Contemporary Clinical Trials 47 (2016) 259–265

Contents lists available at ScienceDirect

Contemporary Clinical Trials

journal homepage: www.elsevier.com/locate/conclintrial

Likelihood ratio meta-analysis: New motivation and approach for an old method

Colin R. Dormuth [a,*], Kristian B. Filion [b,c], Robert W. Platt [d,e]

[a] Department of Anesthesiology, Pharmacology & Therapeutics, University of British Columbia, Vancouver, British Columbia, Canada
[b] Center for Clinical Epidemiology, Lady Davis Research Institute, Jewish General Hospital, Montreal, Quebec, Canada
[c] Department of Medicine, McGill University, Montreal, Quebec, Canada
[d] Departments of Pediatrics and of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Quebec, Canada
[e] Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada

ARTICLE INFO

Article history:
Received 18 August 2015
Received in revised form 20 January 2016
Accepted 23 January 2016
Available online 4 February 2016

Keywords:
Meta-analysis
Likelihood ratio
Meta-analytic methods

ABSTRACT

A 95% confidence interval (CI) in an updated meta-analysis may not have the expected 95% coverage. If a meta-analysis is simply updated with additional data, then the resulting 95% CI will be wrong because it will not have accounted for the fact that the earlier meta-analysis failed or succeeded to exclude the null. This situation can be avoided by using the likelihood ratio (LR) as a measure of evidence that does not depend on type-1 error. We show how an LR-based approach, first advanced by Goodman, can be used in a meta-analysis to pool data from separate studies to quantitatively assess where the total evidence points. The method works by estimating the log-likelihood ratio (LogLR) function from each study. Those functions are then summed to obtain a combined function, which is then used to retrieve the total effect estimate, and a corresponding 'intrinsic' confidence interval. Using as illustrations the CAPRIE trial of clopidogrel versus aspirin in the prevention of ischemic events, and our own meta-analysis of higher potency statins and the risk of acute kidney injury, we show that the LR-based method yields the same point estimate as the traditional analysis, but with an intrinsic confidence interval that is appropriately wider than the traditional 95% CI. The LR-based method can be used to conduct both fixed effect and random effects meta-analyses, it can be applied to old and new meta-analyses alike, and results can be presented in a format that is familiar to a meta-analytic audience.

1. Introduction

The measures of evidence typically provided in a meta-analysis have historically been the 95% confidence interval (CI) and, occasionally, the *p* value. Despite important and in some cases recent advances in meta-analytic methods, such as the use of random-effects modeling and other methods to characterize the heterogeneity of effects across a series of studies [1–6],CIs and *p* values remain problematic for a number of reasons that have been documented and debated by others. [7–11]. A fulsome discussion of the shortcomings of those statistics is beyond the scope of this article, but for the sake of establishing the need for an alternative method of meta-analysis, there are at least two characteristics of *p* and CI worth emphasizing. First, *p* and CI, whether from a fixed effect or random effects meta-analysis, become invalid when an existing meta-analysis is updated with more data. Interpretation of *p* and CI are contingent on type-1 error (α) and sample size values which are supposed to be specified before the meta-analysis. If the null hypothesis is true and a meta-analysis is updated, then the updated 95% CI will be wrong because it will not account for the probability of erroneously excluding the null given the result from the first meta-analysis. Our disregard for the first failure to reject the null means that the total probability of erroneous rejection for the new CI will be well above 0.05, and it will not have the expected 95% coverage probability. The problem of using a CI as a significance test has been emphasized previously [10].

The second characteristic of *p* and CI which bears particular relevance to a meta-analysis is that we choose alpha (usually $\alpha = 0.05/2$) to include the tail densities of the statistical distribution beyond 1.96. Inclusion of the tail densities means that *p* and CI encompass not only the probability that an observed result could have occurred, but also the probability that more extreme results could have occurred. Consider a 95% CI where one of the CI limit values is exactly equal to the null hypothesis. A natural tendency would be to interpret the null hypothesis as being 5% or 1/20th as likely as the observed point estimate (the maximum likelihood estimate). However, the 95% CI values are actually only 1/7th as likely as the point estimate, not 1/20th as likely. The disconnect between our 1/20th perception of relative likelihood and a 1/7th actual relative likelihood is the statistical weight we give to more extreme, unobserved data. Regarding a meta-

* Corresponding author at: Dept. of Anesthesiology, Pharmacology & Therapeutics (APT), UBC Victoria Office, Suite 210, 1110 Government Street, Victoria, BC V8W 1Y2, Canada.
E-mail address: colin.dormuth@ti.ubc.ca (C.R. Dormuth).

http://dx.doi.org/10.1016/j.cct.2016.01.008
1551-7144/

analytic result as statistically significant when the null hypothesis need only be less than 1/7th as likely as the point estimate runs contrary to the notion that meta-analytic results should be very strong because they can affect the treatment of many people [12].

Random effects meta-analysis, by accounting for heterogeneity of effects across a series of studies, produces more conservative CIs than a fixed effect meta-analysis, yet the aforementioned characteristics of *p* and CI lead to an overstatement of the case against the null hypothesis in both fixed and random effects analyses, with the overstatement being most acute for *p* values and CIs that are moderately statistically significant. Such situations are common in meta-analyses, which are sometimes done to achieve greater certainty because the variances from individual studies are sufficiently large so as to prevent precise estimation. Although there are several statistical approaches to estimating a 95% CI, the two limitations described above are applicable regardless of the approach used. CIs and *p* values that are too narrow (small) may also have adverse implications for the power calculations and heterogeneity tests used in meta-analyses. The purpose of this article is to reintroduce and modify an alternative method of using likelihood ratios (LR) to express confidence in meta-analytic results. The core of the method was first advanced by Goodman a quarter century ago [13], and we modify it here for easy, familiar-looking application in lieu of a 95% CI. We discuss how likelihood ratio meta-analysis (LRMA) can overcome the key limitations of the *p* value and the CI. We apply LRMA to a meta-analysis of observational studies that was previously done using CIs. Finally, we discuss the application of LRMA to random effects meta-analysis, Bayesian meta-analysis, and the valuation of future studies.

2. The likelihood function and likelihood ratio

In statistics, a likelihood can be thought of as a mirror image of a conditional probability. Whereas $P(y|\mu_0)$ is the probability of the observed result (y) conditional on the null hypothesis (μ_0) being true, the likelihood $L(\mu|y)$ is the likelihood of a hypothesis μ given the observed result y. An entire likelihood function can be plotted by calculating likelihoods for all possible values of μ, that is, for any and all hypotheses that are not impossible given y. Likelihoods only quantify the information contained in the observed data, and the same data always produce the same likelihood when using the same probability model. This latter feature is useful for meta-analysis because new studies can be added to an existing meta-analysis without altering our interpretation of the likelihood. Likelihoods are not subject to error rate logic, and they are unaltered by our previous research intentions or any long-run assumptions about future unobserved results.

The likelihood function for a normal distribution is denoted

$$L(\mu,\sigma|y) = L = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{y_i-\mu}{\sigma}\right)^2\right\} \quad (1)$$

where μ is the mean and σ is the standard deviation, and where L is jointly sufficient on those two parameters. When L is calculated for two specific hypotheses, such as the null hypothesis (μ_0) and a specific alternative hypothesis (μ_1), then the ratio of the two likelihoods can be expressed as

$$LR = \left[\prod_{i=1}^{n} \frac{1}{\sigma_1\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{y_i-\mu_1}{\sigma_1}\right)^2\right\}\right] \Big/ \left[\prod_{i=1}^{n} \frac{1}{\sigma_0\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{y_i-\mu_0}{\sigma_0}\right)^2\right\}\right], \quad (2)$$

which is interpreted as the relative likelihood of hypothesis μ_0 compared to μ_1, given data y. Eq. (2) can be evaluated for any pair of possible values of μ. Eq. (2) can also be rearranged and simplified to the following form by using logarithms and canceling terms:

$$LR = \exp\left[\frac{(y_i-\mu_1)^2}{2\sigma_1^2} - \frac{(y_i-\mu_0)^2}{2\sigma_0^2} + \ln\left(\frac{\sigma_1}{\sigma_0}\right)\right]. \quad (3)$$

Eq. (3) can be used to evaluate common epidemiologic measures of association, such as the risk difference, or the natural logarithms of multiplicative measures such as the relative risk, rate ratio, or odds ratio. For example, an LR comparing relative risks under the null hypothesis ($\mu_0 = \mathrm{RR_{H0}} = 1.00$) versus some alternative hypothesis ($\mu_1 = \mathrm{RR_{HA}}$) can be expressed using a normal distribution as follows:

$$LR = \exp\left[\frac{(\ln \mathrm{RR_{MLE}} - \ln \mathrm{RR_{HA}})^2}{2\sigma_{\mathrm{HA}}^2} - \frac{(\ln \mathrm{RR_{MLE}} - 0)^2}{2\sigma_{\mathrm{H_0}}^2} + \ln\left(\frac{\sigma_{\mathrm{HA}}}{\sigma_{\mathrm{H_0}}}\right)\right]. \quad (4)$$

The third term in Eq. (4) is equal to zero when $\sigma_{H0} = \sigma_{HA} = \sigma$ for all values of μ, as is typically assumed under a normal distribution. In Eq. (4), the σ^2 terms correspond to the variances of the relative risk, on the natural logarithm scale, under the null and alternative hypotheses. The importance of the third term in Eq. (4) is that relative risk is a function of the cumulative incidences in the exposed and the unexposed arms of a study, and the variance of a binomial variable is $np(1-p)$, which is not constant but varies according to p. We suggest that pragmatic approaches for using Eq. (4) are either to estimate a single intermediate pooled variance in place of σ_{H0} and σ_{HA}, using the null and alternative hypotheses, or to assume that $\sigma_{H0} = \sigma_{HA} = \sigma$, as is done when using the binomial approximation to a normal distribution. The third term in Eq. (4) equals 0 with both options.

