Visualizing Diagnostic Accuracy Measures to Improve Patient Care

Syeda K.B. Zaidi

214686950

Psyc 6135: Data Visualization

Final Term Paper

Submitted May 2nd, 2019

Visualizing Diagnostic Accuracy Measures to Improve Patient Care

**Measures of Diagnostic Accuracy**

A diagnostic accuracy test provides evidence on how well a test correctly identifies disease and informs clinicians about subsequent decisions about treatment for patients, and healthcare providers. The Standards for Reporting of Diagnostic Accuracy Studies (STARD) statement was developed in 2015 to improve the completeness and transparency of reports of diagnostic accuracy studies, to reduce bias and improve comparison between studies (Cohen et al., 2016). This report includes a list of essential items that should be contained within a report of a diagnostic accuracy study, including information about the diagnostic test, the rationale for the cut-off scores, and a cross-tabulation of the distribution of the diagnostic test results. While the STARD statement allows for comprehensive reporting and comparisons between studies, several diagnostic studies do not go further than simply reporting the test result findings. Currently there is a lack of knowledge translation in this field, where diagnostic accuracy studies do not make it past academia and instead remain lost in databases (Reid, Lane, & Feinstein, 1998). As aforementioned, a diagnostic accuracy test should inform clinicians about decisions on patient treatment and care in addition to informing how well a test correctly identifies disease. However, most clinicians do not use the results of these studies and instead continue to use the originally published norms and cut-offs (Mallett, et al., 2012; Reid et al., 1998). This is problematic for several reasons. Firstly, many tests norms and cut-off points are based on outdated data which can be decades old. Some norms are based on very small samples sizes and are not representative of the population that they are being used for. And lastly, several measures are first designed using a specific clinical population (eg. individuals with Parkinson’s disease) but then are adopted to other clinical populations (eg. individuals with dementia) and continue to use the same norms and test properties as the originally published norms. Of course, all these factors limit the diagnostic test’s accuracy in detecting disease. The objective of this paper is to discuss how applying data visualization techniques can facilitate the knowledge translation process to help diagnostic accuracy measures be applicable and informative for clinicians and patient care. Specifically, this paper will discuss common statistics reported in diagnostic accuracy reports, data visualization techniques specific for these statistics, and lastly the potential outcomes of presenting data visually to facilitate the use of diagnostic accuracy measure by clinicians in the health care setting.

**Diagnostic Accuracy Statistics**

Diagnostic accuracy is not a fixed property of a test. A test's accuracy in identifying patients with the target condition typically varies between settings, patient groups and depending on prior testing (Parikh et al., 2008). These sources of variation in diagnostic accuracy are relevant for those who want to apply the findings of a diagnostic accuracy study to answer a specific question about adopting the test in his or her environment. This is important to keep in mind when interpreting any diagnostic accuracy statistic of a test. The most common statistics reported in most diagnostic accuracy studies are the test’s sensitivity, specificity, positive and negative predictive values, and positive and negative likelihood ratios (Cohen et al., 2016; Parikh et al., 2008). If the statistical analysis of receiver operator characteristics (ROC) curves is used, then it is common to report the area under the curve value. Each of these statistical measures is explained in detail in the following section.

**Gold Standard**

The gold standard is the best single test (or a combination of tests) that is considered the current preferred method of diagnosing a particular disease (Parikh et al., 2008). All other methods of diagnosing the disease, including any new test, need to be compared against this ′gold′ standard. The gold standard is different for different diseases. For simplicity, this paper will focus on detection of dementia. The gold standard for diagnosing dementia is a comprehensive neuropsychological assessment with a focus on assessing memory decline. If an individual performs poorly on the neuropsychological assessment, performing below their expected level, then clinicians may have evidence based on the gold standard neuropsychological assessment to diagnose the individual with dementia. However, the gold standard comprehensive assessment is both costly and time consuming, and often not practical for every individual who comes to the clinic with memory complaints. This led to the development of quick cognitive screening measures to assess whether an individual is showing some concern for a referral to a more comprehensive assessment. In order to determine the accuracy of the screening measures, they need to be compared to and validated against the gold standard measure of a neuropsychological assessment. This comparison gives us information about how well the test is performing in accurately identifying the disease.

**Sensitivity and Specificity**

Sensitivity is the ability of a test to correctly classify an individual as ′diseased′, or positive test result. Specificity is the ability of a test to correctly classify an individual as disease- free or negative test result. Sensitivity and specificity are inversely proportional, meaning that as the sensitivity increases, the specificity decreases and vice versa (McGee, 2002; Parikh et al., 2008).

**Predictive Values**

Positive predictive value (PPV) is the percentage of patients with a positive test who actually have the disease, or the number of true positives. Negative predictive value (NPV) is the percentage of patients with a negative test who do not have the disease, or the number of true negatives. Positive and negative predictive values are directly related to the prevalence of the disease in the population. Assuming all other factors remain constant, the PPV will increase with increasing prevalence and NPV decreases with increase in prevalence (McGee, 2002; Parikh et al., 2008).

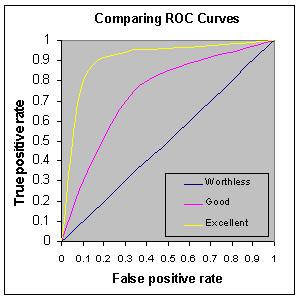
**Likelihood Ratios**

The Likelihood Ratio (LR) of any clinical finding is the probability of that finding in patients with disease divided by the probability of the same finding in patients without disease. Findings with LRs greater than 1 argue for the diagnosis of interest; the bigger the number, the more convincingly the finding suggests that disease. Findings whose LRs lie between 0 and 1 argue against the diagnosis of interest; the closer the LR is to 0, the less likely the disease. Findings whose LRs equal 1 lack diagnostic value (McGee, 2002; Parikh et al., 2008).

**Area Under the Curve**

Accuracy is measured by the area under the ROC curve. An area of 1 represents a perfect test where as an area of .5 represents a worthless test. The area measures discrimination, that is, the ability of the test to correctly classify those with and without the disease (McGee, 2002; Parikh et al., 2008). Figure 1 represents area under the curve of a given test showing excellent, good and worthless values.

**Figure 1.**



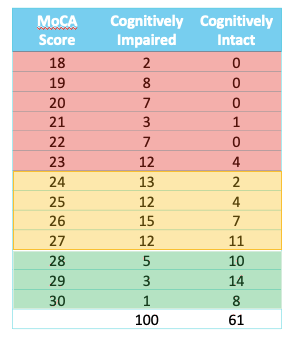
**Visualization to Improve Data Practicality**

These formal recommended quantitative methods for reporting diagnostic accuracy statistics are regularly reviewed in undergraduate textbooks, graduate courses and postgraduate medical examinations, as well as on several certifying examinations for clinicians, practitioners and other health care providers. However, despite the rigorous review of these methods, there is evidence that many clinicians find it difficult to extract usable probabilistic information from diagnostic test accuracy results in the way that they are typically reported (Reid et al., 1998; Steurer, et al., 2002). For example, Reid, Lane, and Feinstein (1998) surveyed the opinion of 300 physicians to assess how often they use the recommended quantitative methods when appraising tests’ diagnosis accuracy including the use of measures of sensitivity, specificity, likelihood ratios, and ROC curves. The authors found that less than 5% of the physicians used formal methods in their practice and 95% reported that this was due to impracticality or non-familiarity with the reported statistics. This is an interesting finding given the amount of exposure to these statistics most clinical professionals have had in their training before being able to practice. As such, Steurer et al., (2002) were interested in examining different forms of presenting test accuracy data to physicians and how this affected their estimates of the probability of disease. They compared presenting a multiple-choice question with straight values of diagnostic accuracy statistics, and then presenting them in a vignette form as a case study. The results found that when the information was presented as a case study 48% of the doctor’s answered the question correctly, in comparison to only 22% when presented as statistical numbers only. This finding suggests that the form in which information is presented to clinicians can drastically impact their understanding of the available data. As such, one possible solution to the lack of data applicability in health care settings could be tackled by presenting data in a visually appealing manner. Allowing for a role of visual aids when presenting diagnostic accuracy information can facilitate better understanding for physicians, increase the use of the data when making decisions, and ultimately improve patient care.

**Visualization of Diagnostic Accuracies**

There are several methods that could be applied to present diagnostic accuracy statistics in a visual way. For example, a simple pie chart can communicate the test’s accuracy for a given population. This could also allow comparison of a single test with different populations. For example, how a cognitive screening test performs with individuals with Parkinson’s disease and then another pie chart to represent how the test performs with individuals with dementia. Using graphs and images to represent data can also be useful in this field to better communicate the message and make it easier to understand for clinicians. The presentation of the distribution of results is also useful and a requirement of the STARD recommendations (Cohen at al., 2016). By showing the distribution of your sample the reader can better understand what each cut-point on a test represents. For example, figure 2 shows the distribution of the number of people who received a certain score on the diagnostic test MoCA, and whether those individuals were impaired or intact on the gold standard neuropsychological battery. This allows to directly compare the performance of the screening test in comparison to the gold standard battery. The colour coded rows represent cut-points with high sensitivity, high specificity and a middle range of scores that are not in the high sensitivity or specificity range.

