HHS Public Access
Author manuscript
Postgrad Med J. Author manuscript; available in PMC 2019 October 01.

Published in final edited form as:
Postgrad Med J. 2018 June ; 94(1112): 364–366. doi:10.1136/postgradmedj-2018-135555.

Rule of 100: a litmus test for informationless diagnostic tests

Alek M Westover, **David Shapiro**, **Michael Brandon Westover**, **Matt T Bianchi**
Harvard Medical School, Massachusetts General Hospital, Boston, Massachusetts, USA

The most common measures of diagnostic test performance, sensitivity and specificity, involve two simplifying assumptions: that disease status is binary (present or absent), and test results are binary (positive or negative). Bayes' theorem tells us that these features must be combined with a third ingredient, the prior or pretest probability (PreTP) of disease, to update the disease probability following the observed test result. As we[1,2] and others[3,4] have written, readers and authors of medical literature often interpret test results in ways that violate Bayes' rule. Publications celebrating diagnostic tests that on close inspection perform near-chance level represent an important example of this interpretation risk.[5,6] In some cases, even frankly paradoxical combinations of sensitivity and specificity are reported, for example, with certain physical examination findings,[7] in which a 'positive' result lowers the disease probability, and 'negative' result increases it.

We recently pointed out a simple litmus test that can be used at a glance to recognise the potential for chance or paradoxical performance, which we term the 'rule of 100'.[6] Any test for which the sensitivity and specificity add to 100% ($Sens + Spec = 1$) does not modify the PreTP of disease, that is, the result (positive or negative) provides a post-test probability (PostTP), identical to the PreTP. Such tests therefore provide no information. Here is a three-line proof:

$$
\begin{aligned}
PostTP &= \frac{PreTP \cdot Sens}{PreTP \cdot Sens + (1 - PreTP) \cdot (1 - Spec)} \\
&= \frac{PreTP \cdot Sens}{PreTP \cdot Sens + (1 - PreTP) \cdot Sens} \\
&= \frac{PreTP \cdot Sens}{Sens} = PreTP
\end{aligned}
$$

The second line of the proof makes use of the fact that $Sens = 1 - Spec$ when $Sens + Spec = 1$. For those comfortable with combining sensitivity and specificity into the likelihood ratio (LR) value, the equivalent fact is that when non-zero values of sensitivity and specificity add to 100%, the LR value=1. Paradoxical results occur when sensitivity and specificity add to less than 100%, or equivalently when the positive LR (LR+) is less than 1, or the negative LR (LR−) is greater than 1.

Correspondence to Dr Matt T Bianchi; thebianchi@gmail.com.
Contributors AMW, DS, MBW and MTB all participated in planning the study, analysing the data and writing the manuscript. MBW submitted the study.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

The Rule of 100 applies no matter *how* sensitivity and specificity combine to produce a total of 100. Thus, it is a fallacy to celebrate performance of tests that have high sensitivity (but low specificity), or high specificity (but low sensitivity), when their sum is 100%. Such tests are no better than tests that have modest values of both sensitivity and specificity (eg, sensitivity=specificity=50%).

METHODS

We performed a PubMed search to identify abstracts reporting sensitivity and specificity of diagnostic tests. The specific search used was '*(((((((sensitivity) AND specificity) AND human) NOT genetic) NOT AUC) NOT ROC) NOT receiver operating curve) AND english [Language] AND 'predictive value' AND diagnostic*'; when run on 31 August 2017, this search returned 133 585 abstracts. We used simple natural language processing (NLP) to identify abstracts in which pairs of sensitivity and specificity were clearly reported. We did not attempt to identify all abstracts reporting sensitivity and specificity, thus the sample on which we report our results is a convenience sample. From the initial pool of 133 585 abstracts, our NLP rules identified 6322 sensitivity-specificity pairs from 5311 abstracts (some had more than one pair of values).

STATISTICAL ANALYSIS

We summarise the proportion of informationless abstracts in our sample using percentages. To provide an estimate of the precision of these percentages given our sample size, we provide 95% CIs (calculated using the normal approximation to the binomial distribution).

