Ophthalmic statistics note 5: diagnostic tests—sensitivity and specificity

Luke J Saunders,1 Haogang Zhu,1,2 Catey Bunce,3 Caroline J Doré,4 Nick Freemantle,5 David P Crabb,1 on behalf of the Ophthalmic Statistics Group

This fifth note from the Ophthalmic Statistics Group illustrates the utility of measurements of sensitivity and specificity in assessing the usefulness of a test for predicting the presence of pathology.

ABSTRACT

This is the fifth statistics note produced by the Ophthalmic Statistics Group (OSG) which is designed to be a simple guide to ophthalmic researchers on a statistical issue with an applied ophthalmic example. The OSG is a collaborative group of statisticians who have come together with a desire to raise the statistical standards of ophthalmic researcher by increasing statistical awareness of common issues.

A typical question facing an ophthalmologist in clinic is: ‘How good is a particular test at diagnosing a particular pathology?’ For example, how useful are the results from a Visual Field Analyser (VFA) in diagnosing glaucoma? Although a positive test result can indicate the presence of a symptom, such as visual field loss, full clinical diagnosis of a disease or condition is considerably more complicated. For instance, in glaucoma, a full clinical examination involving a battery of tests evaluating visual field loss, intraocular pressure, the optic nerve and functional tests such as standard automated perimetry are necessary to make a diagnosis.1 Performing each test is costly and time consuming, and in some countries only a single test may be available for a particular condition. Where multiple tests are available, it is desirable, therefore, to compare their performance in a quantitative way through assessing their diagnostic utility and by examining the sensitivity and specificity with their respective CIs.

Suppose, for example, I have 200 individuals referred for assessment in my glaucoma clinic, 100 of the subjects referred are assessed using a VFA, while the other 100 are assessed using a new imaging technology. Individuals are classified as positive or negative according to whether their test result is positive or negative. The positive and negative refer to presence of an abnormality in visual field or optic disc, rather than the outlook for the patient, though it may seem perverse that a positive test result actually translates to something which can indicate the presence of a symptom, such as visual field loss, full clinical diagnosis of a disease or condition is considerably more complicated. For instance, in glaucoma, a full clinical examination involving a battery of tests evaluating visual field loss, intraocular pressure, the optic nerve and functional tests such as standard automated perimetry are necessary to make a diagnosis.1 Performing each test is costly and time consuming, and in some countries only a single test may be available for a particular condition. Where multiple tests are available, it is desirable, therefore, to compare their performance in a quantitative way through assessing their diagnostic utility and by examining the sensitivity and specificity with their respective CIs.

Sensitivity is the proportion of subjects with a diagnosis who are correctly identified by the test.

\[
\text{Sensitivity} = \frac{\text{True Positives}}{\text{All those with pathology}} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}
\]

\[
= \frac{24}{24 + 5} = 0.96
\]

Table 1 Test results on 200 subjects

| Test          | Test positive (+) | Test negative (−) | Total |
|---------------|-------------------|-------------------|-------|
| VFA           | 42                | 58                | 100   |
| Imaging tech. | 24                | 76                | 100   |
| VFA, VFA test |                   |                   |       |
Specificity is the proportion of subjects without a diagnosis who are correctly identified by the test.

\[
\text{Specificity} = \frac{\text{True Negatives}}{\text{All those without pathology}} = \frac{57}{57 + 18} = 0.76
\]

We can say based on our test results that we expect 96% of patients with glaucoma to have abnormal VFA results (test positive), while 76% of subjects without glaucoma would have normal VFA results (test negative).

If we now consider the imaging technology, we have the following test results: of the 25 glaucoma patients, the imaging technique detects 15 as positive. The sensitivity is therefore 15/25=0.60 (or 60%). Of the 75 subjects without glaucoma, the imaging technique detects 10 as positive. The sensitivity is therefore 10/75=0.13 (or 13%).

