How to use likelihood ratios to interpret evidence from randomized trials

Perneger, Thomas

How to cite

PERNEGER, Thomas. How to use likelihood ratios to interpret evidence from randomized trials. In: Journal of clinical epidemiology, 2021, vol. 136, p. 235–242. doi: 10.1016/j.jclinepi.2021.04.010

This publication URL: https://archive-ouverte.unige.ch/unige:167174
Publication DOI: 10.1016/j.jclinepi.2021.04.010

Journal of Clinical Epidemiology 136 (2021) 235–242

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

How to use likelihood ratios to interpret evidence from randomized trials

Thomas V. Perneger*

Division of Clinical Epidemiology, Geneva University Hospitals, and Faculty of Medicine, University of Geneva, Geneva 1211, Switzerland

Accepted 20 April 2021; Available online 27 April 2021

Abstract

Objective: The likelihood ratio is a method for assessing evidence regarding two simple statistical hypotheses. Its interpretation is simple – for example, a value of 10 means that the first hypothesis is 10 times as strongly supported by the data as the second. A method is shown for deriving likelihood ratios from published trial reports.

Study design: The likelihood ratio compares two hypotheses in light of data: that a new treatment is effective, at a specified level (alternate hypothesis: for instance, the hazard ratio equals 0.7), and that it is not (null hypothesis: the hazard ratio equals 1). The result of the trial is summarised by the test statistic z (ie, the estimated treatment effect divided by its standard error). The expected value of z is 0 under the null hypothesis, and A under the alternate hypothesis. The logarithm of the likelihood ratio is given by $z \cdot A - A^2/2$. The values of A and z can be derived from the alternate hypothesis used for sample size computation, and from the observed treatment effect and its standard error or confidence interval.

Results: Examples are given of trials that yielded strong or moderate evidence in favor of the alternate hypothesis, and of a trial that favored the null hypothesis. The resulting likelihood ratios are applied to initial beliefs about the hypotheses to obtain posterior beliefs.

Conclusions: The likelihood ratio is a simple and easily understandable method for assessing evidence in data about two competing *a priori* hypotheses.

Key Words: Evidence; Clinical trials; Statistical inference; Likelihood ratio; Confidence interval; *P*-value

1. Introduction

Clinical trials are commonly designed around a statistical test of the effectiveness of a new treatment, and the fate of a new treatment often depends on the resultant *P*-value. However, statistical tests and *P*-values are increasingly criticised, partly for epistemological reasons, and partly because these methods are misunderstood by users [1-5]. But alternatives are few. Confidence intervals are used increasingly, with good reason, as they focus on estimating the parameter of interest [6], but they are often treated as substitutes of tests ("is the null value included in the interval or not"). Bayesian analysis [7-9] is an appealing alternative, but this approach has not been widely adopted, possibly because it represents a radical departure from the status quo.

The likelihood ratio may be a tool worth considering for the assessment of evidence from randomized clinical trials [10-12]. The likelihood ratio compares the two hypotheses under consideration in a clinical trial: that the new treatment improves clinical outcomes compared with the old treatment or with placebo (the degree of improvement must be specified), and that it does not. It provides the researcher with a quantitative measure of the strength of evidence, or support, for one hypothesis over the other. It is based on the same data summary (the z statistic) as the *P*-value. It follows from the Bayes theorem, but does not require a full Bayesian analysis. Thus far, this concept has been presented in rather technical documents, and has not been translated into a « how-to » guide for

The study received no specific funding.
No patients or members of the public were involved in the study.
Data availability: not applicable.
Ethics approval was not required.

* Corresponding author. Tel.: +41 22 372 9036, Fax: +41 22 372 9035.
E-mail address: thomas.perneger@hcuge.ch.

https://doi.org/10.1016/j.jclinepi.2021.04.010

What is new?

Key findings

- The likelihood ratio is a useful tool for comparing two competing point hypotheses (eg, the null and the alternate hypotheses specified in a clinical trial) in light of data.
- The likelihood ratio quantifies the support given by the data to one hypothesis over the other.

What this study adds to what was known

- For randomized clinical trials, the likelihood ratio for the alternate hypothesis (used for sample size determination) vs. the null hypothesis can be computed easily from published results (estimate of effectiveness with 95% confidence interval).
- The likelihood ratio can be combined with prior beliefs about treatment effectiveness (eg, even odds of a specific benefit vs. no benefit) by application of the Bayes' theorem.

What is the implication and what should change now

- In addition to reporting the treatment effect and its confidence interval, trial reports should include z statistics for the key trial outcomes, as well as the likelihood ratio for the a priori alternate hypothesis vs. the null.

users of research evidence. The purpose of this paper is to encourage the use of the likelihood ratio by clinical researchers and by users of evidence from clinical trials, by showing how it can be derived from published trial results.

2. Likelihood ratio

2.1. Hypotheses

A clinical trial is concerned with two hypotheses – that the new treatment is ineffective (null hypothesis, H_0) and that it is effective (alternate hypothesis, H_A). Data are collected to produce a measure of effectiveness, such as a ratio of mortality rates. Currently, upon seeing the evidence, the researcher determines which hypothesis to retain, based on the test of H_0.

Instead of seeking to reject or accept the null hypothesis, the researcher may wish to weigh the relative merits of H_0 and H_A in light of data. This can be done by computing a likelihood ratio, provided that both hypotheses are simple hypotheses, that is, that the level of effectiveness is specified in each case. For the null hypothesis this is obvious, the effectiveness is nil. For the alternate hypothesis as well the level of effectiveness must be quantified, such as "on average blood pressure is lowered by 5 mm Hg," or the "relative hazard of death equals 0.70" –composite hypotheses, defined by an inequality, such as "blood pressure is lowered", or "mortality is reduced", cannot be accommodated by this method. Fortunately, in the case of randomized clinical trials, a specific alternate hypothesis is required for the calculation of sample size, and is typically reported.

2.2. Definition of the likelihood ratio

The likelihood of a hypothesis (H) given the data (x), L(H|x), is proportional to the probability of the data under that hypothesis, P(x|H). When two hypotheses are compared, the one which assigns a higher probability to the observed data is deemed the one more strongly supported by the data. Specifically, the strength of support for H_A over H_0 is given by the likelihood ratio (LR):

$$\mathrm{LR} = \mathrm{L}(\mathrm{H_A}|\mathrm{x})/\mathrm{L}(\mathrm{H_0}|\mathrm{x})$$

Because L(H|x) is proportional to P(x|H), this is equivalent to:

$$\mathrm{LR} = \mathrm{P}(\mathrm{x}|\mathrm{H_A})/\mathrm{P}(\mathrm{x}|\mathrm{H_0})$$

that is, the ratio of the probabilities (or probability densities) of the observed result under the two hypotheses. The LR quantifies the support given by the observed data to the specific hypothesis of effectiveness over the hypothesis of no effect. The greater the likelihood ratio, the stronger the support. A likelihood ratio less than 1 indicates that the hypothesis in the numerator is less strongly supported by the data than that in the denominator, and the inverse can be taken to switch their positions.

2.3. Test statistic and likelihood ratio

Many clinical trials use as measure of effectiveness a difference between means (Δ), or a hazard ratio (HR) or another ratio measure (relative risk or odds ratio). Because these measures are subject to sampling error, statistical procedures apply a transformation, namely a division of the quantity of interest – the observed Δ or log(HR) – by its standard error. The result is a test statistic; when the test statistic converges to the standard normal distribution it is called z. The observed value of z contains all the information the trial provides about treatment effectiveness. Both the likelihood ratio and the *P*-value require z. What differs is what is done once z is observed: the frequentist analysis uses the *P*-value as an argument against H_0; in contrast, the likelihood ratio uses z to compare the support that data provide to H_A vs. H_0.

Under the null hypothesis the distribution of z is centered at 0, and under the alternate hypothesis it is centered at the chosen value of effectiveness, which can be called A. To obtain a *P*-value, the test statistic z is computed from

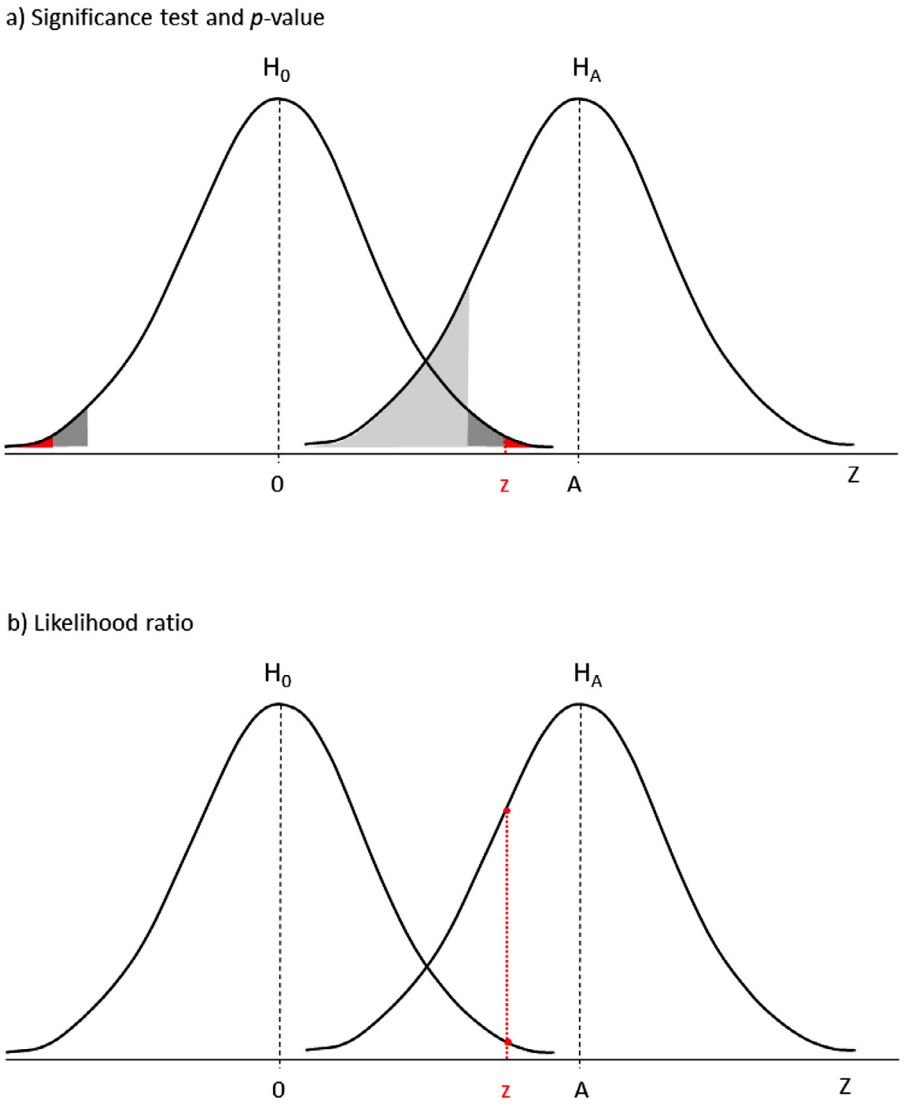

Fig. 1. Interpretation of evidence according to (A) statistical significance testing and *P*-value, and (B) likelihood ratios. The significance test is defined by a two-sided type 1 error rate (dark grey area) under the null hypothesis H_0, and a type 2 error rate (light grey area) under the alternative hypothesis H_A, both defined *a priori*. Once the study result z is observed, a *P*-value is computed (red area). The likelihood ratio is the ratio of the probability densities of the observed result z under the two hypotheses H_A and H_0. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

observed data, then the probability that z may exceed the observed value is obtained under H_0 (Fig.1, upper panel).

To obtain the likelihood ratio, the probability density of z under H_A is divided by its probability density under H_0 (Fig. 1, lower panel). It is a simple ratio of two numbers. Only the observed value of z matters. The two hypotheses are treated equally.

From the definition of the normal density with unit variance, it follows that the natural logarithm of the LR is linearly related to the test statistic z (Box 1). The relationship is given by:

$$\log(\mathrm{LR}) = \mathrm{z} \cdot \mathrm{A} - \mathrm{A}^2/2$$

where A is the expectation (or mean) of z under the alternate hypothesis. More generally, for a contrast between two normal densities centered at A and B:

$$\log(\mathrm{LR}) = \mathrm{z} \cdot (\mathrm{A} - \mathrm{B}) - \left(\mathrm{A}^2 - \mathrm{B}^2\right)/2$$

2.4. How to compute the likelihood ratio from trial results

Thus all that is needed to compute the likelihood ratio are the values of A and z, and a hand-held calculator. The z statistic can be retrieved from the 95% confidence interval for the parameter of interest. For a difference between means (Δ), divide the width of the confidence interval by 3.92 (ie, 2*1.96) to obtain the standard error, then divide the point estimate by this standard error to get z. For a hazard ratio, first take logarithms of the point estimate and of the confidence bounds, then proceed in the same way.

The choice of A is straightforward for randomized clinical trials, since a specific alternate hypothesis has been defined at the design stage of the trial as part of sample size computation. The alternate hypothesis typically represents a clinically relevant difference in the primary outcome variable. To obtain A when the trial seeks to find a difference between means, divide the difference between means assumed under the alternate hypothesis (Δ_A) by the observed standard error, thus $A = \Delta_A/se(\Delta)$. For hazard ratios, first obtain the logarithm of the HR used for the alternate hypothesis (HR_A), and then only divide by the observed standard error of log(HR), thus $A = \log(HR_A)/se(\log(HR))$. When no alternative hypothesis has been specified for the outcome variable of interest, choose a value that is clinically or scientifically relevant.

A caveat: it may be tempting to use for A the observed value of z, which reflects the observed value of Δ or HR (10). The hypothesis A = z receives the highest support from the data and yields the highest value of LR vis-à-vis the null hypothesis. But this is also its main weakness: this "hypothesis" is fully data-driven and is therefore overfitted to the sample at hand. It suffers from the "Texas sharpshooter" fallacy – Joe fires at the barnside, *then* paints the bull's-eye around the bullet hole. With this approach, the LR is always ≥ 1, and a surprising result that favors the null hypothesis over the alternate cannot occur.

Box 1.

Derivation of the likelihood ratio when the test statistic is normally distributed

The probability density function of a normal distribution with expectation μ and variance σ^2 is defined by:

$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$

The z statistic has a normal distribution with parameters $\mu = 0$ and $\sigma = 1$ under the null hypothesis, and $\mu = A$ and $\sigma = 1$ under the alternate hypothesis. Thus the likelihood ratio is given by:

$$LR(A,\ 0; z) = \frac{f_A(z)}{f_0(z)} = \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-A)^2}}{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-0)^2}}$$

and after simplifying the constants,

$$LR(A,\ 0; z) = \frac{e^{-\frac{1}{2}(z-A)^2}}{e^{-\frac{1}{2}z^2}}$$

After taking the logarithm

$$\log(LR(A,0;z)) = -\tfrac{1}{2}(z^2 - 2Az + A^2) - (-\tfrac{1}{2}z^2)$$
$$\log(LR(A,0;z)) = -\tfrac{1}{2}(-2Az + A^2)$$
$$\log(LR(A,0;z)) = Az - \tfrac{1}{2}A^2$$

For a likelihood ratio of hypotheses $\mu = A$ and $\mu = B$

$$\log(LR(A,B;z)) = -\tfrac{1}{2}(z^2 - 2Az + A^2) - (-\tfrac{1}{2}(z^2 - 2Bz + B^2))$$
$$\log(LR(A,B;z)) = -\tfrac{1}{2}(-2Az + A^2 + 2Bz - B^2)$$
$$\log(LR(A,B;z)) = (A - B)z - \tfrac{1}{2}(A^2 - B^2)$$

Once z and A are known, the computation of the logarithm of the likelihood ratio is trivial, as $z{\cdot}A - A^2/2$. To obtain the likelihood ratio itself, exponentiate this quantity.

2.5. Interpretation of the likelihood ratio

The likelihood ratio can be considered on its own: it quantifies how much the data support one hypothesis over another. This is informative in its own right. There isn't yet an arbitrary interpretation guide, such as >10 representing strong evidence, nor is such guideline necessary.

Furthermore, the likelihood ratio can be used to update one's beliefs based on new evidence, by application of the Bayes' theorem:

Posterior odds = LR · Prior odds

This equation shows the relationships between what the researcher believed before the study (prior odds), what the trial results say (likelihood ratio), and what the researcher ought to believe once the results are in (posterior odds) [12].

Many randomized trials start with a state of uncertainty regarding the effectiveness or lack of effectiveness of the new treatment. When only two simple hypotheses are considered, as is required for the likelihood ratio, the uncertainty regarding H_0 and H_A can be represented by prior probabilities of 50% for each, which corresponds to prior odds of 1. In such a case, posterior odds of H_A vs. H_0 would equal the likelihood ratio. The post-trial probability of H_A would reduce to LR/(LR+1), and that of H_0 to 1/(LR+1). Uneven prior odds of the two hypotheses (eg, prior odds of 4, if one put the probability of H_A at 80% and that of H_0 at 20%) are easily accommodated using the formula above.

Note the analogy with the clinical diagnostic process: the clinician formulates prior odds of the patient having a disease (vs. not) based on the history and physical examination, obtains a diagnostic test, and updates the odds of the disease in light of the test result. The likelihood ratio of a positive diagnostic test is sensitivity/(1-specificity), and that of a negative test (1-sensitivity)/specificity [13]. For the clinician, the LR represents the strength of evidence in favor of the disease, vs. its absence; for the trialist, the LR represents the strength of evidence in favor of effectiveness, at the specified level, vs. no effectiveness.

3. Results

Let us consider three examples of application of the likelihood ratio to published trial results.

3.1. Strong evidence

The first trial [14] examined the effect of a community-based intervention, compared to usual care, on the change in systolic blood pressure between baseline and 24 months

Table 1. Procedure to obtain likelihood ratio from published difference in means and its 95% confidence interval

Step	Step by step procedure	Example 1 [14]:
Retrieve results	a) Get difference in means Δ and confidence limits from article	Δ = -5.17 mm Hg, -7.13 to -3.20
Obtain z (normal test statistic)	b) Obtain the width of the confidence interval c) Divide by 3.92 to obtain standard error of Δ d) Divide Δ by its standard error to obtain z	Width = 3.93 SE(Δ) = 1.00 z = -5.17
Select alternate hypothesis	e) Use difference to be detected used for power analysis (Δ_A) f) Compute A as Δ_A divided by the observed standard error	Δ_A = -5 mm Hg A = -5/1.00 = -5
Compute LR vs. null hypothesis	g) Compute log(LR) = z*A – A^2/2 h) Exponentiate to obtain LR in natural units	Log(LR) = 13.35 LR = $6.28*10^5$
Compute LR for A vs. a different hypothesis B	i) Choose difference under hypothesis B (Δ_B) j) Compute log(LR) = z*(A-B) – (A^2-B^2)/2 k) Exponentiate to obtain LR in natural units	Δ_B = -2.5 mm Hg Log(LR_{AB}) = 3.55 LR_{AB} = 34.8

Note: In the calculation of the standard error, 3.93/3.92 = 1.002551… was rounded off to 1.00, to reflect routine practice. Without rounding, the log(LR) becomes 13.28, and the LR is $5.87*10^5$. For A vs. B, the log(LR) is 3.53, and the LR 34.2.

(Table 1). The between-group difference was -5.17 mm Hg, 95% CI -7.13 to -3.20. The width of the confidence interval is divided by 3.92 to obtain the standard error (as luck would have it, it equals 1.00), so z is also -5.17. The difference to be detected was 5 mm Hg, and so was A. The two ingredients, z and A, yield a very large value of log(LR). The observed result supports the hypothesis of an average reduction in systolic blood pressure by 5 mm Hg over the null hypothesis of no difference by a factor of several hundred thousand.

Even someone who was very sceptical about the community intervention, perhaps giving it only one chance in 100 of reducing blood pressure by 5 mm Hg before the trial, should now assign a probability of 99.98% that it reduces blood pressure by 5 mm Hg on average, as opposed to 0 mm Hg. One can also compare other hypotheses: for example, a reduction by 5 mm Hg vs. a reduction by 2.5 mm Hg (Table 1). The trial result supports the former over the latter by a factor of 34.8. If one started with even odds regarding these two hypotheses, after the trial the probability of a reduction by 5 mm Hg would increase to 97.2%.

