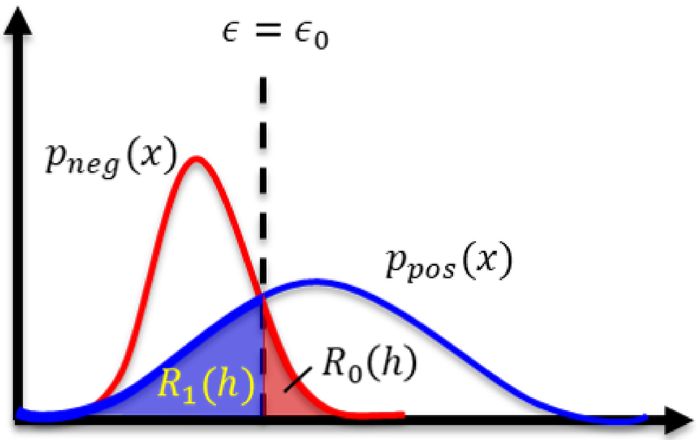


Figure S1: Two probability density functions $P_{pos}(x)$ and $P_{neg}(x)$ representing positive and negative population distributions of an assay and a random variable x representing assay values. The risk of miss-classification with optimal threshold $\epsilon = \epsilon_0$ is represented by shaded area between two distribution overlap. where $R_1(h)$ represents False Negative Rate or type II error and $R_0(h)$ represents False Positive Rate or type I error.

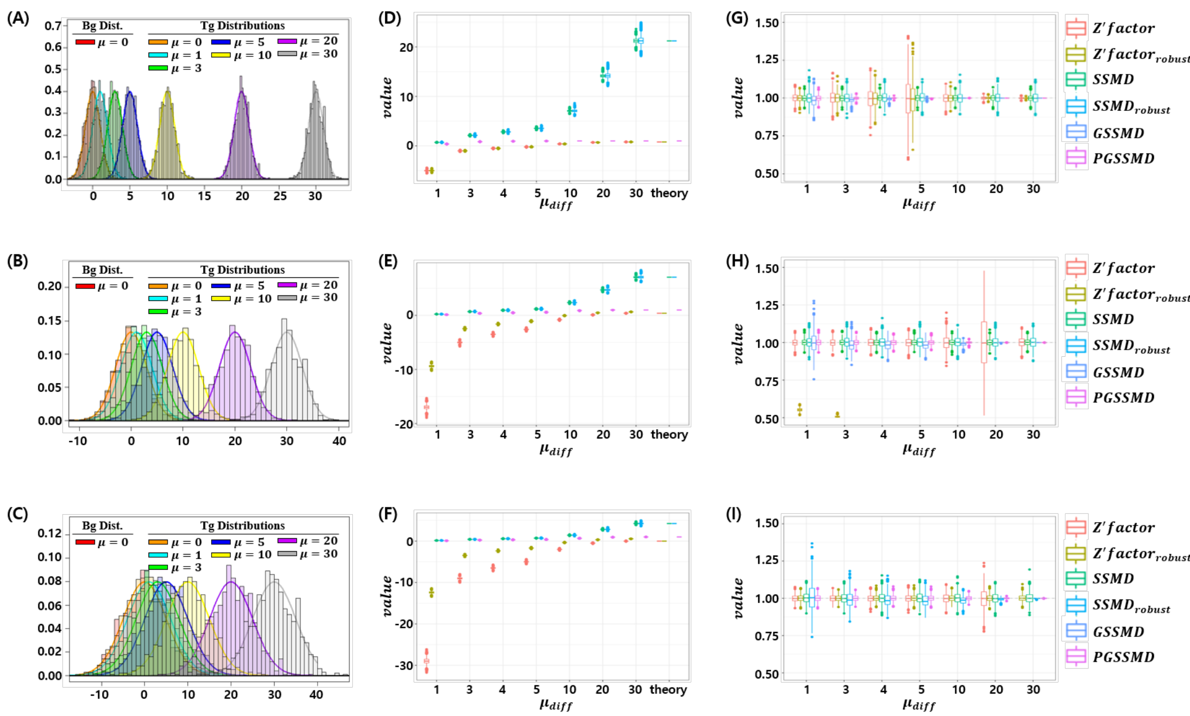


Figure S2: GSSMD provided interpretable and reliable statistical difference estimate in the ideal simulation setup ($\sigma_c = 1,\ 3,\ 5\ ,\ n\ =\ 1,000$ in (A), (B), and (C) respectively). The results in (D) to (I) are acquired from random sampling of $N = 10^6$. (D), (E) and (F) are boxplots of each measure for a given mean difference and standard deviation. (G), (H), and (I) are boxplots of normalized measure for a given mean difference and standard deviation. Here we divided all values by corresponding theoretical value. The standard deviation of the sample SSMD was not sensitive to the mean difference, but the standard deviation of the sample GSSMD decreased as the mean difference increased.