The VFA has higher sensitivity (0.96 vs 0.60) and a lower specificity (0.76 vs 0.88) than the imaging technique.

Sensitivity and specificity are proportions, so CIs can be calculated for them using standard methods.\(^5\) It is important to ensure that the method adopted for calculating the CI is correct for the type of sample that you have. One indication that this is not the case would be when a CI exceeds 100%, which clearly cannot be true because, with a highly sensitive test, there are few false negatives, that is, more patients not being diagnosed, or being diagnosed late on in the disease if the test is repeated. Unsatisfactory as it may be, the answer is: it depends. Consideration needs to be given to the consequences of being a false positive or a false negative. If a false negative error is serious, as would be the case if diagnosing ocular tumours, a high sensitivity would take priority. If there is a need to avoid false positives, as might be the case in the glaucoma example and for other diseases with a low prevalence in the population, then a high specificity would be desirable.

A positive test that has high sensitivity is not a guarantee that the individual has the condition, particularly when specificity is not very high and the condition in question is rare. For instance, in our example scenario, which has a high prevalence of patients with glaucoma (25%), the sensitivity of VFA is high (96%), meaning that a high proportion of individuals with glaucoma would be correctly referred (only 4% would be missed). However, almost half of the positive tests overall are, in fact, false positives (18/42=42.8%), which represents an unacceptable large drain on resources in the context of glaucoma and treatment of many other conditions. It is, therefore, important to bear in mind that, in this context, a positive test requires further confirmation, while a negative test is more likely to be correct because, with a highly sensitive test, there are few false negative results, as seen in table 2. Indeed, if a test is 100% sensitive, there will be no false negatives, so a negative test will always be a true negative.

Choosing a test
So, which is the best test here? On the one hand, using the more sensitive but less specific method (VFA in this fictitious example) will lead to more false positives and thus greater cost and time burdens to clinics and more patients with undue concern, while, on the other hand, using the more specific but less sensitive method (imaging technology) will lead to more false negatives, that is, more patients not being diagnosed, or being diagnosed late on in the disease if the test is repeated. Unsatisfactory as it may be, the answer is: it depends. Consideration needs to be given to the consequences of being a false positive or a false negative. If a false negative error is serious, as would be the case if diagnosing ocular tumours, a high sensitivity would take priority. If there is a need to avoid false positives, as might be the case in the glaucoma example and for other diseases with a low prevalence in the population, then a high specificity would be desirable.

When considering sensitivities and specificities it is also essential to consider the merits of the study used to provide these estimates. Initiatives such as the STARD statement (STAndards for the Reporting of Diagnostic accuracy studies)\(^5\) and QUADAS (QUality Assessment of Diagnostic Accuracy Studies) have been developed in order to aid experimenters and readers to assess the internal and external validity of diagnostic studies, and we would urge readers to look through these checklists to optimise what can be established from these studies. For further information on this topic, we would refer the interested reader to other relevant articles.\(^5\)\(^6\)

### Table 2 VFA test results for subjects with known glaucoma status

| Glaucoma (true status) | Present | Absent | Total |
|------------------------|---------|--------|-------|
| **Test—VFA**           |         |        |       |
| Test positive          | 24      | 18     | 42    |
| True positive          | 15      | 9      | 24    |
| False positive         |         |        |       |
| Test negative          | 1       | 57     | 58    |
| False negative         | 10      | 66     | 76    |
| True negative          | 57      | 75     | 100   |

VFA, Visual Field Analyser.

### Table 3 Imaging technology test results for subjects with known glaucoma status

| Glaucoma (true status) | Present | Absent | Total |
|------------------------|---------|--------|-------|
| **Test—imaging technology** |         |        |       |
| Test positive          | 15      | 9      | 24    |
| True positive          | 15      | 66     | 81    |
| False positive         |         |        |       |
| Test negative          | 10      | 66     | 76    |
| False negative         | 10      | 66     | 76    |
| True negative          | 25      | 75     | 100   |

### Table 4 95% CIs for estimates of sensitivity% and specificity%

|                | VFA                              | Imaging technology |
|----------------|----------------------------------|--------------------|
| Sensitivity    | 96% (88.3% to 100%)              | 60% (40.8% to 79.2%)|
| Specificity    | 76% (66.3% to 85.7%)             | 88% (80.6% to 95.4%)|

VFA, Visual Field Analyser.
As a condition will usually have a low prevalence in the population, a positive test result is likely to require independent confirmation even when using a highly sensitive test, while a negative finding is more likely to be truly negative.

The cut-off used to define a positive test can be selected to obtain an optimal compromise between sensitivity and specificity. A useful strategy for comparing diagnostic test performance is to compare sensitivities at equivalent specificities.

Collaborators Valentina Cipriani, Jonathan Cook, Philippa Cumberland, Gabriela Czanner, Paul Donachie, Andrew Elders, Marta García Finana, Rachel Nash, Neil O’Leary, Toby Prevost, Ana Quartilho, Chris Rogers, Selvaraj Sivasubramaniam, Simon Skene, Irene Stratton, Joana Vasconcelos, Wen Xing.

Contributors LJS, HZ and DPC designed and drafted the paper. LJS, HZ, CB, CJD, NF and DPC reviewed and revised the paper.

Funding CB is partly funded by the National Institute for Health Research (NIHR) Biomedical Research Centre (BRC) at Moorfields Eye Hospital NHS foundation Trust and UCL Institute of Ophthalmology.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non-Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

REFERENCES
1 National Institute for Health and Clinical Excellence. NICE Clinical Guidelines—Glaucoma: Diagnosis and management of chronic open angle glaucoma and ocular hypertension. National Institute for Health and Clinical Excellence, 2009. http://publications.nice.org.uk/glucoma-cg85 (accessed Jun 2014).
2 Altman DG, Bland JM. Diagnostic tests 1: Sensitivity and specificity. Br Med J 1994;308:1552.
3 Gardner MJ, Altman DG. Calculating confidence intervals for proportions and their differences. In: Gardner MJ, Altman DG. eds. Statistics with confidence. London: BMJ Publishing Group, 1989:28–33.
4 Deeks JJ, Altman DG. Sensitivity and specificity and their confidence intervals cannot exceed 100%. Br Med J 1999;318:193.
5 Brown LD, Cai TT, DasGupta A. Interval estimation for a binomial proportion. Stat Sci 2001;16:101–33.
6 Bossuyt PM, Reitsma JB, Bruns DE, et al. Standards for Reporting of Diagnostic Accuracy. Toward complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Standards for Reporting of Diagnostic Accuracy. Ann Intern Med 2003;138:40–4.
7 Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155:529–36.
8 Mallett S, Halligan S, Thompson M, et al. Interpreting diagnostic accuracy studies for clinical care. Br Med J 2012;345:e3999.
9 Linnet K, Bossuyt PMM, Moons KGM, et al. Quantifying the accuracy of a diagnostic test or marker. Clin Chem 2012;58:1292–301.
Ophthalmic statistics note 5: diagnostic tests

—sensitivity and specificity

Luke J Saunders, Haogang Zhu, Catey Bunce, Caroline J Doré, Nick Freemantle, David P Crabb and This fifth note from the Ophthalmic Statistics Group illustrates the utility of measurements of sensitivity and specificity in assessing the usefulness of a test for predicting the presence of pathology.

Br J Ophthalmol published online December 8, 2014

Updated information and services can be found at:
http://bjo.bmj.com/content/early/2014/12/08/bjophthalmol-2014-306055

These include:

References
This article cites 7 articles, 4 of which you can access for free at:
http://bjo.bmj.com/content/early/2014/12/08/bjophthalmol-2014-306055#BIBL

Open Access
This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Topic Collections
Articles on similar topics can be found in the following collections

Open access (178)

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/