Analytical Range is a Poor Metric for Comparing Newborn Screening Assay Performance

The term '*analytical range*' has recently been presented as a metric to compare the performance of tandem mass spectrometry (MS/MS) and fluorimetry, specifically digital microfluidic fluorimetry (DMF), for enzyme activity measurement in newborn screening (NBS) of lysosomal storage disorders (LSD). Here, we use intuitive illustrations to demonstrate why Z-factor or signal-over-noise are superior metrics for the characterization of instrument performance.

> **Signal-over-Background:** *"The assay response for the quality control HIGH sample (typical of a non-affected newborn) divided by the assay response for all processes that are independent of the relevant lysosomal enzyme (i.e., various sources of background)."*[1]
>
> **Measuring Interval:** *"A set of values of quantities of the same kind that can be measured by a given measuring instrument or measuring system with specified instrumental measurement **uncertainty**, under defined conditions."*[2]

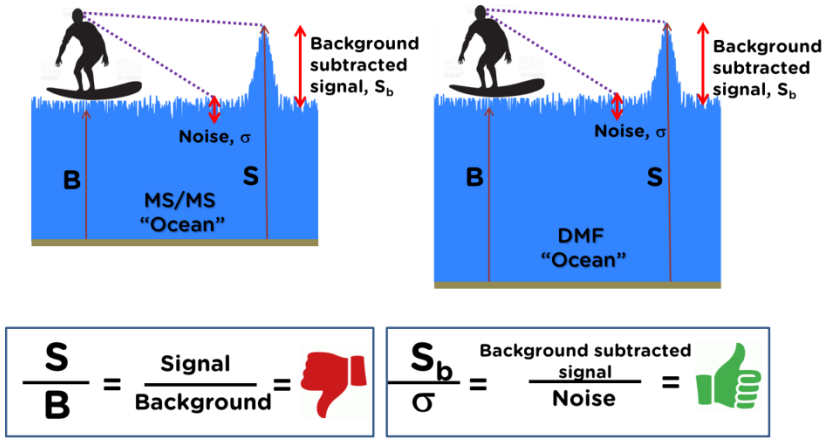
In the context of LSD NBS analytical range has been wrongly defined as **signal-over-background (S/B)** of the analytical method[1]. In practice, this definition elevates the role of assay background while ignoring the importance of 'noise' (or error) inherent to the complete assay process (i.e. sample, assay protocol and instrument) and is therefore a poor comparator for assay platforms. A well-accepted definition of analytical range, per CLSI, is **measuring interval** and must include the measurement uncertainty of each system[2]. Measurement uncertainty includes sample variance (i.e. dried blood spots (DBS) and pre-analytical steps), error introduced by the assay protocol (including pre-processing and analytical steps) and 'noise' inherent to the analyte of interest, all of which are ignored in the analytical range calculations put forth by the MS/MS literature[3,4,5]. The act of introducing a metric that does not factor in measurement uncertainty suggests unfamiliarity with signal detection theory and a disregard for well-established guidelines for assay performance nomenclature[6].

Z-factor has emerged as the gold standard quality metric for measuring high-throughput assay performance[7]. As Zhang *et al.* conclude, Z-factor should be used for comparisons of assay/instrument performance because "[signal-over-background]…does not contain any information regarding data variation; its inappropriateness in evaluation of an assay should be obvious."[7] An assay with a Z-factor value from 0-0.5 is considered marginal and a Z-factor between 0.5 and 1 is excellent.

$$Z = 1 - \frac{3(\sigma_N + \sigma_D)}{|\mu_N - \mu_D|}$$

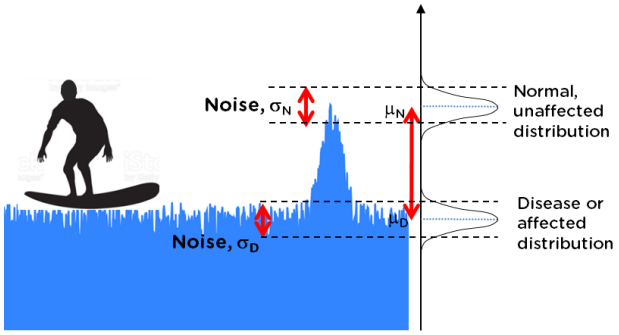
Z factor is calculated as above; Mean (μ) and standard deviation (σ) of the normal (N) and affected/disease (D) specimens are represented by μ_N, σ_N, μ_D and σ_D, respectively.

To illustrate the flawed reasoning in the 'signal-over-background (S/B)' definition for analytical range, we can imagine detecting a single wave (a surfable swell) on a noisy ocean surface. The MS/MS detection system is represented on the left while a fluorimetric system is shown on the right. For argument's sake, let us assume the "background" signal for DMF is higher than MS/MS; however, as illustrated, this background signal has no bearing on the ability to detect the "swell". If one calculates the S/B ratio relative to the MS/MS Ocean floor (lower background) and the S/B ratio for DMF relative to the DMF Ocean floor, in this illustration, we will indeed find that $(S/B)_{MS} > (S/B)_{FL}$. However, as the illustration clearly demonstrates, S/B does not influence the ability to detect that swell – in other words, *does the surfer care how "deep" the ocean is when making this determination*?



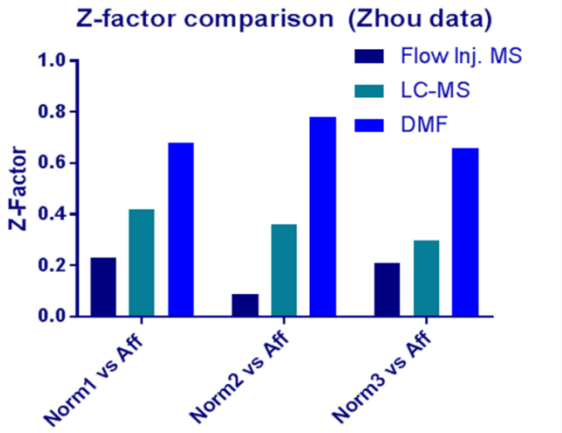
Comparison of 'analytical range' vs Z-factor metrics. "B", background; "S", signal; σ, noise.

The most suitable metric for comparison of assays is the amplitude of the swell over **noise** background; this *signal-over-noise* is fundamentally related to sensitivity in practically all measurement systems. **Z-factor** encompasses both the background subtracted signal and noise related to *both* the signal and background. The same concept of detecting a 'surfable' signal over a background is used to illustrate how Z-factor correctly takes into account the *noise* or spread of both the signal and background. To extend the analogy to enzyme measurement in lysosomal storage disorder screening, we replace the terms signal and background with the "normal" (non-affected) distribution and affected (disease) sample distribution, respectively. These distributions are characterized by the mean values ***and*** by the dispersion/spread.



Definition of Z-factor related to normal and affected distribution.

In a real-world example, the Z-factor formula was applied to data presented by the Centers for Disease Control at a Technical Workshop in 2015[8]. The data represents a repeatability study over 3 different assay and instrument types: flow-injection MS/MS, LC-MS/MS and Digital Microfluidics (DMF). The graph illustrates the calculation of the Z-factor equation from raw data on 3 different proficiency test replicates for "normal" (Norm) acid α-glucosidase (GAA) and one affected (Aff) proficiency test DBS sample. As shown, the Z-factor for the DMF method is significantly higher (>0.5) than both MS/MS methods, indicating that the fluorimetric method produced better repeatability and is better able to distinguish the distributions of normal and Pompe affected samples. Since the captured standard-deviation does not include many dominant pre-analytical noise factors (leukocyte count, pseudodeficiency alleles and sample quality due to DBS punch variation), here Z-factor is more representative of the instrument noise, variation due to extensive analytical steps in MS/MS, and assay performance. In addition, Z-factor is not dependent on the absolute enzyme activity scale since this metric examines the ability for the assay to distinguish two distributions regardless of scale. This is very important because *in vitro* enzyme activity assays (MS/MS vs fluorimetry) are run under very different conditions (buffers, pH, etc.) and thus one cannot make justifiable comparisons using the enzyme activity values themselves.



The goal in newborn screening is to adopt a method that best identifies those at-risk. Because signal-over-background ignores the variation within either the "normal" distribution or the "affected" distribution, analytical range (as defined in MS/MS LSD literature) does not encompass any information regarding the probability of an assay result being assigned to the appropriate distribution. Z-factor has been well-adapted to resolve these distributions, even when the two distributions are not clearly discriminated as is sometimes the case in newborn screening, and therefore is the most appropriate metric for comparing measuring intervals.

---

[1] Gelb MH, Scott CR, Turecek F. *Clinical Chemistry*. 2015; 61:335-346.
[2] Harmonized Terminology Database. http://htd.clsi.org/.
[3] Gelb, MH. *Molecular Genetics and Metabolism*. 2017; 120(1):S50-S51.
[4] Schielen PCJI, Kemper EA, Gelb MH. *International Journal of Neonatal Screening.* 2017; 3(2):6.
[5] Liao H-C, et al. *Clinical Chemistry*. 2017. April 27 epub ahead of print.
[6] McNicol D. A primer of signal detection theory. *Psychology Press*, 2005.
[7] Zhang J-H, Thomas DY, Chung, Oldenburg KR. *Journal of Biomolecular Screening.* 1999. 4(2): 67-73.
[8] Zhou H. 2015. https://www.aphl.org/programs/newborn_screening/Documents/LSDs-1/Zhou-APHL-NewbornScreening-LSD_150415.pdf