Intracranial bruit: Charles Warlow's challenge revisited

Andrew J Larner

Correspondence to
Dr Andrew J Larner, Cognitive Function Clinic, Walton Centre for Neurology and Neurosurgery, Liverpool, UK; a.larner@thewaltoncentre.nhs.uk

Accepted 20 October 2021
Published Online First
1 December 2021

ABSTRACT

Over 20 years ago, Charles Warlow, the founding editor of *Practical Neurology*, offered a copy of his stroke textbook to anyone diagnosing an intracranial arteriovenous malformation by auscultation of the skull alone. This article examines the possible diagnostic value of intracranial bruit in terms of the 2×2 contingency table for diagnostic tests and recounts an historical case.

When performing a neurological examination, do you auscultate for intracranial bruits? Would you do this over the skull? The orbits? Both? Would you do this routinely? Occasionally? Never?

The notional purpose of such an examination is to identify an underlying arteriovenous malformation or dural arteriovenous fistula, turbulent blood flow in the shunt producing a bruit detectable with the stethoscope. But there is little, if any, evidence to recommend this procedure. The founding editor of *Practical Neurology*, Charles Warlow, writing with several colleagues in the second edition of their textbook of stroke medicine, published 20 years ago, opined that:

> The general examination provides rather few clues to the cause of an intracerebral haemorrhage … Auscultating the skull for detecting arteriovenous malformations is useful for impressing naïve readers of textbooks as well as medical students and patients, but is not very rewarding.[1]

In light of this, the following challenge was issued, in the form of an offer:

> We are still waiting for someone, who had no other clues and only by auscultation diagnosed an arteriovenous malformation in an adult, to take up our offer of a free copy of this book.[1]

I have no information as to whether or not this challenge was ever successfully answered and the prize claimed, but I would suspect not. Likewise, I would anticipate that most readers would answer my introductory questions with 'Never'.

Certainly I was surprised when, after publishing with colleagues a case report describing rapid cognitive decline in a patient subsequently diagnosed with a dural arteriovenous fistula,[2] a rare but recognised cause of reversible cognitive impairment,[3] I received communication from the journal editor inviting a response to a correspondent who asked whether or not we had performed skull auscultation.[4]

How might one make a meaningful reply to this enquiry, other than 'No, we did not', when there are no empirical data, not even a single diagnostic test accuracy study, let alone multiple studies suitable for systematic review and meta-analysis?

At risk of transgressing the style of this journal by using terms that the editors might think some readers would find a 'bit opaque', my suggested approach is couched in terms of binary classicism, that is, examining the implications of the 2×2 contingency table for a diagnostic test accuracy study (figure 1).[5] While this approach will be familiar to most readers from the evaluation of investigations, it is equally applicable to clinical signs (although far less often applied).

This assumes that a methodologically robust test accuracy study of skull auscultation could be undertaken. The index test could perhaps be standardised to listening in specified cranial locations and for specified durations and the reference ('gold') standard for diagnosis of arteriovenous malformation or fistula might be a specified modality of brain imaging. Hence, we could construct a 2×2 contingency table.

Making the reasonable assumption that an intracranial bruit is a rare clinical sign, a policy of routinely auscultating the skull in all patients would occasion many true negatives, particularly if there were 'no other clues'. The same would probably also be true even in the presence of clinical

To cite: Larner AJ. *Pract Neurol* 2022;**22**:79–81.