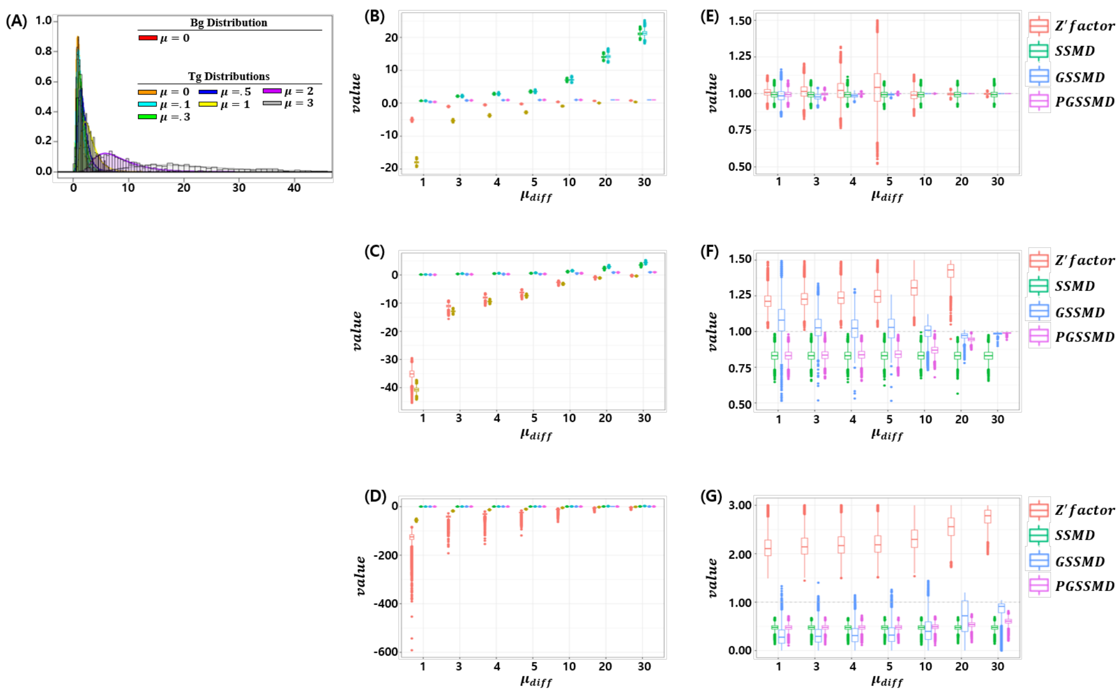


Figure S3: Performance of each measure in log-normal distribution setup. (A) shows log-normal distribution with $\sigma = 0.1$ and various mean differences. The results in (B) to (G) are acquired from random sampling of $N = 10^6$. (B), (C) and (D) are boxplots of each measure for a given mean difference and standard deviation. (E), (F), and (G) are boxplots of normalized (by its corresponding theoretical value) measure for a given mean difference and standard deviation.

