

# Diagnostic uncertainty:

dichotomies are not the answer

Over two decades ago, Alvan Feinstein, a leading figure in the development and evaluation of diagnostic research methods, argued that the binary 'positive-negative' framework used in test accuracy research is not representative of diagnostic decision making in clinical practice.<sup>1</sup> In particular, he called for a move away from the dichotomisation of quantitative test scales (tests that are on a continuous or ordinal scale) and advocated an approach that allows for the explicit recognition of an 'uncertain' diagnostic outcome.

Twenty years on and little progress has been made on this issue; reporting the accuracy of quantitative tests based on a single 'optimal' threshold continues to be common practice. Here we present how improving the methods for evaluating quantitative tests would be of greatest benefit to GPs, providing them with a better evidence-based toolkit of strategies for recognising and handling diagnostic uncertainty head-on.

### DIAGNOSTIC UNCERTAINTY IN PRIMARY CARE

Summerton reported that failures relating to diagnosis account for nearly one-third of GP complaints.<sup>2</sup> Diagnostic uncertainty is particularly rife in general practice due to a number of obstacles intrinsic to the clinical setting.<sup>3</sup> First, the prevalence of serious disease is typically low in a community population, weakening the overall predictive value of diagnostic tests. Secondly, the majority of disease presenting in general practice is commonly in its early stages, when many 'red flag' symptoms are yet to evolve and departures from physiological normality are slight. Furthermore, many of the tests used are not disease-specific; they are typically predictive of a symptom that presents in many diseases. For example, C-reactive protein is a marker of inflammation and is used in the diagnostic work-up of many conditions such as inflammatory bowel disease, bacterial infection, and cardiovascular disease. Predictors of generic symptoms such as inflammation, however, result in weak differential diagnostic capabilities. All of these factors combined leave GPs with a limited selection of useful tools for minimising diagnostic uncertainty.

It could be argued that diagnostic errors are an inevitable consequence of these

fundamental challenges. Yet, general practice may actually be unique in that, due to the typically non-acute nature of clinical presentations, there are a number of 'test of time' strategies,<sup>4</sup> which allow GPs to deal with diagnostic uncertainty more effectively. For example, safety-netting has been proposed as a method of reducing the risk of missing serious disease in patients with an 'uncertain' diagnostic outcome.<sup>5</sup> Research designed around the explicit identification of patients with 'undifferentiated presentations' would facilitate the evaluation of different strategies to handle this subgroup in clinical practice.<sup>6</sup> This would impact significantly on general practice, providing GPs with a set of validated tools to confidently handle diagnostic uncertainty.

### DOWNFALLS OF DICHOTOMISATION

Vital diagnostic information is lost when a quantitative variable is dichotomised.<sup>7</sup> Rifkin *et al*<sup>7</sup> found that the greatest amount of information is lost when disease prevalence is low and the accuracy of the test is poor; a diagnostic scenario typical of general practice.

A key consequence of this information loss is that the predictive value of extreme test results can no longer be differentiated from that of results closer to the dichotomous threshold, even though they may be vastly different. This can make binary accuracy statistics difficult to interpret; modest accuracy may be a result of the test being a poor discriminator across the whole test scale, or it may be that the test is an excellent discriminator towards the extremes of the test scale, but it performs poorly for results close to the threshold. In the latter case, the binary model can cause the discriminatory performance of the test to be understated in accuracy research.<sup>8</sup>

Unless a test is extremely accurate at a single threshold, reducing the interpretation of a quantitative test scale to a simple dichotomy makes it impossible to explicitly identify which test values are

capable of confidently ruling in or ruling out a disease. For example, C-reactive protein is commonly measured to aid the diagnosis of a child presenting with flu-like symptoms. As mentioned before, C-reactive protein is a general inflammatory marker and therefore only extreme values have been found to be effective for ruling in or ruling out serious bacterial infections.<sup>9</sup> Using a single threshold to interpret the results of this test does not tell GPs in which patients serious bacterial infection can be confidently ruled out, which require hospital referral and, most likely, which require the implementation of a 'test of time' strategy because there remains some diagnostic uncertainty. By moving away from dichotomisation and implementing a method that provides a richer interpretation of quantitative test results, we can provide threshold guidance that relates directly to the diagnostic challenges faced in general practice.