2.1. Example 1: CAPRIE Trial

The randomized controlled trial (RCT) of Clopidogrel versus Aspirin in Patients at Risk of Ischaemic Events (CAPRIE) was designed to assess the relative efficacy of clopidogrel (75 mg once daily) to that of aspirin (325 mg once daily) in reducing the risk of a composite primary outcome of ischemic stroke, myocardial infarction (MI), or vascular death [14]. The study population included patients with a history of atherosclerotic vascular disease manifested as either recent ischemic stroke, recent MI, or symptomatic peripheral arterial disease. A total of 9599 clopidogrel and 9586 aspirin patients were followed for 1 to 3 years (mean follow-up = 1.91 years). The CAPRIE Steering Committee reported 939 primary outcome events in clopidogrel patients who were followed for a total of 17,636 person-years and 1021 events in aspirin patients followed for a total of 17,519 person-years (rate ratio 0.914; 95% CI 0.835 to 0.999; p = 0.046).

LRs and CIs for the CAPRIE data are shown in Table 1. Rates such as those in the CAPRIE trial are assumed to follow a Poisson distribution, but for ease of computation, and owing to an ample number of events, we assume that a normal distribution can be used to approximate a Poisson distribution in this example. Further, the Delta Method (Appendix I) has been used to show that the natural logarithm of the rate ratio can also be approximated by a normal distribution, with a variance equal to the inverse of the number of cases in the treatment arm,

Table 1
Likelihood ratios, confidence intervals and intrinsic confidence intervals based on the randomized, blinded, trial of clopidogrel versus aspirin in patients at risk of ischaemic events (CAPRIE) [14].

	Clopidogrel				Aspirin
Patients	9599				9586
Person-years	17,636				17,519
Cases	939				1021
	Ratio	ln(ratio)	ln(LR[a])	LR	1/LR%
Rate Ratio (MLE vs null)	0.914	−0.090	1.998	7.374	13.561
95% CI (lower limit vs null)	0.835	−0.180	0.038	1.039	96.271
95% CI (upper limit vs null)	0.990	−0.010	0.038	1.039	96.271
95% intrinsic CI (lower limit vs null)	0.818	−0.201	−0.992	0.371	269.662
95% intrinsic CI (upper limit vs null)	1.020	0.020	−0.992	0.371	269.662

[a] Compared to null hypothesis; Var(ln(RR)) = 1/939 + 1/1021 = 0.002.

plus the inverse of the number of cases in the control arm ($\mathrm{Var}(\ln(RR)) = 1/939 + 1/1021 = 0.002$). This was the method used to obtain the 95% CI values in Table 1. The 95% CI limits are thus the antilog values $\ln(RR) \pm 1.96(\mathrm{Var}(\ln(RR))^{1/2}$.

The MLE of the rate ratio is 0.914, and the MLE is 7.4 times more likely than the null. The lnLR values for the 95% CI limits are 1.96 units less than the lnLR value for the MLE (0.038 versus to 1.998). Since the upper limit of the 95% CI is close to 1.00, the LR for each confidence limit compared to the null are just above unity (LR = 1.039). Therefore, the CI limit rate ratios of 0.835 and 0.999 are 96.3% as likely under the observed data as the null hypothesis. An LR equal to 1 would have corresponded to the CI limits and the null being exactly as likely. The 95% CI limits are also $1/e^{(1.998\ -\ 0.038)} = 1/7.1 * 100 = 14.1\%$ as likely as the MLE rate ratio of 0.914.

3. The intrinsic confidence interval

An interesting property of the natural logarithm of the LR is that the 95% CI limits are close to the natural logarithm of the likelihood ratio (lnLR) for the maximum likelihood estimate (MLE), minus 1.96 units. The natural antilog of 1.96 is 7.1, which is why the hypotheses at each end of the 95% CI are $1/7.1 * 100 = 14\%$ as likely as the MLE under the observed data, and not 5% as likely as implied in the 95% moniker of the interval. However, a different 95% interval can be specified such that the hypotheses at each of the confidence limits are 5% as likely as the MLE, that is, the likelihood ratio is 20. Such an interval has been called an 'intrinsic' 95% confidence interval (ICI) in order to distinguish it from its 95% CI counterpart [13]. The 95% ICI limits are located at −2.99 units below the MLE on the lnLR function. On the LR function, the 95% ICI limits are $e^{-2.99} = 0.05 = 1/20$th as large as likely as the MLE. Thus, 95% ICIs are wider than 95% CIs, which translates to a lower attribution of certainty and a more conservative inference from the data.

The ICI is described as 'intrinsic' because it is compatible with common perception of a confidence interval. Most readers of medical literature mistakenly think that a p value of 0.05 – one of the 95% CI limit values is equal to the null hypothesis – means that the null hypothesis has a probability of only 5% [15–19]. The corresponding 95% ICI has this interpretation in a relative sense. For example, a risk difference of −0.032 (95% ICI: −0.064 to 0.000) from a hypothetical study would mean that a risk difference of 0 was 5% as likely as a risk difference of −0.032, given the data. Returning to Example 1, the ICI limits in the CAPRIE trial are 0.818 to 1.020. All possible rate ratios inside the ICI are more likely given the data than all rate ratios outside the ICI. Further, the 95% ICI now includes the null hypothesis, and it follows from Table 1 that a rate ratio of 1.00 is more likely (=2.7 times) under the observed CAPRIE data than the rate ratios at the 95% ICI limits. The 95% ICI limits are also 1/20th as likely as the MLE rate ratio of 0.914.

4. Meta-analysis with likelihood ratios

Estimation of the total effect in a meta-analysis usually involves calculating a weighted average of the effects observed in individual studies, where the weight given to each study is typically a function of its variance or sample size. Under the assumption that each study is providing some evidence for the same underlying hypothesis, LRs have an useful property where, for each possible hypothesis, the total meta-analytic effect can be obtained by multiplying together the LRs from each study at each specific value of μ. Equivalently, the lnLRs for each study can be summed at each possible value of μ [13].

4.1. Example 2: statin potency and acute kidney injury

A meta-analysis of observational studies was conducted to assess the association of higher potency statins on the risk of acute kidney injury (AKI) compared to lower potency statins [20]. This meta-analysis, in which two of us participated (CRD and RWP), was conducted by the Canadian Network for Observational Drug Effect Studies (CNODES), which used a common analytical protocol for each of the 9 included studies. The meta-analysis included approximately 2 million patients who initiated statin therapy, with 24,418 patients hospitalized for AKI during the first 2 years of follow-up. In as-treated analyses using a nested case–control approach, current users of high potency statins were 34% more likely than current users of low potency statins to be hospitalized for AKI within 120 days of starting treatment (rate ratio 1.34). The 95% CI for the total rate ratio was 1.25 to 1.43.

A re-analysis of the original meta-analysis using LRs is shown in Fig. 1 with the 95% CIs overlaid on the new ICIs. The 95% CI limits in Fig. 1 were obtained directly from the lnLR function, at minus 1.96 units below the MLE. The SAS program used for this analysis is available on request. The 95% ICI for each study and the total meta-analytic effect were calculated using the same logic as in the CAPRIE example. Point estimates are the same with both types of analysis (because both use maximum likelihood estimation); however, the 95% ICI limit values (the outermost bars) are wider than the 95% CIs (innermost bars). The reason is that the ICI values denote rate ratios which are 5% as likely as the point estimates, given the data, whereas the rate ratios at the 95% CI limits are 14.1% as likely.

Fig. 2 is a graphical depiction of the multiplications behind the statin–AKI LRMA analysis in Fig. 1. The figure was purposely simplified to include only 2 of the original 9 studies (Manitoba and Saskatchewan), but the concept is readily extendable to n studies. Eq. (4) was used to calculate LRs for each province. Each point on the LR curve for the total effect in Fig. 1 is equal to the LR values of the two individual studies multiplied together at the same hypothesis on the horizontal axis. For example, the total LR at a rate ratio of 1.20 is 4.255, which is equal to the corresponding LR = 2.524 from Saskatchewan multiplied by LR = 1.685 from Manitoba. The MLE of the total effect of 1.28 lies between the MLEs for the two studies (1.21 and 1.38), conveying the fact that the total estimate is a weighted average of the parts. All three LR curves are equal to 1 at a rate ratio of 1.00 because every value on each curve is the relative likelihood of its corresponding rate ratio compared to the null hypothesis rate ratio of 1.00. Also, the LR function for the total effect is less dispersed than the individual LR functions, showing graphically how precision is increased by combining the data.

While Fig. 2 provides an alternative format to that of Fig. 1, it may be more practical to plot lnLR functions instead of LR functions in most instances. An advantage of a format like Fig. 2, or a natural logarithm version of it, is that it enables all possible hypotheses and their relative statistical support to be viewed in a single graph. The disadvantages of the format in Fig. 2 are that the graph can quickly become crowded and unwieldy when there are even a handful of studies. The format is also not nearly as recognizable as the common forest plot. Presenting results in the format of Fig. 1 creates a resemblance to a forest plot, which in turn should make for a simpler transition away from 95% CIs. Another advantage of Fig. 1 is that the 95% ICI is compatible with the interpretation that we are tempted to mistakenly give the 95% CI, namely that hypotheses outside the interval are less than 5% as likely as the observed point estimate.

5. Random effects

The LRMA method so far presented is a fixed effect method that assumes each study is estimating the same common effect. A fixed effect approach for Example 2 (statin potency and AKI) can be defended on the grounds that similar populations of patients were used and that each study was conducted according to a common analytical protocol. A random effects analysis may be more suitable when the meta-analysis is thought to include studies that are measuring different effects. In that situation, a random effects analysis can be done by applying

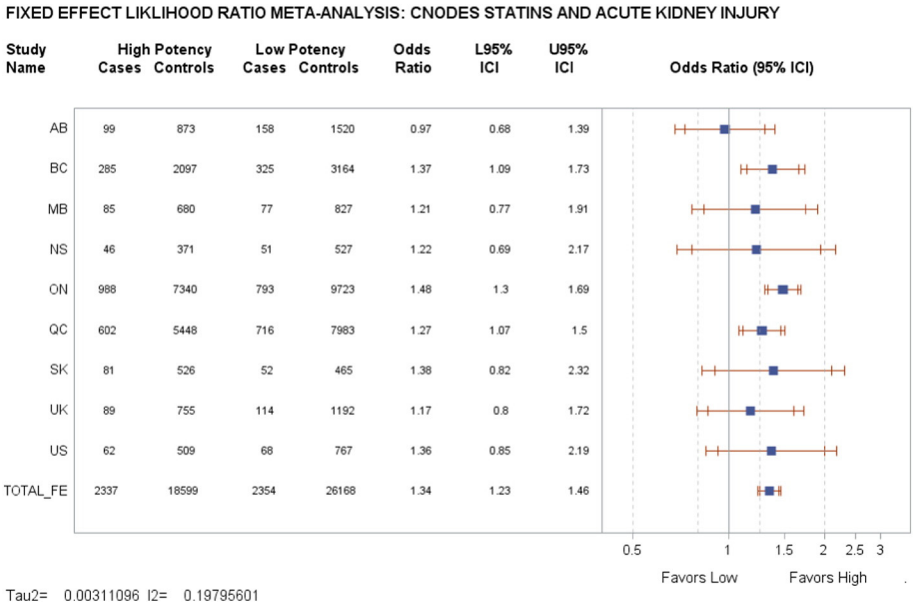

Fig. 1. LRMA meta-analysis of hospitalization for acute kidney injury up to 120 days after cohort entry in patients without chronic kidney disease. Solid bars denote 95% CIs and dashed bars denote 95% ICIs.

a modified version of Eq. (4):

$$\log(LR) = \sum_i \left(\frac{{\sigma_i}^2}{(\sigma_i^2+\tau^2)}\right)\left(\frac{(\ln LR_{\mathrm{MLE}_i}-U_1)^2}{2\sigma_i^2}-\frac{(\ln LR_{\mathrm{MLE}_i}-U_0)^2}{2\sigma_i^2}\right), \tag{5}$$

where U is the grand mean, ${\sigma^2}_i$ is the standard deviation from the ith study, and τ^2 is the estimated variance of underlying effects across studies. As before, ${\sigma^2}_i$ is estimated from the data, as is τ^2. The magnitude of τ^2 is a measure of heterogeneity, with smaller τ^2 denoting less heterogeneity, and τ^2 of zero corresponding to the fixed effect analysis.

The MLEs and ICIs for the individual studies in a random effects analysis are still calculated using Eq. (4) and are the same as in the fixed effect analysis. However, the MLE and 95% ICI for the total effect are now obtained from Eq. (5). When the random effects LRMA method is applied to the data in Example 2, the MLE of the total effect is a rate ratio of 1.31, with a 95% ICI of 1.17 to 1.45. τ^2 is estimated to be 0.003, and the I^2 statistic is 19%. Thus, the fixed effect and random effects analyses are similar in Example 2, which is visually appreciable in Fig. 1, which shows that most of the individual study estimates are rather close.

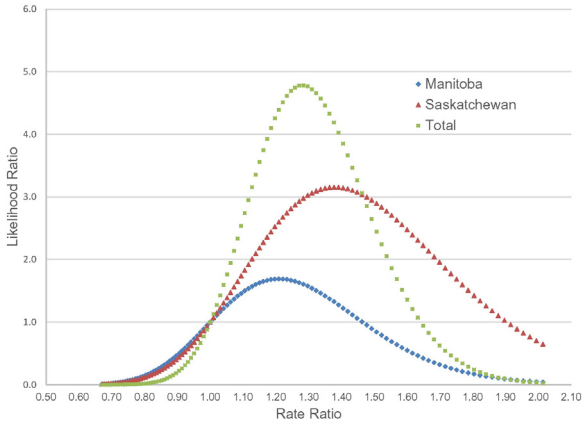

Fig. 2. Plot of LR functions from two separate studies multiplied together.

6. Relation to Bayesian analysis

This section is not an exhaustive treatment of the Bayesian approach. Readers who seek a more thorough discussion of Bayesian analysis and its extension to meta-analysis will find the text book by Spiegelhalter, Abrams and Myles to be a reader-friendly resource [21]. Chapter 8 in particular focuses on evidence synthesis and meta-analysis. The key aspect of the LRMA method worth elucidating here is that LRMA and Bayesian analysis both make use of the LR. In the odds form of Bayes Theorem, the LR is multiplied with the prior odds distribution to obtain the posterior odds distribution. Using the notation in Eq. (4), the odds form of Bayes theorem for a single study can be expressed as

$$\frac{P(\mathrm{RR_{HA}}|\mathrm{RR_{MLE}})}{P(\mathrm{RR_{H_0}}|\mathrm{RR_{MLE}})} = \exp\left[\frac{(\ln \mathrm{RR_{MLE}}-\ln \mathrm{RR_{HA}})^2}{2\sigma_1^2}-\frac{(\ln \mathrm{RR_{MLE}}-0)^2}{2\sigma_0^2}+\ln\left(\frac{\sigma_{\mathrm{HA}}}{\sigma_{\mathrm{H_0}}}\right)\right]\frac{P(\mathrm{RR_{HA}})}{P(\mathrm{RR_{H_0}})}, \tag{6}$$

where the left-hand side of Eq. (6) is the posterior odds of the alternative hypothesis compared to the null hypothesis, and the right-hand side is the LR from Eq. (4) multiplied by the prior odds of $\mathrm{RR_{HA}}$ versus $\mathrm{RR_{H0}}$. In a meta-analysis, the right-hand side would have LRs from each component study, which would be multiplied together along with the prior odds to obtain the posterior odds for the total meta-analytic effect.

Eq. (6) can be used if the desired output is a full statistical distribution and the researcher is willing to declare a prior odds distribution. If so willing, a 95% Bayesian credible interval can be obtained and used instead of the ICI. As a statistical distribution, the posterior odds density function will integrate to 1, and the 95% credible limits will be those odds values that capture 95% of the density. The ICI is calculated from the likelihood ratio function only, which does not integrate to 1. Thus, the 95% ICI does not capture 95% of probability, but instead identifies those values that are 5% as likely as the MLE.

The meta-analysis in Example 2 was motivated in part by adverse event data of renal harm in the Justification for the Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin (JUPITER) trial [22,23]. The introduction of the original CNODES article in Example 2 describes how a non-significant 35% relative increase in the combined

endpoint of acute renal failure (ARF) or a doubling of serum creatinine (SCr×2) was observed in the JUPITER trial [20,22,23]. The trial was stopped early and association with the renal harm remained imprecise; hence the motivation for the CNODES study. The JUPITER data could be used to specify a prior distribution in a Bayesian analysis. That prior could then be multiplied with the LRs from the CNODES studies to obtain a Bayesian posterior and credible interval.

Alternatively, if the researcher wishes to stop short of granting the JUPITER data the status of prior distribution, its LR could still be quantitatively integrated with the CNODES data. Fig. 3 shows what happens when we combine the LR function for the ARF + SCr×2 outcome from the JUPITER trial with the LR functions from the statin–AKI studies done by CNODES. It is immediately noticeable that the MLE from JUPITER trial is compatible with the MLEs from the CNODES studies. It is also apparent that the JUPITER data do not add much additional precision to the total effect estimate. Still, Fig. 3 achieves graphically, and LRMA achieves quantitatively, what we often awkwardly describe in words in our manuscripts when we attempt to place our results in context with other studies.

7. Value of future studies

An exercise similar to adding the JUPITER data to the CNODES data could be done in order to evaluate the value of adding a future study to an existing meta-analysis. Rather than adding data from a previous RCT like we did our last example, we could instead add data from a hypothetical study or multiple hypothetical studies to learn how those additional studies would change the MLE and precision of the total effect estimate of our LRMA. We can experiment with different MLEs, variances, or sample sizes from hypothetical or expected future studies to determine if, for example, another study could potentially provide a useful amount of additional precision. This is particularly relevant to drug safety networks such as CNODES, and other prospective meta-analyses in which the results of all studies may not be available at the same time due to delays in data access or data analysis, or if additional studies are done to overcome low precision.

If a future study comes to fruition, it can easily be added to an existing LRMA because it will not alter the interpretation of the ICI like it would the CI. Adding a new study to an existing meta-analysis because the first meta-analysis was non-statistically significant is not substantially different than the researcher who conducts an experiment, finds the result not quite statistically significant, and subsequently adds more observations in hopes of obtaining a significant result. As discussed earlier, frequentist tests of hypotheses specify the terms of the analysis *before* the data are examined. Pre-specification of the terms of the test is crucial when using type I error rate logic because the α of 0.05 (0.025 two-sided) applies to the pre-specified hypothesis and sample size. If the result after the pre-specified sample size is not statistically significant, then there is no amount of additional data that can change the test result because the test was already completed [24]. Obtaining more data might be perfectly reasonable and even desirable, but the frequentist test paradigm creates problems with how we use that data, and it may be simpler and more feasible to use LRs.

8. Discussion

LRs have a useful application to meta-analyses of RCTs and observational studies in a way that produces familiar and intuitive output. LRMA has numerous advantages over meta-analyses that report the 95% CI as the measure of evidence. The LR in a LRMA is calculated using only the observed data, and unlike measures based on type I errors, decision-making with the LR does not make assumptions about a plethora of future hypothetical studies. The interpretation of any specific LR does not change with multiple looks at the data, and the same data always produces the same LR when using the same probability model. These properties make it perfectly acceptable to apply the LRMA method to new or old meta-analyses alike, to provide a more formal way of comparing competing hypotheses than is allowed by a CI or *p* value. LRMA can be applied to an existing meta-analysis by simply reverse-calculating the point estimate and confidence limits for each study to obtain the standard deviations necessary to calculate the LR. A normal approximation is suitable for most studies of reasonable size, and any discrepancy produced by the distribution is almost surely overshadowed by other sources of distortion inherent to observational studies and RCTs. The advantage of the normal distribution is that the calculations are relatively simple and can be checked in a spreadsheet. In any event, other likelihood functions such as a Poisson or binomial likelihood can be used if needed.

Presenting evidence in the format used in Figs. 1 and 3 is a useful alternative to a forest plot. The format remains basically the same as a forest plot but with wider interval bands. The additional width of an ICI compared to a CI can be large or small depending on the data, but the