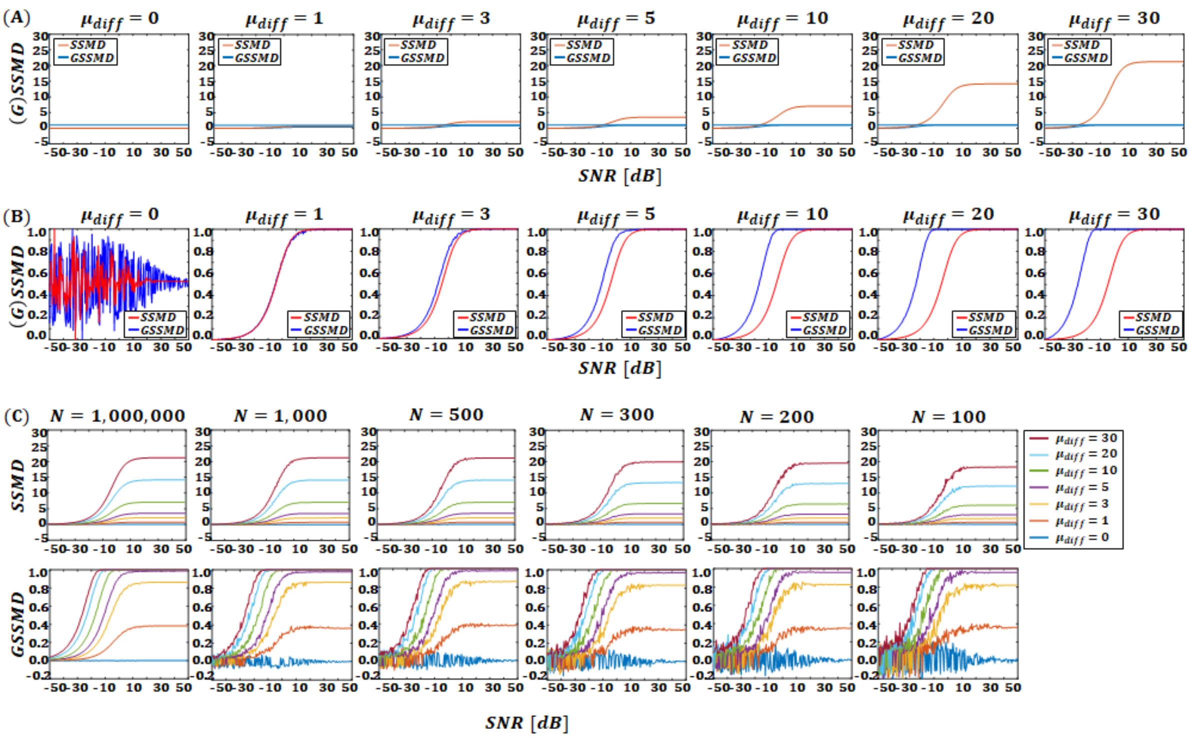


Figure S4: GSSMD is better quality metric for the noisy data sets. (A) The range of SSMD increases with the mean difference, so the meaning of the values can vary depending on the mean difference. Unlike SSMD, GSSMD is limited from 0 to 1 (in case of positive overlap), so the meaning of the value is not affected by the mean difference. (B) Scaled plot for each measure. GSSMD is more sensitive when the mean difference is large. Here we scaled the value from 0 to 1. (C) SSMD and GSSMD with various sample sizes, noise levels and mean difference. GSSMD identifies mean differences even in the high noise regions. x-axis represents SNR in deci-Bell (dB) scale for all figures.

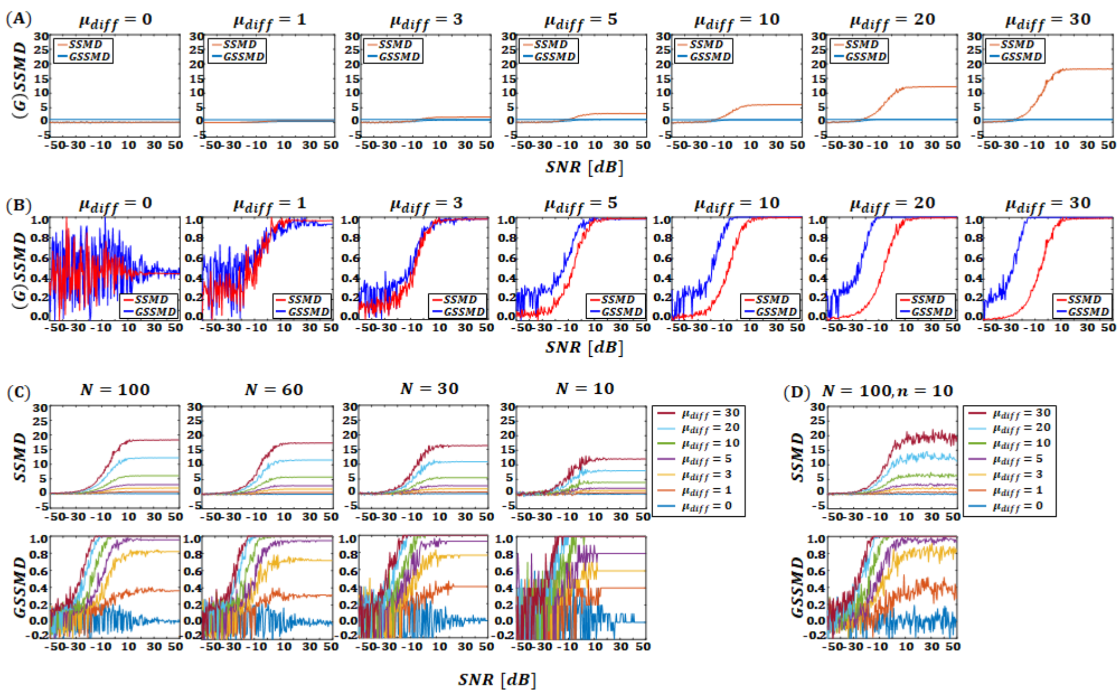


Figure S5: GSSMD can be used for small and noisy data sets. ($N \leq 100$) (A) SSMD and GSSMD with sample size 100. The curve is little noisier but the trend is maintained. (B) Scaled plot for each measure. GSSMD is more sensitive when the mean difference is large, as shown in the figure 4 (B). Here, we adjusted the value from 0 to 1 too. (C) SSMD and GSSMD with sample size less than 100. The signal is too deformed to distinguish mean difference when N = 10. (D) GSSMD and SSMD estimation using resampling. GSSMD and SSMD can be estimated by sampling 10 of the 100 samples. (10 runs) The sensitivity of GSSMD is slightly reduced but still better than SSMD.

