analyses [27]. However, unlike *P*-values and CIs, noninformative posteriors still facilitate inductive inference and report probabilities for a range of clinically relevant treatment effects. Hence, they retain the important Bayesian advantage of shifting the emphasis from statistical to clinical significance [56].

Finally, although we have focused on the role of Bayesian methods for interpreting RCTs, they have other important roles in medical research. Specifically, they can help guide interim analyses in clinical trials [31,39,57], incorporate adaptive randomization methods [31,57], assess for synergy between drugs [31,57], analyze noninferiority trials [58], and guide regulatory approval of drugs or medical devices [31,59,60].

### 4.1. Limitations

Several limitations should be considered when interpreting our findings. First, frequentist inference still dominates the medical literature. Potential explanations include its seeming ''objectivity'' [61], the availability of readily usable computer software to perform frequentist analyses [31], and physicians' generally low level of overall statistical knowledge [62]. Nonetheless, the ''objectivity'' of frequentist statistics is illusory, and readily available software can now perform complex Bayesian analyses problems [41]. The emphasis, therefore, should be to improve readers' understanding of Bayesian inference.

Second, some included studies might still have carefully considered the clinical and statistical implications of frequentist results, to reach similar conclusions as a Bayesian analysis. Nonetheless, Bayesian posterior probabilities would have only enhanced the communication of clinical relevance. In addition, most RCT reports do not systematically discuss results within the context of similar research [16]. In contrast, a Bayesian analysis would make this integration explicit and quantitative. Finally, our re-analysis used only three priors to calculate posterior probabilities. We sought to make these priors clinically sensible within the context of each individual study; specifically, the boundaries of the skeptical and enthusiastic priors were defined by the projected treatment effect. Nonetheless, any future Bayesian analysis of an individual RCT would likely use a wider range of priors than our re-analysis.

## 5. Conclusion

Bayesian inference reports probabilities that are theoretically consistent with the probabilities that clinicians are interested in, can be calculated for a range of clinically

relevant effects, and can be adjusted for differing pre-existing beliefs. It offers advantages to clinicians and researchers: Bayesian posterior probabilities might better inform clinical decision making, and more explicitly interpret study results. Bayesian inference should, therefore, complement existing frequentist methods, and thereby improve RCT reporting.

## Acknowledgment

*Financial support*: Dr Wijeysundera is supported by a Clinician-Scientist Award and Dr Austin by a New Investigator Award from the Canadian Institutes of Health Research. Dr Beattie is the Fraser Elliot Chair of Cardiovascular Anesthesiology at the University Health Network.