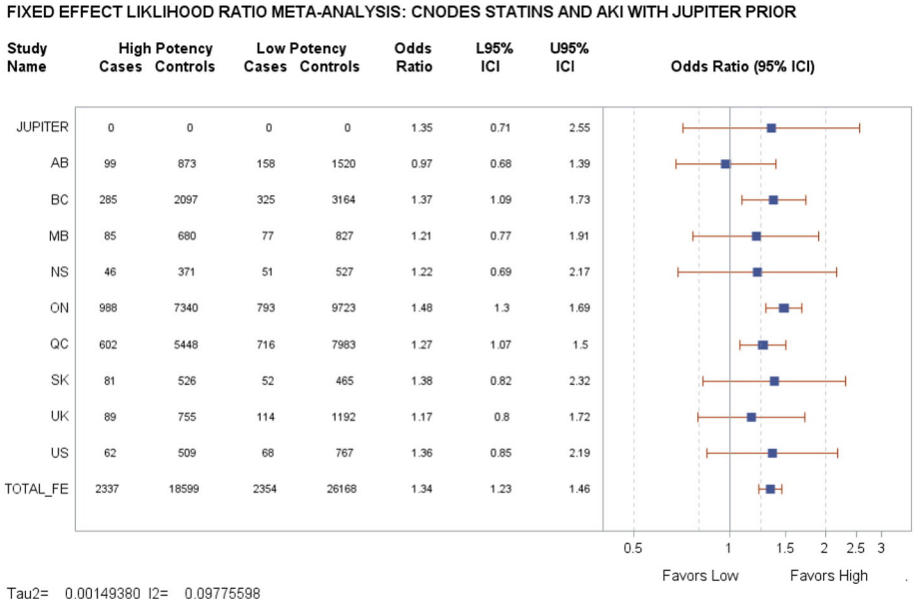

Fig. 3. LRMA meta-analysis of hospitalization for acute kidney injury up to 120 days after cohort entry in patients without chronic kidney disease, with prior data added from the JUPITER trial. Solid bars denote 95% ICIs.

underlying difference in meaning of the ICI is profound. A true evidence-based measure provides a means to quantitatively compare competing hypotheses. All hypotheses outside of the ICI band are less than 5% as likely as the MLE. The ICI is also more compatible with the sentiment that for a meta-analytic effect to be considered real, it should be at least 3 standard errors from the null because the result could influence the treatment of large numbers of patients [12].

LRMA can be used by authors or readers of meta-analyses who are hesitant to relinquish the CI. This can be achieved by providing the 95% ICI and 95% CI together in the same plot. If the ICIs and CIs are plotted together then readers can choose which measure best suits their purpose. A reader who wishes to delineate effect sizes that are less than 5% as likely as the MLE can use the ICI. The CI can be used by readers who instead prefer a decision rule where evidence is not inferred in any given meta-analysis, but after reading many original meta-analyses they will make the right decision 95% of the time.

CNODES Investigators

The Canadian Network for Observational Drug Effect Studies (CNODES) Investigators are: Samy Suissa (Principal Investigator); Colin R. Dormuth (British Columbia); Brenda R. Hemmelgarn (Alberta); Gary F. Teare (Saskatchewan); Patricia Caetano and Dan Chateau (Manitoba); David A. Henry and J. Michael Paterson (Ontario); Jacques LeLorier (Québec); Adrian R. Levy (Nova Scotia); Pierre Ernst (UK Clinical Practice Research Datalink (CPRD)); Robert W. Platt (Methods); and Ingrid S. Sketris (Knowledge Translation).

Funding

This study was financially supported by Canadian Network for Observational Drug Effect Studies (CNODES), which is a collaborating centre of the Drug Safety and Effectiveness Network (DSEN), and is funded by the Canadian Institutes of Health Research (Grant Number DSE-111845). Dr. Filion holds a Canadian Institutes of Health Research New Investigator Award and Dr. Platt holds a National Scholar (Chercheur-national) award (#24919) from the Fonds de Recherche du Québec — Santé (FRQS).

Disclosure and management of competing interests

The following authors report that they have no competing interests: Colin R. Dormuth, Kristian B. Filion, and Robert W. Platt.

Conflicts of interest

The authors have no conflicts of interest to declare.

Appendix I

The variance of a function of a random variable can be derived using the delta method.

Consider an asymptotically normally distributed random variable, Y:

$$Y \sim N(\mu, \sigma^2).$$

A normally distributed function of Y, can be expressed as

$$g(Y) \sim N\left(g(\mu), [g'(\mu)]^2\sigma^2\right).$$

Note that g(Y) must be differentiable and $g'(\mu)$ must not equal zero. If $g'(\mu)$ were zero then the variance of g(Y) would be zero.

The rate ratio can be defined as:

$$RR = \frac{a}{N_1} / \frac{b}{N_0} = \frac{g(a)}{g(b)}$$

where a is the number events in the clopidogrel arm of the CAPRIE trial, b is the number of events in the aspirin arm, and N_1 and N_0 are the total numbers of clopidogrel and aspirin patients, respectively. N_1 and N_0 are constants, fixed at the margin. We wish to show that the variance of the natural logarithm of the rate ratio is:

$$\mathrm{Var}[\ln(RR)] = \frac{1}{a} + \frac{1}{b}.$$

Let

$$h(a) = a/N_1 \text{ and } g(a) = \ln(h(a)); \text{ and}$$

$$h(b) = b/N_0, \text{and } g(b) = \ln(h(b)).$$

The natural logarithm of the rate ratio is

$$\ln(RR) = g(a) - g(b).$$

Applying the delta method to $g(a)$:

$$[g'(a)]^2 s^2 = \left[\frac{d}{da} g(a)\right]^2 s^2 = \left[\frac{1}{h(a)}\right]^2 s^2.$$

a/N_1 can be substituted for $h(a)$, and because a comes from a Poisson distribution, s^2 equals a.

