Expected Monotonicity – A Desirable Property for Evidence Measures?

Susan E. Hodge^a Veronica J. Vieland^b

^aDivision of Epidemiology, NY State Psychiatric Institute, Department of Psychiatry, Columbia School of Physicians and Surgeons, and Division of Statistical Genetics, Department of Biostatistics, Columbia Mailman School of Public Health, New York, N.Y., and ^bBattelle Center for Mathematical Medicine, Research Institute at Nationwide Children's Hospital, and Departments of Pediatrics and Statistics, Ohio State University, Columbus, Ohio, USA

Key Words

Evidence • Likelihoods • Likelihood ratios • Posterior probability of linkage

Abstract

We consider here the principle of 'evidential consistency' – that as one gathers more data, any well-behaved evidence measure should, in some sense, approach the true answer. Evidential consistency is essential for the genome-scan design (GWAS or linkage), where one selects the most promising locus(i) for follow-up, expecting that new data will increase evidence for the correct hypothesis. Earlier work [Vieland, Hum Hered 2006;61:144–156] showed that many popular statistics do not satisfy this principle; Vieland concluded that the problem stems from fundamental difficulties in how we measure evidence and argued for determining criteria to evaluate evidence measures. Here, we investigate in detail one proposed consistency criterion – expected monotonicity (ExpM) – for a simple statistical model (binomial) and four likelihood ratio (LR)-based evidence measures. We show that, with one limited exception, none of these measures displays ExpM; what they do display is sometimes counterintuitive. We conclude that ExpM is *not* a reasonable requirement for evidence measures; moreover, no requirement based on expected values seems feasible. We

demonstrate certain desirable properties of the simple LR and demonstrate a connection between the simple and integrated LRs. We also consider an alternative version of consistency, which is satisfied by certain forms of the integrated LR and posterior probability of linkage.

Copyright © 2010 S. Karger AG, Basel

1. Introduction

We have been interested for some time in the underlying nature of statistical evidence and in what constitutes a good measure of evidence. The very question of what properties an evidence measure should possess is an open one (see classic works such as [1–4], as well as some of our own earlier studies [5–7]). We are interested both in determining what those properties should be and in finding a measure of evidence that does possess them.

One critical issue concerns the concept of 'evidential consistency,' that is, the principle that as one gathers more data, one's evidence measure should, in some sense, move toward the true answer. Evidential consistency represents an essential underpinning of the genome-scan design, in which one performs a genome-wide search for linkage and/or association and then selects the locus with the strongest evidence for follow-up or replication. Investiga-

tors implicitly assume that if they start with a dataset showing evidence for a hypothesis and collect a second dataset that also shows evidence for the hypothesis, then the combined evidence from the two datasets should be greater than the original evidence. Without this assumption, interpreting follow-up studies in which the evidence appears to go down is moot.

This general issue has been explored by many authors in many contexts. However, Vieland [6] raised the question in a new form, with specific relevance to human genetics. She proposed that a well-behaved measure of evidence should be ‘consistent’ (still in a sense to be defined), not just asymptotically, and moreover that such consistency should hold regardless of how the initial data are selected for follow-up. In particular, she proposed that when a dataset is selected for follow-up specifically because it provides unusually large support for one hypothesis or the other, the addition of a new set of data should still change the evidence measure in the correct direction, and not tend to ‘regress to the mean.’ She went on to propose one specific form of consistency, ‘expected monotonicity,’ which we define below. In this paper, we examine several commonly used evidence measures and determine whether they display expected monotonicity.

One can see the importance of this concept of consistency in particular when one considers that workhorse of human genetics, the genome scan. In this situation, investigators perform an initial GWAS or a genome-wide linkage study and identify the most ‘promising’ markers or locations, that is, those thought to be showing the strongest evidence for association and/or linkage. But when they go to follow up these most promising findings, typically the positive evidence appears to melt away. This phenomenon is sometimes referred to as ‘the winner’s curse,’ and it has been observed in countless data analyses over the past two decades. Vieland and colleagues [8, 9] analyzed it in some detail and showed that, in the presence of underlying heterogeneity in the data, ‘evidence’ as measured by maximum lod scores [10] or so-called model-free statistics can systematically diminish as additional data are considered, even when the initial finding is a true positive (also see [11, 12]).

Usually, when a signal diminishes on follow-up, investigators conclude that the initial finding was a false positive. But if loci selected for showing the strongest evidence in an initial scan are precisely those most likely to diminish on follow-up, regardless of whether they represent true or false positives, then this could be entirely the wrong conclusion. That is, initial findings that seemed so promising but then disappeared might well have been

true positives in the first place, and discarding them could represent a huge inflation of type II error rates. This is why some form of monotonicity is important: it is essential that we have a measure of evidence that behaves correctly – in a sense to be defined – regardless of how we select which initial experiment (marker) to follow up. As Vieland [6] put it in her ‘temperature’ analogy: when using a thermometer to measure temperature, if we stoke the boiler in a room that is already warm, we expect the new thermometer reading to indicate a *higher* temperature, not a lower one. Similarly, evidence measures per se should not be subject to the winner’s curse.

While this form of the winner’s curse is generally accepted as inevitable, Vieland [6] argued that it reflects an underlying problem with the way we measure evidence, rather than an inherent liability of the genome-scan design, and she proposed some specific properties that evidence measures should have. Key among these was a property she termed ‘expected monotonicity’ (ExpM). Informally, ExpM says that, when new data are considered, if the hypothesis H of interest (e.g. ‘linkage’) is true, then on average, the evidence measure should increase, whereas if H is false (‘no linkage’), then on average the evidence measure should decrease.

More formally, let T be any proposed evidence measure, where T_n is its value after n observations; and consider the value of T after the addition of m new observations. Define the conditional expected value $CondEV = E[T_{n+m} | T_n]$. Here, we are considering evidence for or against H , and without loss of generality we have assumed that T is scaled such that increasing values indicate increased evidence for H . Then ExpM implies:

$$CondEV \begin{cases} > T_n & \text{if } H \text{ is true} \\ < T_n & \text{if } H \text{ is false} \end{cases} \quad (1)$$

The stipulation that it is the expected value (EV) of T that should behave appropriately, rather than T itself, acknowledges the stochastic nature of the underlying data, such that even in the presence of linkage, excess recombinations can occur in any given sample.

In this paper, we investigate the concept of ExpM at its most basic level, considering a simple statistical model (the binomial distribution) and four types of likelihood ratio (LR)-based evidence measures. We show that, with one limited exception, none of the measures we consider displays ExpM, and indeed, that what they do display is sometimes surprising and counterintuitive. We conclude that ExpM is *not* a reasonable requirement to impose upon evidence measures, and we briefly consider the implications of this fact.

2. Framework of This Study: Definitions and Assumptions

Table 1 lists the abbreviations we use throughout the paper.

2.1. Binomial Distribution

We limit attention to the binomial distribution with parameter θ , and we restrict our primary investigation to distributions in which θ is in $[0, 0.5]$. We consider only LRs involving one hypothesis (H_0) that specifies a single value $\theta = \theta_0$ on the border ($\theta_0 = 0.5$). The other hypothesis (H_1) either specifies a single alternative value $\theta = \theta_1$, where $\theta_1 \in [0, 0.5]$ (for the simple LR), or represents the compound hypothesis $\theta < 0.5$. This model corresponds to linkage analysis of fully informative gametes (FIGs) [13], i.e., gametes that unambiguously represent either a recombination event or a nonrecombination event. We therefore refer to the observed data as FIGs and consider θ to be the recombination fraction. However, we also briefly discuss what happens for full binomial distributions, where the parameter θ can range over $[0, 1]$ and θ_0 is not on the border ($0 < \theta_0 < 1$). Finally, we discuss possible alternative criteria that investigators might consider, rather than ExpM.

Notation: We have a sample of n FIGs, of which x are recombinant. We always start with a dataset consisting of n observations; this is our ‘initial’ dataset. We then add either one new observation, or m new observations, to form our ‘updated’ dataset.

We will examine four types of statistics that one can potentially use as measures of evidence: the simple LR (sLR); the maximized LR (maxLR), and the posterior probability of linkage (PPL), along with the closely related quantity, the integrated LR (intLR). In the next subsection (2.2.) we define what we mean by each measure and give the formulation of each measure under the one-sided binomial distribution; in 2.3., we show a useful property shared by all four types of measures. Then, in section 3., we turn explicitly to expected monotonicity.

2.2. Four Likelihood-Based Measures of Evidence

2.2.1. The Simple LR

Let $L(\theta)$ represent the likelihood function of θ . To use the sLR, one specifies a simple alternative hypothesis, i.e., $H_0: \theta = 0.5$ versus $H_1: \theta = \theta_1$, where $\theta_1 \in [0, 0.5]$. The sLR is defined as

$$sLR = \frac{L(\theta_1)}{L(0.5)}, \quad 0 \leq \theta_1 < 0.5. \quad (2)$$

Table 1. Abbreviations used throughout the paper

FIG	fully informative gamete
EV	expected value
CondEV	conditional expected value
ExpM	expected monotonicity
LR	likelihood ratio
sLR	simple likelihood ratio
maxLR	maximized likelihood ratio
PPL	posterior probability of linkage
intLR	integrated likelihood ratio

A value of $sLR > 1$ supports H_1 (linkage at that value of θ_1) over H_0 , whereas a value < 1 supports H_0 over H_1 , in the sense that the ‘supported’ hypothesis is the one that makes the data more probable.

For a dataset of n FIGs, of which x are recombinants, the expression in (2) becomes

$$sLR = \frac{\theta_1^x (1 - \theta_1)^{n-x}}{(0.5)^n} = 2^n \theta_1^x (1 - \theta_1)^{n-x}, \quad 0 \leq \theta_1 < 0.5.$$

When considering two sets of data, one of size n and with x recombinants, the other of size m , with y recombinants, the sLR for the combined data is simply

$$sLR = 2^{n+m} \theta_1^{x+y} (1 - \theta_1)^{n+m-x-y}. \quad (3)$$

This follows directly from the first principles of probability theory.

2.2.2. The Maximized LR

Now the alternative hypothesis is compound, i.e., $H_0: \theta = 0.5$ versus $H_1: \theta < 0.5$. One maximizes the LR over the parameter and takes that maximum value as the measure of evidence:

$$\max LR = \max_{\theta \in [0, 0.5]} \frac{L(\theta)}{L(0.5)}.$$

Unlike in the case of the sLR, where it was intuitively clear that a $sLR > 1$ or < 1 supported one hypothesis or the other, here it is not so clear, since the maxLR is always ≥ 1 . The general consensus is that, in order to support H_1 , the maxLR must be ‘enough’ greater than unity, according to some agreed-upon criterion (see, e.g., [2]).

For the linkage situation, since θ cannot exceed 0.5, we have

$$\hat{\theta} = \begin{cases} \frac{x}{n} & \text{if } x \leq n/2, \\ 0.5 & \text{if } x > n/2 \end{cases},$$

and therefore

$$\max LR_n(x) = \begin{cases} \frac{2^n x^x (n-x)^{n-x}}{n^n} & \text{if } x \leq n/2. \\ 1 & \text{if } x > n/2 \end{cases} \quad (4)$$

When considering two sets of data, the maxLR is given by

$$\max LR_{n+m}(x+y) = \begin{cases} \frac{(x+y)^{x+y} (n+m-x-y)^{n+m-x-y}}{\left(\frac{n+m}{2}\right)^{n+m}} & \text{if } x+y \leq (n+m)/2. \\ 1 & \text{if } x+y > (n+m)/2 \end{cases} \quad (5)$$

That is, the data are combined and the maxLR is re-computed at the maximum value based on all the observations. (The alternative, maximizing the LR separately in the two sets, is not a viable course of action; also see section 5.2. below.)

2.2.3. The Posterior Probability of Linkage and the Integrated LR

Here we consider the PPL introduced by Vieland [14] as a measure of evidence. Consider again $H_0: \theta = 0.5$ (no linkage) versus $H_1: \theta < 0.5$ (linkage). The PPL approach posits a prior probability distribution for θ , denoted $\pi(\theta)$, defined on $\theta \in [0, 0.5]$. This function assigns a point mass ρ to $\theta = 0.5$ (representing the prior probability of ‘no linkage’) and distributes the rest of the probability $(1 - \rho)$ over $\theta \in [0, 0.5]$. Then the PPL is defined as

$$PPL = \frac{\int_{0 \leq \theta < 0.5} L(\theta) \pi(\theta) d\theta}{\int_{0 \leq \theta \leq 0.5} L(\theta) \pi(\theta) d\theta}. \quad (6)$$

In this study, we consider the special case where $\pi(\theta)$ is constant on $\theta \in [0, 0.5]$:

$$\pi(\theta) = \begin{cases} 2(1-\rho) & \text{for } 0 \leq \theta < 0.5 \\ \rho & \text{for } \theta = 0.5 \end{cases}.$$

Now $\pi(\theta)$ can come out of the integration, so the PPL in (6) can be expressed as

$$\begin{aligned} PPL &= \frac{2(1-\rho) \int_{0 \leq \theta < 0.5} L(\theta) d\theta}{2(1-\rho) \int_{0 \leq \theta < 0.5} L(\theta) d\theta + \rho L(\theta=0.5)} = \frac{\left(\frac{1-\rho}{\rho}\right)^R}{\left(\frac{1-\rho}{\rho}\right)^{R+1}} \\ &= \frac{(1-\rho)R}{(1-\rho)R + \rho}, \end{aligned} \quad (7)$$

where R is defined as the intLR:

$$\text{intLR} = R \equiv 2 \int_{0 \leq \theta < 0.5} L(\theta) d\theta / L(0.5).$$

Note that R corresponds to the Bayes’ Factor [15] in this simple case. Eq. (7) shows that for a given value of ρ , the PPL and R are simple rescalings of each other.

Since the prior probability of linkage is $1 - \rho$, it follows that $PPL > 1 - \rho$ implies support for $\theta < 0.5$, i.e., for linkage, whereas $PPL < 1 - \rho$ supports $\theta = 0.5$, i.e., lack of linkage. ($PPL = 1 - \rho$ implies no information in the data about linkage.) Equivalently, $R > 1$ supports linkage, $R < 1$ supports lack of linkage, and $R = 1$ implies no information, just as with the sLR.

Here is the formulation of the intLR, $R_n(x)$, for a sample of n FIGs with x recombinants:

$$R_n(x) = \frac{2 \int_0^{0.5} \theta^x (1-\theta)^{n-x} d\theta}{(0.5)^n} = 2^{n+1} \int_0^{0.5} \theta^x (1-\theta)^{n-x} d\theta. \quad (8)$$

This integral in (8) corresponds to an Incomplete Beta function and can be expressed as a finite sum, as follows:

$$R_n(0) = \frac{2^{n+1} - 1}{n+1},$$

and

$$R_n(x) = \sum_{i=0}^x (-1)^i \binom{x}{i} 2^{x-i} \frac{2^{n-x+1+i} - 1}{n-x+1+i}, \quad x > 0. \quad (9)$$

Details are in Appendix D (The Integrated LR). Thus, (9) gives the formulation for $R_n(x)$; for the PPL, substitute the expressions in (9) into (7), yielding:

$$PPL_n(x) = \frac{(1-\rho)R_n(x)}{(1-\rho)R_n(x) + \rho}. \quad (10)$$

We will examine the behavior of both the PPL in (10) and $R_n(x)$ in (8) for expected monotonicity.

When considering two sets of data, we will assume that $R_{n+m}(x+y)$ is computed similarly to maxLR, that is, by ‘pooling’ the $n+m$ observations together and recalculating the intLR. The ‘pooled’ PPL is then computed from this updated value of $R_{n+m}(x+y)$. We note, however, that in the case of $R_{n+m}(x+y)$ (and the corresponding PPL), we also have another option for obtaining the combined evidence, namely, via Bayesian sequential updating (section 5.2. in the Discussion).

2.3. A Feature of All the Measures

Here we point out a feature common to all the measures (and indeed, a feature that seems reasonable, but also see section 5.3. in the Discussion on the question of reasonableness): given an initial dataset and its corresponding evidence measure, if the next observation is a recombinant, the measure of the new dataset is less than

before, whereas if the next observation is a nonrecombinant, the measure of the new dataset is greater than before. That is, let $T_n(x)$ be an evidence measure for a dataset of n observations, of which x are recombinant. Then

$$T_{n+1}(x) > T_n(x) > T_{n+1}(x+1). \quad (11)$$

The CondEV after addition of a single observation to an existing dataset is given by

$$\begin{aligned} \text{CondEV}_\theta &= \theta T_{n+1}(x+1) + (1-\theta)T_{n+1}(x) = \\ &T_{n+1}(x) - [T_{n+1}(x) - T_{n+1}(x+1)]\theta, \end{aligned} \quad (12)$$

where θ in (12) represents the true value of the recombination fraction. The quantity in square brackets in (12) is positive, from (11), and therefore the CondEV is a monotonically decreasing linear function of (the true) θ .

We will be interested in the behavior of the CondEV when the true value of θ is 0.5:

$$\text{CondEV}_{\theta=0.5} = (0.5)[T_{n+1}(x+1) + T_{n+1}(x)], \quad (13)$$

since that corresponds to ‘no linkage,’ i.e., to the ‘H is false’ part of (1). In some situations considered below, we will examine this quantity directly; in others, it will be convenient instead to examine θ_{cutoff} . This is defined as that value of θ for which the CondEV, after one additional observation, equals the initial value of the evidence measure. I.e., letting A represent the initial value, θ_{cutoff} is that value of θ satisfying

$$E_\theta[T_{n+1}(x) | T_n = A] = A.$$

Therefore, from (12),

$$\theta_{\text{cutoff}} = \frac{T_{n+1}(x) - A}{T_{n+1}(x) - T_{n+1}(x+1)}. \quad (14)$$

The quantity θ_{cutoff} does not inherently interest us, but provides a useful tool for examining the behavior of the CondEV in some of the situations we consider below. Here is the connection between the CondEV in (13) and the cutoff θ in (14): since the linear function of θ in (12) has a negative slope, it follows that the $\text{CondEV}_{\theta=0.5}$ (after one additional observation) is $>A$, $=A$, or $<A$, if, and only if, θ_{cutoff} is >0.5 , $=0.5$, or <0.5 , respectively. Moreover, whatever the value of the CondEV (after one additional observation) when the true value of $\theta = 0.5$, that CondEV increases linearly as the true value of θ decreases toward zero. In particular, for those cases where $0 < \theta_{\text{cutoff}} < 0.5$, we have:

$$\text{CondEV} = E_\theta[T_{n+1}(x)|T_n(x)] \begin{cases} > A \text{ for } 0 \leq \theta < \theta_{\text{cutoff}} \\ = A \text{ for } \theta = \theta_{\text{cutoff}} \\ < A \text{ for } \theta_{\text{cutoff}} < \theta \leq 0.5 \end{cases}. \quad (15)$$

In some cases below, it will suffice to examine what happens when we add a single new observation, in which case (12)–(15) apply. However, in general we are interested in what happens when we add m new observations, which will require additional derivation.

3. Do We Have Expected Monotonicity?

In this section, we examine each statistic (and in most cases, its log as well) defined in section 2. to determine whether it displays expected monotonicity, as defined in (1).

3.1. Does the Simple LR Display ExpM?

In this subsection we will show that the sLR of section 2.2.1. displays a kind of ‘half’ expected monotonicity, whereas its log displays expected monotonicity exactly as expressed in (1).

Let $LR_n(x)$ denote the sLR for n FIGs and x recombinants. Note two features of the sLR: first, these LR's have a convenient property of ‘factorization’; that is, if one combines sLRs from two independent datasets, the new sLR is the product of the two individual sLRs:

$$\text{Factorization: } sLR_n(x) \cdot sLR_m(y) = sLR_{n+m}(x+y). \quad (16)$$

The proof follows directly from the definition of likelihood. Equivalently, adding new data to already observed data involves taking the sLR from the initial data, then multiplying that by the sLR based on the new data.

The second feature concerns the EV of the sLR. It is well known that if the hypothesis in the denominator is true, EV equals unity. In contrast, if the numerator's hypothesis is true, EV exceeds unity (proof in Appendix A, The Simple LR). Thus, we have:

$$E[sLR_n(x)] \begin{cases} = 1 \text{ if } H_0 \text{ true} \\ > 1 \text{ if } H_1 \text{ true} \end{cases}. \quad (17)$$

These results in (17) seem somewhat bizarre; one would ‘want’ the first EV to be less than unity, not equal unity. In other words, if H_0 is true (i.e., there is no linkage), the EV of the sLR remains unity, no matter how much data we collect.

Now we can derive the CondEV we are interested in, i.e., the EV of the sLR after $n+m$ observations, given the value of the sLR after the first n observations. The results in (16) and (17) together imply:

$$E[sLR_{n+m}(x)|sLR_n(x) = A] \begin{cases} = A \text{ if } H_0 \text{ true} \\ > A \text{ if } H_1 \text{ true} \end{cases} \quad (18)$$

(proof in Appendix A, The Simple LR). Note that this result holds for all $m \geq 1$.

In other words, the sLR displays ‘half’ ExpM, in that when the hypothesis in the numerator is true, the conditional EV of the sLR is greater than the previous observed value of the sLR, as we would want. However, when the hypothesis in the denominator is true, this conditional EV *equals* the previously observed value, rather than decreasing, no matter how many additional observations we make. Moreover, if we interchange numerator and denominator in the sLR, the relationships in (17) are reversed, which seems no less bizarre.

In contrast, the log of the sLR behaves more symmetrically and intuitively:

$$E[\log sLR_n(x)] \begin{cases} < 0 \text{ if } H_0 \text{ true} \\ > 0 \text{ if } H_1 \text{ true} \end{cases} \quad (19)$$

(proof in Appendix A, The Simple LR). In other words, a simple monotonic transformation can greatly affect the behavior of EVs; also see section 4.1. below for further discussion of this point. From (19) it follows that

$$E[\log sLR_{n+m}(x) | \log sLR_n(x) = A] \begin{cases} < A \text{ if } H_0 \text{ true} \\ > A \text{ if } H_1 \text{ true} \end{cases} \quad (20)$$

(proof in Appendix A, The Simple LR). Thus, the log of the sLR, in contrast to the sLR itself, fully displays ExpM. This result also holds for all $m \geq 1$.

3.2. Does the Maximized LR Display ExpM?

In this section we show that the maxLR in section 2.2.2. does not display ExpM, and neither does its log. In fact, for the maxLR, the ConDEV is always greater than the initial value, even when the true $\theta = 0.5$. Thus, the maxLR’s behavior is even worse than the ‘half’ ExpM we saw with the sLR: not only does the ConDEV not decrease when the null hypothesis is true, but it actually increases. We can see this even when we add just one new observation, i.e.,

$$E_\theta[\max LR_{n+1}(x) | \max LR_n = A] > A \text{ for all } \theta \leq 0.5. \quad (21)$$

Proof is in Appendix B (The Maximized LR). In other words, the ConDEV always goes *up*, even when there is no linkage, which violates expected monotonicity. (The reader might think this is obvious, since the maxLR is always ≥ 1 , but note it is not true for the log maxLR below.)

We also examined the log of the maxLR to see if it was ‘better’ behaved, as the log of the simple LR was better behaved than the sLR itself. For this task, it was useful to use the concept of the θ_{cutoff} (section 2.3.). We determined

Table 2. θ_{cutoff} values for selected values of n and x , for log maxLR, PPL and log intLR (log $R_n(x)$)

n	x	log maxLR ^a	PPL, with $\rho = 0.98^b$	log intLR ^b
5	0	0.256	0.436	0.296
5	1	0.441	0.491	0.368
5	2	<u>0.704</u>	0.498	0.411
10	0	0.207	0.170	0.253
10	1	0.318	0.408	0.316
10	2	0.392	0.480	0.363
10	3	0.466	0.495	0.399
10	4	<u>0.577</u>	0.498	0.425
20	0	0.172	0.046	0.212
20	1	0.249	0.095	0.260
20	7	0.468	0.496	0.421
20	8	<u>0.514</u>	0.487	0.436

^a For the log maxLR, values of $\theta_{cutoff} > 0.5$ are underlined.

^b For the PPL and the log intLR, θ_{cutoff} always < 0.5 .

that for the log maxLR, θ_{cutoff} is sometimes < 0.5 , sometimes > 0.5 (see Appendix B, The Maximized LR, section B.2.). So, log maxLR does not display ExpM either, and we cannot even say consistently whether the ConDEV goes up or down when there is no linkage. Note that log max LR is the usual form of the maximum lod score, hence this particular result is perhaps the most immediately relevant to ordinary linkage analysis. We therefore consider it in more detail:

Table 2 (first column of results) shows some examples of θ_{cutoff} values for log maxLR as functions of n and x . For example, in an initial sample of 10 FIGs, of which none are recombinant ($n = 10, x = 0$), θ_{cutoff} equals 0.207. This means that the ConDEV increases if the true θ is less than about 0.207, which is good for ExpM. However, the ConDEV decreases, not only when there is no linkage ($\theta = 0.5$), but also for values of θ that do represent linkage (e.g. $\theta = 0.21$). Thus, even in this very simple case, we can see how the winner’s curse might lead the maximum lod score to diminish when adding new data to an original data set, even when the linkage is real.

On the other hand, if in that same initial sample of $n = 10$, there are 4 recombinants, then $\theta_{cutoff} = 0.577$, which is greater than 0.5; i.e., the ConDEV increases, even if the true value of θ is 0.5 – similar to the behavior we saw for the maxLR itself.

Apparently, θ_{cutoff} when less than 0.5 represents some kind of ‘midpoint’ between the null hypothesis ($\theta = 0.5$)

Table 3. Values of $R_n(x)$ for $n = 1$ to 16, $x = 0$ to n

$n \backslash x$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1.500	0.5000														
2	2.333	0.6667	0.3333													
3	3.750	0.9167	0.4167	0.2500												
4	6.200	1.300	0.5333	0.3000	0.2000											
5	10.50	1.900	0.7000	0.3667	0.2333	0.1667										
6	18.14	2.857	0.9429	0.4571	0.2762	0.1905	0.1429									
7	31.87	4.411	1.304	0.5821	0.3321	0.2202	0.1607	0.1250								
8	56.78	6.972	1.849	0.7579	0.4063	0.2579	0.1825	0.1389	0.1111							
9	102.3	11.26	2.689	1.009	0.5063	0.3063	0.2095	0.1556	0.1222	0.1000						
10	186.1	18.51	4.002	1.376	0.6433	0.3694	0.2433	0.1758	0.1354	0.1091	0.0909					
11	341.2	30.93	6.086	1.918	0.8338	0.4527	0.2861	0.2005	0.1510	0.1197	0.0985	0.0833				
12	630.1	52.42	9.441	2.732	1.103	0.5645	0.3410	0.2312	0.1699	0.1322	0.1072	0.0897	0.0769			
13	1,170	89.94	14.91	3.974	1.490	0.7166	0.4124	0.2696	0.1928	0.1470	0.1174	0.0971	0.0824	0.0714		
14	2,184	156.0	23.92	5.896	2.053	0.9265	0.5066	0.3183	0.2209	0.1646	0.1293	0.1055	0.0886	0.0762	0.0667	
15	4,096	273.0	38.93	8.906	2.885	1.221	0.6324	0.3808	0.2558	0.1860	0.1433	0.1152	0.0957	0.0815	0.0708	0.0625

Shading separates R values ≥ 1 from those < 1 .

and the ‘observed’ value of θ (i.e., the value of θ supported by the observations in the initial dataset). Consider the example above with $n = 10$ and $x = 0$, so $\theta_{\text{cutoff}} = 0.207$. The initial dataset supports a low value of θ (in fact, the MLE would be zero). Apparently, if the true value of θ is anything over 0.207, the ConDEV of log maxLR views that value as being ‘closer’ to 0.5 than to zero, so it lowers the log maxLR from its initial value. On the other hand, for $n = 10$ and $x = 4$, θ_{cutoff} is > 0.5 . Now the initial dataset supports a high value of θ (MLE is 0.4), so apparently, no matter what the true value of θ , the ConDEV views additional data as supporting linkage.

3.3. Do the PPL and the Integrated LR Display ExpM?

The intLR, $R_n(x)$, functions as the ‘building block’ of the PPL, and for this simple case we are considering, the intLR contains all the information in the data, as noted in section 2.2.3. We begin (in section 3.3.1.) by showing a useful property of the intLR. Only then do we tackle the question of ExpM for the PPL (section 3.3.2.) and for the intLR (section 3.3.3.).