Figure 1 illustrates the distribution of the sum of these two values across the 6322 sensitivity-specificity pairs. Approximately 3.5% (95% CI 3% to 3.9%) of the abstracts had results that were at or near-chance level, defined as having a sum of ≤110%. We reviewed the 219 abstracts manually to assess whether authors recognised the low information provided in the reported tests. We prespecified three categories for scoring authors' interpretation of sensitivity and specificity values: recognised as poor performance (n=118; 54%), claimed as useful (n=41; 19%) or unclear/not described (n=60; 27%). Most LR values for tests meeting the <110% criteria are between 0.5 and 2.0. If we broaden our definition of 'informationless' to include any LR+<2, or any LR−>0.5, then n=769 abstracts (12%) meet one or both of these criteria. Figure 2 shows the mapping of sensitivity and specificity onto LR+ and LR− values to illustrate the relationship and to serve as a nomogram for assessing test performance in a simple visual manner. While any cut-off value proposed to define what is 'near' an LR value of 1 will be arbitrary, one can appreciate two key points from the LR landscape across a range of sensitivity and specificity values. First, the combinations that yield LR values at or near 1 can be easily visualised to help avoid the fallacy of valuing a test that satisfies the Rule of 100. Second, for certain extreme asymmetries of sensitivity and specificity, the figure helps to correctly interpret tests that have reasonable LR values despite having either sensitivity or specificity values that seem 'too low'.

Bayes' rule tells us that two tests with the same LR value will modify the disease probability equally, regardless of the combination of sensitivity and specificity values that yielded that

LR. In figure 2, consider an LR+ value of 6: this can be obtained by a test with sensitivity of 30% and specificity of 95%, or by one with sensitivity of 90% and specificity of 85%. One should not conclude that the first test had a sensitivity that was ‘too low’, at least with regard to the interpretation of a positive result. We could, by contrast, conclude it was ‘too low’ with respect to a negative result, because the LR− value is only 0.74.

We conclude that many published abstracts report diagnostic test performance that is essentially informationless, and this poor performance may go unrecognised. Our estimate represents a lower bound because our search was not exhaustive, and not all papers reporting sensitivity and specificity values at or near 100% list those values in the abstract.[8] Repeating this kind of analysis using full text of all published manuscripts would shed light on the full distribution of test performance results.

Fortunately, the Rule of 100 is easy to use, though full protection against informationless tests should incorporate the LR-based approach shown in figure 2. Another sure fire method to protect against such paradoxes is to evaluate test performance with receiver operating curves, where performance near chance is visually obvious. Finally, even for tests that have reasonable sensitivity and specificity (and LR) values, PreTP is always needed to provide context to any test result.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

REFERENCES

1. Westover MB, Westover KD, Bianchi MT. Significance testing as perverse probabilistic reasoning. BMC Med 2011;9:20. [PubMed: 21356064]
2. Bianchi MT. Evidence that home apnea testing does not follow AASM practice guidelines--or Bayes’ theorem. J Clin Sleep Med 2015;11:189. [PubMed: 25580609]
3. Phelps MA, Levitt MA. Pretest probability estimates: a pitfall to the clinical utility of evidence-based medicine? Acad Emerg Med 2004;11:692–4. [PubMed: 15175211]
4. Cahan A, Gilon D, Manor O, et al. Probabilistic reasoning and clinical decision-making: do doctors overestimate diagnostic probabilities? QJM 2003;96:763–9. [PubMed: 14500863]
5. Laratta CR, Tsai WH, Wick J, et al. Validity of administrative data for identification of obstructive sleep apnea. J Sleep Res 2017;26:132–8. [PubMed: 27761958]
6. Bianchi MT. Bayes’ theorem and the rule of 100: a commentary on ‘Validity of administrative data for identification of obstructive sleep apnea’. J Sleep Res 2017;26:401. [PubMed: 28425149]
7. McGee S. Evidence-Based Physical Diagnosis. Philadelphia, PA: Saunders, 2001.
8. Patt BT, Jarjoura D, Lambert L, et al. Prevalence of obstructive sleep apnea in patients with chronic wounds. J Clin Sleep Med 2010;6:541–4. [PubMed: 21206743]

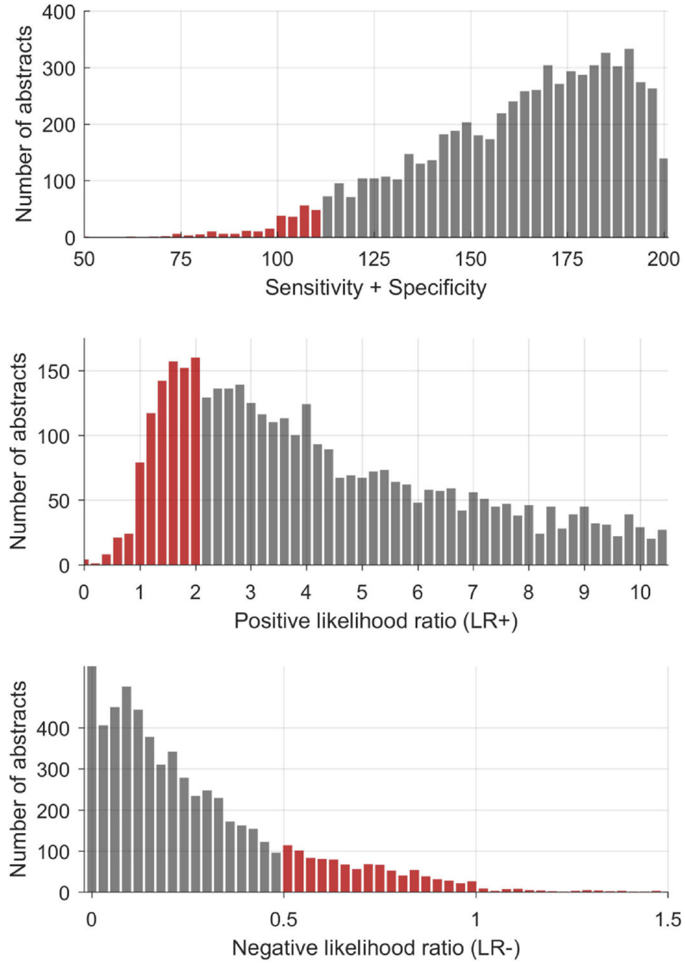

Figure 1.
Distribution of diagnostic test characteristics. In each panel, a histogram illustrates the distribution of diagnostic test characteristics according to the sum of sensitivity and specificity (A), the LR+ (B) and the LR− (C).

A

Positive likelihood ratio (LR+)

Specificity (%) \ Sensitivity (%)	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95
5	0.05	0.11	0.16	0.21	0.26	0.32	0.37	0.42	0.47	0.53	0.58	0.63	0.68	0.74	0.79	0.84	0.89	0.95	1
10	0.06	0.11	0.17	0.22	0.28	0.33	0.39	0.44	0.50	0.56	0.61	0.67	0.72	0.78	0.83	0.89	0.94	1	1.06
15	0.06	0.12	0.18	0.24	0.29	0.35	0.41	0.47	0.53	0.59	0.65	0.71	0.76	0.82	0.88	0.94	1	1.06	1.12
20	0.06	0.13	0.19	0.25	0.31	0.38	0.44	0.50	0.56	0.63	0.69	0.75	0.81	0.88	0.94	1	1.06	1.13	1.19
25	0.07	0.13	0.20	0.27	0.33	0.40	0.47	0.53	0.60	0.67	0.73	0.80	0.87	0.93	1	1.07	1.13	1.20	1.27
30	0.07	0.14	0.21	0.29	0.36	0.43	0.50	0.57	0.64	0.71	0.79	0.86	0.93	1	1.07	1.14	1.21	1.29	1.36
35	0.08	0.15	0.23	0.31	0.38	0.46	0.54	0.62	0.69	0.77	0.85	0.92	1	1.08	1.15	1.23	1.31	1.38	1.46
40	0.08	0.17	0.25	0.33	0.42	0.50	0.58	0.67	0.75	0.83	0.92	1	1.08	1.17	1.25	1.33	1.42	1.50	1.58
45	0.09	0.18	0.27	0.36	0.45	0.55	0.64	0.73	0.82	0.91	1	1.09	1.18	1.27	1.36	1.45	1.55	1.64	1.73
50	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1	1.10	1.20	1.30	1.40	1.50	1.60	1.70	1.80	1.90
55	0.11	0.22	0.33	0.44	0.56	0.67	0.78	0.89	1	1.11	1.22	1.33	1.44	1.56	1.67	1.78	1.89	2.00	2.11
60	0.13	0.25	0.38	0.50	0.63	0.75	0.88	1	1.13	1.25	1.38	1.50	1.63	1.75	1.88	2.00	2.13	2.25	2.38
65	0.14	0.29	0.43	0.57	0.71	0.86	1	1.14	1.29	1.43	1.57	1.71	1.86	2.00	2.14	2.29	2.43	2.57	2.71
70	0.17	0.33	0.50	0.67	0.83	1	1.17	1.33	1.50	1.67	1.83	2.00	2.17	2.33	2.50	2.67	2.83	3.00	3.17
75	0.20	0.40	0.60	0.80	1	1.20	1.40	1.60	1.80	2.00	2.20	2.40	2.60	2.80	3.00	3.20	3.40	3.60	3.80
80	0.25	0.50	0.75	1	1.25	1.50	1.75	2.00	2.25	2.50	2.75	3.00	3.25	3.50	3.75	4.00	4.25	4.50	4.75
85	0.33	0.67	1	1.33	1.67	2.00	2.33	2.67	3.00	3.33	3.67	4.00	4.33	4.67	5.00	5.33	5.67	6.00	6.33
90	0.50	1	1.50	2.00	2.50	3.00	3.50	4.00	4.50	5.00	5.50	6.00	6.50	7.00	7.50	8.00	8.50	9.00	9.50
95	1	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00	10.0	11.0	12.0	13.0	14.0	15.0	16.0	17.0	18.0	19.0

B

Negative likelihood ratio (LR-)

Specificity (%) \ Sensitivity (%)	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95
5	19.0	18.0	17.0	16.0	15.0	14.0	13.0	12.0	11.0	10.0	9.00	8.00	7.00	6.00	5.00	4.00	3.00	2.00	1
10	9.50	9.00	8.50	8.00	7.50	7.00	6.50	6.00	5.50	5.00	4.50	4.00	3.50	3.00	2.50	2.00	1.50	1	0.50
15	6.33	6.00	5.67	5.33	5.00	4.67	4.33	4.00	3.67	3.33	3.00	2.67	2.33	2.00	1.67	1.33	1	0.67	0.33
20	4.75	4.50	4.25	4.00	3.75	3.50	3.25	3.00	2.75	2.50	2.25	2.00	1.75	1.50	1.25	1	0.75	0.50	0.25
25	3.80	3.60	3.40	3.20	3.00	2.80	2.60	2.40	2.20	2.00	1.80	1.60	1.40	1.20	1	0.80	0.60	0.40	0.20
30	3.17	3.00	2.83	2.67	2.50	2.33	2.17	2.00	1.83	1.67	1.50	1.33	1.17	1	0.83	0.67	0.50	0.33	0.17
35	2.71	2.57	2.43	2.29	2.14	2.00	1.86	1.71	1.57	1.43	1.29	1.14	1	0.86	0.71	0.57	0.43	0.29	0.14
40	2.38	2.25	2.13	2.00	1.88	1.75	1.63	1.50	1.38	1.25	1.13	1	0.88	0.75	0.63	0.50	0.38	0.25	0.13
45	2.11	2.00	1.89	1.78	1.67	1.56	1.44	1.33	1.22	1.11	1	0.89	0.78	0.67	0.56	0.44	0.33	0.22	0.11
50	1.90	1.80	1.70	1.60	1.50	1.40	1.30	1.20	1.10	1	0.90	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10
55	1.73	1.64	1.55	1.45	1.36	1.27	1.18	1.09	1	0.91	0.82	0.73	0.64	0.55	0.45	0.36	0.27	0.18	0.09
60	1.58	1.50	1.42	1.33	1.25	1.17	1.08	1	0.92	0.83	0.75	0.67	0.58	0.50	0.42	0.33	0.25	0.17	0.08
65	1.46	1.38	1.31	1.23	1.15	1.08	1	0.92	0.85	0.77	0.69	0.62	0.54	0.46	0.38	0.31	0.23	0.15	0.08
70	1.36	1.29	1.21	1.14	1.07	1	0.93	0.86	0.79	0.71	0.64	0.57	0.50	0.43	0.36	0.29	0.21	0.14	0.07
75	1.27	1.20	1.13	1.07	1	0.93	0.87	0.80	0.73	0.67	0.60	0.53	0.47	0.40	0.33	0.27	0.20	0.13	0.07
80	1.19	1.13	1.06	1	0.94	0.88	0.81	0.75	0.69	0.63	0.56	0.50	0.44	0.38	0.31	0.25	0.19	0.13	0.06
85	1.12	1.06	1	0.94	0.88	0.82	0.76	0.71	0.65	0.59	0.53	0.47	0.41	0.35	0.29	0.24	0.18	0.12	0.06
90	1.06	1	0.94	0.89	0.83	0.78	0.72	0.67	0.61	0.56	0.50	0.44	0.39	0.33	0.28	0.22	0.17	0.11	0.06
95	1	0.95	0.89	0.84	0.79	0.74	0.68	0.63	0.58	0.53	0.47	0.42	0.37	0.32	0.26	0.21	0.16	0.11	0.05

Figure 2.

Likelihood ratio values as a function of sensitivity and specificity values. The LR+ values (A) and LR− values (B) are given for combinations of sensitivity (x-axis) and specificity (y-axis). In each panel, the diagonal indicates when the sum=100%, which maps to an LR value of 1 (black shading). We highlighted two approaches to define ‘close to 1’: when the sum of sensitivity and specificity adds to <110% (dark grey shading); and when the LR+ value is <2, and when the LR− value is >0.5 (light grey shading). Paradoxical LR values, indicated by dotted lines, are found above the diagonal in each panel.