The Nonequivalent "Equivalent Cutoff"

This white paper reviews the validity of the recently proposed "equivalent cutoff" and the critical shortcoming in the effectiveness of this idea, including that equivalent cutoff ignores the cardinal rule of newborn screening: to minimize the chance of a false negative result.

Newborn screening of lysosomal storage disorders (LSDs) can be performed using different methods, workflows, screening algorithms, assay protocols, reagents/kits and instruments, all of which impact the output of the test. To determine the relative risk of disease, screening results are compared to a cutoff, which must be established empirically by each screening program based on the performance of their chosen assay system, the referral capacity of follow-up centers and, of course, the desire to avoid false negatives. **"Equivalent cutoff" has been proposed as a method to normalize screening results across LSD screening platforms, but unfortunately falls far short of this goal.** We will examine the basis for this unusual metric and highlight the biological and mathematical fallacies of this approach.

Equivalent cutoff[1,2] compares LSD newborn screening results, obtained from different programs, by applying a fixed percentage of the daily mean enzyme activity of presumed normal individuals. This cutoff has been theoretically applied to the reported enzyme activity values; all enzyme activity screening results that fall below this cutoff are described as screen positive results. In reality, there are a myriad of false assumptions and other shortcomings that invalidate the equivalent cutoff approach as a method for comparison.

As a percentage of the mean, "equivalent cutoff" accounts for differences in reported enzyme activities between different methods – scale differences. If scale differences are, in fact, the only difference between two programs running the same LSD assay, then the application of the same equivalent cutoff value should result in a comparable screen positive rate. However, this assumption is **highly flawed for the following reasons**. Distributions of enzyme activities from both normal and affected samples are inherently dependent on the assay protocol and conditions, the specific reagent kit used, the screening algorithm, preanalytical steps and multiple other sample conditions (especially for dried blood spots; DBS). An excellent example of the fallacy of equivalent cutoff can be found in two published distributions of tandem mass spectrometry (MS/MS) IDUA assay results obtained using either a 3-plex or a 6-plex reagent assay, both developed by the same group and analyzed using flow injection tandem mass spectrometry[3,4]. In their discussion, the authors highlight the many improvements of the 6-plex assay over their previous 3-plex assay and define very different optimal cutoffs of 32% and 10% of the daily mean for the 3-plex and 6-plex assays, respectively.

In recommending two highly different cutoffs for their two assays, the authors implicitly prove that the probability density function, PDF (shape of the distributions), can be different as the result, solely, of alterations to an assay method, without even considering all the other reasons mentioned above. The PDFs for the non-affected cohort in the 3-plex and 6-plex assays (both of which use the same flow-injection MS/MS within the same lab with the same population) differ not only by a scale factor (note the different mean values), which would have been accounted for by equivalent cutoff, but also by the overall shape of the curve.

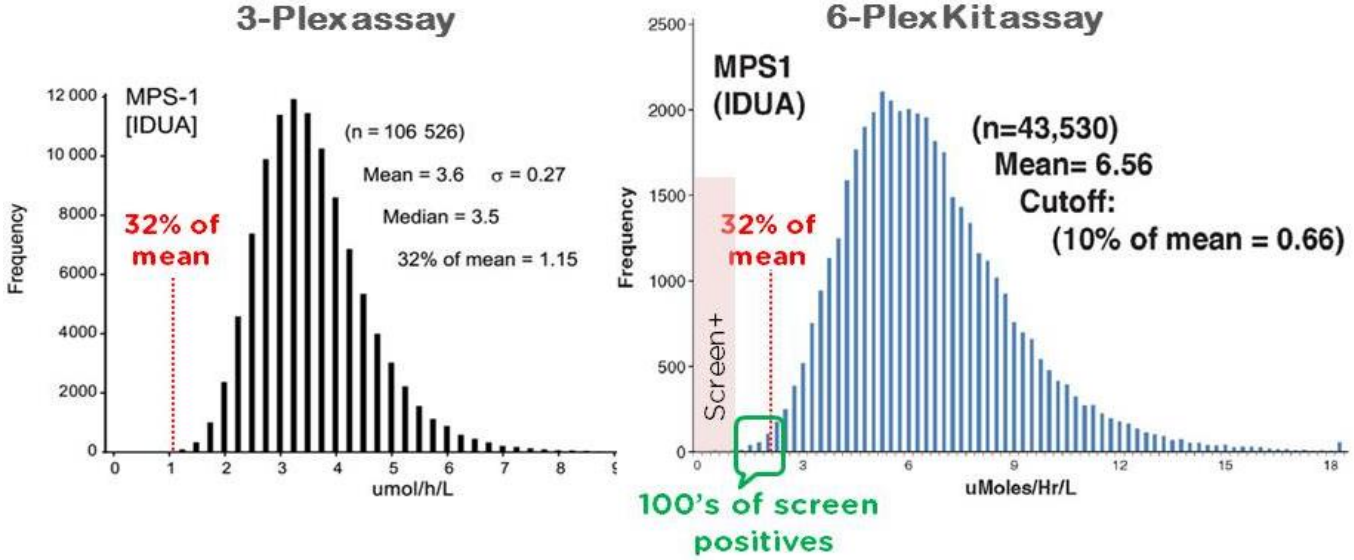

Figure 1: Applying author-optimized 32% of the mean cutoff (derived from a 3-plex MS/MS IDUA assay) to the 6-plex MS/MS IDUA assay results in hundreds of screen positive results. Adapted from Scott *et al.* (2013) and Elliott *et al.* (2016).