3.3.1. A Property of the Integrated LR

We denote the intLR by $R_n(x)$, as defined in (8). A useful property is that when sample size is increased by one

observation, the amount by which $R_n(x)$ goes up if the additional observation is a nonrecombinant equals the amount by which $R_n(x)$ goes down if the additional observation is a recombinant. That is, $R_n(x)$ is the mean of the two possible values of R_{n+1} :

$$R_n(x) = \frac{R_{n+1}(x) + R_{n+1}(x+1)}{2}. \quad (22)$$

Below we will use Δ to indicate that amount by which $R_n(x)$ goes up or down; i.e.,

$$\Delta \equiv R_{n+1}(x) - R_n(x) = R_n(x) - R_{n+1}(x+1), \quad (23)$$

where Δ is positive. Proof of (22) is in Appendix D, The Integrated LR, section D.2.

Table 3 shows some representative $R_n(x)$ values. Any given $R_n(x)$ value can be converted to a posterior probability as a function of the chosen prior probability of non-linkage, ρ . For example, with $\rho = 0.98$ (i.e., prior probability of linkage taken to be 2%), a dataset of 10 FIGs and zero recombinants yields $R_{10}(0) = 186.1$, and this in turn yields a PPL of

$$\frac{(0.02)(186.1)}{(0.02)(186.1) + (0.98)} = 0.79,$$

i.e., much higher than the prior probability. On the other hand, if the 10 FIGs include four recombinants, $R_{10}(4) = 0.6433$, so the PPL becomes

$$\frac{(0.02)(0.6433)}{(0.02)(0.6433) + (0.98)} = 0.013,$$

i.e., lower than the prior probability.

3.3.2. ConDEV of the PPL

We begin with the ConDEV of the PPL itself. For convenience, we express the PPL in terms of the intLR, $R_n(x)$, as in (7). Let A denote the initial value of $R_n(x)$, and B the corresponding initial value of the PPL, i.e.,

$$B = \frac{(1-\rho)A}{(1-\rho)A + \rho}.$$

Thus,

$$\text{ConDEV} = E_\theta[\text{PPL}_{n+1}(x) \mid \text{PPL}_n(x) = B].$$

Again, as with the maxLR, we use θ_{cutoff} . In this situation, we can solve for θ_{cutoff} explicitly:

$$\theta_{\text{cutoff}} = 0.5 - \frac{(1-\rho)\Delta}{2[(1-\rho)A + \rho]}, \quad (24)$$

where A is the initial value of $R_n(x)$, ρ is the prior probability of ‘no linkage,’ and Δ is the amount by which $R_n(x)$ goes up if the next single observation is a nonrecombinant, or goes down if the next single observation is a recombinant (see (23) above). Derivation of (24) is in Appendix C (The Posterior Probability of Linkage).

Unlike the situation for maxLR and log maxLR, here for the PPL, θ_{cutoff} is *always* less than 0.5, even if all n of the initial observations were recombinant. (For example, for $n = 10$, $x = 10$, $\theta_{\text{cutoff}} = 0.4999$.) Table 2 (second column of results) shows some sample θ_{cutoff} values for the PPL.

3.3.3. ConDEV of the Integrated LR, $R_n(x)$

Since the intLR functions as the ‘building block’ of the PPL, and since the intLR contains all the information in the data for this simple case we are considering, we also examined its behavior alone: does $R_n(x)$ display ExpM? Interestingly, we observed a peculiarity analogous to that observed with the simple LR, as in (18): when the true value of θ is less than 0.5 – i.e., when there is linkage – then the ConDEV of $R_n(x)$ is greater than the previous observed value, as we would want under ExpM. However, when the true state of nature is no linkage ($\theta = 0.5$), this ConDEV does not go down but *equals* the previous observed value:

$$E[R_{n+m}(x) \mid R_n(x) = A] \begin{cases} = A & \text{if } \theta = 0.5 \\ > A & \text{if } \theta < 0.5 \end{cases} \quad (25)$$

Moreover, this result in (25) holds for all m , not just $m = 1$. That is, if there is no linkage, the ConDEV never goes down, no matter how many additional observations we make. Proof of (25) is given in Appendix D (The Integrated LR, in section D.3.).

Finally, we examined the log of the intLR. Here, as with the PPL, we demonstrated that there is a cutoff value, $0 < \theta_{\text{cutoff}} < 0.5$, such that when the true value of θ is less than that cutoff, the ConDEV increases after one observation, and when the true value is greater, the ConDEV decreases. See section D.4. of Appendix D (The Integrated LR). That is, the ConDEV does go down when the true value of θ is 0.5 (no linkage), but it also goes down for a range of values of $\theta < 0.5$. Table 2 (third column of results) shows some sample values.

4. Rethinking Monotonicity

4.1. Thoughts about Expected Values

These preliminary results revealed peculiarities of EVs, leading to the conclusion that ‘expected monotonicity’ in its original formulation is not an appropriate criterion for an evidence measure. On one level, EVs seem appealing, corresponding as they do to the commonsense concept of ‘average.’ As Vieland [6] wrote, ‘Thus, the key property is really monotonic increase (or decrease) of the value of the measurement *on average*...’ (italics hers; p 148, left). However, EVs are actually ill-behaved quantities in several respects. For example, there is the well-known principle that, ‘The expected value of the function does not equal the function of the expected value’ (except when the function is linear). If this were not the case, all variances would equal zero, because $E[X^2]$ would always equal $E^2[X]$. It also follows more or less directly from this axiom that for f , a monotonic increasing function, $E[X] < E[Y]$ does not necessarily ensure $E[f(X)] < E[f(Y)]$. (Counterexample: let $X = 4$ or 6 , with equal probabilities $1/2$, and $Y = 1$ or 10 , with equal probabilities $1/2$. Then $E[X] = 5.0$, which is less than $E[Y] = 5.5$, but $E[\log_{10}(X)] \approx 0.69$, which is greater than $E[\log_{10}(Y)] = 0.5$.)

These counterintuitive facts arise directly from the nature of EVs: ordinarily, one thinks of the EV as weighting the $f(X)$ by the probabilities of each X . However, we suggest instead viewing the situation the other way around, and thinking of the EV as weighting the *probabilities* by the $f(X)$ values. Then one can see why weighting the prob-

abilities by, e.g., $\log(X)$ will have a very different effect than weighting by X^2 , since the former function ‘flattens’ X , whereas the latter does the opposite.

Equations (17) and (19) highlight several peculiarities of the sLR: (a) When H_1 is true, $E[L(H_1)/L(H_0)] > 1$, but when H_0 is true, $E[L(H_1)/L(H_0)] = 1$ and it never decreases, even though $L(H_1)/L(H_0)$ does decrease as $n \rightarrow \infty$. (b) If we invert the numerator and denominator in the sLR, we observe the exact opposite behavior. (c) In contrast, if we examine a straightforward monotonically increasing transform of the LR, namely the log LR, then the EV does behave as we would intuit; i.e., it is < 0 when H_0 is true and > 0 when H_1 is true.

Thus, focusing on EVs can lead to the awkward situation where whether a statistic has a desirable property varies, depending on whether we work with the statistic in original form or with its log or with its inverse, etc. It seems intuitively appealing to consider the average behavior of evidence measures, since evidence is measured in the context of stochastic variation. But at least in the current setting, this approach proves to be mathematically problematic.

4.2. Broadening the Search: Evidential Consistency

Expected monotonicity does not seem to be working, so what alternative criterion might we try instead? Recall that all of the measures considered here have the nice property exemplified in (11), namely, that when we add a single nonrecombinant (or recombinant) event, each measure goes up (or down) appropriately. This too represents a type of monotonicity, but without EVs. Perhaps we should broaden our search and think in more general terms of ‘evidential consistency.’ One possible alternative formulation: let us bypass EVs altogether and extend this simpler notion of consistency embodied in (11) to the case of general m . Here is the generalization of (11):

‘If data are incremented by new data that, taken alone, ‘support’ one hypothesis over the other, the statistic T should rise or fall accordingly.’ (26)

According to this property, if the evidence measure is $T_n(x)$ for our initial dataset, and if the evidence measure $T_m(y)$ for m new observations supports either H_1 or H_0 , then the combined evidence for the updated dataset, $T_{n+m}(x+y)$, should be accordingly larger or smaller than the initial $T_n(x)$. This property is intuitively appealing; it is also deterministic and avoids EVs. Moreover, it is invariant to monotonically increasing transformations: in particular, if the property holds for some proposed evidence measure, it also holds for the log of that measure.

Property (26) leads to three conditions. Think of two datasets, one with n observations and x recombinants, and the other with m observations and y recombinants, with respective evidence measures $T_n(x)$ and $T_m(y)$, where higher values support linkage and lower values support nonlinkage.

- Condition 1: if both evidence measures support linkage, then the evidence measure for the combined dataset should be larger than either of the two initial measures: $T_{m+n}(x+y) > \max\{T_n(x), T_m(y)\}$.
- Condition 2: if both support nonlinkage, then the combined measure should be less than either of the two initial ones: $T_{m+n}(x+y) < \min\{T_n(x), T_m(y)\}$.
- Condition 3: if one supports nonlinkage (w.l.o.g. say this is the first one) and the other supports linkage, then the combined measure should fall between the two initial ones: $T_n(x) < T_{m+n}(x+y) < T_m(y)$.

4.2.1. Simple LR

It is straightforward to show that the sLR does satisfy (26): $LR > 1$ implies support for H_1 and $LR < 1$ implies support for H_0 . If new data, considered alone, contribute a value of the LR that is greater than unity, then by Factorization (16), the new value of the total LR will be greater than the previous one; and vice-versa if the new data yield a LR less than unity. QED. The same is true of the log of the sLR, since as mentioned above, (26) is invariant to monotonically increasing transformations.