$= \frac{N_1^2}{a^2} s^2$, s^2 for $h(a)$ is $\hat{\mathrm{Var}}(ka) = k^2 \hat{\mathrm{Var}}(a) = \left(\frac{1}{N_1}\right)^2 \cdot a$

$= \frac{N_1^2}{a^2} \cdot \frac{1}{(N_1)^2} \cdot a$, where the constant $k = 1/N_1$

$= \frac{1}{a}$.

The delta method can be used again to obtain the variance of rate in the aspirin control group ($= 1/b$). Adding the two results together yields the variance of the natural logarithm of the rate ratio.

References

[1] R. DerSimonian, N. Laird, Meta-analysis in clinical trials, Control. Clin. Trials 7 (1986) 177–188.
[2] D.B. Petitti, Approaches to heterogeneity in meta-analysis, Stat. Med. 20 (2001) 3625–3633.
[3] R. DerSimonian, N. Laird, Meta-analysis in clinical trials revisited, Control. Clin. Trials 45 (Pt A) (Nov. 2015) 139–145.
[4] J.P.A. Ioannidis, N.A. Patsopoulos, E. Evangelou, Uncertainty in heterogeneity estimates in meta-analyses, BMJ 335 (2007) 914–916.
[5] J.P. Higgins, S.G. Thompson, D.J. Spiegelhalter, A re-evaluation of random-effects meta-analysis, J. R. Stat. Soc. Ser. A 172 (1) (Jan. 2009) 137–159.

[6] J.P. Higgins, S.G. Thompson, Quantifying heterogeneity in a meta-analysis, Stat. Med. 21 (11) (Jun. 15, 2002) 1539–1558.
[7] A.M. Walker, Reporting the results of epidemiologic studies, Am. J. Public Health 76 (5) (May 1986) 556–558.
[8] J.L. Fleiss, Significance tests have a role in epidemiologic research: reactions to A. M. Walker, Am. J. Public Health 76 (5) (May 1986) 559–560.
[9] W.D. Thompson, Statistical criteria in the interpretation of epidemiologic data, Am. J. Public Health 77 (2) (Feb. 1987) 191–194.
[10] C. Poole, Beyond the confidence interval, Am. J. Public Health 77 (2) (Feb. 1987) 195–199.
[11] S.N. Goodman, Toward evidence-based medical statistics. 1: the P value fallacy, Ann. Intern. Med. 130 (12) (Jun. 15 1999) 995–1004.
[12] R. Peto, Why do we need systematic overviews of randomized trials? Stat. Med. 6 (3) (Apr.–May 1987) 233–244.
[13] S.N. Goodman, Meta-analysis and evidence, Control. Clin. Trials 10 (2) (Jun. 1989) 188–204 (Erratum in: Control. Clin. Trials Dec. 1989; 10(4):435).
[14] CAPRIE Steering Committee, A randomised, blinded, trial of clopidogrel versus aspirin in patients at risk of ischaemic events (CAPRIE), Lancet 348 (9038) (Nov. 16, 1996) 1329–1339.
[15] S.N. Goodman, Toward evidence-based medical statistics. 1: the P value fallacy, Ann. Intern. Med. 130 (12) (Jun. 15, 1999) 995–1004.
[16] W. Browner, T. Newman, Are all significant P values created equal? The analogy between diagnostic tests and clinical research, J. Am. Med. Assoc. 257 (1987) 2459–2463.
[17] G.A. Diamond, J.S. Forrester, Clinical trials and statistical verdicts: probable grounds for appeal, Ann. Intern. Med. 98 (1983) 385–394.
[18] R.J. Lilford, D. Braunholtz, For debate: the statistical basis of public policy: a paradigm shift is overdue, BMJ 313 (1996) 603–607.
[19] P.R. Freeman, The role of p-values in analysing trial results, Stat. Med. 12 (1993) 1442–1552.
[20] C.R. Dormuth, B.R. Hemmelgarn, J.M. Paterson, M.T. James, G.F. Teare, C.B. Raymond, J.P. Lafrance, A. Levy, A.X. Garg, P. Ernst, Canadian Network for Observational Drug Effect Studies. Use of high potency statins and rates of admission for acute kidney injury: multicenter, retrospective observational analysis of administrative databases, BMJ 346 (Mar. 18, 2013) f880.
[21] D.J. Spiegelhalter, K.R. Abrams, J.P. Myles, Chapter 8, in Bayesian Approaches to Clinical Trials and Health-Care Evaluation, John Wiley & Sons, Ltd, Chichester, UK, 2003.
[22] P.M. Ridker, E. Danielson, F.A. Fonseca, J. Genest, A.M. Gotto Jr., J.J. Kastelein, et al., JUPITER Study Group. Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein, N. Engl. J. Med. 359 (2008) 2195–2207.
[23] M.D. Roberts, CRESTOR (Rosuvastatin calcium) NDA 21–366 JUPITER, United States Food and Drug Administration Endocrinologic and Metabolic Drugs Advisory Committee, 2009 (www.fda.gov/downloads/advisorycommittees/committeesmeetingmaterials/drugs/endocrinologicandmetabolicdrugsadvisorycommittee/ucm194918.pdf).
[24] J. Cornfield, Sequential trials, sequential analysis and the likelihood principle, Am. Stat. 20 (2) (1966) 18–23.