What happens if we apply the "equivalent cutoff" of 32%, as established for IDUA in the 3-plex MS/MS assay, to the 6-plex IDUA assay results? It should be recognized that neither of these studies are prospective newborn screening studies and therefore there may be an inherent bias towards low false positives as the true rate of false negatives will never be known. However, as shown in Figure 1, **100's of screen positive results would be found**! The high screen positive rate for the 6-plex IDUA assay that is found using the 32% cutoff may not imply that the 6-plex (or the 3-plex) is an inferior method. It simply reinforces what is already known in the newborn screening community: that the determination of an optimal cutoff requires meticulous examination of each unique assay to fit the individual needs of each newborn screening program (accomplished through pilot studies), and that a one-size-fits-all equivalent cutoff in no way accomplishes the goal of removing lab-to-lab cutoff bias from the overall performance of a given newborn screening system.

We have shown that applying the optimal 3-plex cutoff to the 6-plex assay results in an overabundance of screen positives; **what happens if the converse calculation is applied – if we apply the 6-plex optimal cutoff to the 3-plex assay?** As shown in Figure 2, a 10% optimal cutoff, as determined by the authors, produces a threshold far to the left of the normal (presumed normal) distribution for the 6-plex data. While not shown in these histograms, the disease PDF would be situated to the left of the normal distribution. If this same cutoff, however, was applied to the 3-plex assay, the threshold would land far to the left shoulder of the data around 0.36 µmol/h/L and **would have missed all the MPS-I affected samples as per the published affected enzyme activity values[3].**

According to supporting reasoning put forth for an "equivalent cutoff", methods with a high number of screen positive results are considered inferior to those with low number of screen positives without any regard to false positives. This reasoning is based on the flawed premise that differing cutoffs between laboratories are the main bottleneck to assessing the performance across laboratories/screening approaches.

One further cautionary tale in interpreting screen positive rates derived from "equivalent cutoffs" is that the screen positive rate ignores other (more important) screening results: positive predictive value and false negative rate. As shown in the latter example (Figure 2) where a 10% cutoff is applied to the 3-plex MS/MS IDUA assay, the screen positive rate for the 3-plex assay would be near zero. However, as illustrated in the histogram, there is a strong risk that false negatives would result. The screen positive rate alone does not provide a comprehensive profile of performance for an assay system.

The factors underlying the inherent differences in the PDF (shape) of these distributions (histograms) -- even in cases where the same assay system is in use -- will be explored in the near future.

These findings highlight that percentage of the mean as an "equivalent cutoff" in no way can produce an "apples-to-apples" comparison between assay systems.

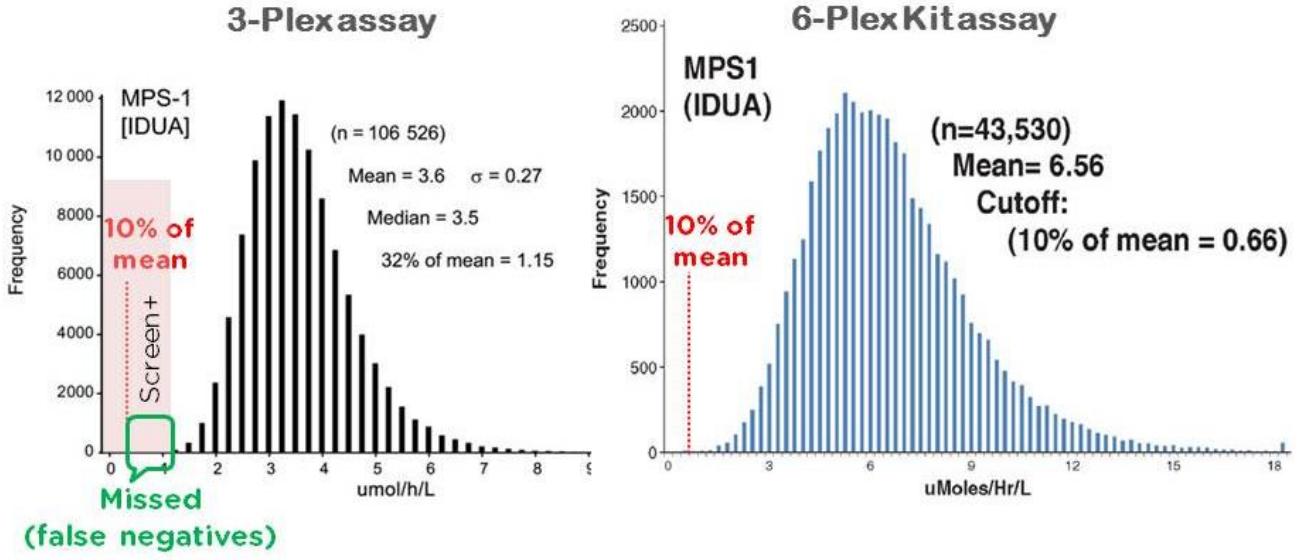

Figure 2: Applying the author-optimized 10% of the mean cutoff (derived in a 6-plex MS/MS IDUA assay) to the 3-plex MS/MS IDUA assay results in potential false negative results. Adapted from Scott et al. (2013) and Elliott et al. (2016).

[1] Schielen PC, Kemper EA, Gelb MH. *International Journal of Neonatal Screening*. 2017; 3(2):6.
[2] Gelb MH, Scott CR, Turecek F, *et al*. *Molecular Genetics and Metabolism Reports*. 2017; 12:80-81.
[3] Scott CR, Elliott S, Buroker N, *et al*. *The Journal of Pediatrics*. 2013; 163(2):498-503.
[4] Elliott S, Buroker N, Cournoyer JJ, *et al*. *Molecular Genetics and Metabolism*. 2016; 118(4):304-309.