## References

[1] Moher D, Schulz KF, Altman D, for the CONSORT Group. The CONSORT statement: revised recommendations for improving the quality of reports for parallel-group randomized trials. JAMA 2001;285:1987–91.
[2] Kane RL, Wang J, Garrard J. Reporting in randomized clinical trials improved after adoption of the CONSORT statement. J Clin Epidemiol 2007;60:241–9.
[3] Altman DG. Why we need confidence intervals. World J Surg 2005;29:554–6.
[4] Fisher RA. Statistical methods and statistical inference. Edinburgh, UK: Oliver & Boyd; 1959.
[5] Goodman SN. Towards evidence-based medical statistics. 1: the P value fallacy. Ann Intern Med 1999;130:995–1004.
[6] Neyman J, Pearson ES. On the use of and the interpretation of certain test criteria of purpose of statistical inference: parts I and II. Biometrika 1928;20:175–240. 263–94.
[7] Sterne JAC, Davey Smith G. Sifting the evidence-what's wrong with significance tests? BMJ 2001;322:226–31.
[8] Diamond GA, Forrester JS. Clinical trials and statistical verdicts: probable grounds for appeal. Ann Intern Med 1983;98:385–94.
[9] Freedman L. Bayesian statistical methods. BMJ 1996;313:569–70.
[10] Wulff HR, Andersen B, Brandenhoff P, Guttler F. What do doctors know about statistics? Stat Med 1987;6:3–10.
[11] Davidoff F. Standing statistics right side up. Ann Intern Med 1999;130:1019–21.
[12] Salsburg DS. The religion of statistics as practiced in medical journals. Am Stat 1985;39:220–3.
[13] Upshur REG. The ethics of alpha: reflections on statistics, evidence and values in medicine. Theor Med 2001;22:565–76.
[14] Browner WS, Newman TB. Are all significant P values created equal? The analogy between diagnostic tests and clinical research. JAMA 1987;257:2459–63.
[15] Austin PC, Mamdani MM, Juurlink DN, Hux JE. Testing multiple statistical hypotheses resulted in spurious associations: a study if astrological signs and health. J Clin Epidemiol 2006;59:964–9.
[16] Clarke M, Alderson P, Chalmers I. Discussion sections in reports of controlled trials published in general medical journals. JAMA 2002;287:2799–801.
[17] Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimum clinically important difference. Control Clin Trials 1989;10:407–15.
[18] Austin PC, Brunner LJ, Hux JE. Bayeswatch: an overview of Bayesian statistics. J Eval Clin Pract 2002;8:277–86.
[19] Lang JM, Rothman KJ, Cann CI. That confounded P-value. Epidemiology 1998;9:7–8.
[20] Brophy JM, Joseph L. Placing trials in context using Bayesian analysis. GUSTO revisited by Reverend Bayes. JAMA 1995;273:871–5.
[21] Diamond GA, Kaul S. Prior convictions: Bayesian approaches to the analysis and interpretation of clinical megatrials. J Am Coll Cardiol 2004;43:1929–39.
[22] Gardner MJ, Altman DG. Confidence intervals rather than P values: estimation rather than hypothesis testing. BMJ 1986;292:746–50.
[23] Annals of Internal Medicine. Information for authors. Available at: http://www.annals.org/shared/author_info_stats.html. Accessed 20 March, 2007.
[24] Chan KB, Man-Son-Hing M, Molnar FJ, Laupacis A. How well is the clinical importance of study results reported? An assessment of randomized controlled trials. CMAJ 2001;165:1197–202.
[25] Man-Son-Hing M, Laupacis A, O'Rourke K, Molnar FJ, Mahon J, Chan KB, et al. Determination of the clinical importance of study results. J Gen Intern Med 2002;17:469–76.
[26] Feinstein AR. P-values and confidence intervals: two sides of the same unsatisfactory coin. J Clin Epidemiol 1998;51:355–60.
[27] Burton PR. Helping doctors to draw appropriate inferences from the analysis of medical studies. Stat Med 1994;13:1699–713.
[28] Altman D, Bland JM. Confidence intervals illuminate absence of evidence. BMJ 2004;328:1016–7.
[29] Barnard GA, Bayes T. Studies in the history of probability and statistics: IX. Thomas Bayes' essay towards solving a problem in the doctrine of chances. Biometrika 1958;45:293–315.
[30] Louis TA. Introduction to Bayesian methods II: fundamental concepts. Clin Trials 2005;2:291–4. discussion 301–4, 364–78.
[31] Berry DA. Bayesian clinical trials. Nature Rev Drug Discov 2006;5:27–36.
[32] Greenland S. Bayesian perspectives for epidemiological research: I. Foundations and basic methods. Int J Epidemiol 2006;35:765–75.
[33] Efron B. Why isn't everyone a Bayesian? Am Stat 1986;40:1–5.
[34] Boden WE, O'Rourke RA, Teo KK, Hartigan PM, Maron DJ, Kostuk WJ, et al. Optimal medical therapy with or without PCI for stable coronary disease. N Engl J Med 2007;356:1503–16.
[35] Hochman JS, Steg PG. Does preventive PCI work? N Engl J Med 2007;356:1572–4.
[36] Goodman SN. Toward evidence-based medical statistics. 2: the Bayes factor. Ann Intern Med 1999;130:1005–13.
[37] Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian approaches to clinical trials and health-care evaluation. Chichester, UK: John Wiley & Sons; 2004.
[38] Spiegelhalter DJ, Freedman LS, Parmar MKB. Bayesian approaches to randomized trials. J R Statist Soc A 1994;157:357–416.
[39] Parmar MK, Griffiths GO, Spiegelhalter DJ, Souhami RL, Altman DG, van der Scheuren E, et al. Monitoring of large randomised clinical trials: a new approach with Bayesian methods. Lancet 2001;358:375–81.
[40] Christensen R. Testing Fisher, Neyman, Pearson, and Bayes. Am Stat 2005;59:121–6.
[41] MRC Biostatistics Unit. The BUGS Project: Bayesian inference using Gibbs sampling. Available at: http://www.mrc-bsu.cam.ac.uk/-bugs. Accessed 26 March, 2007.
[42] Parmar MK, Ledermann JA, Colombo N, du Bois A, Delaloye JF, Kristensen GB, et al. Paclitaxel plus platinum-based chemotherapy versus conventional platinum-based chemotherapy in women with relapsed ovarian cancer: the ICON4/AGO-OVAR-2.2 trial. Lancet 2003;361:2099–106.
[43] Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. Stat Med 1998;17:2815–34.
[44] R Development Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available at: www.R-project.org. Accessed 1 July, 2007.

[45] Spiegelhalter DJ, Freedman LS, Parmar MKB. Applying Bayesian ideas in drug development and clinical trials. Stat Med 1993;12: 1501−11.
[46] Sung L, Hayden J, Greenberg ML, Koren G, Feldman BM, Tomlinson GA. Seven items were identified for inclusion when reporting a Bayesian analysis of a clinical study. J Clin Epidemiol 2005;58:261−8.
[47] Johnson SR, Feldman BM, Pope JE, Tomlinson GA. Shifting our thinking about uncommon disease trials: the case of methotrexate in scleroderma. J Rheumatol 2008;in press.
[48] Lilford RJ, Braunholtz D. The statistical basis of public policy: a paradigm shift is overdue. BMJ 1996;313:603−7.
[49] Parmar MK, Ungerleider RS, Simon R. Assessing whether to perform a confirmatory randomized clinical trial. J Natl Cancer Inst 1996;88: 1645−51.
[50] Lamas GA, Pfeffer MA, Hamm P, Wertheimer J, Rouleau JL, Braunwald E. Do the results of randomized clinical trials of cardiovascular drugs influence medical practice? The SAVE Investigators. N Engl J Med 1992;327:241−7.
[51] Majumdar SR, Inui TS, Gurwitz JH, Gillman MW, McLaughlin TJ, Soumerai SB. Influence of physician specialty on adoption and relinquishment of calcium channel blockers and other treatments for myocardial infarction. J Gen Intern Med 2001;16:351−9.
[52] Wood S. Medical therapy takes COURAGE: no benefit of PCI over optimal drugs for preventing events in stable CAD. Available at: http://www.theheart.org/article/779535.do. Accessed 29 March, 2007.
[53] Stafford RS, Furberg CD, Finkelstein SN, Cockburn IM, Alehegn T, Ma J. Impact of clinical trial results on national trends in alpha-blocker prescribing, 1996−2002. JAMA 2004;291:54−62.
[54] Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. Curr Opin Rheumatol 2002;14:109−14.
[55] Kirwan JR. Minimum clinically important difference: the crock of gold at the end of the rainbow? J Rheumatol 2001;28:439−44.
[56] Burton PR, Gurrin LC, Campbell MJ. Clinical significance not statistical significance: a simple Bayesian alternative to p values. J Epidemiol Community Health 1998;52:318−23.
[57] Berry DA. Introduction to Bayesian methods III: use and interpretation of Bayesian tools in design and analysis. Clin Trials 2005;2: 295−300. discussion 301−4, 364−78.
[58] Kaul S, Diamond GA, Weintraub WS. Trials and tribulations of non--inferiority: the ximelagatran experience. J Am Coll Cardiol 2005;46: 1986−95.
[59] Berry DA. Bayesian statistics. Med Decis Making 2006;26:429−30.
[60] Campbell G. The experience in the FDA's Center for Devices and Radiological Health with Bayesian strategies. Clin Trials 2005;2: 359−63. discussion 364−78.
[61] Whitehead J. The case for frequentism in clinical trials. Stat Med 1993;12:1405−13. discussion 1415−9.
[62] Windish DM, Huot SJ, Green ML. Medicine residents' understanding of the biostatistics and results in the medical literature. JAMA 2007;298:1010−22.

## Appendix

Spiegelhalter et al. [37] provide detailed descriptions of the methods outlined.

### *Expressing a dichotomous outcome as a normally distributed likelihood*

Consider a hypothetical randomized controlled trial with the following outcomes for the two arms

| | Study arm | |
|---|---|---|
| Event | Intervention | Control |
| Outcome | a | b |
| No outcome | c | d |

In this case, the likelihood is expressed as the natural logarithm (log) of the odds ratio (OR) and follows a normal distribution. The mean ($\theta$) and standard deviation ($s$) are calculated as follows:

$$\theta = \log\left[\frac{\left(a+\frac{1}{2}\right)\left(d+\frac{1}{2}\right)}{\left(b+\frac{1}{2}\right)\left(c+\frac{1}{2}\right)}\right]$$

$$s = \sqrt{\frac{1}{a+\frac{1}{2}} + \frac{1}{b+\frac{1}{2}} + \frac{1}{c+\frac{1}{2}} + \frac{1}{d+\frac{1}{2}}}$$

### *Expressing a time-to-event outcome as a normally distributed likelihood*

Consider a hypothetical randomized controlled trial (RCT) where the treatment effect was a hazard ratio (HR) with lower (LCI) and upper 95% confidence interval (UCI).

In this case, the likelihood is expressed as log HR, and follows a normal distribution. The mean ($\theta$) and standard deviation ($s$) for this distribution are calculated as follows:

$$\theta = \log\left(\mathrm{HR}\right)$$

$$s = \frac{\log\left(\mathrm{UCI}\right) - \log\left(\mathrm{HR}\right)}{1.96}$$

### *Calculating a posterior from a prior and likelihood*

Assume that the prior is normally distributed with mean ($\theta_{\mathrm{prior}}$) and standard deviation ($s_{\mathrm{prior}}$), whereas the likelihood is normally distributed with mean ($\theta_L$) and standard deviation ($s_{\mathrm{L}}$). The mean ($\theta_L$) and standard deviation ($s_{\mathrm{post}}$) for the normally distributed posterior is then calculated as follows:

$$\theta_{post} = \frac{\left[\frac{\theta_{\mathrm{prior}}}{s^2_{\mathrm{prior}}} + \frac{\theta_L}{s^2_L}\right]}{\left[\frac{1}{s^2_{\mathrm{prior}}} + \frac{1}{s^2_L}\right]}$$

$$s_{\mathrm{post}} = \sqrt{\frac{1}{\frac{1}{s^2_{\mathrm{prior}}} + \frac{1}{s^2_L}}}$$

### *List of included studies*

The search results are presented in the Fig. A1. The included trials are listed below:


[1] Alexandrov AV, Molina CA, Grotta JC, Garami Z, Ford SR, Alvarez-Sabin J, et al. Ultrasound-enhanced systemic thrombolysis for acute ischemic stroke. N Engl J Med 2004; 351:2170–8.

[2] Alloueche A, Bailey W, Barton S, Bwika J, Chimpeni P, Falade CO, et al. Comparison of chlorproguanil-dapsone with sulfadoxine-pyrimethamine for the treatment of uncomplicated falciparum malaria in young African children: double-blind randomised controlled trial. Lancet 2004; 363:1843–8.

[3] Anand KJ, Hall RW, Desai N, Shephard B, Bergqvist LL, Young TE, et al. Effects of morphine analgesia in ventilated preterm neonates: primary outcomes from the NEOPAIN randomised trial. Lancet 2004; 363:1673–82.

[4] Anderson GL, Limacher M, Assaf AR, Bassford T, Beresford SA, Black H, et al. Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: the Women's Health Initiative randomized controlled trial. JAMA 2004; 291:1701–12.

[5] Andre T, Boni C, Mounedji-Boudiaf L, Navarro M, Tabernero J, Hickish T, et al. Oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment for colon cancer. N Engl J Med 2004; 350:2343–51.

[6] Andrews DW, Scott CB, Sperduto PW, Flanders AE, Gaspar LE, Schell MC, et al. Whole brain radiation therapy with or without stereotactic radiosurgery boost for patients with one to three brain metastases: phase III results of the RTOG 9508 randomised trial. Lancet 2004; 363:1665–72.

[7] Ardissino D, Cavallini C, Bramucci E, Indolfi C, Marzocchi A, Manari A, et al. Sirolimus-eluting vs uncoated stents for prevention of restenosis in small coronary arteries: a randomized trial. JAMA 2004; 292:2727–34.

[8] Arriagada R, Bergman B, Dunant A, Le Chevalier T, Pignon JP, Vansteenkiste J. Cisplatin-based adjuvant chemotherapy in patients with completely resected non-small-cell lung cancer. N Engl J Med 2004; 350:351–60.

[9] Barwell JR, Davies CE, Deacon J, Harvey K, Minor J, Sassano A, et al. Comparison of surgery and compression with compression alone in chronic venous ulceration (ESCHAR study): randomised controlled trial. Lancet 2004; 363:1854−9.

[10] Bergman GJ, Winters JC, Groenier KH, Pool JJ, Meyboom-de Jong B, Postema K, et al. Manipulative therapy in addition to usual medical care for patients with shoulder dysfunction and pain: a randomized, controlled trial. Ann Intern Med 2004; 141:432−9.

[11] Bernier J, Domenge C, Ozsahin M, Matuszewska K, Lefebvre JL, Greiner RH, et al. Postoperative irradiation with or without concomitant chemotherapy for locally advanced head and neck cancer. N Engl J Med 2004; 350:1945−52.

[12] Bjornson CL, Klassen TP, Williamson J, Brant R, Mitton C, Plint A, et al. A randomized trial of a single dose of oral dexamethasone for mild croup. N Engl J Med 2004; 351:1306−13.

[13] Braunwald E, Domanski MJ, Fowler SE, Geller NL, Gersh BJ, Hsia J, et al. Angiotensin-converting-enzyme inhibition in stable coronary artery disease. N Engl J Med 2004; 351:2058−68.

[14] Brower RG, Lanken PN, MacIntyre N, Matthay MA, Morris A, Ancukiewicz M, et al. Higher versus lower positive end-expiratory pressures in patients with the acute respiratory distress syndrome. N Engl J Med 2004; 351:327−36.

[15] Caporali R, Cimmino MA, Ferraccioli G, Gerli R, Klersy C, Salvarani C, et al. Prednisone plus methotrexate for polymyalgia rheumatica: a randomized, double-blind, placebo-controlled trial. Ann Intern Med 2004; 141:493−500.

[16] Carrat F, Bani-Sadr F, Pol S, Rosenthal E, Lunel-Fabiani F, Benzekri A, et al. Pegylated interferon alfa-2b vs standard interferon alfa-2b, plus ribavirin, for chronic hepatitis C in HIV-infected patients: a randomized controlled trial. JAMA 2004; 292: 2839−48.

[17] Chintu C, Bhat GJ, Walker AS, Mulenga V, Sinyinza F, Lishimpi K, et al. Co-trimoxazole as prophylaxis against opportunistic infections in HIV-infected Zambian children (CHAP): a double-blind randomised placebo-controlled trial. Lancet 2004; 364: 1865−71.

[18] Chung RT, Andersen J, Volberding P, Robbins GK, Liu T, Sherman KE, et al. Peginterferon Alfa-2a plus ribavirin versus interferon alfa-2a plus ribavirin for chronic hepatitis C in HIV-coinfected persons. N Engl J Med 2004; 351:451−9.

[19] Colhoun HM, Betteridge DJ, Durrington PN, Hitman GA, Neil HA, Livingstone SJ, et al. Primary prevention of cardiovascular disease with atorvastatin in type 2 diabetes in the Collaborative Atorvastatin Diabetes Study (CARDS): multicentre randomised placebo-controlled trial. Lancet 2004; 364:685−96.

[20] Coombes RC, Hall E, Gibson LJ, Paridaens R, Jassem J, Delozier T, et al. A randomized trial of exemestane after two to three years of tamoxifen therapy in postmenopausal women with primary breast cancer. N Engl J Med 2004; 350:1081−92.

[21] Cooper JS, Pajak TF, Forastiere AA, Jacobs J, Campbell BH, Saxman SB, et al. Postoperative concurrent radiotherapy and chemotherapy for high-risk squamous-cell carcinoma of the head and neck. N Engl J Med 2004; 350:1937−44.

[22] Corey L, Wald A, Patel R, Sacks SL, Tyring SK, Warren T, et al. Once-daily valacyclovir to reduce the risk of transmission of genital herpes. N Engl J Med 2004; 350:11−20.

[23] Cunningham D, Humblet Y, Siena S, Khayat D, Bleiberg H, Santoro A, et al. Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. N Engl J Med 2004; 351:337−45.

[24] D'Amico AV, Manola J, Loffredo M, Renshaw AA, DellaCroce A, Kantoff PW. 6-month androgen suppression plus radiation therapy vs radiation therapy alone for patients with clinically localized prostate cancer: a randomized controlled trial. JAMA 2004; 292:821−7.

[25] de Lemos JA, Blazing MA, Wiviott SD, Lewis EF, Fox KA, White HD, et al. Early intensive vs a delayed conservative simvastatin strategy in patients with acute coronary syndromes: phase Z of the A to Z trial. JAMA 2004; 292:1307−16.

[26] DeBusk RF, Miller NH, Parker KM, Bandura A, Kraemer HC, Cher DJ, et al. Care management for low-risk patients with heart failure: a randomized, controlled trial. Ann Intern Med 2004; 141:606−13.

[27] Diener HC, Bogousslavsky J, Brass LM, Cimminiello C, Csiba L, Kaste M, et al. Aspirin and clopidogrel compared with clopidogrel alone after recent ischaemic stroke or transient ischaemic attack in high-risk patients (MATCH): randomised, double-blind, placebo-controlled trial. Lancet 2004; 364:331−7.

[28] Donta ST, Engel CCJ, Collins JF, Baseman JB, Dever LL, Taylor T, et al. Benefits and harms of doxycycline treatment for Gulf War veterans' illnesses: a randomized, double-blind, placebo-controlled trial. Ann Intern Med 2004; 141:85−94.

[29] Esteban A, Frutos-Vivar F, Ferguson ND, Arabi Y, Apezteguia C, Gonzalez M, et al. Noninvasive positive-pressure ventilation for respiratory failure after extubation. N Engl J Med 2004; 350:2452−60.

[30] Ferguson JJ, Califf RM, Antman EM, Cohen M, Grines CL, Goodman S, et al. Enoxaparin vs unfractionated heparin in high-risk patients with non-ST-segment elevation acute coronary syndromes managed with an intended early invasive strategy: primary results of the SYNERGY randomized trial. JAMA 2004; 292:45−54.

[31] Fernandez-Aviles F, Alonso JJ, Castro-Beiras A, Vazquez N, Blanco J, Alonso-Briales J, et al. Routine invasive strategy within 24 hours of thrombolysis versus ischaemia-guided conservative approach for acute myocardial infarction with ST-segment elevation (GRACIA-1): a randomised controlled trial. Lancet 2004; 364:1045−53.

[32] Finfer S, Bellomo R, Boyce N, French J, Myburgh J, Norton R. A comparison of albumin and saline for fluid resuscitation in the intensive care unit. N Engl J Med 2004; 350:2247−56.

[33] Gale EA, Bingley PJ, Emmett CL, Collier T. European Nicotinamide Diabetes Intervention Trial (ENDIT): a randomised controlled trial of intervention before the onset of type 1 diabetes. Lancet 2004; 363:925−31.

[34] Gournay V, Roze JC, Kuster A, Daoud P, Cambonie G, Hascoet JM, et al. Prophylactic ibuprofen versus placebo in very premature infants: a randomised, double-blind, placebo-controlled trial. Lancet 2004; 364:1939−44.

[35] Guerin C, Gaillard S, Lemasson S, Ayzac L, Girard R, Beuret P, et al. Effects of systematic prone positioning in hypoxemic acute respiratory failure: a randomized controlled trial. JAMA 2004; 292:2379−87.

[36] Harper DM, Franco EL, Wheeler C, Ferris DG, Jenkins D, Schuind A, et al. Efficacy of a bivalent L1 virus-like particle vaccine in prevention of infection with human papillomavirus types 16 and 18 in young women: a randomised controlled trial. Lancet 2004; 364:1757−65.

[37] Harrison TW, Oborne J, Newton S, Tattersfield AE. Doubling the dose of inhaled corticosteroid to prevent asthma exacerbations: randomised controlled trial. Lancet 2004; 363:271−5.

[38] Hohnloser SH, Kuck KH, Dorian P, Roberts RS, Hampton JR, Hatala R, et al. Prophylactic use of an implantable cardioverter-defibrillator after acute myocardial infarction. N Engl J Med 2004; 351:2481−8.