### SOLUTIONS FOR TEST ACCURACY RESEARCH

Feinstein argues that clinicians recognise that diagnostic tests are rarely capable of ruling in or ruling out disease and typically interpret test results as either 'positive', 'negative' or 'uncertain'. To bridge the gap between diagnostic decision making in clinical practice and diagnostic research, he proposes that two thresholds should be identified on a quantitative test scale to represent this trichotomous interpretation. Furthermore, this would allow for the proportion of the target population that continue to have an 'uncertain' disease state post-test to be reported.<sup>1</sup>

Irwig and Glasziou dispute that if the test scale is continuous in nature, then forcing test values into three categories would still result in a significant loss of information and, if categorisation is necessary, at least five to seven categories should be identified and likelihood ratios calculated for each category.<sup>10</sup> However, despite numerous calls for the use of multilevel likelihood ratios to navigate large zones of diagnostic

---

*"Vital diagnostic information is lost when a quantitative variable is dichotomised."*

---

*“General practice is an ideal platform for research on patients with an ‘uncertain’ diagnostic outcome.”*

uncertainty, they are still yet to be frequently adopted in diagnostic research.<sup>11</sup>

Any degree of categorisation will result in some information loss, nevertheless typically quantitative test scales are categorised to make it easier for clinicians to interpret results. Nowadays, with test results typically being communicated electronically, it is possible to convey complex diagnostic probability calculations in a graphical format that could be easily understood by clinicians,<sup>12</sup> potentially reducing the need for any categorisation.

## CONCLUSION

Although the reduction of a quantitative test scale to a dichotomy may appear to provide a means of simplifying the interpretation of test results, it fails to allow for the explicit recognition of diagnostic uncertainty. New ways of communicating the true discriminatory capabilities of diagnostic tests are needed so that clinicians have the necessary information available to them to incorporate test results into their decision making effectively.

General practice is an ideal platform for research on patients with an ‘uncertain’ diagnostic outcome due to the high volume of diagnostically-challenging cases.<sup>6</sup> The explicit identification of this subgroup would facilitate a focused stream of diagnostic and prognostic research into the efficacy of different strategies used to manage these patients.

Diagnostic uncertainty is unavoidable; it is only when we start to recognise that it exists that we can invest the time and research effort into identifying the best strategies to overcome it.

### Bethany Shinkins,

DPhil Student in Primary Health Care, Department of Primary Care Health Sciences, University of Oxford, Oxford.

### Rafael Perera,

Head of Statistics, Department of Primary Care Health Sciences, University of Oxford, Oxford.

### Acknowledgements

The authors thank Richard Stevens for his ongoing contribution to this project.

### Funding

Bethany Shinkins is funded by the National School for Primary Care Research — Capacity Building Award. Rafael Perera receives funding from the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research programme (RP-PG-0407-10347 and RP-PG-0407-10338). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

### Competing interests

The authors have declared no competing interests.

### Provenance

Commissioned; not externally peer reviewed.

DOI: 10.3399/bjgp13X664090

## ADDRESS FOR CORRESPONDENCE

### Bethany Shinkins

Department of Primary Care Health Sciences, University of Oxford, New Radcliffe House, 2nd floor, Walton Street, Jericho, Wolfson College, Oxford, OX2 6NW, UK.

E-mail: [bethany.shinkins@phc.ox.ac.uk](mailto:bethany.shinkins@phc.ox.ac.uk)

## REFERENCES

1. Feinstein AR. The inadequacy of binary models for the clinical reality of three-zone diagnostic decisions. *J Clin Epidemiol* 1990; **43(11)**: 109–113.
2. Summerton N. Diagnosis and general practice. *Br J Gen Pract* 2000; **50(461)**: 995–1000.
3. McCowan C, Fahey T. Diagnosis and diagnostic testing in primary care. *Br J Gen Pract* 2006; **56(526)**: 323–324.
4. Heneghan C, Glasziou P, Thompson M, *et al*. Diagnostic strategies used in primary care. *BMJ* 2009; **338**: b946.
5. Almond S, Mant D, Thompson M. Diagnostic safety-netting. *Br J Gen Pract* 2009; **59(568)**: 872–874.
6. Green C, Holden J. Diagnostic uncertainty in general practice. *Eur J Gen Pract* 2003; **9(1)**: 13–15.
7. Rifkin RD. Maximum Shannon information content of diagnostic medical testing. Including application to multiple non-independent tests. *Med Decis Making* 1985; **5(2)**: 179–190.
8. Bowden SC, Loring DW. The diagnostic utility of multiple-level likelihood ratios. *J Int Neuropsychol Soc* 2009; **15(5)**: 769–776.
9. Van den Bruel A, Thompson MJ, Haj-Hassan T, *et al*. Diagnostic value of laboratory tests in identifying serious infections in febrile children: systematic review. *BMJ* 2011; **342**: d3082.
10. Irwig L, Glasziou P. Trichotomous decisions do not imply trichotomous tests. *J Clin Epidemiol* 1991; **44(11)**: 1279–1280.
11. Grimes DA, Schulz KF. Refining clinical diagnosis with likelihood ratios. *Lancet* 2005; **365(9469)**: 1500–1505.
12. Tandberg D, Deely JJ, O'Malley AJ. Generalized likelihood ratios for quantitative diagnostic test scores. *Am J Emerg Med* 1997; **15(7)**: 694–699.