Peer Review File

Article information: http://dx.doi.org/10.21037/jlpm-21-3

Reviewer A

Comment 1: Table 1: I wonder whether the author can also add the relative % of adverse events (compared to the total number of tests performed) beyond the absolute numbers of adverse events?

Response 1: Usage data required for percentages is generally unavailable. However, I have added an example for continuous glucose meters on page 8 with usage data.

Comment 2: To demonstrate the principle that acceptability should be put into the total context of diagnostic and healthcare (i.e. taking into account user errors and other pre- and postanalytical errors) can perhaps also be illustrated by referencing to a publication on badly performed COVID-tests (COVID testing by non-professionals at border controls, clinical sensitivity may decline from 80% to 30).
Only FDA regulatory standards are referred to. What about the EU IVDR?

Response 2: Adding the suggested COVID or EU IVDR discussion is beyond the scope of this manuscript.

Reviewer B

Comment 1: Introduction, Line 3, "… compared to a reference measurement procure."

Response 1: I have made the change.

Comment 2: A hierarchy among performance standards, Line 7, to this reviewer's knowledge CLSI documents do not prescribe standards but rather provides guidance on protocols and procedures. The acceptability criteria is at the discretion on the user.

Response 2: I believe the text as written is ok. Note that I state (line 8-9 page 4) "EP21 suggests the protocol and analysis for method comparison studies but does not proscribe acceptable limits."

Comment 3: Similarly, as noted by the author himself, ISO 15189 does not prescribe acceptability criteria and seemed misplaced under "Hierarchy Among Performance Standards".

Response 3: ISO 15189 is included because of the measurement uncertainty

requirement: Clause 5.5.1.4 states that laboratories "shall determine measurement uncertainty for each measurement procedure in the examination phase used to report measured quantity values." It also states that "Upon request, the laboratory shall make its estimates of measurement uncertainty available to laboratory users. While acceptability per se is not required, measurement uncertainty reports the interval captured by 95% of the data.

Comment 4: On the other hand, the Milan consensus is a "Hierarchy Among Performance Standards" recommended by a professional body.

Response 4: The goal of this section is to present a hierarchy of *regulatory* acceptance standards. The Milan conference, while recommended by a professional body, has no regulatory status.

Comment 5: The Error Grid describes clinical risk and does not directly link to clinical outcome. It is driven by clinical surveys.

Response 5: The commentator is correct in that no test result is directly linked to clinical outcomes. Only a treatment decision is linked to a clinical outcome. In the Milan text, I say, "the effect of measurement performance on clinical outcomes" Hence, it is commonly accepted to associate test results with outcomes, so I have left the error grid text unchanged.

Comment 6: The argument against 95% acceptance limit can similarly be applied to the Error Grid analysis, where it is hard to consider a result falling close to high risks zones won't cause harm – given the zones are demarcated by clinical surveys.

Response 6: I agree with this comment but remind the commentator that error grids have multiple zones unlike the dichotomous 95% limits. Hence in the glucose meter error grid, a point on the zone A B intersection may cause slight harm. A zone on the C D intersection will cause greater harm. I also make this point on page 5 lines 10-12.

Comment 7: It should be noted that the Error Grid also has its own shortcomings compared to more conventional method evaluation: https://care.diabetesjournals.org/content/29/8/1805.

Response 7: The Clarke error grid in this reference has been replaced by the Parkes error grid or the surveillance error grid.

Comment 8: The section on "The problem with the method comparison protocols" provides good food for thought.

Response 8: Thank you.

Comment 9: It could be considered that the use of sigma matrix attempts to express total error

in frequency of unacceptable error.

Response 9: I have added a short section on Six Sigma.

Comment 10: While any loss of life due to erroneous results is tragic, one has to put things in perspective and consider the number of lives saved by having timely glucose results. Additionally, comparing the death rates to other iatrogenic causes of death (e.g. medication error, adverse side effects, erroneous diagnosis/ treatment), the performance of glucometers should be considered favourably.

Response 10: I believe I address this comment with the heading: The problem when assays fail acceptance criteria starting on page 8.

Comment 11: The example of interference causing error is different from comparing them against a reference method (acceptability criteria). It is generally an unexpected interaction with an endogenous substance of the patient. It is hard to imagine a system being absolutely perfect under every scenarios, particularly when it involves the biology of a human. Nonetheless, it is agreed that the effect of interference on laboratory measurement is under-reported by the manufacturer, yet at the same time, one must ask the realistic question of how can the manufacturer obtain enough representative interfering substances (including representative variety of interference antibodies) to provide such information?

Response 11: One can have acceptance criteria for interfering substances as not all interfering substances cause huge errors: example the effect of hematocrit on glucose meter results. I agree with the commentator that some interferences may be unanticipated and mention in the conclusions that not all errors can be eliminated.

Comment 12: Interestingly, a recent paper on FMEA was published in this journal (Asking why: moving beyond error detection to failure mode and effects analysis) and made similar comments as the authors.

Response 12: I have added this reference.