4.2.2. Maximized LR

By contrast, the maxLR does not satisfy (26). In fact, it is arguable whether the maxLR will satisfy any reasonable formulation of evidential consistency, since, as mentioned, this statistic is always ≥ 1 . For example, let the initial sample consist of 10 observations, of which one is recombinant; thus,

$$\max LR = \frac{1^1 9^9}{5^{10}} = 39.67.$$

Three more observations are made, of which one is recombinant; these new observations taken alone yield

$$\max LR = \frac{1^1 2^2}{(1.5)^3} = 1.185.$$

Arguably, since in both samples the proportion of recombinants is $< 1/2$, it seems appropriate to consider them both as supporting H_1 . However, the updated dataset now consists of 13 observations, of which 2 are recombinant, yielding updated

$$\max LR = \frac{2^2 11^{11}}{(6.5)^{13}} = 30.87,$$

which is *less* than the maxLR for the initial dataset. The same pattern occurs with log maxLR, a one-to-one monotonic increasing transformation. (These outcomes occur despite the fact that the ConDEV of the maxLR, after one additional observation, always increases, as does the ConDEV of the log maxLR for some but not all true values of θ , as we saw in section 3.2.).

4.2.3. PPL and Integrated LR

We begin with the intLR, $R_n(x)$. Evidential consistency does not hold here, either. (But see section 5.2. in the Discussion for an important caveat regarding the use of sequential updating in the PPL framework.) Here, too, $R > 1$ supports H_1 and $R < 1$ supports H_0 (see section 2.2.3.). From table 3, here are examples violating each of the three conditions:

- Violates Condition 1 above: $R_4(0) = 6.20$ and $R_{10}(3) = 1.38$, but $R_{14}(3) = 5.90$, which is not greater than the maximum of the two initial values.
- Violates Condition 2 above: $R_3(1) = 0.92$ and $R_4(2) = 0.53$, but $R_7(3) = 0.58$, which is not less than the minimum of the two initial values.
- Violates Condition 3 above: $R_{12}(4) = 1.10$ and $R_3(1) = 0.92$, but $R_{15}(5) = 1.22$, which does not fall between the two values.

The same pattern occurs with the PPL itself, as well as with $\log(R)$, since both are one-to-one monotonically increasing transformations of R .

Thus, we can have two data subsets, each of which supports the alternative hypothesis, yet when combined they provide less support than one of the datasets alone (violates Condition 1); we can have two datasets, each of which supports the null hypothesis, but taken together they provide less support than one of the datasets alone (violates Condition 2), etc. Put another way, we can start with a dataset that supports one hypothesis, collect a second dataset that supports that same hypothesis, yet find that our combined support for that hypothesis from the two datasets has gone *down* from what we had from the first dataset alone.

5. Discussion

5.1. Summary of Results

We began with a property – ExpM – posited by Vieland [6] as a requirement for any well-behaved measure of statistical evidence. ExpM requires that the ConDEV of the evidence measure should go up or down appropriately, depending on the true state of nature. We have investi-

gated the behavior of the ConDEV for several different possible evidence measures: the sLR, the maxLR, and the PPL, along with its closely related quantity, the intLR, $R_n(x)$. We determined that ExpM is *not* a feature of any of the LR-based evidence measures we considered (except for the log of the sLR). Table 4 summarizes those results.

Indeed, EVs are sufficiently peculiar in their behavior that it seems clear in retrospect that basing a requirement on them is a non-starter.

Nevertheless, it seems obvious that a good evidence measure should display some kind of consistency, in the sense of going up as more evidence accumulates for one hypothesis, and down as more accumulates for the other, even in settings such as those giving rise to the winner's curse. In this connection, we also considered as a potential alternative the property in (26), in section 4.2., which states simply that if new data taken alone support H_1 , then the evidence in the updated data should be higher than it was before the new data were added, and vice-versa for H_0 . This proposed property has the advantage of being deterministic and not involving EVs with all their vagaries. Notably, we found that whereas the sLR and its log do satisfy this property, neither the maxLR nor the PPL (and intLR) as defined above, nor their logs, satisfy it.

5.2. Updating the PPL

Extending our discussion from expected monotonicity to the more general concept of evidential consistency raises an important point in connection with the PPL. Recall that in section 2.2.3., we explicitly assumed that when we consider new data, we compute the PPL by doing the following: we combine (pool) the initial n observations and the new m observations; we recompute (re-integrate) the intLR; and we use this new intLR as the basis for transformation onto the PPL scale. As originally described [14], this is the way in which the ordinary two-point PPL is computed over two sets of data, so it was appropriate to examine this method in the context of the one-sided binomial distribution considered in this paper. We refer to this as updating by 'pooling and re-integrating.' It is analogous to what is standard practice for the sLR (3) and the maxLR (5).

However, there are situations in which the updating process is handled differently. Specifically, the multipoint PPL is not computed this way; nor are non-location parameters (especially, parameters of the trait model) handled in this way within the actual PPL framework. Rather, in order to compute the multipoint PPL, the LR at each genomic position is integrated over all (trait) parameters of the model, yielding a single value of the intLR, $R_n(x)$.

Table 4. Summary of results for conditional expected values

In paper	Evidence measure	Results for	
		measure itself	log of measure
Sec. 2.2.1., Sec. 3.1., App. A	sLR	CondeV = A for $\theta = 0.5$ > A for $\theta < 0.5$ (for all m)	CondeV < A for H_0 > A for H_1 (for all m)
Sec. 2.2.2., Sec. 3.2., App. B	maxLR	CondeV > A for all θ (for $m = 1$)	for some datasets, $\exists \theta_{cutoff} \in (0, 0.5) \ni$ CondeV < A for $\theta < \theta_{cutoff}$ > A for $\theta > \theta_{cutoff}$ for other datasets, CondeV > A for all θ (for $m = 1$)
Sec. 2.2.3., Sec. 3.3., App. C	PPL	$\exists \theta_{cutoff} \in (0, 0.5) \ni$ CondeV < B for $\theta < \theta_{cutoff}$ > B for $\theta > \theta_{cutoff}$ (for $m = 1$)	–
Sec. 2.2.3., Sec. 3.3., App. D	intLR	CondeV = A for $\theta = 0.5$ > A for $\theta < 0.5$ (for all m)	$\exists \theta_{cutoff} \in (0, 0.5) \ni$ CondeV < log A for $\theta < \theta_{cutoff}$ > log A for $\theta > \theta_{cutoff}$ (for $m = 1$)

A denotes the initial value of that evidence measure, except for PPL and log intLR, where A denotes the initial value of the intLR. The PPL's initial value is denoted by B . Sec. = Section; App. = Appendix.

This integration is performed separately in each subset of the data, and then the resultant R s are multiplied together. Thus, if we have n families in the initial dataset and m in the new dataset, the combined evidence across the two datasets at a given location is measured via $R_{n+m}(x+y) = R_n(x) \times R_m(y)$. This is updating by ‘multiplying.’ That is, rather than pooling and re-integrating, we simply multiply the two R s. Then the updated PPL is computed as in (10), substituting $R_{n+m}(x+y)$ into the formula, to replace $R_n(x)$.

It is trivial to see that this form of the PPL does indeed have the evidential consistency property in (26), since we know from section 2.2.3. that $R > 1$ means that the data support H_1 , whereas $R < 1$ means that the data support H_0 . Thus, when the new data support H_1 , R_m is greater than 1, and the PPL goes up; when the new data support H_0 , R_m is less than 1, and the PPL goes down. This property holds even in the settings giving rise to the winner's curse, in which we select a location for follow-up based on the size of the initial PPL.

The relationship between $R_n(x)$ and which is the better-supported hypothesis is closely connected to the corresponding relationship for sLR, but is not quite so intuitive.

Under the assumption of a uniform prior on θ , $R_n(x)$ can be viewed as the arithmetic average of all possible sLRs. In other words, for given data, $R_n(x)$ equals the sLR evaluated at that value of θ_1 which produces the average value of all possible sLRs.

This also helps explain why the CondeV of the intLR behaves, for better and for worse, so much like the CondeV of the sLR; see (18) and (25): $R_n(x)$ is in effect an sLR, albeit one in which the corresponding value of θ_1 is dictated by the data. (The maxLR also has the property that its θ_1 is dictated by the data; it is the use of maximization, rather than integration, to obtain θ_1 , that leads to the maxLR's problematic properties.) It has been argued elsewhere that *only* the sLR has the requisite properties to serve as a measure of statistical evidence [1–3, 5, 7]. But the sLR has the obvious limitation that in the presence of any complexity whatsoever it requires forcing all parameters to specific (and in general, presumably wrong) values, with no way to measure the evidence independently of the assumed values. Insofar as the intLR allows us to handle unknown parameters in an effective manner while still maintaining key properties of the sLR, it emerges as a compelling alternative to the sLR in this context.

5.3. Conclusions and Future Work

By focusing on one very simple statistical model, we have tried to illustrate fundamental issues with some familiar approaches to evidence measurement. More complex models are expected to raise more, not fewer, issues, and we see no reason to believe that the problems we have pointed out above could be resolved simply by moving to, e.g., higher-dimensional likelihoods or continuous distributions. Thus, without making any attempt to extrapolate to other models our specific findings regarding the behavior of LR_n in the context of the one-sided binomial model, we believe that our results nevertheless provide a key starting point for consideration of more general applications. It is noteworthy also in this context that the underlying issue (non-monotonicity) examined here has already been shown to arise in connection with so-called model-free linkage statistics [9] and with both variance-components and regression-based quantitative trait linkage statistics [16], and it is easily seen that exactly the same problem arises in the context of association studies.

We conclude with three take-home messages:

(1) In our search for an evidence measure, we cannot necessarily trust our ideas of what look like ‘reasonable’ properties to require of a measure. For example, it seems ‘reasonable’ that any valid evidence measure should satisfy the property in (11), that the evidence measure should decrease or increase, if the next observation is a recombinant or nonrecombinant, respectively. And indeed all the evidence measures examined here do satisfy that property. However, it also seems ‘reasonable’ that this property should be generalizable, i.e., that if we add new data of more than a single observation that, taken alone, support one hypothesis or the other, then the updated evidence measure should accordingly rise or fall. Yet, this is precisely the consistency property we proposed in (26), and we have seen in section 4.2. that this property is not so straightforward after all.

(2) The intLR has intriguing properties vis-à-vis the sLR, and the ‘multipoint’ form of the standard PPL may serve as a template for further research into evidence measures having the consistency property in (26).

(3) The topic remains topical! These questions are important because the phenomenon of strong evidence melting away with the accumulation of new data represents an ongoing critical problem for investigators in this age of GWAS and genome-wide linkage analyses. Also, more generally, with the accumulation of enormous quantities of genetic data from the Human Genome Project, from microarrays and chips, and potentially soon

from whole-genome sequencing, it is more important than ever to determine exactly what statistical evidence is and how we ought to be measuring it.

Acknowledgments

This work was supported in part by NIH grants DK-31813 and MH-48858 (S.E.H.) and MH086117 (V.J.V.) and by the New York State Psychiatric Institute.

Appendix A. The Simple LR

A.1. Expected Value of Simple LR

Here we show (17) in the text:

$$E[LR_n(x)] \begin{cases} = 1 & \text{if } H_0 \text{ true} \\ > 1 & \text{if } H_1 \text{ true} \end{cases} \quad (\text{A.1})$$

Part 1. Recapitulate the well-known proof that

$$\begin{aligned} E_0 \left[\frac{L_1(\theta)}{L_0(\theta)} \right] &= 1: \\ E_0 \left[\frac{L_1(\theta)}{L_0(\theta)} \right] &= \int_{\text{all } x} \frac{L_1(\theta)}{L_0(\theta)} f_0(x; \theta) dx = \int_{\text{all } x} \frac{f_1(x; \theta)}{f_0(x; \theta)} f_0(x; \theta) dx \\ &= \int_{\text{all } x} f_1(x; \theta) dx = 1, \end{aligned}$$

where $f_i(x; \theta)$ represents the probability density function of x when H_i is true. For a discrete p.d.f., the proof is identical; just replace the integral signs with summation signs.

Part 2. Show that

$$E_1 \left[\frac{L_1(\theta)}{L_0(\theta)} \right] > 1:$$

For the purposes of this proof, let Y denote

$$\frac{L_0(\theta)}{L_1(\theta)}.$$

Rewrite the EV from Part 1, but interchange subscripts 0 and 1, to yield

$$E_1(Y) = E_1 \left[\frac{L_0(\theta)}{L_1(\theta)} \right] = 1. \quad (\text{A.2})$$

From Jensen’s Inequality, $E[g(Y)] > g(E[Y])$ for a convex function $g(y)$, i.e., a function whose second derivative is positive. The function $g(y) = 1/y$ is such a function; therefore,

$$E[1/Y] > \frac{1}{E[Y]}.$$

In our case, $E_1[Y] = 1$, from (A.2), and therefore,

$$E_1[1/Y] = E_1 \left[\frac{L_1(\theta)}{L_0(\theta)} \right] > 1.$$

This proves (A.1) QED.

A.2. Conditional Expected Value of Simple LR

Here we show (18) in the text:

$$E[LR_{n+m}(x)|LR_n(x)=A] \begin{cases} = A & \text{if } H_0 \text{ true} \\ > A & \text{if } H_1 \text{ true} \end{cases}$$

From the Factorization property, (16) in the text,

$$E[LR_{n+m}(x+y) | LR_n(x)=A] = E[LR_n(x) \cdot LR_m(y) | LR_n(x)=A] = A \cdot E[LR_m(y)]. \quad (\text{A.3})$$

From (A.1) above, $E[LR_m(y)]$ is equal to or greater than unity, according as H_0 or H_1 is true, and therefore, the quantity in (A.3) is, correspondingly, equal to or greater than A . QED.

A.3. Expected Value of Log of Simple LR

Here we show (19) in the text:

$$E[\log LR_n(x)] \begin{cases} < 0 & \text{if } H_0 \text{ true} \\ > 0 & \text{if } H_1 \text{ true} \end{cases} \quad (\text{A.4})$$

For the purposes of this proof, let W denote

$$\frac{L_1(\theta)}{L_0(\theta)}.$$

From Jensen's Inequality, $E[g(W)] < g(E[W])$ for a concave function $g(w)$, i.e., a function whose second derivative is negative. The function $g(w) = \log w$ is such a function; therefore,

$$E_0[\log W] < \log E_0[W]. \quad (\text{A.5})$$

We know from (A.1) that

$$E_0[W] = E_0\left[\frac{L_1(\theta)}{L_0(\theta)}\right] = 1. \quad (\text{A.6})$$

From (A.5) and (A.6), $E_0[\log W] < \log E_0[W] = \log(1) = 0$. This proves that $E[\log LR_n(x)] < 0$ when H_0 is true. The result can also be written as

$$E[\log L_1(\theta) - \log L_0(\theta)] < 0, \text{ i.e., } E[\log L_1(\theta)] < E[\log L_0(\theta)] \text{ when } H_0 \text{ is true.} \quad (\text{A.7})$$

To prove that this ExpV is >0 when H_1 is true, simply interchange subscripts 1 and 0 in (A.7). QED.

A.4. Conditional Expected Value of Log of Simple LR

Here we show (20) in the text:

$$E[\log LR_{n+m}(x)|\log LR_n(x)=A] \begin{cases} < A & \text{if } H_0 \text{ true} \\ > A & \text{if } H_1 \text{ true} \end{cases} \quad (\text{A.8})$$

Proof: From the Factorization property, (16) in the text, $\log LR_{n+m}(x)$ can be broken down into a sum of its value for the first n observations plus its value for the second m observations. The ExpV of that first sum equals A , since that is what we are conditioning on. The ExpV of the second sum is independent of what happened in the first n observations and follows (A.4), i.e., is less than zero when H_0 is true, greater than zero when H_1 is true:

$$\begin{aligned} & E[\log LR_{n+m}(x)|\log LR_n(x)=A] \\ &= E[\log LR_n(x)|\log LR_n(x)=A] + E[\log LR_m(x)|\log LR_n(x)=A] \\ &= A + \begin{cases} < 0 & \text{if } H_0 \text{ true} \\ > 0 & \text{if } H_1 \text{ true} \end{cases} \end{aligned}$$

which proves (A.8). QED.

Appendix B. The Maximized LR

B.1. Conditional Expected Value of Maximized LR

Here we prove (21) in the text, i.e.,

$$E_\theta[\max LR_{n+1}(x) | \max LR_n = A] > A \text{ for all } \theta \leq 0.5. \quad (\text{B.1})$$

We begin with a dataset of n observations, of which x are recombinants. (We stipulate that

$$x \leq \frac{n-1}{2},$$

simply so that we do not have to deal with $\hat{\theta}$ being up against the 0.5 boundary, either in the initial observation or if the additional observation is a recombinant.)

From section 2.3. in the text, it is sufficient to show that the ConDEV evaluated at $\theta = 0.5$ is greater than the initial value of the maxLR. That is, from (4) and (13) in the text, we need to show

$$\begin{aligned} (0.5) & \left[\frac{2^{n+1} (x+1)^{x+1} (n-x)^{n-x}}{(n+1)^{n+1}} + \frac{2^{n+1} x^x (n+1-x)^{n+1-x}}{(n+1)^{n+1}} \right] \\ & > \frac{2^n}{n^n} x^x (n-x)^{n-x}. \end{aligned} \quad (\text{B.2})$$

Multiply both sides of (B.2) by the positive quantity

$$\frac{(n+1)^{n+1}}{2^n x^x (n-x)^{n-x}},$$

yielding:

$$\frac{(x+1)^{x+1}}{x^x} + \frac{(n+1-x)^{n+1-x}}{(n-x)^{n-x}} > \frac{(n+1)^{n+1}}{n^n}. \quad (\text{B.3})$$

Showing (B.3) will demonstrate (B.1). At this point, it is useful to define the following function $f(u)$:

$$f(u) = \frac{(u+1)^{u+1}}{u^u}, \quad u \geq 0, \quad (\text{B.4})$$

Note that $f(u)$ is monotonically increasing and is concave, i.e.,

$$\begin{aligned} f'(u) &> 0 \\ f''(u) &< 0. \end{aligned} \quad (\text{B.5})$$

(Details not shown.) Also note that

$$\lim_{u \rightarrow 0} f(u) = 1,$$

so define

$$f(0) = 1 \quad (\text{B.6})$$

to preserve continuity.

Thus, showing (B.3) is equivalent to showing

$$f(x) + f(n-x) > f(n). \quad (\text{B.7})$$

To prove (B.7), note that for $x_1 < x_2$ and ε any positive increment,

$$f(x_1 + \varepsilon) - f(x_1) > f(x_2 + \varepsilon) - f(x_2), \quad (\text{B.8})$$

from $f(x)$'s properties in (B.5); i.e., for the same increment ε , $f(x)$ decreases less as x increases. In (B.8), let $x_1 = 0$, $\varepsilon = x$, and $x_2 = n - x$, yielding

$$f(x) - f(0) > f(n) - f(n - x). \quad (\text{B.9})$$

Since $f(0) = 1$, from (B.6), (B.9) implies

$$f(x) - 1 > f(n) - f(n - x) \Rightarrow f(x) + f(n - x) > f(n) + 1 > f(n).$$

Thus,

$$f(x) + f(n - x) > f(n).$$

This proves (B.7) and therefore demonstrates (B.3), and therefore (B.1). QED.

B.2. Conditional Expected Value of log of Maximized LR

Here we prove that for log maxLR, the CondEV is sometimes less than, sometimes greater than the initial value of the statistic, as described in section 3.2. We investigate this using θ_{cutoff} . From (4) in the text,

$$\log \text{maxLR} = x \log x + (n - x) \log(n - x) + n \log 2 - n \log n,$$

and this is our 'A' in (14) in the text. From (14), then,

$$\theta_{\text{cutoff}} = \frac{\text{Num}}{\text{Den}},$$

where

$$\text{Num} = x \log x + (n + 1 - x) \log(n + 1 - x) + (n + 1) \log 2 - (n + 1) \log(n + 1) - [x \log x + (n - x) \log(n - x) + n \log 2 - n \log n].$$

This quantity can be re-expressed as $\text{Num} = \log f(n - x) - \log f(n) + \log 2$, where f is the function defined above in eq. (B.4). Similarly, the denominator can be expressed as $\text{Den} = \log f(n - x) - \log f(x)$. So

$$\theta_{\text{cutoff}} = \frac{\log f(n - x) - \log f(n) + \log 2}{\log f(n - x) - \log f(x)}. \quad (\text{B.10})$$

Numerical calculations of (B.10) reveal that θ_{cutoff} is sometimes less than 0.5, sometimes greater; see table 2, first column of results. Therefore, the CondEV of log maxLR is sometimes less than, sometimes greater than the initial value of the log maxLR, based on the reasoning in section 2.4. (In fact, θ_{cutoff} is always < 0.5 for $x = 0$ and then increases as x approaches $n/2$; details not shown.)

Appendix C. The Posterior Probability of Linkage

Here we show that for the PPL

$$\theta_{\text{cutoff}} = 0.5 - \frac{(1 - \rho)\Delta}{2[(1 - \rho)A + \rho]}, \quad (\text{C.1})$$

i.e., (24) in the text. Our initial dataset of n observations, of which x are recombinants, yields $R_n(x) = A$, so the initial value of the PPL is

$$\text{PPL}(n, x) = \frac{(1 - \rho)A}{(1 - \rho)A + \rho}.$$

Also note

$$R_{n+1}(x) = A + \Delta \text{ and } R_{n+1}(x + 1) = A - \Delta,$$

where $\Delta \equiv R_{n+1}(x) - R_n(x) = R_n(x) - R_{n+1}(x + 1)$, as in (23) in the text, and as proved below in section D.2. Thus, from (14) in the text,

$$\theta_{\text{cutoff}} = \frac{\text{PPL}_{n+1}(x) - \text{PPL}_n(x)}{\text{PPL}_{n+1}(x) - \text{PPL}_{n+1}(x + 1)} = \frac{\text{Num}}{\text{Den}}, \quad (\text{C.2})$$

where

$$\begin{aligned} \text{Num} &= \frac{(1 - \rho)(A + \Delta)}{(1 - \rho)(A + \Delta) + \rho} - \frac{(1 - \rho)A}{(1 - \rho)A + \rho} \\ &= \frac{(1 - \rho)\Delta\rho}{[(1 - \rho)(A + \Delta) + \rho] \cdot [(1 - \rho)A + \rho]} \end{aligned}$$

and

$$\begin{aligned} \text{Den} &= \frac{(1 - \rho)(A + \Delta)}{(1 - \rho)(A + \Delta) + \rho} - \frac{(1 - \rho)(A - \Delta)}{(1 - \rho)(A - \Delta) + \rho} \\ &= \frac{2(1 - \rho)\Delta\rho}{[(1 - \rho)(A + \Delta) + \rho] \cdot [(1 - \rho)(A - \Delta) + \rho]}. \end{aligned}$$

Substitute Num and Den back into (C.2) and cancel appropriate terms to yield (C.1). QED. (In fact, θ_{cutoff} is always between 0 and 0.5; details not shown.) Table 2, second column of results, gives sample values.

Appendix D. The Integrated LR

D.1. Formulation of $R_n(x)$

The intLR, $R_n(x)$, from (8) in the text, is:

$$R_n(x) = 2^{n+1} \int_{0.5 < \theta < 1} \theta^x (1 - \theta)^{n-x} d\theta = 2^{n+1} \int_{0.5 < \varphi \leq 1} (1 - \varphi)^x \varphi^{n-x} d\varphi. \quad (\text{D.1})$$

The change of variables from θ to $\varphi = 1 - \theta$ makes the integral easier to work with, since x is generally less than $n/2$.

For $x = 0$ and $x = 1$, $R_n(x)$ becomes, respectively,

$$R_n(0) = 2^{n+1} \int_{0.5 < \varphi \leq 1} \varphi^n d\varphi = \frac{2^{n+1} - 1}{n + 1}$$

and

$$R_n(1) = 2^{n+1} \int_{0.5 < \varphi \leq 1} (1 - \varphi)\varphi^{n-1} d\varphi = 2 \cdot \frac{2^n - 1}{n} - R_n(0), \quad \forall n \geq 1. \quad (\text{D.2})$$

For general x , the integral in (D.1) is the Incomplete Beta function. To evaluate it, substitute the Binomial Expansion,

$$(1 - \varphi)^x = \sum_{i=0}^x (-1)^i \binom{x}{i} \varphi^i,$$

into (D.1):

$$\begin{aligned} R_n(x) &= 2^{n+1} \int_{0.5 < \varphi \leq 1} (1 - \varphi)^x \varphi^{n-x} d\varphi = 2^{n+1} \sum_{i=0}^x (-1)^i \binom{x}{i} \cdot \int_{0.5 < \varphi \leq 1} \varphi^{i+n-x} d\varphi \\ &= 2^{n+1} \sum_{i=0}^x (-1)^i \binom{x}{i} \frac{1 - (0.5)^{i+n-x+1}}{n - x + 1 + i} = \sum_{i=0}^x (-1)^i \binom{x}{i} 2^{x-i} \frac{2^{n-x+1+i} - 1}{n - x + 1 + i}. \end{aligned} \quad (\text{D.3})$$

D.2. Recursive Calculation of $R_n(x)$

As an alternative to calculating $R_n(x)$ directly from (D.3), one can calculate it recursively, based on this simple relationship (this is (22) in the text):

$$(0.5)[R_{n+1}(x) + R_{n+1}(x+1)] = R_n(x). \quad (D.4)$$

Proof: From the formulation of $R_n(x)$ in (8) or (D.1),

$$\begin{aligned} (0.5)[R_{n+1}(x) + R_{n+1}(x+1)] &= \\ 2^{n+1} \int_{0 \leq \theta < 0.5} \theta^x (1-\theta)^{n+1-x} d\theta + 2^{n+1} \int_{0 \leq \theta < 0.5} \theta^{x+1} (1-\theta)^{n-x} d\theta \\ &= 2^{n+1} \int_{0 \leq \theta < 0.5} \theta^x (1-\theta)^{n-x} [(1-\theta) + \theta] d\theta \\ &= 2^{n+1} \int_{0 \leq \theta < 0.5} \theta^x (1-\theta)^{n-x} d\theta = R_n(x). \end{aligned}$$

QED.

A consequence of (D.4) is:

$$R_n(x) = 2R_{n-1}(x-1) - R_n(x-1). \quad (D.5)$$

To calculate tables of R values, begin with $R_1(0) = 1.5$ and $R_1(1) = 0.5$. We also have all $R_n(0)$ values, from (D.2). From these and (D.5), all subsequent $R_n(x)$ values can be calculated recursively.

Another consequence of (D.4) is that R_{n+1} terms can be expressed as follows:

$$\begin{aligned} R_{n+1}(x) &= R_n(x) + \Delta \\ R_{n+1}(x+1) &= R_n(x) - \Delta, \text{ where } \Delta > 0. \end{aligned} \quad (D.6)$$

We used this in section 3.3.1. in the text, eq. (23), and in Appendix 3 above and will use it below in section D.4.

D.3. Conditional Expected Value of the Integrated LR, i.e., $R_n(x)$

Here we show (25) in the text:

$$\text{CondEV} = E_\theta [R_{n+m}(x) | R_n(x) = A] \begin{cases} = A & \text{if } \theta = 0.5 \\ > A & \text{if } \theta < 0.5 \end{cases} \quad (D.7)$$

and, moreover, the CondEV is monotonically decreasing in θ as θ goes from 0 to 0.5.

D.3.1. A Useful Identity

We will need the identity in (D.9) below. First, note that

$$\begin{aligned} E_\theta [R_{n+m}(x) | R_n(x) = A] &= E_\theta [R_{n+m}(x) | x_n = a] \\ &= \sum_{i=0}^m \binom{m}{i} (1-\theta)^{m-i} \theta^i R_{n+m}(a+i), \end{aligned} \quad (D.8)$$

where a represents the value of x_n that corresponds to $R_n(x) = A$. This is because after m additional observations, there are $m+1$ possible outcomes: if every observation is a nonrecombinant, the new value of R will be $R_{n+m}(a)$, and this happens with the probability $(1-\theta)^m$; if exactly one observation is a recombinant, the new value will be $R_{n+m}(a+1)$, and this can happen in m different ways, so the corresponding probability is $m(1-\theta)^{m-1}\theta$, and so on. Then (D.8) can be broken down into the weighted average of two sums of similar form but lower order:

$$\begin{aligned} E_\theta [R_{n+m}(x) | x_n = a] &= \\ &= (1-\theta) \sum_{i=0}^{m-1} \left\{ \binom{m-1}{i} (1-\theta)^{m-1-i} \theta^i R_{n+m}(a+i) \right\} \\ &+ \theta \sum_{i=0}^{m-1} \left\{ \binom{m-1}{i} (1-\theta)^{m-1-i} \theta^i R_{n+m}(a+1+i) \right\}. \end{aligned} \quad (D.9)$$

That is, the original sum from 0 to m can be expressed as a weighted average of two similar sums that go from 0 to $m-1$. Derivation of (D.9) uses the Pascal's Triangle identity

$$\binom{m}{i} = \binom{m-1}{i-1} + \binom{m-1}{i};$$

algebraic details not shown. We will use this identity in the proofs in sections D.3.2. and D.3.3. below.

D.3.2. Show First Part of (D.7) – What Happens when $\theta = 0.5$ Show

$$\begin{aligned} \text{CondEV} &= E_{\theta=0.5} [R_{n+m}(x) | R_n(x) = A] = A, \text{ i.e.,} \\ \text{show } E_{\theta=0.5} [R_{n+m}(x) | x_n = a] &= R_n(a). \end{aligned} \quad (D.10)$$

We show this by Induction. Again, a represents the value of x_n that corresponds to $R_n(x) = A$:

Part 1. Show (D.10) for $m = 1$.

Proof:

$$E_{\theta=0.5} [R_{n+1}(x) | x_n = a] = (0.5)[R_{n+1}(a) + R_{n+1}(a+1)] = R_n(a), \quad (D.11)$$

from (13) in the text and (D.4) above. QED.

Part 2. Show (D.10) for $m > 1$.

$$\text{Given: } E_{\theta=0.5} [R_{n+m-1}(x) | x_n = a] = R_n(a). \quad (D.12)$$

$$\text{Prove: } E_{\theta=0.5} [R_{n+m}(x) | x_n = a] = R_n(a). \quad (D.13)$$

Proof: In the identity (D.9), use the same reasoning as we used for (D.8): the first sum represents the CondEV of R after $n+m-1$ observations, given that the $(n+1)$ -th observation equaled a ; and the second sum represents the CondEV after the same number of observations, given that the $(n+1)$ -th observation equaled $a+1$. Thus (D.9) can be written as follows, with $\theta = 0.5$:

$$\begin{aligned} E_{\theta=0.5} [R_{n+m}(x) | x_n = a] &= (0.5)\{E_{\theta=0.5} [R_{n+m}(x) | x_{n+1} = a] \\ &+ E_{\theta=0.5} [R_{n+m}(x) | x_{n+1} = a+1]\}. \end{aligned} \quad (D.14)$$

Let ν represent $n+1$; then (D.14) can be rewritten

$$\begin{aligned} E_{\theta=0.5} [R_{n+m}(x) | x_n = a] &= (0.5)\{E_{\theta=0.5} [R_{\nu-1+m}(x) | x_\nu = a] \\ &+ E_{\theta=0.5} [R_{\nu-1+m}(x) | x_\nu = a+1]\}. \end{aligned} \quad (D.15)$$

By assumption (D.12), the first and second EVs on the RHS of (D.15) can be rewritten, respectively, as:

$$\begin{aligned} E_{\theta=0.5} [R_{\nu-1+m}(x) | x_\nu = a] &= R_\nu(a) = R_{n+1}(a) \\ E_{\theta=0.5} [R_{\nu-1+m}(x) | x_\nu = a+1] &= R_\nu(a+1) = R_{n+1}(a+1). \end{aligned} \quad (D.16)$$

Substitute the two expressions in (D.16) back into (D.14):

$$E_{\theta=0.5} [R_{n+m}(x) | x_n = a] = (0.5)[R_{n+1}(a) + R_{n+1}(a+1)],$$

which equals $R_n(a) = A$, from (D.11). This proves (D.13). QED.

D.3.3. Show Second Part of (D.7) – What Happens when $\theta < 0.5$
Show

$$\text{CondEV} = E[R_{n+m}(x) | R_n(x) = A] > A \text{ if } \theta < 0.5, \quad (\text{D.17})$$

and this is a monotonically decreasing function of θ , as θ goes from 0 to 0.5. Again, a represents the value of x_n that corresponds to $R_n(x) = A$.

It is straightforward to show (D.17) for $m = 1$: from (D.11) we know $E[R_{n+1}(x) | R_n(x) = A] = A$ when $\theta = 0.5$, and from (12) in the text we know that the CondEV after a single observation is a monotonically decreasing linear function of θ . That proves (D.17) for $m = 1$. QED.

For $m > 1$, we turn to the identity in (D.9). For ease of notation, let C_1 and C_2 represent, respectively, the two sums, i.e.,

$$C_1 = \sum_{i=0}^{m-1} \left\{ \binom{m-1}{i} (1-\theta)^{m-1-i} \theta^i R_{n+m}(a+i) \right\}$$

and

$$C_2 = \sum_{i=0}^{m-1} \left\{ \binom{m-1}{i} (1-\theta)^{m-1-i} \theta^i R_{n+m}(a+1+i) \right\}$$

Then (D.9) can be rewritten:

$$E_\theta[R_{n+m}(x) | x_n = a] = (1-\theta)C_1 + \theta C_2 = C_1 - \theta(C_1 - C_2).$$

This is a linear function in θ , and all we need to do, in order to prove that it is monotonically decreasing, is to show that $C_1 - C_2$ is positive, where:

$$C_1 - C_2 = \sum_{i=0}^{m-1} \left\{ \binom{m-1}{i} (1-\theta)^{m-1-i} \theta^i [R_{n+m}(a+i) - R_{n+m}(a+1+i)] \right\}. \quad (\text{D.18})$$

The fact $R_{n+m}(a+1+i)$ is less than $R_{n+m}(a+i)$ follows from, for example, eq. (D.6). Therefore, each term in square brackets in (D.18) is positive, and therefore, the sum is positive. Thus, the expression in (D.17) is monotonically decreasing in θ , and since that expression equals A when $\theta = 0.5$, this proves (D.17). QED.

D.4. Conditional Expected Value of $\log R_n(x)$

We show that when there is no linkage ($\theta = 0.5$), the CondEV of $\log R$ after one observation is less than the initial value:

$$E_{\theta=0.5}[\log R_{n+1}(x) | \log R_n(x)] < \log R_n(x),$$

so that θ_{cutoff} is less than 0.5, as stated in section 3.3.3.

Proof:

$$\begin{aligned} E_{\theta=0.5}[\log R_{n+1}(x) | \log R_n(x)] \\ = (0.5)[\log R_{n+1}(x) + \log R_{n+1}(x+1)], \end{aligned} \quad (\text{D.19})$$

from (13) in the text. We want to show that the quantity in (D.19) is less than $\log A$, where A denotes the initial value of $R_n(x)$. From (D.6), (D.19) can be rewritten

$$\begin{aligned} E_{\theta=0.5}[\log R_{n+1}(x) | \log R_n(x)] &= (0.5)[\log(A+\Delta) + \log(A-\Delta)] \\ &= (0.5)\log[(A+\Delta)(A-\Delta)] = (0.5)\log(A^2 - \Delta^2). \end{aligned}$$

Clearly, $(0.5)\log(A^2 - \Delta^2) < (0.5)\log(A^2) = \log A$. QED.

As for the value of θ_{cutoff} , this is given by

$$\theta_{\text{cutoff}} = \frac{\log(A+\Delta) - \log A}{\log(A+\Delta) - \log(A-\Delta)}.$$

(In fact, θ_{cutoff} is always between 0 and 0.5; details not shown.) Table 2, third column of results, gives sample values.

References

- Hacking I: Logic of Statistical Inference. Cambridge, Cambridge University Press, 1965.
- Edwards AWF: Likelihood. Expanded Edition. Baltimore, Johns Hopkins Press, 1992.
- Royall R: Statistical Evidence: A Likelihood Paradigm. London, Chapman & Hall, 1997.
- Taper ML, Lele SR: The Nature of Scientific Evidence: Statistical, Philosophical, and Empirical Considerations. Chicago, University of Chicago Press, 2004.
- Vieland VJ, Hodge SE: Review of 'Statistical Evidence: A Likelihood Paradigm', by R Royall. Am J Hum Genet 1998;63:283–289.
- Vieland VJ: Thermometers: something for statistical geneticists to think about. Hum Hered 2006;61:144–156.
- Strug LJ, Hodge SE: An alternative foundation for the planning and evaluation of linkage analysis. I. Decoupling 'error probabilities' from 'measures of evidence'. Hum Hered 2006;61:166–188.
- Vieland VJ, Wang K, Huang J: Power to detect linkage based on multiple sets of data in the presence of locus heterogeneity: comparative evaluation of model-based linkage methods for affected sib pair data. Hum Hered 2001;51:199–208.
- Huang J, Vieland VJ: Comparison of 'model-free' and 'model-based' linkage statistics in the presence of locus heterogeneity: single data set and multiple data set applications. Hum Hered 2001;51:217–225.
- Morton NE: Sequential tests for the detection of linkage. Am J Hum Genet 1955;7:277–318.
- Suarez B, Hampe CL, van Eerdewegh P: Problems of replicating linkage claims in psychiatry; in Gershon ES, Cloninger CR (eds): Genetic Approaches to Mental Disorders. Washington, DC, American Psychiatric Press, 1994, pp 23–46.
- Gorroochurn P, Hodge SE, Heiman GA, Durner M, Greenberg DA: Non-replication in association studies: 'Pseudo-failures' to replicate? Genet Med 2007;9:325–331.
- Kidd KK, Ott J: Power and sample size in linkage studies (abstract). Cytogenet Cell Genet 1984;36:510–511.
- Vieland VJ: Bayesian linkage analysis, or: How I learned to stop worrying and love the posterior probability of linkage. Am J Hum Genet 1998;63:947–954.
- Kass RE, Raftery AE: Bayes factors. J Am Stat Assoc 1995;90:773–795.
- Bartlett CW, Vieland VJ: Accumulating quantitative trait linkage evidence across multiple datasets using the posterior probability of linkage. Genet Epi 2006;31:91–102.