**Figure 2.**



In addition, some authors have recommended the use of interactive tools for diagnostic accuracy studies to help visualize and better understand the data (Fanshawe et al., 2018). The tools are created with the use of the Shiny application in R Studio and are titled ‘Test Accuracy’ (<https://micncltools.shinyapps.io/TestAccuracy)> and ‘Clinical Accuracy and Utility’ (<https://micncltools.shinyapps.io/ClinicalAccuracyAndUtility)>. The first of these provides a clear interface for displaying and visualizing measures of diagnostic accuracy, such as, sensitivity and specificity. It does so by showing the natural frequencies of true positive, true negatives, false positives and false negatives that would result for a given prevalence and sample size. Whereas the second tool is designed to help users to interpret pre-test and post-test probabilities of disease in relation to clinical decision thresholds. Additionally, predictive probabilities are shown across the full range of possible prevalences from 0% to 100% to show the user the relationship between these two parameters. While these interactive tools may not be useful in informing clinicians about quick decisions during a check-up with a patient, they are practically useful for clinicians’ own use in better understanding the applicability of the statistics (Fanshawe et al., 2018).

**Usefulness of Visual Aids**

From several studies examining the use of visual data, researchers conclude that visual aids tend to be most effective when they are transparent—when their elements are well defined and they accurately and clearly represent the relevant information by making part-to-whole relationships in the data visually available (Puhan, Steurer, Bachmann, & Ter Riet, 2005). This is particularly important to keep in mind when the audience is academic clinicians who likely have a good understanding of some of the elements that are presented. Appropriately designed visual aids can improve comprehension of risks associated with different medical treatments, screenings, and lifestyles. It has also been found that risk information presented visually is judged as easier to understand and recall, and requires less viewing time than the same information presented numerically ([Feldman-Stewart, Brundage, & Zotov, 2007](https://journals-sagepub-com.ezproxy.library.yorku.ca/doi/full/10.1177/0963721413491570); [Gaissmaier et al., 2012](https://journals-sagepub-com.ezproxy.library.yorku.ca/doi/full/10.1177/0963721413491570)). While most doctors and clinicians are numerical and graph literate, the use of visual aids can help them remember the information so they can readily apply this in their everyday practice.

**Conclusion**

Quantifying diagnostic accuracy is an important first step in assessing whether a new diagnostic device is suitable for implementation into clinical practice. To many clinicians and researchers, statistical measures of diagnostic accuracy are not intuitive, cumbersome to remember, or may seem impractical when trying to make judgements in their practice. To overcome this concern, the use of data that is visually presented and visual aids should be readily used in studies and reports of diagnostic accuracies. The use of visual aids is judged as easier to understand and recall and has shown to requires less cognitive demand when interpreting data. In addition, past research demonstrates that more meaningful formats of presenting information can drastically reduce errors made by physicians when judging probabilities of disease outcomes. This addition of visual aids in clinical training and practice can then help to facilitate knowledge translation from research to clinic and help clinicians use diagnostic accuracy statistics to inform their everyday clinical judgements.

**References**

Cohen, J. F., Korevaar, D. A., Altman, D. G., Bruns, D. E., Gatsonis, C. A., Hooft, L., ... & Bossuyt, P. M. (2016). STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open*, *6*(11), e012799.

Fanshawe, T. R., Power, M., Graziadio, S., Ordóñez-Mena, J. M., Simpson, J., & Allen, J. (2018). Interactive visualisation for interpreting diagnostic test accuracy study results. *BMJ evidence-based medicine*, *23*(1), 13-16.

Feldman-Stewart, D., Brennenstuhl, S., McIsaac, K., Austoker, J., Charvet, A., Hewitson, P., Whelan, T. (2007). A systematic review of information in decision aids. *Health Expectations*, 10, 46–61.

Gaissmaier, W., Wegwarth, O., Skopec, D., Müller, A., Broschinski, S., Politi, M. C. (2012). Numbers can be worth a thousand pictures: Individual differences in understanding graphical and numerical representations of health-related information. *Health Psychology*, 31, 286–296.

McGee, S. (2002). Simplifying likelihood ratios. *Journal of general internal medicine*, *17*(8), 647-650.

Mallett, S., Halligan, S., Thompson, M., Collins, G. S., & Altman, D. G. (2012). Interpreting diagnostic accuracy studies for patient care. *Bmj*, *345*, e3999.

Parikh, R., Mathai, A., Parikh, S., Sekhar, G. C., & Thomas, R. (2008). Understanding and using sensitivity, specificity and predictive values. *Indian journal of ophthalmology*, *56*(1), 45.

Puhan, M. A., Steurer, J., Bachmann, L. M., & Ter Riet, G. (2005). A randomized trial of ways to describe test accuracy: the effect on physicians' post-test probability estimates. *Annals of internal medicine*, *143*(3), 184-189.

Reid M. C., Lane, D. A., & Feinstein, A. R. (1998). Academic calculations versus clinical judgments: practicing physicians’ use of quantitative measures of test accuracy. *The American journal of medicine*, *104*(4), 374-380.

Steurer, J., Fischer, J. E., Bachmann, L. M., Koller, M., & ter Riet, G. (2002). Communicating accuracy of tests to general practitioners: a controlled study. *Bmj*, *324*(7341), 824-826.