3.2. Weak evidence

The second trial tested pembrolizumab vs. standard chemotherapy in patients with head and neck cancer [15] (Table 2). The alternate hypothesis for overall survival was HR = 0.70. The analysis yielded a reduction of overall mortality by 20% (HR = 0.80), and the resulting likelihood ratio was 4.3 for the planned alternate hypothesis (HR_A = 0.70) over the null hypothesis (HR_0 = 1). This reflects much weaker evidence. The probability of a relative mortality reduction by 30% (vs. none) would increase from 50% (even odds) to only 81.1%. For a strong proponent of this treatment, who would set the a priori probability of effectiveness (ie, HR_A = 0.7) to 80%, the result would increase this probability to 94.5%, but for a sceptic, the probability of effectiveness (at HR_A = 0.7) might move from 20% to 51.8%. Both the proponent and the sceptic apply the same likelihood ratio, but they end up with different conclusions (or posterior odds) because they start from different initial beliefs (prior odds).

3.3. Evidence in favor of the null hypothesis

The third trial compared hydroxychloroquine plus standard of care to standard of care alone, in patients hospitalized with coronavirus disease 2019 [16]; the primary outcome was the conversion to negative nasopharyngeal swabs for the virus (Table 2, last column). The alternate hypothesis was HR = 1.43 (formulated as HR = 0.7 for standard of care vs. hydroxychloroquine, which translates to 1/0.7 = 1.43 for hydroxychloroquine vs. standard of care). The observed hazard ratio was 0.85 for hydroxychloroquine vs. standard of care. The conclusion was that “hydroxychloroquine did not result in a significantly higher probability of negative conversion”. This is of course correct, but the results are stronger than this statement of “absence of evidence” suggests. The LR was only 0.036. Taking the inverse yields a LR of 27.6 for the null hypothesis over the alternate, a rather strong support for no effect over the hypothesized HR_A = 1.43. For someone who assigned even odds to these 2 hypotheses before the trial, the probability that hydroxychloroquine

Table 2. Procedure to obtain likelihood ratios from published hazard ratios and confidence intervals, and to apply Bayes theorem to the results

Step	Step by step procedure	Example 2 [15]:	Example 3 [16]
Retrieve results	a) Get HR and confidence limits from article	HR 0.80; 95% CI 0.65 to 0.98	HR 0.85 95% CI 0.58 to 1.23
Obtain z (normal test statistic)	b) Take logarithm of 95% confidence bounds of HR	-0.4308 and -0.0202	-0.5447 and 0.2070
	c) Obtain the width of the confidence interval	Width = 0.4106	Width = 0.7517
	d) Divide by 3.92 to obtain standard error of log(HR)	SE(log(HR)) = 0.1047	SE(log(HR)) = 0.1918
	e) Take natural logarithm of HR	Log(HR) = -0.2231	Log(HR) = -0.1625
	f) Divide log(HR) by its standard error to obtain z	z = -2.1309	Z = -0.8472
Select alternate hypothesis	g) Use value of HR_A used for power analysis	$HR_A = 0.70$	$HR_A = 1.43$
	h) Take logarithm of HR_A	Log (HR_A) = -0.3567	Log (HR_A) = 0.3577
	i) Divide by the observed standard error of log(HR)	A = -3.4066	A = 1.8648
Compute LR for alternate hypothesis vs. null hypothesis	j) Compute $\log(LR) = z^*A - A^2/2$	Log(LR) = 1.4567	Log(LR) = -3.3186
	k) Exponentiate to obtain LR in natural units	LR = 4.29	LR = 0.0362
Compute LR for null hypothesis vs. alternate hypothesis	l) Take inverse of LR; $LR_{0A} = 1/LR_{A0}$		$LR_{0A} = 27.6$
Apply Bayes theorem to initial even odds	m) Under even odds, compute post-test odds O = LR	O = 4.29	O = 27.6
	n) Post-test probability P = O/(O+1)	$P_A = 81.1\%$	$P_0 = 96.5\%$

treatment leads to faster negativation of virus detection tests by a factor 1.43 should now decrease to 3.5%.

4. Discussion

4.1. Desirable properties of the likelihood ratio

4.1.1. Theoretical foundation

Likelihood plays a central role in statistics [11]. Most statistical estimation procedures yield the parameter value that has the highest likelihood, given the data, and classic statistical tests are based on the likelihood ratio. The idea of choosing between hypotheses on the basis of their likelihoods is therefore natural. The reasoning required to apply a likelihood ratio to 2 hypotheses may be a stepping stone to a full Bayesian analysis.

4.1.2. Interpretability

The likelihood ratio is a simple procedure, and this may help avoid misinterpretation, a problem that plagues *P*-values and statements of statistical significance. The two hypotheses under consideration are treated equitably; only the choice of the numerator vs. denominator is arbitrary, and easily reversed. The likelihood ratio only depends on the observed result, and not on possible results that have not occurred. It has a non-technical interpretation: how much do the results of the trial support one simple hypothesis over another? Furthermore, many health professionals are familiar with likelihood ratios applied to diagnostic testing, where the goal is to assess the probability of the absence or presence of disease, given a diagnostic test result [7,13].

The likelihood ratio complements, but does not replace, descriptions of substantive results and estimation procedures. The difference in means or the hazard ratio, with a confidence interval, tells the researcher what the most likely effect is based on the data, knowing that the data are subject to random variation and that the estimator is by definition overfitted. The likelihood ratio captures what the data say about the relative merits of two prespecified hypotheses. When the selected hypotheses are scientifically or clinically meaningful, this adds a useful insight.

4.1.3. Integration of evidence

The likelihood ratio enables readers to update their beliefs about treatment effectiveness by application of the Bayes' theorem. This mechanism also allows the combination of trial results with other sources of knowledge. This is particularly important in the era of precision medicine [17], as one difficulty lies in integrating knowledge of molecular mechanisms (which can be represented by strong prior odds) with the need for empirical confirmation (represented by the likelihood ratio).

The likelihood ratio also allows the pooling of evidence from several trials. If one trial yields a LR of 5, and a second independent trial produces a LR of 3, then the combined LR is the product, 15. This is a direct consequence of the Bayes' theorem. The evidence as represented by log(LR) is additive. This is a particularly useful property for meta-analysis of clinical trials. It also simplifies the interpretation of trial results when an interim analysis was performed: the log(LR) obtained for the first

part of the trial is simply added to the log(LR) obtained in the second part, without any adjustment for multiplicity.

4.2. Weaknesses

4.2.1. Two hypotheses

The likelihood ratio assesses the weight of evidence regarding two point hypotheses. For clinical trials that have pre-specified a null and an alternate hypothesis this approach is natural. But for observational research, selecting two point hypotheses may be challenging. Researchers and readers may hold different opinions as to which hypotheses should be compared, and indeed wonder why only two deserve consideration. In such situations a Bayesian analysis that starts with a full prior distribution for the parameter of interest will be indicated. Of note, any pair of simple hypotheses can be compared using the likelihood ratio once z is known.

Even in the case of clinical trials, limiting possible options to two simple hypotheses is an over-simplification of reality. Noone believes that, say, HR = 1 and HR = 0.7 are the only possible parameter values. But these values represent two scenarios, two paradigmatic states of the world, one in which the new treatment doesn't work, and the other in which the treatment is clinically beneficial. The likelihood ratio identifies the option that receives stronger support from the data. This is valuable, but does not replace other statistical procedures, such as the estimation of the treatment effect.

In some cases, the binary simplification required by the likelihood ratio may represent fairly the nature of the underlying uncertainty. Many drugs have a known molecular mechanism – a receptor is blocked, or an enzyme inhibited – and the question addressed by the clinical trial is whether this mechanism is causally related to clinical outcomes. The two simple hypotheses stand in for causality (H_A) and its absence (H_0).

4.2.2. No immunity from bias, confounding, or poor practice

Like other statistical procedures, the likelihood ratio can be affected by various methodologic problems, such as biased sampling, missing data, non-compliance, imprecise measurement, or model mis-specification. These still need to be taken into account when interpreting the results of a trial. For instance, while the evidence obtained in the trial of hypertension management [14] supported the hypothesis that the new program reduced blood pressure by 5 mm Hg rather than not at all, one might still remain doubtful if there had been large losses to follow-up in the intervention arm, or if the measurement procedures had been biased (neither were the case).

Statistical tests are sometimes unfairly blamed for what is just poor scientific practice, including unjustified dichotomization of results as "significant" or not [5], data dredging for small *P*-values, or publication bias [18]. The more intuitively understandable nature of the likelihood ratio may protect it from such misuse.

4.2.3. Extra work

This paper shows that the computation of likelihood ratios is reasonably easy for readers of trial reports, without waiting for this method to be adopted by trial statisticians. Nevertheless, this requires extra effort. But there is no reason why z statistics and likelihood ratios couldn't be provided directly in trial reports, alongside the substantive result (ie, difference in means or hazard ratios, with confidence intervals). This process would be facilitated if reporting guidelines such as CONSORT required it.

5. Conclusion

The likelihood ratio measures the relative support given to two simple hypotheses by the results of a clinical trial. It can be computed from information routinely available in trial reports. Requiring authors of trial reports to publish relevant likelihood ratios (eg, for the specified alternate hypothesis vs. the null) and z statistics may enhance the understanding and facilitate the application of trial results.

CRediT authorship contribution statement

Thomas V. Perneger: Conceptualization, Methodology, Formal analysis, Resources, Writing - original draft, Writing - review & editing, Visualization, Project administration.

References

[1] Goodman SN. Toward evidence-based medical statistics, 1: the P value fallacy. Ann Intern Med 1999;130:995–1004.
[2] Sterne JAC, Davey Smith G. Sifting the evidence - what's wrong with significance tests? BMJ 2001;322:226–31.
[3] Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol 2016;31:337–50.
[4] Wasserstein RL, Lazar NA. The ASA's statement on p-values: context, process, and purpose. Am Stat 2016;70:129–33.
[5] Wasserstein RL, Schirm AL, Lazar NA. Moving to a world beyond "p<0.05. Am Stat 2019;73(Suppl 1):1–19.
[6] Rothman KJ. A show of confidence. N Engl J Med 1978;299:1362–3.
[7] Spiegelhalter DJ, Friedman LS, Parmar MKB. Bayesian approaches to trials. J Royal Stat Soc A 1994;157:357–87.
[8] Berry DA. Bayesian clinical trials. Nat Rev Drug Discov 2006;5:27–36.
[9] Lee JJ, Chu CT. Bayesian clinical trials in action. Stat Med 2012;31:2955–72.
[10] Goodman SN. Toward evidence-based medical statistics, 2: the Bayes factor. Ann Intern Med 1999;130:1005–13.
[11] Edwards AWF. Likelihood. Expanded edition. Baltimore MD: Johns Hopkins University Press; 1992.
[12] Royall R. Statistical Evidence - A Likelihood Paradigm. London: Champan & Hall; 1997.
[13] Grimes DA, Schulz KF. Refining clinical diagnosis with likelihood ratios. Lancet 2005;365:1500–5.

[14] Jafar TH, Gandhi M, de Silva A, Jehan I, Naheed A, Finkelstein EA, et al. A community-based intervention for managing hypertension in rural South Asia. N Engl J Med 2020;382:717–26.

[15] Cohen EEW, Soulières D, Le Tourneau C, Dinis J, Licitra L, Ahn MJ, et al. Pembrolizumab versus methotrexate, docetaxel, or cetuximab for recurrent or metastatic head-and-neck squamous cell carcinoma (KEYNOTE-040): a randomised, open-label, phase 3 study. Lancet 2019;393:156–67.

[16] Tang W, Cao Z, Han M, Wang Z, Chen J, Sun W, et al. Hydroxychloroquine in patients with mainly mild to moderate coronavirus disease 2019: open label, randomised controlled trial. BMJ 2020;369:m1849.

[17] Tonelli MR, Shirts BH. Knowledge for precision medicine. Mechanistic reasoning and methodological pluralism. JAMA 2017;3218:1649–50.

[18] Perneger TV, Combescure C. The distribution of p-values in medical research articles suggested selective reporting associated with statistical significance. J Clin Epidemiol 2017;87:70–7.