



Published in final edited form as:

*Clin Chem.* 2009 January ; 55(1): 15–17. doi:10.1373/clinchem.2008.114694.

## Cut-points in clinical chemistry: time for fundamental reassessment

Andrew J. Vickers and Hans Lilja

Memorial Sloan-Kettering Cancer Center

When prostate-specific antigen (PSA) testing started to become widespread in late 1980s, it became apparent that different assays yielded discrepant values. Unraveling of the various molecular forms of PSA in the blood and other bodily fluids in early 1990s contributed critical information as to how to avoid these assay-related biases {PMID: 1716536}. The subsequent design of a novel PSA-calibrator by Stamey et al provided another important advance, which later was endorsed by the WHO {PMID: 9126223}. In this issue of *Clinical Chemistry* {reference}, Jansen and colleagues examine the re-calibration of one widely used PSA assay with the WHO-endorsed calibrator. They report that calibration affects both the likelihood that a man will be biopsied and, moreover, the probability that if he does undergo biopsy, cancer will be found. The paper is a timely reminder that a man's PSA level, as given on a laboratory report, is not a simple statement of a true biological state, but is affected by subtle details of laboratory technique. A man whose PSA rises from 1.9 to 2.4 ng / ml over the course of a year might well panic that this sudden rise in PSA is over a widely discussed "PSA velocity" cut-point of 0.35 ng / ml / year {PMID: 17077354} and is thus indicative of cancer. As Jansen et al. show, however, such a rise might be largely explained by a change in assay.

Jansen et al.'s main results focus on a different and more widely used cut-point, an absolute value for PSA of 3 or 4 ng / ml as an indication for biopsy. This reflects contemporary urologic practice and the authors are right to state that their analyses have important practical implications. Here we want to suggest an additional recommendation: why not abandon the cut-point of 3 or 4 ng / ml altogether? This reflects our view that cut-points in clinical chemistry are problematic. We believe that we need a new way of thinking about the relationship between biomarker measures and clinical decision-making.

This is not trivial: cut-points could not be more central to routine clinical chemistry. Almost all of the many millions of clinical laboratory reports produced yearly include both the absolute value of each analyte and an indicator as to whether it is above or below a cut-point. A PSA above 4 ng / ml is starred, or bolded, or marked in red, as is, for example, a fasting blood glucose above 5.8 mmol/L or hemoglobin below 13 g/ dL. We see numerous problems with the uncritical use of such cut-points.

### 1) The rationale for many cut-points is unknown

Much as we would like to think that medicine is based solely on clear evidence, the origin of many cut-points is unclear. The cut-point of 4 ng / ml for PSA is one example. In the past two years, we have conducted numerous enquiries as to the origin of this cut-point, contacting clinical chemists, urologists who were in practice when the PSA test was first introduced, and those in industry. Most had no idea of where 4 ng / ml came from; others referred us to early landmark studies {PMID: 1707140} {PMID: 7525995} or reports we could not find. However, the empirical studies we were referred to could not provide empirical validation of this cut-point as they used 4 ng / ml as criterion for biopsy. The best source, appears to be an internal industry report selecting 4 ng / ml on the basis of a reference range.

## 2) Cut-points are often chosen using irrational methodologies

Reference ranges are the most common source of clinical chemistry cut-points: a blood marker is measured in a group of individuals who report no disease; the range of values that includes 95% of the population is defined as normal; abnormal is defined as values in the top or bottom 2.5%. The essential problem with reference ranges is that they are a statistical construct, entirely disconnected from any consideration of health or illness. For a start, a reference range defines a fixed proportion of the population (e.g. top or bottom 2.5%) as abnormal, irrespective of the incidence of disease. Using a reference range for a laboratory screening test for say, sarcoma in children, would define about 2 million US children per year as worthy of additional work-up, when fewer than 1,000 cases are diagnosed each year. Moreover, reference ranges are naturally highly dependent on the population studied. As a simple thought experiment, consider a 95% reference range for body mass index calculated on the contemporary US population, compared to the US population in 1975, or an African population.

We propose that a viable alternative to reference ranges must include relationship(s) between the laboratory value and a clinical outcome. This is exactly what has been done with body mass index, which was correlated with longevity (thus explaining why obesity cut-points are not rising over time). Yet given a relationship between a laboratory measure and a clinical outcome, we still need a statistical methodology to obtain a cut-point. The most common approach is to choose a cut-point that maximizes sensitivity plus specificity, often stated as choosing the point closest to the top left of the ROC curve {PMID: 18435502}. This approach weights sensitivity and specificity equally and is thus, in our view, not optimal. Take the case of a test for infection where the disease is often fatal and the treatment a highly effective, safe and well tolerated antibiotic. A cut-point associated with 70% sensitivity and 70% specificity would be closer to the top left of the ROC curve than a cut-point with 100% sensitivity and 30% specificity, yet it is the latter cut-point that we would choose. We do so because the clinical consequences of a missed diagnosis are grave in comparison to the minimal harms associated with this unnecessary treatment. Other clinical settings may involve an entirely different balance between disease severity versus treatment efficacy and side effects, leading to different choices for the importance of sensitivity compared to specificity.

## 3) Cut-points are invariant to patient preference

Imagine that a man tells us that he is very anxious about invasive medical procedures and tolerates them poorly. Accordingly, he is willing to undergo prostate biopsy only when the doctor is quite convinced of that he may need it. One might imagine that in such a case a doctor might forgo biopsy if the man's PSA is only slightly raised above the commonly used cut-point of 4 (e.g. 4.5 ng / ml) but be more insistent for a larger elevation (such as 17 ng / ml). A similar approach might be used with respect to elevated cholesterol in a patient who experiences side-effects from a statin: whether the patient is advised to continue medication despite the side-effects will depend on the degree to which cholesterol is elevated.

The problem is that a doctor has no rational way to reset a cut-point in the light of patient preference. A PSA cut-point of 4 ng / ml might be too low for the anxious patient, but who is to say whether 6 ng / ml, 10 ng / ml or 15 ng / ml is the new cut-point? Conversely, this cut-point may be far too high for very risk averse younger men. Similarly, if 240 mg/dL cholesterol is an insufficiently high threshold for a patient with low statin tolerance, how much higher would be "high enough" for a doctor to insist on treatment?

## 4) Cut-points cannot include multiple pieces of information

Medical decisions are rarely based on a single clinical finding, symptom, or laboratory measure. With respect to our statin intolerant patient, the physician's advice is likely to take into account

other risk factors such as blood pressure, smoking, diabetes and family history. In the case of prostate cancer, some urologists would consider ordering both a total PSA level and a free-to-total PSA ratio {18337732}, which tends to be lower in men with cancer. This raises the question of how the doctor should integrate these two measures. Is biopsy indicated if PSA is elevated but free-to-total PSA ratio normal? If not, would there be some very high level of PSA that would override the normal free-to-total PSA ratio?

### Use of risk prediction in place of cut-points

A simple, rational alternative to the traditional use of clinical chemistry cut-points is to use laboratory measures to calculate the probability of clinically relevant states or events. For example, instead of the laboratory report stating whether total PSA and free-to-total PSA ratio are in “normal” range, a probability of prostate cancer would be given on the basis of a statistical model including age and total PSA, or both variables. Similarly, instead of marking cholesterol as out of range, the laboratory report could incorporate other laboratory values, such as hemoglobin A1C, and clinical information, such as blood pressure, to give an estimate of a patient’s risk of a cardiovascular event within 10 years.

The advantages of such an approach mirror each of the problems described above. First, creation of a statistical risk prediction model is a complex scientific procedure that would need careful documentation in the peer-reviewed literature. This will likely include reference to the sort of calibration issues reported by Jansen et al. {reference} Second, and importantly, the choice of any cut-point would be rationally based on clinical consequences. To return to two of our previous examples, patients at risk of the fatal infection would likely be willing to take an antibiotic even if they had only a 1% risk of disease; in contrast, many men are likely to require a 20% or higher risk of prostate cancer to be willing to undergo an uncomfortable procedure such as a prostate biopsy. Third, use of risk prediction can easily incorporate patient preference, in part, because information is presented in intuitive terms (compare “Your cholesterol is 240 mg/dL” to “You have a 1 in 5 chance of a heart attack in the next ten years”). Fourth, risk prediction can easily incorporate multiple items of information.

In our view, the use of clinical chemistry cut-points is a holdover from a time when bedside calculation of probabilities was impractical for the practicing clinician. It used to be easier to define a patient as normal or abnormal and treat accordingly, than to calculate a probability on the basis of a statistical prediction model. With the widespread availability of information technology, this is no longer true: one can envisage computer systems that would seamlessly integrate multiple laboratory reports, imaging and clinical data to provide the treating physicians and the patient with clinically relevant and properly validated risk predictions. Development of such systems would require both clinical research — to develop and validate by independent replication appropriate risk prediction models — and new platforms for bioinformatics. As such, it is likely that these systems remain several years in the future. In the meantime, however, we must not forget the very real limitations of our current approaches: cut-points certainly make life simpler, but rarely reflect complex biological systems adequately, and changes in calibration can importantly influence widely accepted cut-points as Janssen et al report in this issue {reference}.