[39] Hommes OR, Sorensen PS, Fazekas F, Enriquez MM, Koelmel HW, Fernandez O, et al. Intravenous immunoglobulin in secondary progressive multiple sclerosis: randomised placebo-controlled trial. Lancet 2004; 364:1149−56.

[40] Homs MY, Steyerberg EW, Eijkenboom WM, Tilanus HW, Stalpers LJ, Bartelsman JF, et al. Single-dose brachytherapy versus metal stent placement for the palliation of dysphagia from oesophageal cancer: multicentre randomised trial. Lancet 2004; 364:1497−504.

[41] Huddleston JM, Long KH, Naessens JM, Vanness D, Larson D, Trousdale R, et al. Medical and surgical comanagement after elective hip and knee arthroplasty: a randomized, controlled trial. Ann Intern Med 2004; 141:28−38.

[42] Hughes KS, Schnaper LA, Berry D, Cirrincione C, McCormick B, Shank B, et al. Lumpectomy plus tamoxifen with or without irradiation in women 70 years of age or older with early breast cancer. N Engl J Med 2004; 351:971−7.

[43] Jocham D, Richter A, Hoffmann L, Iwig K, Fahlenkamp D, Zakrzewski G, et al. Adjuvant autologous renal tumour cell vaccine and risk of tumour progression in patients with renal-cell carcinoma after radical nephrectomy: phase III, randomised controlled trial. Lancet 2004; 363:594−9.

[44] Julius S, Kjeldsen SE, Weber M, Brunner HR, Ekman S, Hansson L, et al. Outcomes in hypertensive patients at high cardiovascular risk treated with regimens based on valsartan or amlodipine: the VALUE randomised trial. Lancet 2004; 363:2022−31.

[45] Kadish A, Dyer A, Daubert JP, Quigg R, Estes NA, Anderson KP, et al. Prophylactic defibrillator implantation in patients with nonischemic dilated cardiomyopathy. N Engl J Med 2004; 350:2151−8.

[46] Kastrati A, Mehilli J, Schuhlen H, Dirschinger J, Dotzer F, ten Berg JM, et al. A clinical trial of abciximab in elective percutaneous coronary intervention after pretreatment with clopidogrel. N Engl J Med 2004; 350:232−8.

[47] Kato H, Ichinose Y, Ohta M, Hata E, Tsubota N, Tada H, et al. A randomized trial of adjuvant chemotherapy with uracil-tegafur for adenocarcinoma of the lung. N Engl J Med 2004; 350:1713−21.

[48] Kaul R, Kimani J, Nagelkerke NJ, Fonck K, Ngugi EN, Keli F, et al. Monthly antibiotic chemoprophylaxis and incidence of sexually transmitted infections and HIV-1 infection in Kenyan sex workers: a randomized controlled trial. JAMA 2004; 291:2555−62.

[49] Kelly D, Jara P, Rodeck B, Lykavieris P, Burdelski M, Becker M, et al. Tacrolimus and steroids versus ciclosporin microemulsion, steroids, and azathioprine in children undergoing liver transplantation: randomised European multicentre trial. Lancet 2004; 364: 1054−61.

[50] Koblin B, Chesney M, Coates T. Effects of a behavioural intervention to reduce acquisition of HIV infection among men who have sex with men: the EXPLORE randomised controlled study. Lancet 2004; 364:41−50.

[51] Landolfi R, Marchioli R, Kutti J, Gisslinger H, Tognoni G, Patrono C, et al. Efficacy and safety of low-dose aspirin in polycythemia vera. N Engl J Med 2004; 350:114−24.

[52] Leung KL, Kwok SP, Lam SC, Lee JF, Yiu RY, Ng SS, et al. Laparoscopic resection of rectosigmoid carcinoma: prospective randomised trial. Lancet 2004; 363:1187−92.

[53] Liaw YF, Sung JJ, Chow WC, Farrell G, Lee CZ, Yuen H, et al. Lamivudine for patients with chronic

hepatitis B and advanced liver disease. N Engl J Med 2004; 351:1521–31.

[54] Lux AL, Edwards SW, Hancock E, Johnson AL, Kennedy CR, Newton RW, et al. The United Kingdom Infantile Spasms Study comparing vigabatrin with prednisolone or tetracosactide at 14 days: a multicentre, randomised controlled trial. Lancet 2004; 364:1773–8.

[55] McFalls EO, Ward HB, Moritz TE, Goldman S, Krupski WC, Littooy F, et al. Coronary-artery revascularization before elective major vascular surgery. N Engl J Med 2004; 351:2795–804.

[56] Merten GJ, Burgess WP, Gray LV, Holleman JH, Roush TS, Kowalchuk GJ, et al. Prevention of contrast-induced nephropathy with sodium bicarbonate: a randomized controlled trial. JAMA 2004; 291:2328–34.

[57] Milpied N, Deconinck E, Gaillard F, Delwail V, Foussard C, Berthou C, et al. Initial treatment of aggressive lymphoma with high-dose chemotherapy and autologous stem-cell support. N Engl J Med 2004; 350:1287–95.