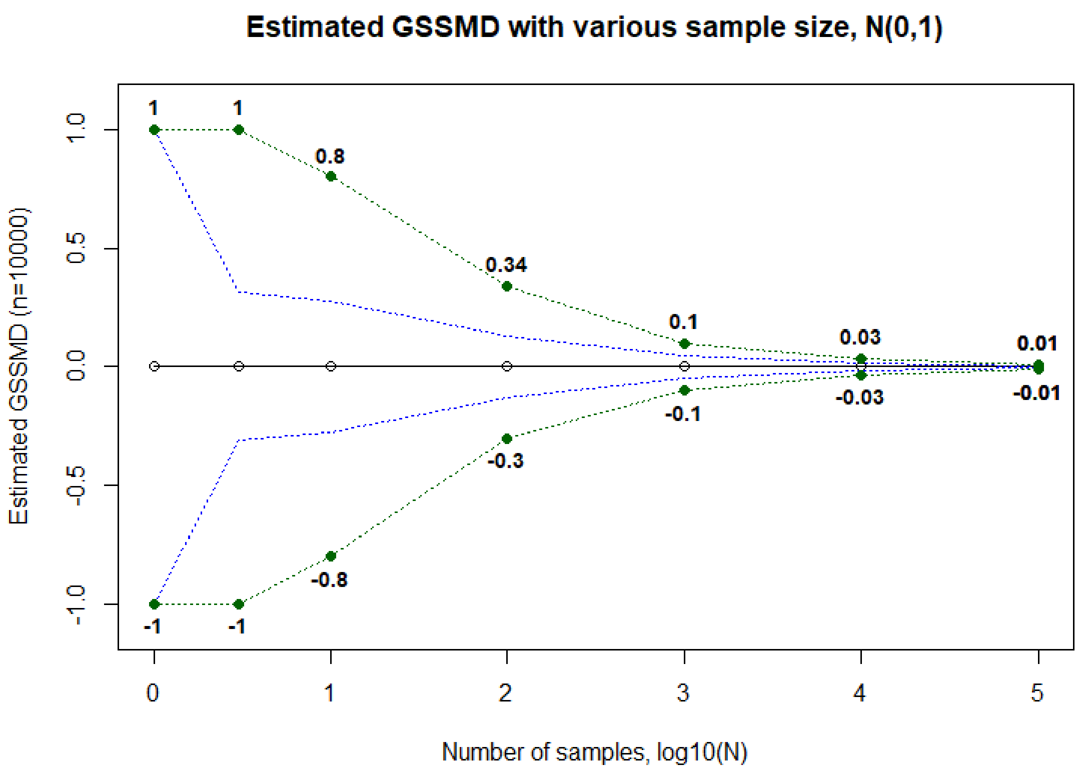


Figure S6: Effect of sample size on GSSMD when normal distributions overlap completely ($\mathcal{N}(0,1)$). The blue line represents the standard deviation and the green line represents the maximum and minimum of the estmated GSSMD. The number of independent trials is $n = 10^4$

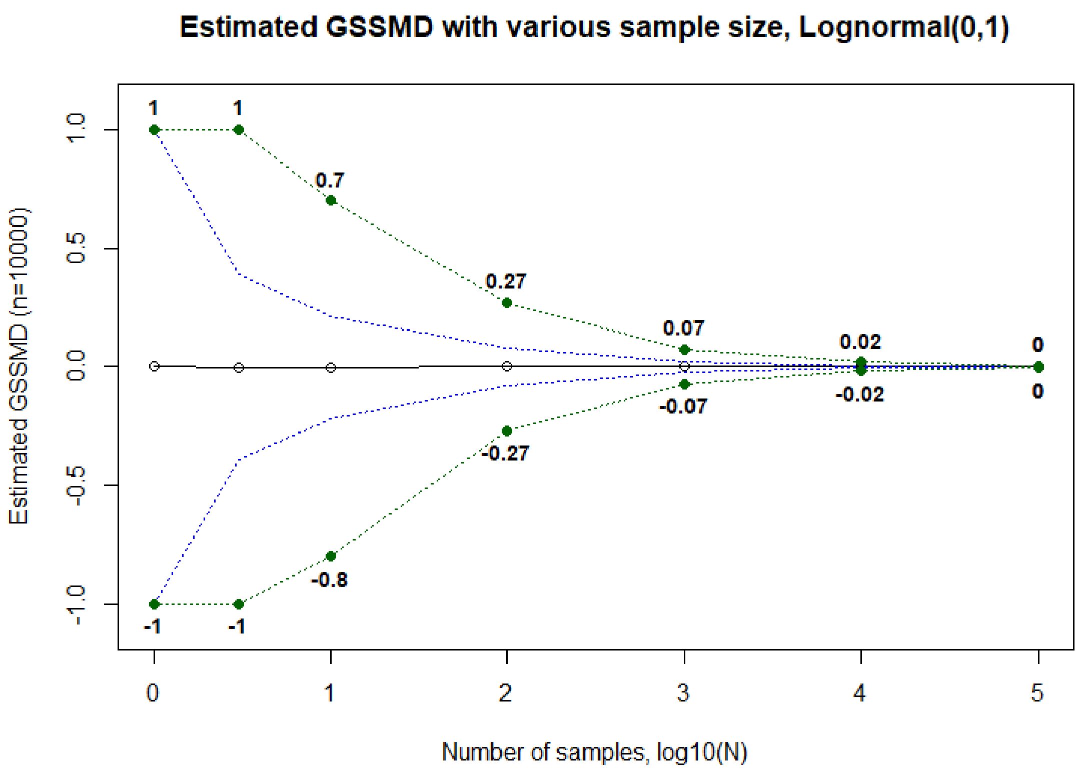
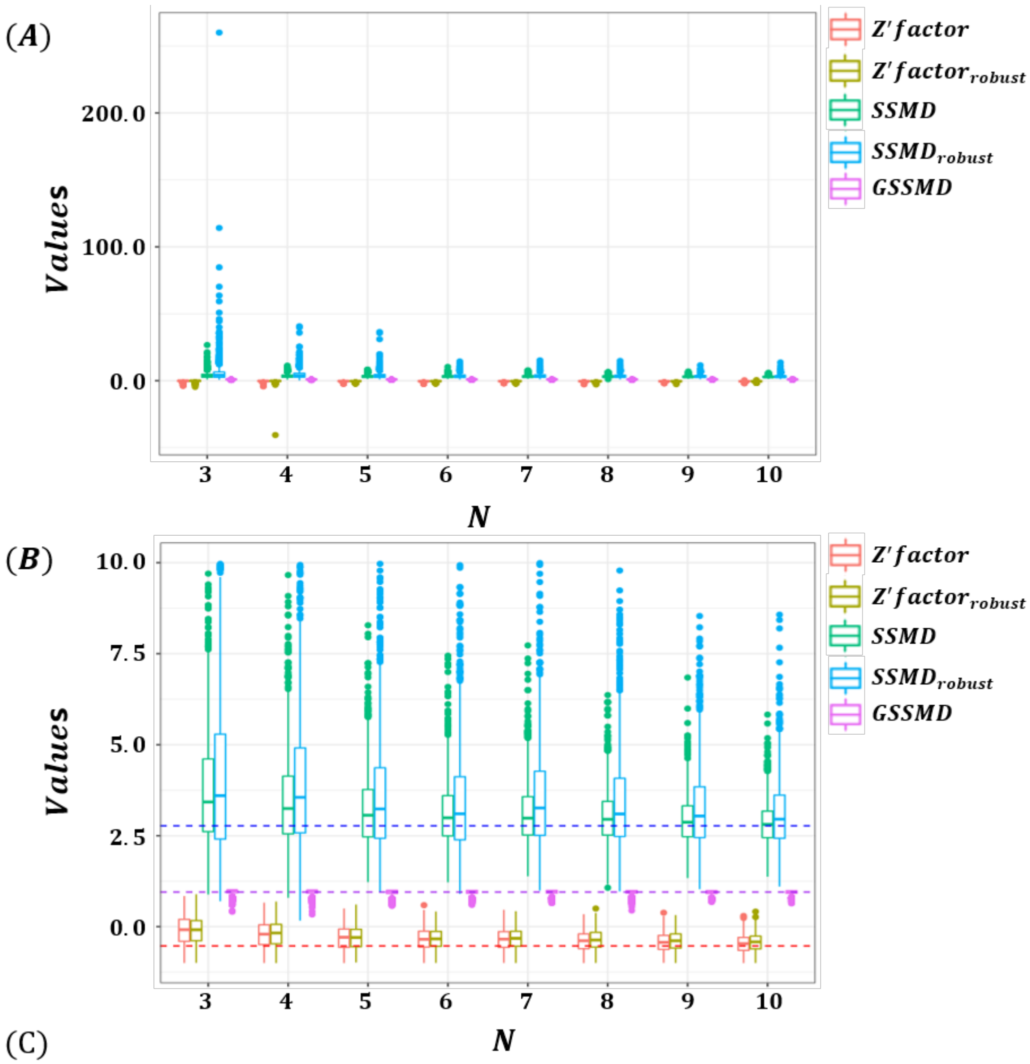


Figure S7: Effect of sample size on GSSMD when log-normal distributions overlap completely ($Log\mathcal{N}(0,1)$). The blue line represents the standard deviation and the green line represents the maximum and minimum of the estmated GSSMD. The number of independent trials is $n = 10^4$



(C)

N	3	4	5	6	7	8	9	10	*Theory*
SSMD	4.085 ±2.405	3.512 ±1.376	3.247 ±1.067	3.133 ±0.947	3.126 ±0.873	3.030 ±0.759	2.939 ±0.670	2.860 ±0.606	2.77
GSSMD	0.950 ±0.083	0.946 ±0.081	0.941 ±0.077	0.942 ±0.072	0.947 ±0.065	0.942 ±0.069	0.941 ±0.061	0.938 ±0.062	0.95

Figure S8: Parametric estimation of GSSMD showed least variance in the estimation of true effect size. $\mu_0 = 0, \mu_1 = 3.92, \sigma = 1$ for all sampling distributions. (A) Original scale plot showed large variance in SSMD and robust SSMD estimation. (B) y scale is modified to provide mode detailed plot. (C) Calculated values of SSMD and GSSMD from the plot (A) and (B).

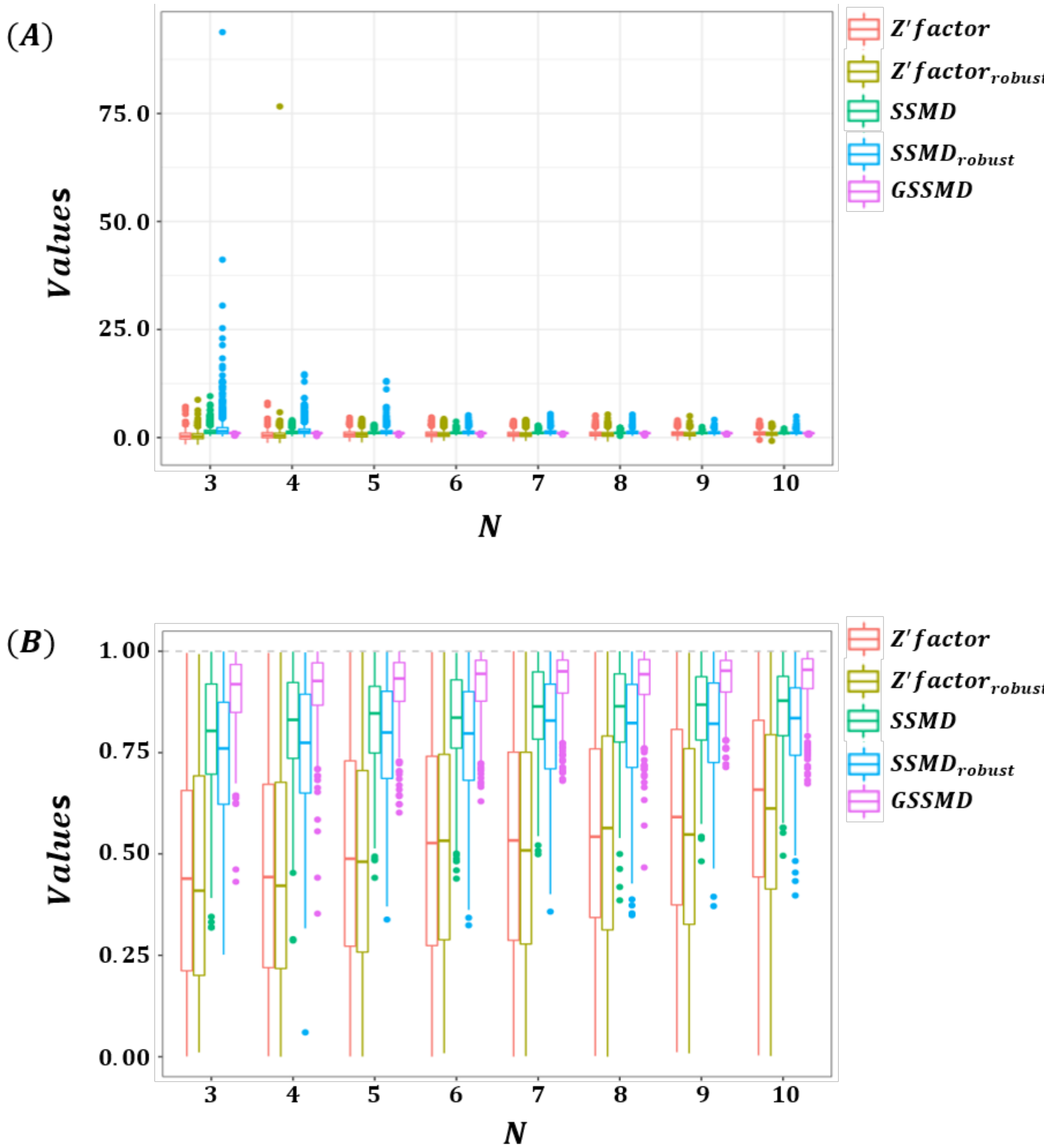


Figure S9: Parametric estimation of GSSMD showed least variance in the estimation of true effect size. (Scaled Plot) Here we scaled all values by its theoretical values. (A) Original scale plot showed large variance in SSMD and robust SSMD estimation. (B) y scale is modified to provide mode detailed plot. In (B), We can see that GSSMD is the best estimator of true population effect size even though we have very small sample size ($N = 3 \rightarrow 10$). Therefore, we can use this parametric version of GSSMD for the very small sample size problem.

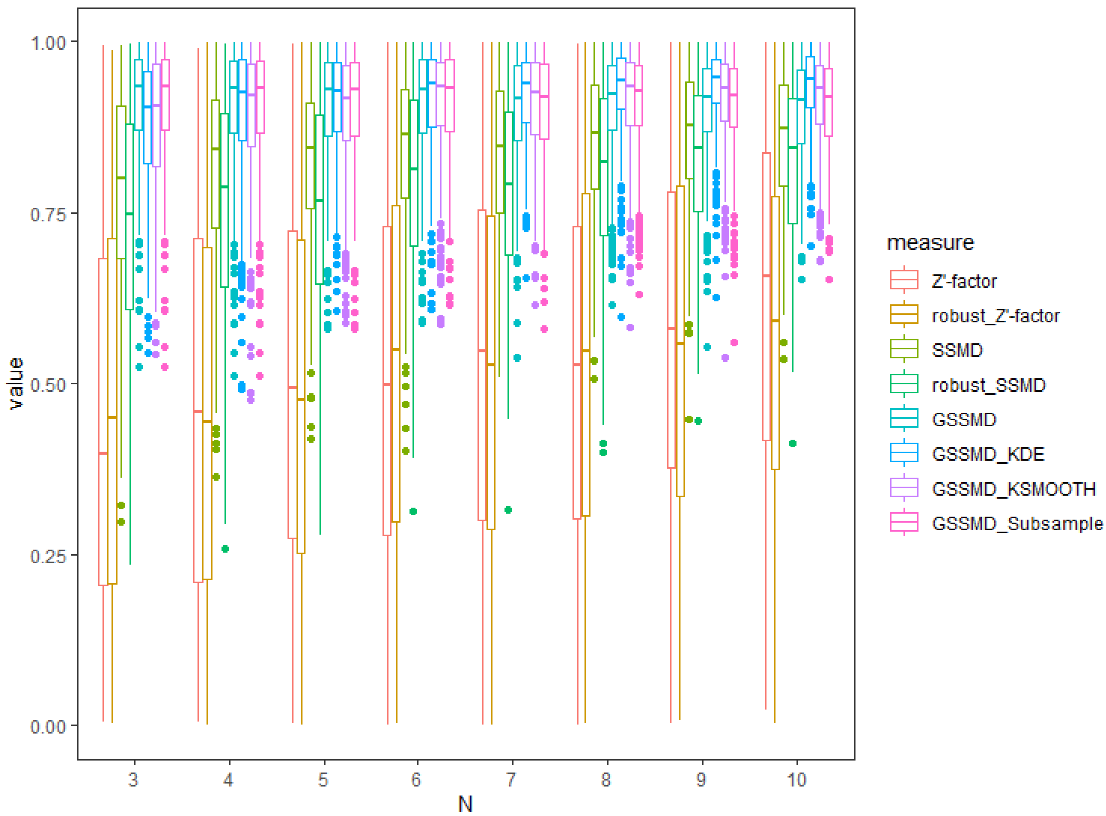


Figure S10: GSSMD are calculated using various methods such as historgram, KDE, KSMOOTH, and subsampling methods. These values are compared with Z'-factor and SSMD. The simulation setting is the same as figure S6 and S7.

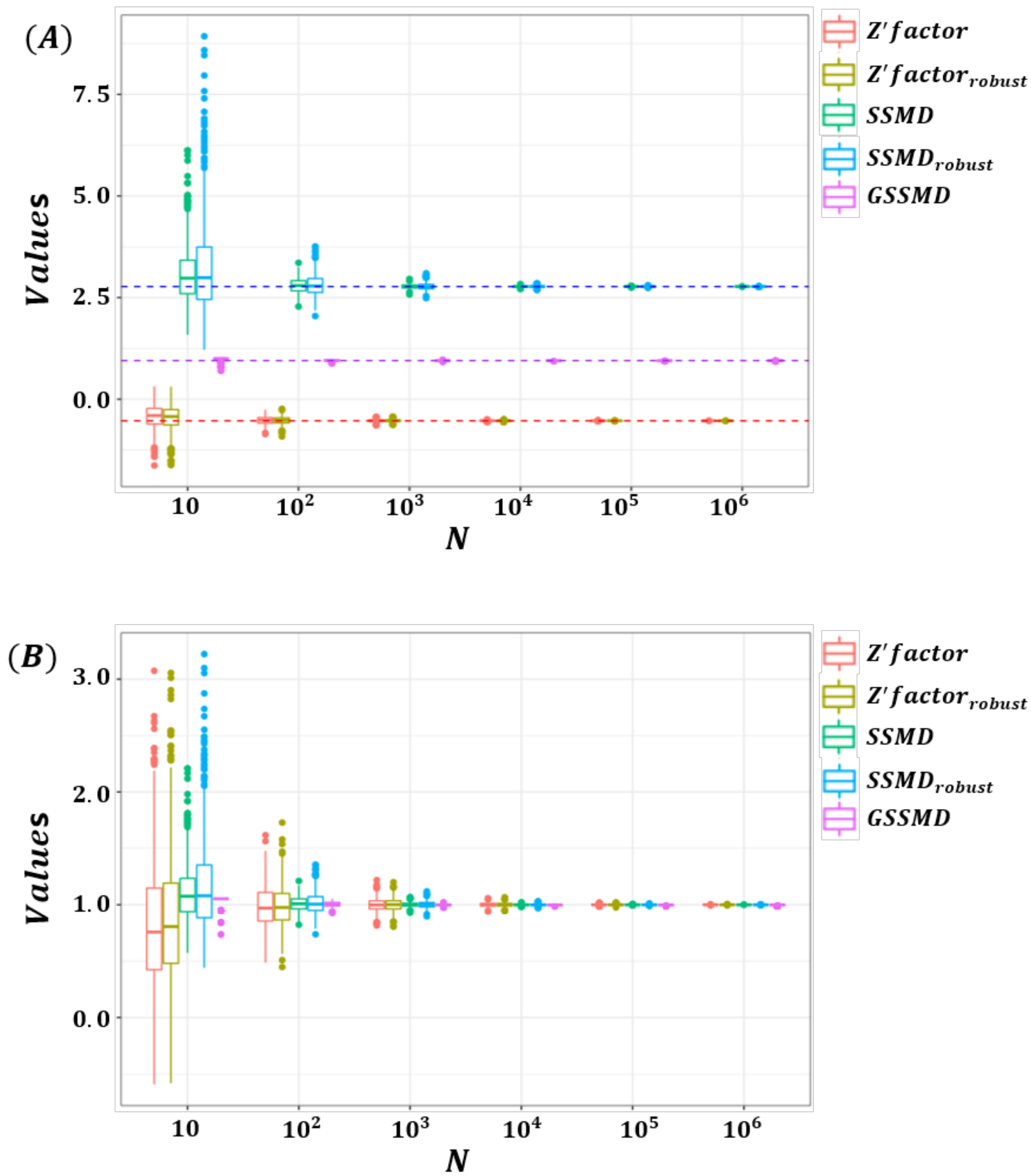


Figure S11: Estimation of GSSMD in large sample size. (A) GSSMD showed least variance in the estimation of true effect size where dashed lines illustrate true population effect size conditioned by the simulation. (B) We normalized each estimated measure by its theoretical value (SSMD $= 2.77$, GSSMD $= 0.95$, Z'-factor $= 0.5$). Similarly, as shown in the figure, GSSMD showed least variance compared to all measures.

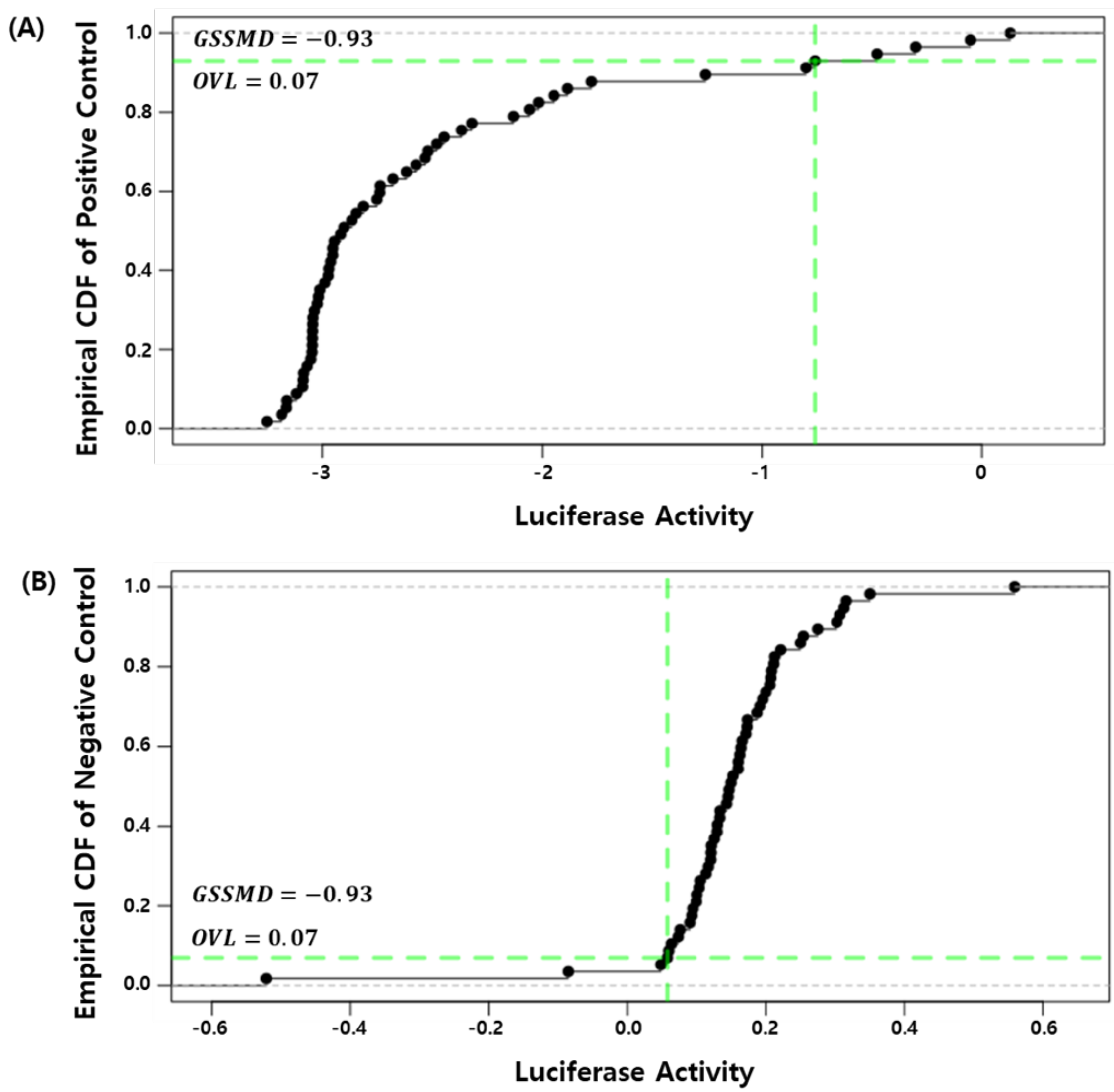


Figure S12: Selection of upper and lower bounds of threshold for hit selection. (A) Empirical CDF of positive control distribution. The horizontal green line represent 7% of ECDF corresponding to the total OVL area between positive negative distributions. (B) Empirical CDF of negative control distribution.