		Reference Standard: Condition present	Reference Standard: Condition absent
Index Test Outcome	Positive	True positive [TP]	False positive [FP]
	Negative	False negative [FN]	True negative [TN]

Sensitivity	=	TP/(TP + FN)
Specificity	=	TN/(FP + TN)
Positive predictive value	=	TP/(TP + FP)
Negative predictive value	=	TN/(FN + TN)
Accuracy	=	(TP + TN)/(TP + FP + FN + TN)
Critical success index (CSI)	=	TP/(TP + FP + FN)
F measure	=	2TP/(2TP + FP + FN) = 2CSI/(1 + CSI)

Figure 1 Standard 2×2 contingency table for diagnostic or screening test accuracy studies and formulae for selected test measures (all range from 0 to 1, with higher values better).

pointers, which might perhaps include suspected intracranial haemorrhage or subacute cognitive decline. Likewise, very few false positives (ie, hearing a bruit in the absence of an arteriovenous malformation or fistula) would be anticipated. Hence, with many true negatives and very few false positives, one would anticipate very high test specificity for intracranial bruit. Following the 'SpPin' heuristic[6]—that for a highly Specific test, a Positive test rules the diagnosis in—any bruit would likely be diagnostic.

Considering test sensitivity, any assumptions are a little more tentative. Nevertheless, one might suspect that in patients with a proven arteriovenous malformation or fistula, false negatives (no bruit heard) would be more common than true positives if the skull is not a good transmitter (ie, is an effective filter) of the sound of a bruit. If this assumption was correct, the result would be low test sensitivity. With very few true positives and even fewer false positives, positive predictive value might be quite high, but with wide confidence intervals. Box 1 provides a worked example based on these various assumptions.

Since true negatives also feature in the numerator of both negative predictive value and correct classification accuracy (figure 1), one might anticipate that, because of the overwhelming preponderance of true negatives, these values would be inflated (box 1). Indeed, this is a situation in which a more meaningful metric to assess the value of a diagnostic test might be one that ignores true negatives, such as the Critical Success Index (or threat score) or the F measure.[7] The F measure is the harmonic mean of sensitivity and positive predictive value and is widely used in information retrieval and machine learning contexts but is infrequently used in the clinical literature.

Another option might be to construct a receiver operating characteristic curve, which plots sensitivity (true positive rate) against false positive rate (=1−specificity), with the area under the curve used as the measure of diagnostic accuracy. However, because the sign under investigation is a binary classifier, the 'curve' would in fact be a single point, a receiver operating characteristic dot rather than plot, with the test accuracy measure given by the area under a triangle rather than area under a curve, hence ½(sensitivity+specificity),[8] sometimes called 'balanced accuracy'.

As chance would have it, since writing my response to the journal, I have come across an account that might be viewed as the exception that proves the rule (or, more correctly, the regularity) and hence the wisdom of Warlow's challenge. It emanates from no less an authority than a Nobel Prize winner (in Medicine or Physiology in 1981), David H. Hubel (1926–2013), who qualified in medicine before pursuing his career in visual neurophysiology, mostly in collaboration

Box 1 Worked example of the evaluation of intracranial bruit for the diagnosis of arteriovenous fistula using the standard 2×2 contingency table for test accuracy studies

Assume that for a 'high' prevalence population (eg, those presenting with rapid cognitive decline), the prevalence of arteriovenous fistula with an intracranial bruit is 1% (probably an overestimate).

If 1000 patients are tested, TP+FN=10; TN+FP=990.

Also assume (see text) that TN>>>FN>TP>FP.

Assume the outcome of the proposed diagnostic test accuracy study produces the following 2×2 contingency table:

		Reference Standard: AVF present	Reference Standard: AVF absent
Index Test Outcome	Positive: bruit	True positive [TP] = 4	False positive [FP] = 2
	Negative: no bruit	Fales negative [FN] = 6	True negative [TN] = 988

Hence,
Sensitivity=4/(4+6)=0.4.
Specificity=988/(2+988)=0.999.
Positive predictive value=4/(4+2)=0.67.
Negative predictive value=988/(6+988)=0.994.
Accuracy=(4+988)/1000=0.992.
Critical Success Index=4/(4+2+6)=0.33.
F measure=2×0.33/(1+0.33)=0.5.

with Torsten N. Wiesel. Recalling his days as a junior doctor on the neurology service at the Johns Hopkins Hospital in Baltimore, circa 1954, Hubel reports:

> The Chief of Neurology, Jack Magladery, cultivated an eccentric bedside manner. He always began his neurological examination of a patient by listening to the eye with a stethoscope, never hearing anything but always impressing onlookers. One night I was prowling around the medical wards looking for interesting cases when one of the medical interns brought me to see a patient who was a major puzzle, with a hemiparesis and an assortment of other symptoms that I can't remember, and in any case made no sense to anyone. I began examining him, and because a few house staff were looking on I started by listening to the eyes. To my amazement, from one of the eyes came a noise like a pulsating fire hose, a bruit the likes of which I had never heard before, and which made it immediately clear that this man's problem was a cerebral arteriovenous shunt. So suddenly there were crowds of interns and residents around the bed, and I had been catapulted to instant fame. That cancelled out any number of previous blunders.[9]

Interestingly, Magladery did publish on cerebrovascular haemorrhage, but made no mention of auscultation or intracranial bruits.[10] Maybe he used auscultation to create some time for the purpose of gathering his thoughts (thanks to the peer reviewer for this suggestion).

Key points

- Neurological signs may be evaluated in the same way as other investigations by entering the findings of a diagnostic test accuracy study in a 2×2 contingency table.
- In the absence of empirical data, we may make some inferences about the value of signs based on assumptions about their prevalence.

Further reading

- Larner AJ. The 2×2 matrix. Contingency, confusion and the metrics of binary classification. London, Springer, 2021. In press.
- Palace J, Fuller G. Don't do the blood* test!!. *Pract Neurol* 2020;20:428–9.

Whether Hubel's example would (prospectively) meet Warlow's challenge is moot. Would the hemiparesis count as a clue? Certainly, it appears from the (retrospective) account that diagnosis was done 'only by auscultation'. Nevertheless, this clinical anecdote provides no evidence against, and indeed further evidence in favour of, Charles Warlow's view that 'auscultating the skull for detecting arteriovenous malformations is useful for impressing naïve readers of textbooks as well as medical students and patients, but is not very rewarding'.[1]

Contributors AJL is the sole author of this paper.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned. Externally peer reviewed by Tom Hughes, Cardiff, UK.

ORCID iD
Andrew J Larner http://orcid.org/0000-0003-0128-8010

REFERENCES

1 In. ISBN: Warlow C, Dennis M, van Gijn J, et al.. *Stroke a practical guide to management*. 2 ed. Oxford: Blackwell Publishing, 2001: 360.
2 Randall A, Ellis R, Hywel B, *et al*. Rapid cognitive decline: not always Creutzfeldt-Jakob disease. *J R Coll Physicians Edinb* 2015;45:209–12.
3 Wilson M, Doran M, Enevoldson TP, *et al*. Cognitive profiles associated with intracranial dural arteriovenous fistula. *Age Ageing* 2010;39:389–92.
4 Lowenthal MN. Auscultation of the skull. *J R Coll Physicians Edinb* 2016;46:214.
5 Larner AJ. Auscultation of the skull: author's reply. *J R Coll Physicians Edinb* 2016;46:214.
6 Strauss SE, Glasziou P, Richardson WS. Evidence-based medicine. In: *How to practice and teach EBM*. 5th ed. Edinburgh: Elsevier, 2019: 165–6.
7 Larner AJ. Assessing cognitive screeners with the critical success index. *Prog Neurol Psychiatry* 2021;25 (3):33–7.
8 Mbizvo GK, Larner AJ. Receiver operating characteristic plot and area under the curve with binary classifiers: pragmatic analysis of cognitive screening instruments. *Neurodegener Dis Manag* 2021;11:353–60.
9 Hubel DH, Wiesel TN. *Brain and visual perception. The story of a 25-year collaboration*. Oxford: Oxford University Press, 2005: 16.
10 Magladery JW. The natural course of cerebrovascular hemorrhage. *Clin Neurosurg* 1963;9:106–13.