[58] Myles PS, Leslie K, McNeil J, Forbes A, Chan MT. Bispectral index monitoring to prevent awareness during anaesthesia: the B-Aware randomised controlled trial. Lancet 2004; 363:1757–63.

[59] Neumayer L, Giobbie-Hurder A, Jonasson O, Fitzgibbons RJ, Dunlop D, Gibbs J, et al. Open mesh versus laparoscopic mesh repair of inguinal hernia. N Engl J Med 2004; 350:1819–27.

[60] Pappone C, Manguso F, Santinelli R, Vicedomini G, Sala S, Paglino G, et al. Radiofrequency ablation in children with asymptomatic Wolff-Parkinson-White syndrome. N Engl J Med 2004; 351:1197–205.

[61] Perondi MB, Reis AG, Paiva EF, Nadkarni VM, Berg RA. A comparison of high-dose and standard-dose epinephrine in children with cardiac arrest. N Engl J Med 2004; 350:1722–30.

[62] Poole-Wilson PA, Lubsen J, Kirwan BA, van Dalen FJ, Wagener G, Danchin N, et al. Effect of long-acting nifedipine on mortality and cardiovascular morbidity in patients with stable angina requiring treatment (ACTION trial): randomised controlled trial. Lancet 2004; 364:849–57.

[63] Prandoni P, Lensing AW, Prins MH, Frulla M, Marchiori A, Bernardi E, et al. Below-knee elastic compression stockings to prevent the post-thrombotic syndrome: a randomized, controlled trial. Ann Intern Med 2004; 141:249–56.

[64] Prinssen M, Verhoeven EL, Buth J, Cuypers PW, van Sambeek MR, Balm R, et al. A randomized trial comparing conventional and endovascular repair of abdominal aortic aneurysms. N Engl J Med 2004; 351:1607–18.

[65] Pryor KO, Fahey T, Jr, Lien CA, Goldstein PA. Surgical site infection and the routine use of perioperative hyperoxia in a general surgical population: a randomized controlled trial. JAMA 2004; 291:79–87.

[66] Raghu G, Brown KK, Bradford WZ, Starko K, Noble PW, Schwartz DA, et al. A placebo-controlled trial of interferon gamma-1b in patients with idiopathic pulmonary fibrosis. N Engl J Med 2004; 350:125–33.

[67] Remuzzi G, Lesti M, Gotti E, Ganeva M, Dimitrov BD, Ene-Iordache B, et al. Mycophenolate mofetil versus azathioprine for prevention of acute rejection in renal transplantation (MYSS): a randomised trial. Lancet 2004; 364:503–12.

[68] Roberts I, Yates D, Sandercock P, Farrell B, Wasserberg J, Lomas G, et al. Effect of intravenous corticosteroids on death within 14 days in 10008 adults with clinically significant head injury (MRC CRASH trial): randomised placebo-controlled trial. Lancet 2004; 364:1321–8.

[69] Rose PG, Nerenstone S, Brady MF, Clarke-Pearson D, Olt G, Rubin SC, et al. Secondary surgical cytoreduction for advanced ovarian carcinoma. N Engl J Med 2004; 351:2489–97.

[70] Saag MS, Cahn P, Raffi F, Wolff M, Pearce D, Molina JM, et al. Efficacy and safety of emtricitabine vs stavudine in combination therapy in antiretroviral-naive patients: a randomized trial. JAMA 2004; 292:180–9.

[71] Sauer R, Becker H, Hohenberger W, Rodel C, Wittekind C, Fietkau R, et al. Preoperative versus postoperative chemoradiotherapy for rectal cancer. N Engl J Med 2004; 351:1731–40.

[72] Senat MV, Deprest J, Boulvain M, Paupe A, Winer N, Ville Y. Endoscopic laser surgery versus serial amnioreduction for severe twin-to-twin transfusion syndrome. N Engl J Med 2004; 351:136–44.

[73] Silverstein M, Mack C, Reavis N, Koepsell TD, Gross GS, Grossman DC. Effect of a clinic-based referral system to head start: a randomized controlled trial. JAMA 2004; 292:968–71.

[74] Sobel JD, Wiesenfeld HC, Martens M, Danna P, Hooton TM, Rompalo A, et al. Maintenance fluconazole therapy for recurrent vulvovaginal candidiasis. N Engl J Med 2004; 351:876–83.

[75] Stone GW, Ellis SG, Cox DA, Hermiller J, O'Shaughnessy C, Mann JT, et al. A polymer-based, paclitaxel-eluting stent in patients with coronary artery disease. N Engl J Med 2004; 350:221–31.

[76] Taha TE, Kumwenda NI, Hoover DR, Fiscus SA, Kafulafula G, Nkhoma C, et al. Nevirapine and zidovudine at birth to reduce perinatal transmission of HIV in an African setting: a randomized controlled trial. JAMA 2004; 292:202–9.

[77] Tannock IF, de Wit R, Berry WR, Horti J, Pluzanska A, Chi KN, et al. Docetaxel plus prednisone or mitoxantrone plus prednisone for advanced prostate cancer. N Engl J Med 2004; 351:1502–12.

[78] Thwaites GE, Nguyen DB, Nguyen HD, Hoang TQ, Do TT, Nguyen TC, et al. Dexamethasone for the

treatment of tuberculous meningitis in adolescents and adults. N Engl J Med 2004; 351:1741–51.
[79] To MS, Alfirevic Z, Heath VC, Cicero S, Cacho AM, Williamson PR, et al. Cervical cerclage for prevention of preterm delivery in women with short cervix: randomised controlled trial. Lancet 2004; 363: 1849–53.
[80] Toole JF, Malinow MR, Chambless LE, Spence JD, Pettigrew LC, Howard VJ, et al. Lowering homocysteine in patients with ischemic stroke to prevent recurrent stroke, myocardial infarction, and death: the Vitamin Intervention for Stroke Prevention (VISP) randomized controlled trial. JAMA 2004; 291:565–75.
[81] Tran TH, Dolecek C, Pham PM, Nguyen TD, Nguyen TT, Le HT, et al. Dihydroartemisinin-piperaquine against multidrug-resistant Plasmodium falciparum malaria in Vietnam: randomised clinical trial. Lancet 2004; 363:18–22.
[82] van Koningsveld R, Schmitz PI, Meche FG, Visser LH, Meulstee J, van Doorn PA. Effect of methylprednisolone when added to standard treatment with intravenous immunoglobulin for Guillain-Barre syndrome: randomised trial. Lancet 2004; 363:192–6.
[83] Van Overmeire B, Allegaert K, Casaer A, Debauche C, Decaluwe W, Jespers A, et al. Prophylactic ibuprofen in premature infants: a multicentre, randomised, double-blind, placebo-controlled trial. Lancet 2004; 364: 1945–9.
[84] Verrier ED, Shernan SK, Taylor KM, Van de Werf F, Newman MF, Chen JC, et al. Terminal complement blockade with pexelizumab during coronary artery bypass graft surgery requiring cardiopulmonary bypass: a randomized trial. JAMA 2004; 291:2319–27.
[85] Verweij J, Casali PG, Zalcberg J, LeCesne A, Reichardt P, Blay JY, et al. Progression-free survival in gastrointestinal stromal tumours with high-dose imatinib: randomised trial. Lancet 2004; 364:1127–34.
[86] Wenzel V, Krismer AC, Arntz HR, Sitter H, Stadlbauer KH, Lindner KH. A comparison of vasopressin and epinephrine for out-of-hospital cardiopulmonary resuscitation. N Engl J Med 2004; 350:105–13.
[87] Wertheim HF, Vos MC, Ott A, Voss A, Kluytmans JA, Vandenbroucke-Grauls CM, et al. Mupirocin prophylaxis against nosocomial *Staphylococcus aureus* infections in nonsurgical patients: a randomized study. Ann Intern Med 2004; 140:419–25.
[88] Wong BC, Lam SK, Wong WM, Chen JS, Zheng TT, Feng RE, et al. Helicobacter pylori eradication to prevent gastric cancer in a high-risk region of China: a randomized controlled trial. JAMA 2004; 291:187–94.

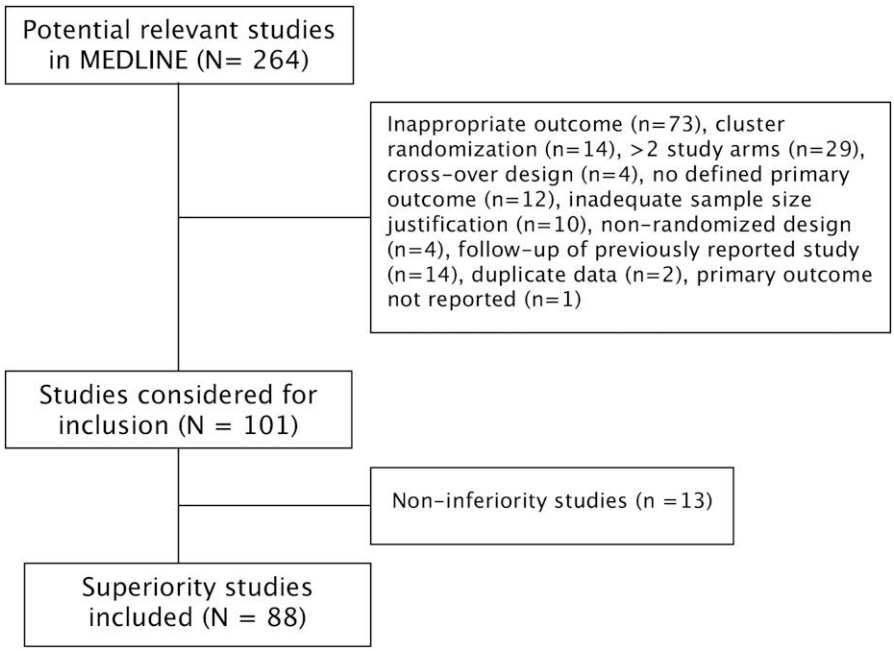


Fig. A1. Literature search for selected studies.