As in many other medical scientific disciplines, achievements and conclusions in veterinary clinical research are deduced from observations; made on clinical, experimental animal material and/or through ex vivo/in vitro experiments. To make the step from gathering, visualising and exploring observations to actually have them support conclusions and statements that will enhance the collected knowledge within the specific research area, the data will need to be assessed by means of well-defined relevant tests and criteria. For objective and unbiased use of such criteria they will need to be defined a priori to data assessment. All of this is the backbone of hypothesis-based research and thus common knowledge for researchers. The application of basic statistical tests to assess statistical probability of differences and the traditional choice of $P<0.05$ as a criterion for acceptable level of risk for statistical type I error is so deeply rooted that many hardly give it a thought when applying these. Unfortunately, this tendency for automation can lead to answering the wrong question in diagnostic studies through choosing sub-optimal methods for testing or in other cases where the chain for proper deduction is broken by the lack of appropriate a priori criteria to test the question asked. As indicated by the title, the latter is in focus here.

The world of validation of biomarkers for diagnostic purposes in veterinary medicine is used for exemplification. For clarity this process is often divided into four phases
were aspects found critical to the final diagnostic application is assessed in each phase (consult review for details):

Phase I: Analytical performance: Assessment of analytical and practicability characteristics.

Phase II: Overlap performance: Assessment of value in discriminating health from disease.

Phase III: Clinical performance: Evaluation of diagnostic sensitivity and specificity in the clinical settings of interest.

Phase IV: Outcome and utility performance: Assessment of whether the individual (or community) gain advantage from the test.

For the first three phases useful guidelines exist for study design to objectively obtain representative and valid data, e.g. 2-4, where my personal experience is that these are easily followed by authors, and imprecision of a new test is validly assessed to be e.g. 6%. However, often the first versions of a manuscript submitted for review fail to apply appropriate clinically relevant criteria to assess the clinical acceptability of this result, in my opinion mainly because clinical acceptability is specific to clinical settings and biomarkers, so no one-size-fits-all. But the lack of criteria will not allow the authors to support clear conclusions and often lead to extended passages in the ‘discussion’ in an attempt to approach a conclusion through discussion. Clearly stated criteria for acceptability in ‘Materials and Methods’, e.g. maximum allowable imprecision 5%, enable authors to focus discussion on concluding and aspects related to acceptability or non-acceptability. Thus another beneficial effect of introducing relevant criteria of acceptability is often that the discussion is substantially shortened and focussed to the clear benefit of the reader.
Guidelines exist for defining criteria for clinical acceptability, e.g. a consensus agreement on ranking of methods to define analytical performance criteria\(^5\), which is a mix of objective and subjective methods, with a focus on increasing clinical relevance\(^6\). The authors can use such guidelines to identify and argue for their choice of criteria, a process veterinary journals may consider to support/encourage the authors to focus on performing \textit{a priori} to data evaluation to maintain the scientific process intact.

So along with other checkpoints prior to publication of veterinary diagnostic studies it is recommended to check:

- Were clinically relevant questions asked?
- Were these tested appropriately?
- And not least were clinically relevant criteria set to allow the tests and data to actually answer the questions and support conclusions?

Without this trilogy intact the supposed clinically relevant conclusions drawn are at risk to be unsupported, irrelevant or even misleading. It would be beneficial for the process of publication of veterinary clinical research if the frequency of studies lacking appropriate use of assessment criteria could be minimised by informing researchers about the necessity of the \textit{a priori} establishment of such criteria.

References:


