Reviewer’s report

Title: A simple, step-by-step guide to interpreting decision curve analysis

Version: 0 Date: 29 Mar 2019

Reviewer: Michael Power

Reviewer's report:

I do not think that this paper should be published without an extensive revision.

The paper adds little or nothing to the authors' previous publications on "decision curve analysis". I will make some suggestions below on how this deficiency could be addressed.

The authors should take the aspiration of fully informed, shared clinical decision making much more seriously.

They should also view the provision of information on uncertainties (statistical and structural) as an ethical imperative when supporting informed clinical decision making.

I suspect that at least some of the authors are fluent in R. If this is the case, they could easily develop a Shiny app that clinicians and patients could use to explore the parameters and uncertainties that affect clinical decision making after receiving the result of a diagnostic test.

The motivation for this paper is that, despite the original papers on decision curve analysis being fairly often cited, their concept is poorly understood. The approach the authors take is to rehash their previous papers. I suggest that they carefully consider why their previous papers on clinical decision curve analysis are poorly understood.

I found the present paper incomprehensible, although I probably have a deeper knowledge of the evaluation of clinical tests and of clinical decision making than the main target audience of the authors.

The reason, I would suggest, is that the language (and thinking) is fuzzy, imprecise, oversimplistic, and often simply wrong.

The authors would benefit from studying a book such as Stephen Pinker's "The Sense of Style: The Thinking Person's Guide to Writing in the 21st Century", especially chapter 3 on "The curse of knowledge".

The curse of knowledge is that, when you have intimate knowledge and experience of an issue, it is very difficult to take account of the ignorance of other people when communicating with them.
There are so many examples of fuzzy, imprecise, over-simplistic language that I will only give a few examples.

Evaluation of a test (the technology) is often confused with evaluation of a test result (the clinical meaning for a patient).

In the original paper, the figure purporting to show a decision tree, does not label the decision and chance nodes, and the value nodes hide a large amount of detail essential to understanding the difference between decision analysis and decision curve analysis.

The caption to figure 1 in this paper says that "a, b, c and d give, respectively, the value of true positive, false positive, false negative and true negative", but the text says "a − c, [is] the value of a true-positive result".

In the BMJ paper, tests are said to provide positive and negative predictive values. This is wrong. Predictive values depend on both the test result and the prevalence. The authors approach to "decision curve analysis" appears to overlook this fact, which is of crucial importance when using a test in situations with different prevalences.

Also, in the BMJ paper, the example chosen has two tests with a minuscule difference between the two decision curve analyses.

Any experienced clinician (or health economist) would know intuitively that this difference is dwarfed by the range of statistical (let alone structural) uncertainties, and would recommend that other factors be taken into account when deciding between the adoption of the two tests.

The misunderstandings in the BMJ paper are carried through into the present paper.

The present paper states in the abstract "We argue that many of the difficulties with interpreting decision curves can be solved by relabeling the y-axis as "benefit" and the x-axis as "preference".

But figures 2 and 3 do not do this.

In Figure 1, the x axis is labelled "Preference" and the extremes are labelled "I am worried about disease" and "I am worried about biopsy".

However, "I am worried about disease" and "I am worried about biopsy" are not preferences.

A preference implies that one thing is compared to another. The labels should be explicit about the comparison and the units of valuation.

If a figure has threshold probability as a label on an axis, the caption (and quite possibly the text as well) should be explicit about what the probability is, what the threshold is for doing what, where the threshold comes from.
The analogy with international trading (in the present paper and the BMJ paper) is confused and therefore confusing. It suggests that the authors understand neither the evaluation of clinical utility nor how international trading accounts work.

If the accounting of international trading is used as an analogy, the authors should explicitly map the components of clinical test evaluation to the components of international trading accounts, e.g.:  

- clinical outcome \(\leftrightarrow\) commercial product
- measure of clinical outcome \(\leftrightarrow\) unit price in local currency
- commensurable value of clinical outcome (e.g. QALY, willingness to pay, ...) \(\leftrightarrow\) exchange rate between local currency and currency for accounts

The paper that I would like to see would:

1) Review the conventional evaluations of accuracy and clinical utility of diagnostic tests. This would avoid silly straw man arguments about full pathway clinical utility not being equivalent to diagnostic accuracy.

The review should include TPR, FPR, FNR, TNR, sensitivity, and specificity. It should discuss positive and negative predictive values and how these depend on prevalence. It should discuss ROC curves, how they represent the tradeoff between sensitivity and specificity, and the challenge of identifying the optimum threshold for regarding a test result as positive or negative. It should discuss the AUROC as a measure of test accuracy, and how very different shapes of ROC can have the same AUROC (an analogy could be drawn with Anscombe's Quartet).

It should discuss the role of post-test probability in relation to the thresholds for acting or not acting on positive and negative results. (The book they cite may provide some guidance on this.)

And it should discuss how clinical utility can be quantified to allow arithmetical comparison of different outcomes (commensurable values).

2) Review conventional decision analysis in sufficient detail for it to be compared and contrasted with decision curve analysis.

It should be illustrated with a figure that visualises and names concepts such as decision nodes, chance nodes, processes, probabilities, outcomes, values of outcomes.
3) Review decision curve analysis in detail. Since the "math is trivial", give the math, define the terms of the equations, and explain how the values of the parameters are derived.

Explain why true negative results and false negative results are either not valued or are hidden in the math.

Avoid statements such as "benefit is net benefit". This is as helpful as "Brexit means Brexit".

4) Compare and contrast conventional decision analysis with decision curve analysis (DCA), explaining the simplifications of DCA, pros and cons of the two approaches, when each approach would be indicated, and how DCA could support (a) shared clinical decision making and (b) confront uncertainties head on.

Level of interest
Please indicate how interesting you found the manuscript:

An article of importance in its field

Quality of written English
Please indicate the quality of language in the manuscript:

Not suitable for publication unless extensively edited

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?
If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests.

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal.