Mathematical relationships between control group variability and assay quality metrics

Andrew Lim[1]

[1]*The ALBORADA Drug Discovery Institute, University of Cambridge*

Corresponding Author:

Andrew Lim, The ALBORADA Drug Discovery Institute, University of Cambridge, Island Research Building, Cambridge Biomedical Campus, Hills Road, Cambridge CB2 0AH, UK.

Email: astl3@cam.ac.uk

Abstract

Assay quality metrics have been used in various high-throughput screening (HTS) campaigns to indicate assay quality. Z'-factor has become one of the most widely used metrics, along with other metrics such as standardized mean difference (SSMD). In using these metrics, it is important to understand how these metrics can be impacted by the separation between control groups (indicated by the HZ ratio) and the coefficient of variation (CV) within each control group. In this paper, several mathematical equations have been derived to understand the relationship between assay quality metrics (such as Z'-factor and SSMD) and control group datasets (summarised by CV and HZ). These equations increase our understanding of the factors that improve assay quality metrics, thus providing a quantitative means to visualise how affecting control groups can impact assay quality metrics.

Keywords: assay quality metrics, Z'-factor, coefficient of variation, high-throughput screening, data analysis

Graphical Abstract

Let $CV_H = \frac{\sigma_H}{\mu_H}$, $CV_Z = \frac{\sigma_Z}{\mu_Z}$ and $HZ = \frac{\mu_H}{\mu_Z}$, where
σ_H = Standard deviation of the HPE group
σ_Z = Standard deviation of the ZPE group
μ_H = Mean of the HPE group
μ_Z = Mean of the ZPE group
HPE - 100% effect, ZPE - 0% effect

$$\mathbf{CV_H}\left(\frac{\mathbf{HZ}}{\mathbf{|HZ-1|}}\right)+\mathbf{CV_Z}\left(\frac{\mathbf{1}}{\mathbf{|HZ-1|}}\right)=\frac{\mathbf{1-Z'}}{\mathbf{3}} \qquad \mathbf{CV_Z}=-\mathbf{CV_H HZ}+\frac{\mathbf{1-Z'}}{\mathbf{3}}(\mathbf{|HZ-1|})$$

$$\mathbf{SSMD}=\frac{\mathbf{HZ-1}}{\sqrt{(\mathbf{CV_H HZ})^2+\mathbf{CV_Z}^2}} \qquad \mathbf{|SSMD|}=\frac{\mathbf{3(HZ+1)}}{\mathbf{(1-Z')}(\sqrt{\mathbf{HZ^2+1}})}$$

Where Z' = required Z'-factor and SSMD = standardized mean difference

1. Introduction

Maintenance of assay quality in assay development is crucial in ensuring success in high-throughput screening (HTS) campaigns in hit discovery programmes. The use of assay quality metrics is important to ensure that the assay is capable of providing robust information on the compound activity. Consequently, a high-quality assay should enable the user to identify compounds with significant biological activity ('hits') confidently. This is important as compounds are usually tested once in a primary screening stage in HTS campaigns.

There are several assay metrics which have been used historically in HTS campaigns. Previously, signal-to-noise (SN) ratio and signal-to-background (SB) ratios have been used to determine assay quality. Although SN and SB are used interchangeably, there is a subtle

difference between the two. SN originally refers to the ratio of the difference between the mean signal and mean background to the standard deviation of the background. In the context of positive and negative control groups, SN can refer to the ratio of the difference of the mean difference of two control groups to the standard deviation of the negative control group. SB, on the other hand, refers to the ratio of the mean signal to the mean background. While these metrics are simple, SN only considers the variability in the background, while there is no consideration of variability in SB.

Other metrics have been proposed to measure assay quality more comprehensively by considering the variability of the control groups and the extent of separation between the control groups. Sittampalam et al. proposed the concept of signal window (SW) which takes the ratio between the distance from the upper noise limit of one group to the lower noise limit of another group and the standard deviation of the negative control group.[1] Zhang et al. developed the concept of a signal window by dividing the distance between the two groups (now defined as separation band) by the difference of means between the two groups, thus introducing the concept of the Z-factor.[2] The Z'-factor is an extension of the Z-factor by utilizing the means and standard deviations of the positive control group ($\mu_{c+}$and σ_{c+} respectively) and negative control group ($\mu_{c-}$and σ_{c-} respectively):

$$Z' = 1 - \frac{3(\sigma_{c+} + \sigma_{c-})}{|\mu_{c+} - \mu_{c-}|}$$

Since its conception, the Z'-factor has been used extensively as an assay quality metric in HTS campaigns. The utility of the Z'-factor has been studied, with analysis suggesting that Z'-factor is a good assay metric compared to SW.[3] Although it is widely accepted that a Z'-factor of 0.5 is required for an assay to be deemed acceptable, the value of 0.5 was

determined by convention rather than a theoretical basis. As such, there is an argument for the possibility of accepting assays with a Z'-factor less than 0.5 under certain circumstances.[4]

The Z'-factor is strictly an assay quality parameter, hence it does not quantitatively identify the number of hit compounds in a HTS campaign.[5] To overcome this point, Sui and Wu link power calculation with Z'-factor.[6] Additionally, another assay quality metric has been proposed known as strictly standardized mean difference (SSMD).[7] Originally used in RNA interference (RNAi) HTS campaigns, this metric was argued to be more theoretically sound than Z'-factor.

Regardless of the choice of the assay quality metric, it is important to understand how variables can impact the chosen metric. It is safe to assume that changing the variance in each control group and the magnitude of separation between the two controls will affect the assay metric. However, the exact impact of such change has not yet been studied extensively. So far, the Z-factor and Z'-factor are found to be directly linked to several assay variables, for example, SB and coefficient of variation (CV)[6] and newly-defined assay coefficient of variation (CV_A)[3]. The use of CV_A which is defined as the ratio of the standard deviation of the maximum signal control to the difference of the mean between the two controls summarises the combined CV of both control groups, but cannot be broken down into its constituent CVs for each control group.

Understanding the impact of each control group on assay quality metrics is beneficial in the assay development process, where the user might be interested to know if changing the variance of one control group over the other would impact the assay quality metric more. In

addition, it is also worth knowing to what extent an increase in the separation window between the two control groups improves the assay quality metric.

To do so requires a breakdown of assay quality metrics into their assay parameters for each control group. This includes the CVs of each control group and a measure of signal separation between the two control groups. In this study, I begin by clarifying some terms and defining some assay variables, and then I proceed to express two assay metrics, namely Z'-factor and SSMD, using these variables while considering two different scenarios – equal CVs for both groups and different CVs for each group.

2. Methods and results

2.1 Definition of terms and concepts

There are several concepts and terms which require definitions and clarifications. First, I would like to introduce the concepts of one hundred per cent effect (HPE) and zero per cent effect (ZPE). Most of the current discussion around assay design revolves around the terms positive control (PC) and negative control (NC). For the clarity of the argumentation, I would like to use the terms HPE and ZPE to denote the set of controls that gives a hundred per cent effect and zero per cent effect respectively in the assay. The said effect is linked to the aims of the assay. For example, in an inhibitor discovery project, the HPE refers to the fully inhibited controls while the ZPE refers to the untreated controls. On the other hand, in an agonist discovery project, the HPE refers to the fully stimulated controls while again the ZPE refers to the untreated controls. By using HPE and ZPE, one can surmise that ZPE would almost always refer to the untreated controls while HPE refers to controls treated pharmacologically with a tool compound.

One concept that can be taken for granted in the discussion of assay quality is the ratio between the two sets of controls. While SB refers to the ratio of the mean signal to the mean background, in assay development this is often translated to the ratio between the mean of the positive controls (as mean signal) and the mean of the negative controls (as mean background). However, it should be noted that in most cases, especially in engineering, the SB ratio usually implies that the signal has a greater magnitude than the background noise and therefore SB ratio is greater than one. This notion of the signal being greater than the background noise may lead to the conception (or misconception) that the SB ratio is simply the ratio between a large number and a small number and as such can only be greater than one.

However, in assay design, the assumption that HPE (or positive) control has a greater signal than ZPE (or negative) control is not necessarily true. Previously, it was suggested that assays can be classified based on the raw signal used to characterise biological activity, for example, an activation assay whereby a compound increases the raw signal and an inhibition assay whereby a compound decreases the raw signal.[1] However, the notion that the increase in raw assay signal is exclusively linked to an activation assay and vice versa is not necessarily true.

The 'directionality' of the assay can go either way, increasing or decreasing, independent of the biological activity. For example in a kinase inhibitor project, one can design an assay based on substrate depletion (monitoring consumption of ATP) or product formation (monitoring formation of ADP). The HPE and ZPE controls in both assay designs are the same, whereby the HPE controls refer to controls treated with a tool compound (for example, a known kinase inhibitor) and the ZPE controls refer to untreated controls (DMSO controls).

Even though the HPE and ZPE controls are the same in both assay design scenarios, the 'directionality' of the assay differs, and as such, depending on the 'directionality' of the assay (increasing signal or decreasing signal), the ratio between HPE controls and ZPE controls can be less than one when ZPE controls have a larger raw signal which decreases upon pharmacological treatment. HPE and ZPE are fixed to the objectives of the assay, and are therefore independent of the 'directionality' of the assay, as Table 1 summarises.

Table 1: The definition of HPE and ZPE is independent of the ratio between HPE and ZPE, using an example of a kinase inhibitor project and its possible assays.

ATP + Substrate → ADP + Phosphorylated substrate

Assay kit to measure activity	Kinase-Glo® (luminescence)	ADP-Glo™ (luminescence)	Transcreener® ADP2 FP (TR-FRET)
Mode of measurement	ATP depletion (substrate depletion)	ADP formation (product formation)	ADP formation (product formation)
Relative magnitude of HPE (100% inhibition) signal	High	Low	High
Relative magnitude of ZPE (untreated) signal	Low	High	Low
Ratio HPE/ZPE	> 1	< 1	> 1

To avoid creating confusion on the concept of SB ratio and whether the SB ratio must be greater than one, I introduce a similar ratio called HPE to ZPE ratio (abbreviated as HZ in this article). HZ is defined as

$$\mathrm{HZ} = \frac{\mu_H}{\mu_Z}$$

where μ_H is the mean of HPE controls and μ_Z is the mean of ZPE controls. By using this definition, HZ can either be bigger than one (when $\mu_H > \mu_Z$) or less than one (when $\mu_Z > \mu_H$). A key assumption in the formulation of the HZ ratio is that both means are positive values, thus the HZ ratio is a non-zero positive value.

The coefficient of variation (CV) is the ratio of the standard deviation to the means:

$$\mathrm{CV} = \frac{\sigma}{\mu}$$

The CV has been used as a measure of variability to the mean of the group. CVs for HPE and ZPE groups, respectively known as $\mathrm{CV_H}$ and $\mathrm{CV_Z}$, are defined based on their respective means and standard deviations of each group:

$$\mathrm{CV_H} = \frac{\sigma_H}{\mu_H} \text{ and } \mathrm{CV_Z} = \frac{\sigma_Z}{\mu_Z}$$

The Z'-factor is an assay quality metric used commonly in high-throughput screening (HTS) campaigns based on the relationship between the means and standard deviations of the HPE and ZPE groups:

$$Z' = 1 - \frac{3(\sigma_H + \sigma_Z)}{|\mu_H - \mu_Z|}$$

It is worth noting that the equation considers the absolute difference between the means of the controls, thus eliminating any information on the 'directionality' of the assay.

2.2 Expressing Z'-factor as a function of CV and HZ

Part 1: Assumption of equal CVs in HPE and ZPE

By using the definitions of HZ, CV and Z'-factor, a relationship between Z'-factor and both CV and HZ can be derived so that the user can observe how CV and HZ can impact Z'-factor. For clarification, this is not an attempt to support the use of HZ or CV as a sole assay quality metric, especially when HZ does not measure variability. Instead, this exercise provides insight into how impactful modifying HZ and/or CV can be on improving the Z'-factor in the process of assay development, especially when an initial pilot study does not indicate a good Z'-factor and efforts are required to improve the Z'-factor. In that scenario, which is more important or impactful in affecting and improving Z'-factor – HZ or CV?

First, by rearranging the Z'-factor equation, we get:

$$\frac{\sigma_H + \sigma_Z}{|\mu_H - \mu_Z|} = \frac{1 - Z'}{3}$$

Assume that the CV is the same for HPE and ZPE, i.e. $\mathrm{CV} = \frac{\sigma_H}{\mu_H} = \frac{\sigma_Z}{\mu_Z}$

$$\frac{\mathrm{CV}\mu_H + \mathrm{CV}\mu_Z}{|\mu_H - \mu_Z|} = \frac{1 - Z'}{3}$$

$$\frac{\mathrm{CV}(\mu_H + \mu_Z)}{|\mu_H - \mu_Z|} = \frac{1 - Z'}{3}$$

Rearranging the equation for the CV in terms of the control means:

$$\mathrm{CV} = \frac{1 - Z'}{3}\left(\frac{|\mu_H - \mu_Z|}{\mu_H + \mu_Z}\right)$$

Using the definition of HZ, it goes that $\mu_H = \mathrm{HZ} \cdot \mu_Z$. Replacing this relationship into the right-hand side (RHS) of the equation and simplifying it gives:

$$\mathrm{CV} = \frac{1 - Z'}{3}\left(\frac{|\mathrm{HZ} \cdot \mu_Z - \mu_Z|}{\mathrm{HZ} \cdot \mu_Z + \mu_Z}\right)$$

$$\mathrm{CV} = \frac{1 - Z'}{3}\left(\frac{|\mu_Z||\mathrm{HZ} - 1|}{\mu_Z\,(\mathrm{HZ} + 1)}\right)$$

With the assumption that ZPE control means are positive values so that $|\mu_Z| = \mu_Z$, the variables can cancel each other out and this leads to the final equation:

$$\mathbf{CV} = \frac{\mathbf{1} - \boldsymbol{Z'}}{\mathbf{3}}\left(\frac{|\mathbf{HZ} - \mathbf{1}|}{\mathbf{HZ} + \mathbf{1}}\right) \qquad (\mathbf{1})$$

Equation 1 provides an insight into the relationship between Z'-factor and CV and HZ (Figure 1A), confirming previous findings by Sui and Wu[6]. Additionally, equation 1 can be plotted as two profile plots (Figures 1B and 1C) to observe the impact of CV or HZ on Z'-factor whilst the other variable remains constant.

In Figure 1A, the graph can be divided into two segments, the first segment representing HZ > 1 (where HPE control groups have larger signal than ZPE control groups) and the second segment representing HZ < 1 (where HPE control groups have smaller signal than ZPE control groups). In the first segment, as HZ increases (denoting a greater separation between HPE and ZPE, with HPE > ZPE), the CV increases to the asymptote limit of $\frac{1-Z'}{3}$. On the other hand, in the second segment, CV increases to $\frac{1-Z'}{3}$ as HZ decreases (denoting a greater

separation between HPE and ZPE, with ZPE > HPE). Even though the line crosses at the y-intercept of $\frac{1-Z'}{3}$, in practice, it is not possible as it would require HZ to be zero.

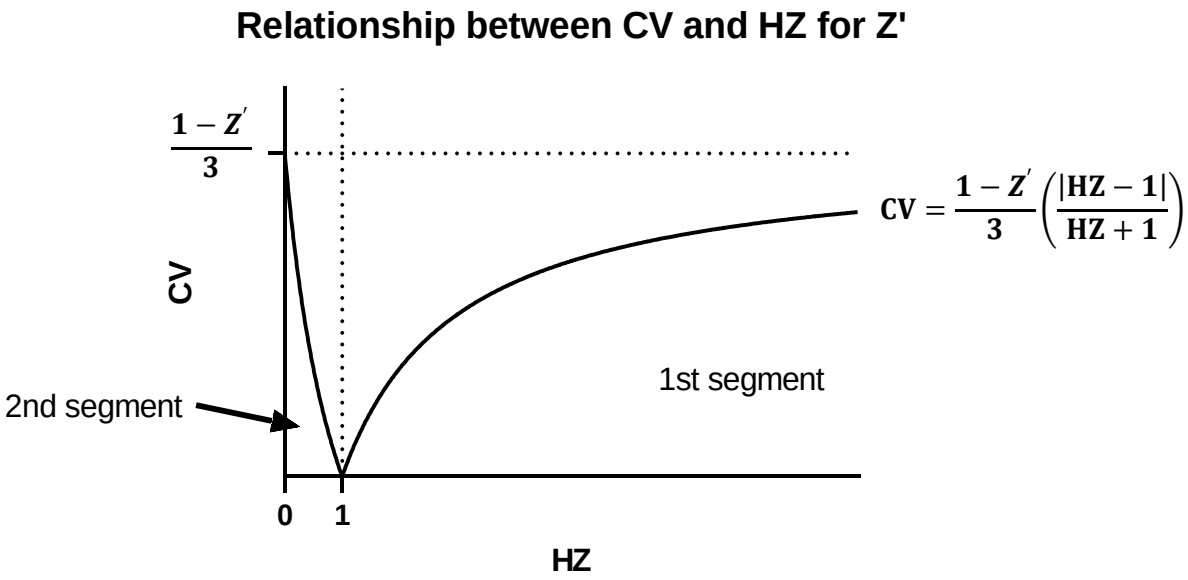
In both cases, one can infer from Equation 1 that the limit of the CV is defined by $\frac{1-Z'}{3}$, with Z' being the Z'-factor threshold determined by the user. For example, with a Z'-factor threshold of 0.50, which is a commonly used minimum threshold to describe an excellent assay, one could use Equation 1 to determine that the maximum CV limit is (1 – 0.50)/3 = 0.167, or 16.7%. In other words, the maximum CV tolerable to satisfy the 0.50 Z'-factor threshold is 0.167, and that value decreases with a lower degree of separation between HPE and ZPE. For example, in an assay with a small degree of separation such as a two-fold difference between HPE and ZPE (i.e. HZ of either 2 or 0.5), the maximum tolerable CV to satisfy the Z'-factor minimum threshold of 0.50 decreases to approximately 0.056 (or 5.6%). We arrive at the same conclusions when equation 1 is visualized in both profile plots (Figure 1B and 1C). Note that the maximum limit of the CV is impacted by the level of separation between the two control groups and not by the assay 'directionality' (denoted as to whether HZ > 1 or HZ < 1). Proof of this is noted in Appendix 1.

With that, Equation 1 can be reformulated as an inequality to determine the range of CV values to satisfy the Z'-factor threshold *k*, so that Z' ≥ *k*:

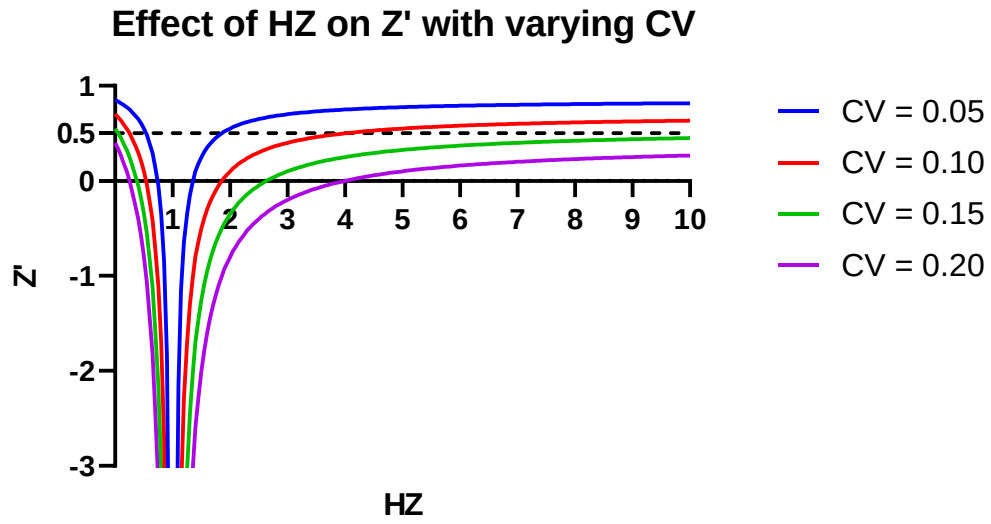
$$\mathbf{CV} \leq \frac{\mathbf{1}-\boldsymbol{k}}{\mathbf{3}}\left(\frac{|\mathbf{HZ}-\mathbf{1}|}{\mathbf{HZ}+\mathbf{1}}\right) \qquad (\mathbf{1b})$$

The inequality 1b and the shaded region that satisfies the inequality are shown in Figure 1B.

A


Relationship between CV and HZ for Z'
$\frac{1-Z'}{3}$
CV
$CV = \frac{1-Z'}{3}\left(\frac{|HZ-1|}{HZ+1}\right)$
2nd segment
1st segment
0
1
HZ

B


Effect of HZ on Z' with varying CV
1
0.5
0
-1
-2
-3
Z'
1
2
3
4
5
6
7
8
9
10
HZ
CV = 0.05
CV = 0.10
CV = 0.15
CV = 0.20

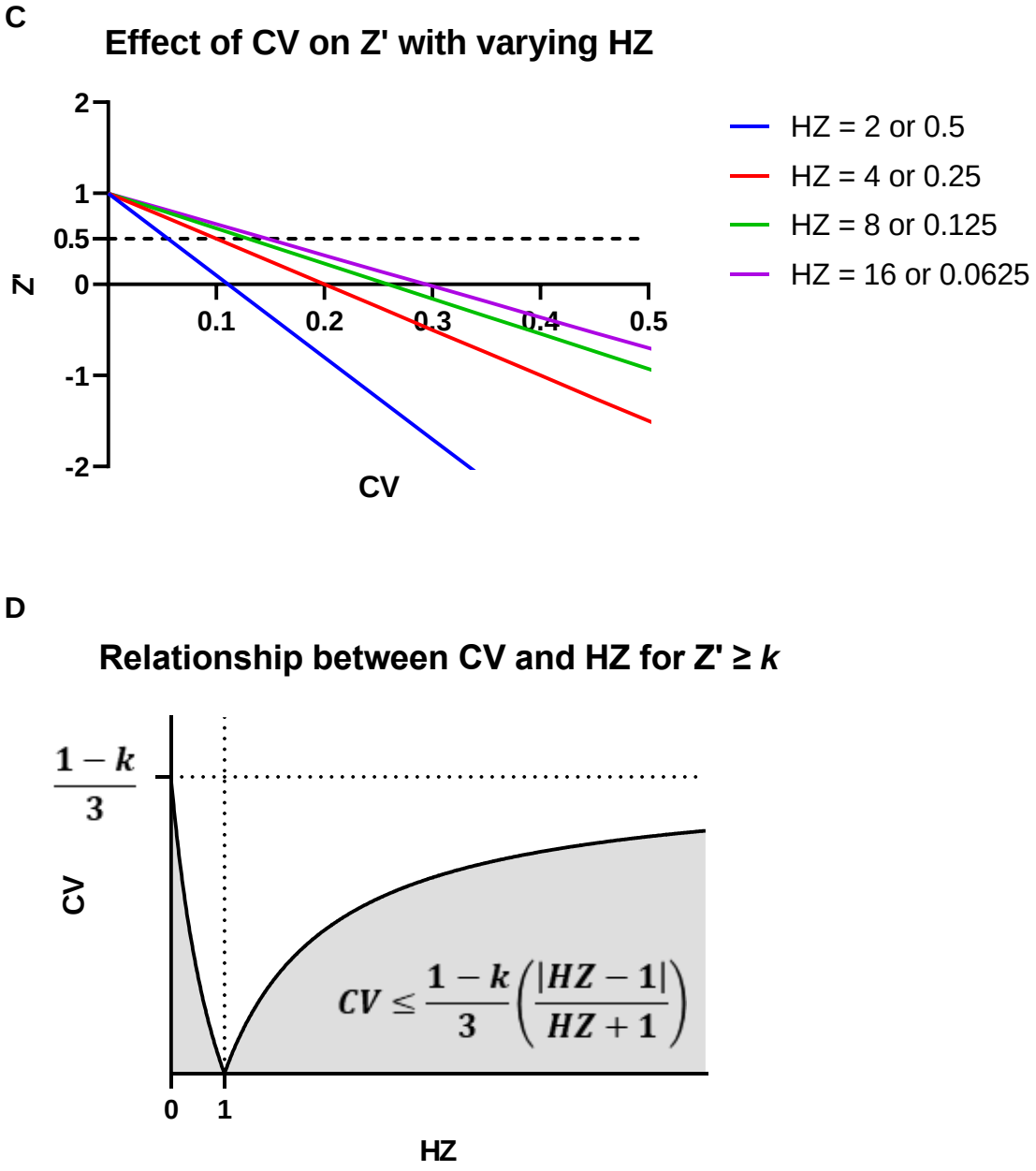


Figure 1. A. Graphical representation of Equation 1, with the two bounded segments labelled. B. Graphical representation of the effect of HZ on Z'-factor under various values of CV. C. Graphical representation of the effect of CV on Z'-factor under various values of HZ. Note that the equation is independent of the assay directionality, as shown in Appendix 1. Additionally, as HZ increases, the CV threshold to satisfy the required Z'-factor approaches $\frac{1-Z'}{3}$. D. Graphical representation of the inequality 1b. The shaded region satisfies the condition where Z'-factor ≥ *k*. All graphs simulated using GraphPad Prism 9.

Part 2: Assumption of unequal CVs in HPE and ZPE

As much as Equation 1 provides insight into how Z'-factor can be influenced by CV and HZ, it is a gross simplification as both HPE and ZPE are assumed to have an equal CV.

As described in part 1, by rearranging the Z'-factor equation, we get:

$$\frac{\sigma_H + \sigma_Z}{|\mu_H - \mu_Z|} = \frac{1 - Z'}{3}$$

If the CV is different for HPE and ZPE, i.e. $\mathrm{CV_H} = \frac{\sigma_H}{\mu_H}$ and $\mathrm{CV_Z} = \frac{\sigma_Z}{\mu_Z}$, then:

$$\frac{\mathrm{CV_H}\mu_H + \mathrm{CV_Z}\mu_Z}{|\mu_H - \mu_Z|} = \frac{1 - Z'}{3}$$

Using the definition of HZ, it goes that $\mu_H = \mathrm{HZ} \cdot \mu_Z$. Replacing this relationship into the left-hand side (LHS) of the equation and simplifying it gives:

$$\frac{\mathrm{CV_H} \cdot \mathrm{HZ} \cdot \mu_Z + \mathrm{CV_Z}\mu_Z}{|\mathrm{HZ} \cdot \mu_Z - \mu_Z|} = \frac{1 - Z'}{3}$$

$$\frac{\mu_Z\,(\mathrm{CV_H HZ} + \mathrm{CV_Z})}{|\mu_Z||\mathrm{HZ} - 1|} = \frac{1 - Z'}{3}$$

With the assumption that ZPE control means are positive values so that $|\mu_Z| = \mu_Z$, the variables can cancel each other out, giving:

$$\frac{\mathrm{CV_H HZ} + \mathrm{CV_Z}}{|\mathrm{HZ} - 1|} = \frac{1 - Z'}{3}$$

Upon rearrangement, we can derive two equations to understand the relationship between $\mathrm{CV_H}$ and $\mathrm{CV_Z}$ in determining the Z'-factor within the boundary of the HZ:

$$\mathbf{CV_H}\left(\frac{\mathbf{HZ}}{|\mathbf{HZ}-\mathbf{1}|}\right)+\mathbf{CV_Z}\left(\frac{\mathbf{1}}{|\mathbf{HZ}-\mathbf{1}|}\right)=\frac{\mathbf{1}-\boldsymbol{Z'}}{\mathbf{3}} \quad (\mathbf{2})$$

$$\mathbf{CV_Z}=-\mathbf{CV_H}\mathbf{HZ}+\frac{\mathbf{1}-\boldsymbol{Z'}}{\mathbf{3}}(|\mathbf{HZ}-\mathbf{1}|) \quad (\mathbf{3})$$

It is important to note that equations 2 and 3 cannot be defined when HZ = 1. That would be a reasonable caveat as HZ = 1 implies that HPE and ZPE groups have equal means that would render assay development using these controls impossible.

Once again, equations 2 and 3 can be reformulated as an inequality to determine the range of CV values to satisfy the Z'-factor threshold *k*, so that Z' ≥ *k*:

$$\mathbf{CV_H}\left(\frac{\mathbf{HZ}}{|\mathbf{HZ}-\mathbf{1}|}\right)+\mathbf{CV_Z}\left(\frac{\mathbf{1}}{|\mathbf{HZ}-\mathbf{1}|}\right)\leq\frac{\mathbf{1}-\boldsymbol{k}}{\mathbf{3}} \quad (\mathbf{2b})$$

$$\mathbf{CV_Z}\leq-\mathbf{CV_H}\mathbf{HZ}+\frac{\mathbf{1}-\boldsymbol{k}}{\mathbf{3}}(|\mathbf{HZ}-\mathbf{1}|) \quad (\mathbf{3b})$$

Equation 2 can be illustrated as a series of 3-dimensional profile plots to observe how the three variables CV_H, CV_Z and HZ impact Z'-factor (Supplementary Information 1). In addition, Equation 2 summarises the impact of CV_H and CV_Z on the overall CV limit defined by $\frac{1-Z'}{3}$. The weightage and therefore the impact of CV_H and CV_Z on the overall CV is defined by HZ. Inequality 2b also states that the weighted average of CV_H and CV_Z should not exceed the overall CV limit (e.g. 0.167 for *k* = 0.50) to surpass the minimum Z'-factor threshold *k*.

It is interesting to note the contribution of CV_H and CV_Z in each assay 'directionality'. In the first case where HPE > ZPE (i.e., HZ > 1), as HZ increases to infinity, indicating an

increasing window of separation between control groups, $\lim_{HZ\to\infty} \frac{HZ}{|HZ-1|} = 1$ and $\lim_{HZ\to\infty} \frac{1}{|HZ-1|} = 0$, and as such, inequality 2b can be simplified to $CV_H \leq \frac{1-k}{3}$, implying that CV_H defines the overall CV limit to satisfy the minimum Z'-factor threshold *k*. In that scenario, controlling CV_Z would bring little benefit to either satisfy or increase the minimum Z'-factor threshold.

However, a different conclusion can be drawn from assays with ZPE > HPE (i.e., HZ < 1). As the window of separation between control groups increases, HZ decreases and approaches zero and therefore $\lim_{HZ\to 0} \frac{HZ}{|HZ-1|} = 0$ and $\lim_{HZ\to 0} \frac{1}{|HZ-1|} = 1$. As such, inequality 2b can be further simplified to $CV_Z \leq \frac{1-k}{3}$, implying that when ZPE > HPE and HZ < 1, CV_Z defines the overall CV limit to satisfy the minimum Z'-factor threshold *k*.

Equation 3 and its inequality 3b show the relationship between CV_H and CV_Z in varying HZ ratios (Figure 2A). From equation 3, there is a linear relationship between CV_H and CV_Z. More notably, the x- and y-intercepts define the theoretical limits of CV_H and CV_Z respectively. It is interesting to note that these limits can be defined using the overall CV limit ($\frac{1-Z'}{3}$) and a multiplicative factor based on HZ.

The slope of this linear equation is -HZ and as such, the straight line becomes steeper with increasing HZ and vice versa (Figure 2B). Increasing the window of separation between control groups (indicated by increasing HZ for HZ > 1 and decreasing HZ for HZ < 1) can impact the theoretical limit of CV_H and CV_Z and therefore provides an idea of the amount of leeway one has in controlling these CVs to achieve good Z'-factor in an assay.

When HZ > 1, increasing the window of separation between the controls (i.e., increasing HZ) causes a massive increase in the CV_Z limit but not so much in the CV_H limit. This means that it is more important to control and/or reduce the CV in the HPE group in assays where HZ > 1 to meet the minimum Z'-factor threshold defined by the user.

On the other hand, when HZ < 1, increasing the window of separation between the controls (i.e., decreasing HZ) causes a massive increase in the CV_H limit but not so much in the CV_Z limit. This implies that it is far more important to control and/or reduce the CV in the ZPE group in assays where HZ < 1 to meet the minimum Z'-factor threshold defined by the user.

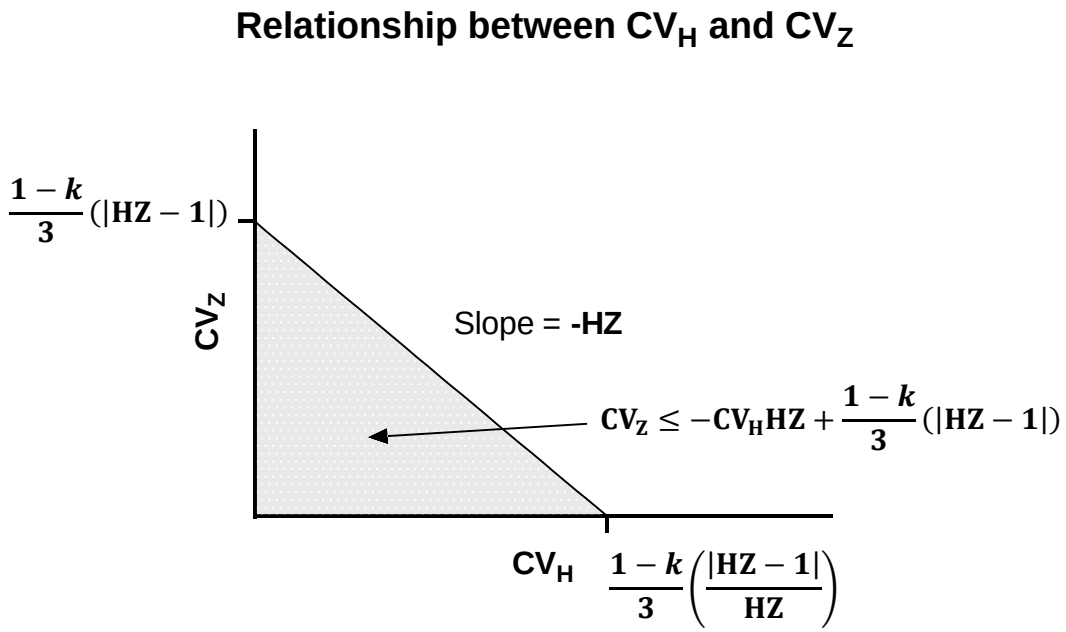
Nonetheless, in both forms of assay 'directionality', regardless of whether HPE is larger than ZPE or not, increasing the window of separation between the two control groups improves the theoretical limit of both CVs. This implies that a greater separation between the two control groups gives a better chance to satisfy the predetermined Z'-factor threshold as there is greater room for variability within each control group as represented by the theoretical limit of CV.

In a special case where both CV_H and CV_Z are similar, that is $CV = CV_H = CV_Z$, equation 2 can be simplified to equation 1 as such:

$$\mathrm{CV}\left(\frac{\mathrm{HZ}}{|\mathrm{HZ}-1|}\right)+\mathrm{CV}\left(\frac{1}{|\mathrm{HZ}-1|}\right)=\frac{1-Z'}{3}$$

$$\mathrm{CV}\left(\frac{\mathrm{HZ}+1}{|\mathrm{HZ}-1|}\right)=\frac{1-Z'}{3}\Rightarrow \mathbf{CV}=\frac{\mathbf{1}-\boldsymbol{Z'}}{\mathbf{3}}\left(\frac{|\mathbf{HZ}-\mathbf{1}|}{\mathbf{HZ}+\mathbf{1}}\right)$$

A



B

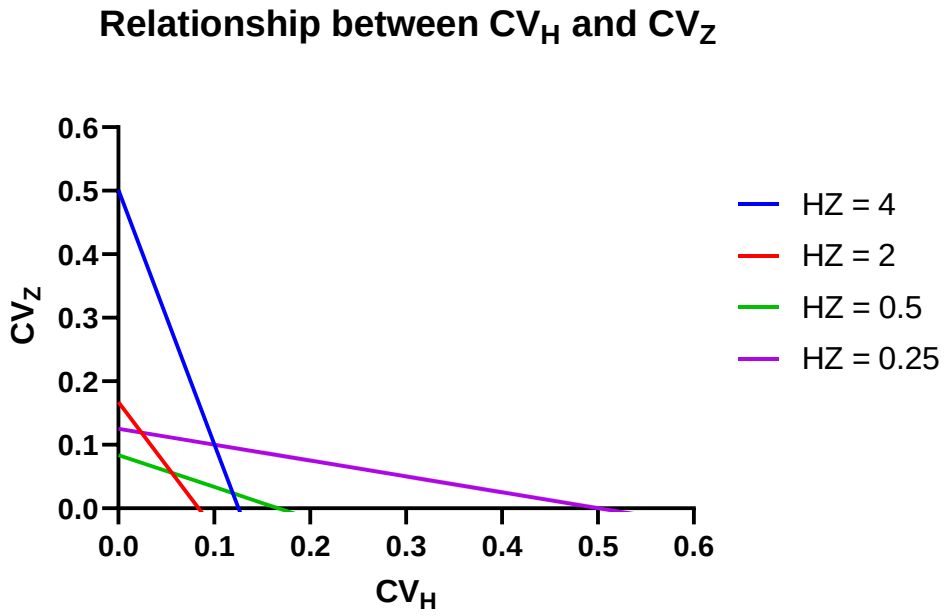


Figure 2. A. Relationship between CV_H and CV_Z, with the x- and y-intercepts labelled with their respective theoretical limits. The shaded region refers to the region which satisfies inequality 3b and therefore the region in which its CV_H and CV_Z values satisfy the required Z'-factor threshold *k*. B. A simulation of equation 3 using various values of HZ to highlight

the trend in CV_H and CV_Z theoretical limits in response to HZ. In this scenario, the Z'-factor threshold *k* is set to 0.5. Graphs simulated using GraphPad Prism 9.

2.3 Expressing SSMD as a function of CV and HZ

SSMD is the ratio of the mean to the standard deviation of the difference between two groups. In the context of two control groups, HPE and ZPE groups, we can express SSMD as such:

$$\text{SSMD} = \frac{\mu_H - \mu_Z}{\sqrt{\sigma_H^2 + \sigma_Z^2}}$$

When ZPE > HPE, the difference of means is a negative number, and as such SSMD is negative under such circumstances. Hence there are two aspects to consider when interpreting SSMD. The positive or negative symbol denotes the assay 'directionality' whilst the absolute value denotes the magnitude of the difference between the two control groups.

Using the definitions of HZ, CV_H and CV_Z, we can express SSMD in terms of these variables so that the user can observe how CV (both CV_H and CV_Z) and HZ can affect Z'-factor. First, we express the standard deviations in terms of CV and mean.

$$\text{SSMD} = \frac{\mu_H - \mu_Z}{\sqrt{(\text{CV}_\text{H}\mu_H)^2 + (\text{CV}_\text{Z}\mu_Z)^2}}$$

Next, we express the HPE mean in terms of HZ and ZPE mean.

$$\text{SSMD} = \frac{\text{HZ} \cdot \mu_Z - \mu_Z}{\sqrt{(\text{CV}_\text{H} \cdot \text{HZ} \cdot \mu_Z)^2 + (\text{CV}_\text{Z}\mu_Z)^2}}$$

Factorisation and taking out ZPE mean as the common term, gives

$$\text{SSMD} = \frac{\mu_Z(\text{HZ} - 1)}{\mu_Z\sqrt{(\text{CV}_\text{H}\text{HZ})^2 + \text{CV}_\text{Z}^{\,2}}}$$

Simplification then gives equation 4:

$$\mathbf{SSMD} = \frac{\mathbf{HZ - 1}}{\sqrt{(\mathbf{CV_H HZ})^2 + \mathbf{CV_Z}^{\mathbf{2}}}} \qquad (\mathbf{4})$$

If both CV_H and CV_Z are similar, that is $CV = CV_H = CV_Z$, equation 4 can be simplified to:

$$\mathbf{SSMD} = \frac{\mathbf{HZ - 1}}{\mathbf{CV}\sqrt{\mathbf{HZ^2 + 1}}} \qquad (\mathbf{5})$$

The relationship between both CVs and HZ in determining SSMD is slightly more complex than that in determining Z'-factor. Nonetheless, there is still some insight to be taken by examining both equations 4 and 5.

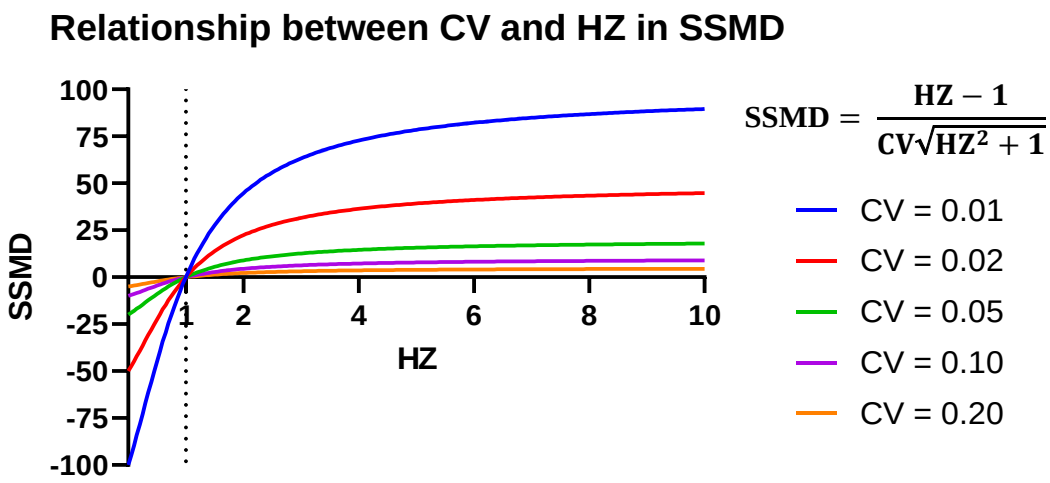
First, it is noted that the CV in general is inversely correlated with the absolute SSMD value, which implies that the greater the variation within each group, the smaller the absolute SSMD value (Figure 3A). This observation is true regardless of the 'directionality' of the assay, be it HZ > 1 or HZ < 1. The HZ determines the upper and lower boundary/limit of the SSMD. The greater the separation between the two control groups (as HZ approaches infinity for HZ > 1 or approaches zero for HZ < 1), the larger the absolute SSMD value.

If we attempt to break down the CV into its constituent CV_H and CV_Z, we can observe how CV_H and CV_Z individually influence SSMD. Increasing the CV_H while maintaining a

constant CV_Z value leads to a marked decrease in the absolute SSMD value at HZ > 1 (Figure 3B). This change in absolute SSMD is not prominent at HZ < 1. This implies that in assays where HPE > ZPE and HZ > 1, improving the CV_H by reducing it would markedly improve the absolute SSMD value.

On the other hand, increasing CV_Z while maintaining a constant CV_H value leads to a marked decrease in the absolute SSMD value at HZ < 1 (Figure 3C). This change in absolute SSMD is not prominent at HZ > 1. This implies that in assays where ZPE > HPE and HZ < 1, improving the CV_Z by reducing it would markedly improve the absolute SSMD value.

A



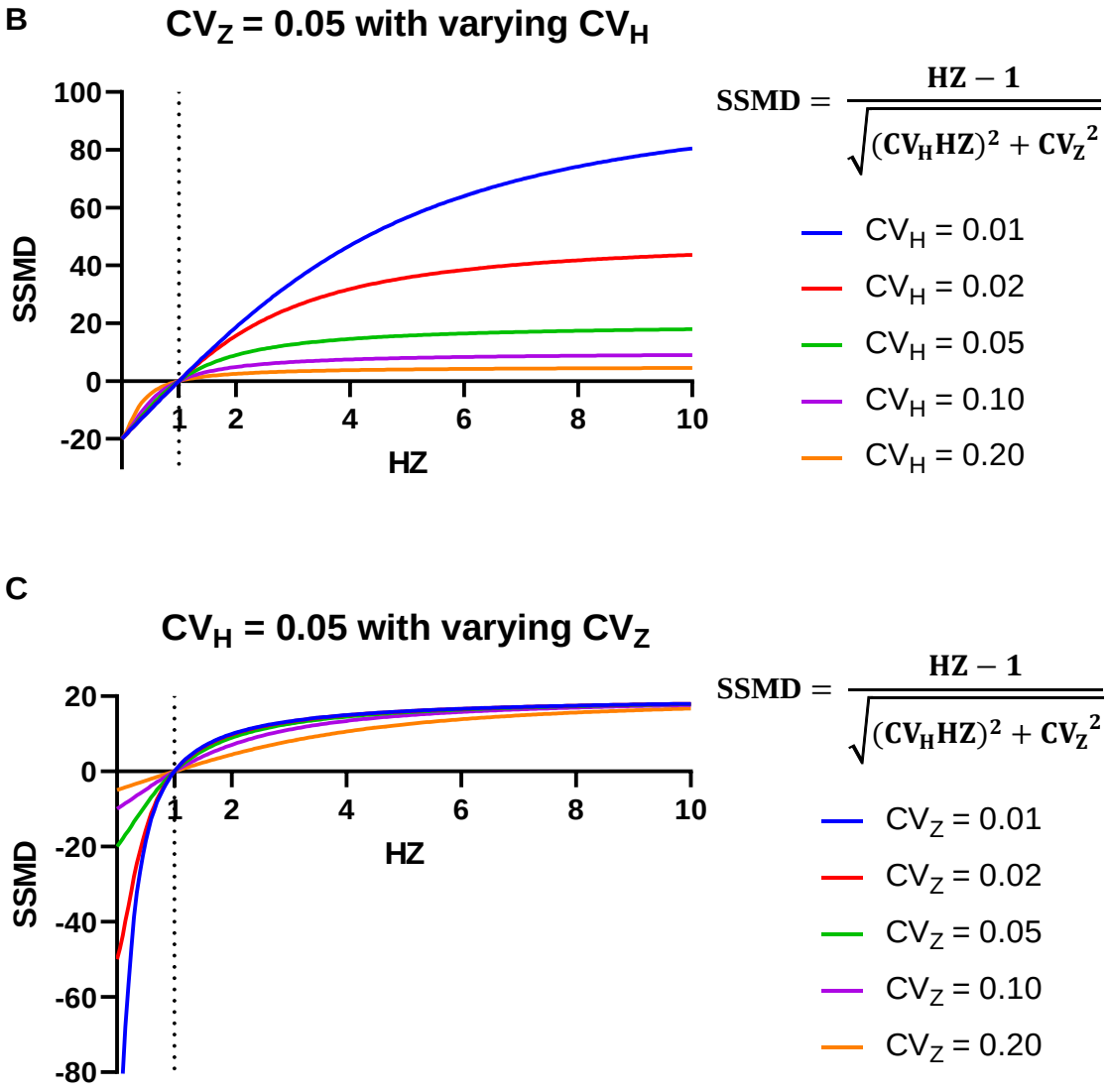


Figure 3. A. The relationship between CV and HZ in determining SSMD according to equation 4. B. The impact of CV_H on SSMD according to equation 5. C. The impact of CV_Z on SSMD according to equation 5. Graphs simulated using GraphPad Prism 9.

2.4 The relationship between Z'-factor and SSMD

In both analyses of Z'-factor and SSMD, one can observe that a similar trend can be observed with regards to how CV_H, CV_Z and HZ can influence these assay metrics. It is also possible

to show that Z'-factor and SSMD can be related to each other, especially in the case where CVs in HPE and ZPE are equal.

First, we can define the absolute SSMD value from equation 5 and derive the CV as a function of HZ and the absolute value of SSMD.

$$|\mathrm{SSMD}| = \left|\frac{\mathrm{HZ}-1}{\mathrm{CV}\sqrt{\mathrm{HZ}^2+1}}\right| = \frac{|\mathrm{HZ}-1|}{\left|\mathrm{CV}\sqrt{\mathrm{HZ}^2+1}\right|}$$

As CV > 0 and HZ > 0, the denominator will always be a positive number. Hence:

$$|\mathrm{SSMD}| = \frac{|\mathrm{HZ}-1|}{\mathrm{CV}\sqrt{\mathrm{HZ}^2+1}} \Rightarrow \mathrm{CV} = \frac{|\mathrm{HZ}-1|}{|\mathrm{SSMD}|\sqrt{\mathrm{HZ}^2+1}}$$

Combining with equation 1 gives:

$$\frac{|\mathrm{HZ}-1|}{|\mathrm{SSMD}|\sqrt{\mathrm{HZ}^2+1}} = \frac{1-\mathrm{Z}'}{3}\left(\frac{|\mathrm{HZ}-1|}{\mathrm{HZ}+1}\right)$$

Simplification and rearranging the terms give:

$$\frac{3|\mathrm{HZ}-1|(\mathrm{HZ}+1)}{|\mathrm{SSMD}|(\sqrt{\mathrm{HZ}^2+1})(|\mathrm{HZ}-1|)} = 1-\mathrm{Z}'$$

$$\mathbf{1} - \frac{\mathbf{3(HZ+1)}}{\mathbf{|SSMD|}(\sqrt{\mathbf{HZ^2+1}})} = \mathbf{Z'} \Rightarrow \mathbf{|SSMD|} = \frac{\mathbf{3(HZ+1)}}{\mathbf{(1-Z')}(\sqrt{\mathbf{HZ^2+1}})} \qquad (\mathbf{6})$$

By comparing the original definition of the Z'-factor, one can see that in the case where CVs in HPE and ZPE are equal:

$$\frac{\boldsymbol{\sigma_H + \sigma_Z}}{\boldsymbol{|\mu_H - \mu_Z|}} = \frac{\mathbf{(HZ+1)}}{\mathbf{|SSMD|}(\sqrt{\mathbf{HZ^2+1}})} = \frac{\mathbf{1-Z'}}{\mathbf{3}} \qquad (\mathbf{7})$$

It is worth noting that $\mathrm{HZ} + 1 > \sqrt{\mathrm{HZ}^2 + 1}$ for HZ > 0, and as such:

$$\frac{\mathrm{HZ} + 1}{\sqrt{\mathrm{HZ}^2 + 1}} > 1$$

$$\Rightarrow \frac{3(\mathrm{HZ} + 1)}{|\mathrm{SSMD}|(\sqrt{\mathrm{HZ}^2 + 1})} > \frac{3}{|\mathrm{SSMD}|}$$

$$\Rightarrow 1 - \frac{3(\mathrm{HZ} + 1)}{|\mathrm{SSMD}|(\sqrt{\mathrm{HZ}^2 + 1})} < 1 - \frac{3}{|\mathrm{SSMD}|}$$

$$\Rightarrow \mathrm{Z}' < 1 - \frac{3}{|\mathrm{SSMD}|} \text{ or } |\mathrm{SSMD}| > \frac{3}{1 - \mathrm{Z}'}$$

This implies that the commonly-used Z'-factor threshold of 0.5 would require an absolute SSMD value of more than 6. In other words, in such an assay, the difference of the means between the two control groups is at least six times greater than the sum of the standard deviations of the two groups. This reinstates Zhang's point that the Z'-factor threshold is stricter than the SSMD threshold.[7]

Equation 6 can be represented graphically to show the relationship between the Z'-factor and the absolute SSMD value. From Figure 4, it can be seen that the curve shifts to the left upon an increase in the separation window. This is true regardless of the 'directionality' of the assay, in situations where HPE > ZPE (HZ > 1) or ZPE > HPE (HZ < 1). This relationship implies that to achieve the desired Z'-factor threshold, assays with a poor separation window between HPE and ZPE (HZ value close to 1) would need to be compensated by a stronger differentiation between the two control groups (high absolute SSMD value).

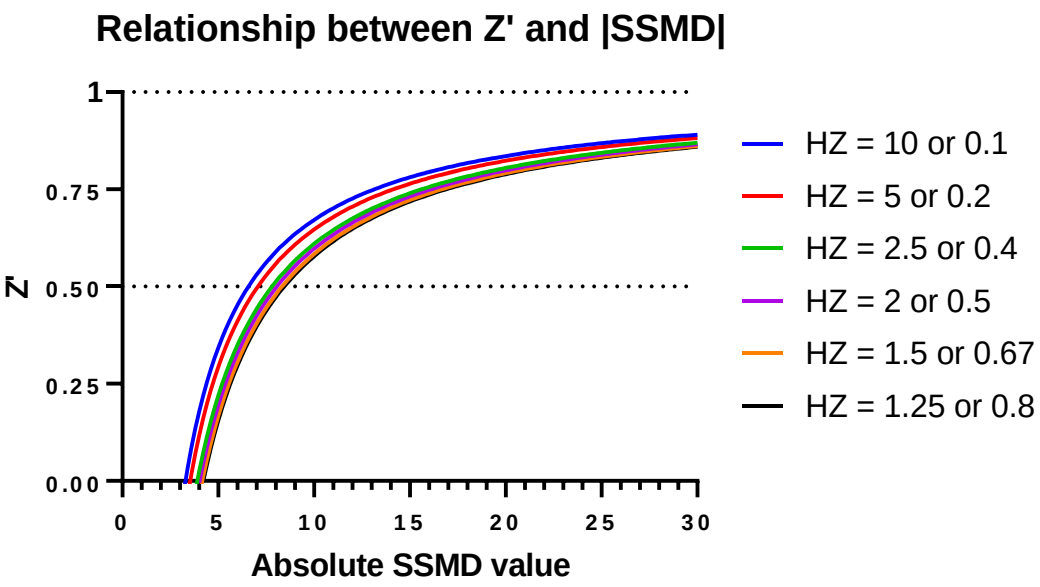


Figure 4. The relationship between Z'-factor and SSMD with varying HZ ratios. Note that different HZ values which give a similar fold difference between HPE and ZPE (e.g. HZ value of 10 or 0.1) will give the same curve. Proof of this is noted in Appendix 2. Graph simulated using GraphPad Prism 9.

From equation 7, one can observe that the theoretical overall CV limit of the assay is represented by the $\frac{\sigma_H+\sigma_Z}{|\mu_H-\mu_Z|}$ in the Z'-factor equation. More interestingly, that coefficient can now be expressed as a function of the absolute value of SSMD and HZ. This implies that there is a formalised relationship between two parameters Z'-factor and SSMD, expressed in equation 6. It has been mentioned previously that the value of SSMD can be interpreted from a probabilistic viewpoint, but not for Z'-factor.[7] From equation 6, the Z'-factor can be converted to its corresponding absolute SSMD value to understand its probabilistic value, and such conversion requires the knowledge of the HZ independent of the CVs of each control group.

2.5 Application of equations using simulated examples

To illustrate these points in practice, various simulated examples of initial Z'-factor values with the corresponding assay variables (CV_H, CV_Z and HZ) are presented in Table 2.

Table 2: Simulated Z'-factor values based on various assay variables (CV_H, CV_Z and HZ).

Example	Mean HPE	CV_H	Mean ZPE	CV_Z	HZ	Z'-factor
Case 1: Large separation between control groups						
1	10000	0.2	1000	0.2	10	0.27
2	10000	0.1	1000	0.2	10	0.60
3	10000	0.2	1000	0.1	10	0.30
4	10000	0.1	1000	0.1	10	0.63
5	20000	0.2	1000	0.2	20	0.33
Case 2: Small separation between control groups, HZ > 1						
6	15000	0.2	5000	0.2	3	-0.20
7	15000	0.1	5000	0.2	3	0.25
8	15000	0.2	5000	0.1	3	-0.05
9	30000	0.2	5000	0.2	6	0.16
Case 3: Small separation between control groups, HZ < 1						
11	2000	0.2	8000	0.2	0.25	0
12	2000	0.1	8000	0.2	0.25	0.10
13	2000	0.2	8000	0.1	0.25	0.40
14	1000	0.2	8000	0.2	0.125	0.23

In Case 1, it can be seen that in assays with large separation between control groups, further increasing the HZ by two-fold in this scenario (Table 2, example 5) gives a small improvement to the Z'-factor. However, if the CV of the control group with the greater raw signal (CV_H in this scenario) is halved (Table 2, example 2), then one can see a marked improvement on the Z'-factor. Improving the remaining CV_Z only also gives modest improvement on the Z'-factor (Table 2, example 3).

In Case 2, where there is relatively small separation between control groups, increasing the HZ (Table 2, example 9) does improve the Z'-factor more substantially than in Case 1. As with in Case 1 and based on Equation 2, one can see that for HZ > 1, reducing the CV_H rather than CV_Z gives greater improvement on the Z'-factor (Table 2, examples 7 and 8). The reverse is true in Case 3, where the assay directionality is changed, giving HZ < 1. In this scenario, reducing the CV_Z in the ZPE (i.e. negative control group) actually improves the Z'-factor markedly (Table 2, example 13), as predicted in Equations 2 and 3.

3. Discussion and conclusions

Assay quality metrics can be used to measure the process of improvement in assay development. Users can use these metrics to guide their decision-making process on the next steps in improving the assay. For example, users can choose to further increase the separation between the two control groups by changing the pharmacological tool compound or by increasing the assay time to allow greater separation between the two groups, assuming that the biochemical reaction has not yet gone to completion. In addition, the user can also attempt to reduce the variability in each control group, for example, by improving its instrumentation or by controlling the biological variation in control groups.

Occasionally, users may find that controlling one aspect of the assay parameter (e.g. signal separation or variability) is simpler or more practical than the other. For example, cell-based assays tend to have a larger biological variation in their controls (especially untreated controls) than biochemical assays. As such, it is worthwhile to understand how variability and signal separation contribute to the assay quality metric so that users can efficiently improve the assay.

Prior literature has provided some insight on the impact of CV and/or separation between two control groups (indicated by SB) in common assay metrics such as Z'-factor and SSMD. However, prior analyses used a general CV or a modified CV that looked into the overall assay rather than its constituent control groups. Here we derive several equations which incorporate the CV of each control group and the HZ. We demonstrate through these equations that these variables can impact assay metrics such as Z'-factor and SSMD. Coincidentally, we show that the relationship between Z'-factor and SSMD is dependent on the extent of separation between the two control groups.

To improve assay quality metrics (Z'-factor or SSMD), controlling the CV of the control group with the higher signal magnitude is more important than the other group, especially as the separation between the two control groups increases. Noticeably, the control group with the higher signal implies a different biological group depending on the assay 'directionality'. In assays where HZ > 1, this refers to the HPE group, usually corresponding to the group treated pharmacologically, while in assays where HZ < 1, this refers to the ZPE group, usually the untreated group. In assays with a greater separation between the two control groups, it would be more worthwhile to focus on reducing the variability in the control

groups with the higher signal magnitude. As the separation between the two control groups decreases, it would be just as important to control the variability in the remaining control group. We have derived equations 2 and 3 to formalize the relationship between these two CVs and guide the user in deciding which control groups to improve for their variability.

The equations derived here take into consideration the assay 'directionality', hence preventing any confusion regarding the need to identify the 'top' or 'bottom' signal. The use of HPE and ZPE nomenclature, and subsequently the HZ ratio, should clarify any confusion, as the assignment of the HPE and ZPE group remains constant within the context of the assay while disregarding the magnitude and direction of the signal.

It is hoped that these equations (Figure 5) can provide a quantitative measure to help users to recognise assay variables (CV and/or HZ) that matter more and respond more quickly in improving the assay quality, thus making the assay development process more efficient and effective. By identifying assay variables that matter more, users can select the appropriate intervention to improve the require assay variables in the assay development process. Although not discussed in the article, these equations can be reformulated into their robust equivalents, using median, median absolute deviation and robust coefficient of variation as robust estimators of mean, standard deviation and CV respectively.

Let $CV_H = \frac{\sigma_H}{\mu_H}$, $CV_Z = \frac{\sigma_Z}{\mu_Z}$ and $HZ = \frac{\mu_H}{\mu_Z}$, where
σ_H = Standard deviation of the HPE group
σ_Z = Standard deviation of the ZPE group
μ_H = Mean of the HPE group
μ_Z = Mean of the ZPE group
HPE - 100% effect, ZPE - 0% effect

$$\mathbf{CV_H}\left(\frac{\mathbf{HZ}}{\mathbf{|HZ-1|}}\right)+\mathbf{CV_Z}\left(\frac{\mathbf{1}}{\mathbf{|HZ-1|}}\right)=\frac{\mathbf{1-Z'}}{\mathbf{3}} \qquad \mathbf{CV_Z}=-\mathbf{CV_H HZ}+\frac{\mathbf{1-Z'}}{\mathbf{3}}(\mathbf{|HZ-1|})$$

$$\mathbf{SSMD}=\frac{\mathbf{HZ-1}}{\sqrt{(\mathbf{CV_H HZ})^2+\mathbf{CV_Z}^2}} \qquad \mathbf{|SSMD|}=\frac{\mathbf{3(HZ+1)}}{\left(\mathbf{1-Z'}\right)\left(\sqrt{\mathbf{HZ^2+1}}\right)}$$

Where Z' = required Z'-factor and SSMD = standardized mean difference

Figure 5. Summary of the mathematical equations discussed in this article.

4. Appendix

Appendix 1

Proof that equation 1 is not affected by the assay 'directionality'

If one control group has a *h*-fold signal over the other control group, then the HZ can be either *h* (if HPE > ZPE) or 1/*h* (if ZPE > HPE).

If HZ = *h*

$$\mathrm{CV}=\frac{1-Z'}{3}\left(\frac{|h-1|}{h+1}\right)$$

If HZ = 1/*h*

$$\mathrm{CV}=\frac{1-Z'}{3}\left(\frac{\left|\frac{1}{h}-1\right|}{\frac{1}{h}+1}\right)$$

$$\mathrm{CV} = \frac{1 - Z'}{3}\left(\frac{\left|\frac{1-h}{h}\right|}{\frac{1+h}{h}}\right)$$

$$\mathrm{CV} = \frac{1 - Z'}{3}\left(\frac{(\frac{1}{h})|1-h|}{(\frac{1}{h})(1+h)}\right) \Rightarrow \frac{1 - Z'}{3}\left(\frac{|h-1|}{h+1}\right) \text{ [NB: } |1-h| = |h-1|]$$

As such, CV is the same for HZ = *h* or 1/*h* and thus equation 1 is not affected by the assay 'directionality'.

Appendix 2

Proof that equation 6 is not affected by the assay 'directionality'

If one control group has a *h*-fold signal over the other control group, then the HZ can be either *h* (if HPE > ZPE) or 1/*h* (if ZPE > HPE).

If HZ = *h*

$$Z' = 1 - \frac{3(h+1)}{|\mathrm{SSMD}|(\sqrt{h^2+1})}$$

If HZ = 1/*h*

$$Z' = 1 - \frac{3\left(\frac{1}{h}+1\right)}{|\mathrm{SSMD}|(\sqrt{(\frac{1}{h})^2+1})}$$

$$Z' = 1 - \frac{3\left(\frac{1+h}{h}\right)}{|\mathrm{SSMD}|(\sqrt{\frac{1+h^2}{h^2}})}$$

$$Z' = 1 - \frac{3\left(\frac{1}{h}\right)(h+1)}{|\mathrm{SSMD}|\left(\frac{1}{h}\right)\left(\sqrt{h^2+1}\right)} \Rightarrow 1 - \frac{3(h+1)}{|\mathrm{SSMD}|(\sqrt{h^2+1})}$$

As such, Z'-factor is the same for HZ = *h* or 1/*h* and thus equation 6 is not affected by the assay 'directionality'.

Appendix 3

Proof that equations 2, 3 and 4 are affected by the assay 'directionality'

If one control group has a *h*-fold signal over the other control group, then the HZ can be either *h* (if HPE > ZPE) or 1/*h* (if ZPE > HPE).

For equation 2, if HZ = *h*

$$\mathrm{CV_H}\left(\frac{h}{|h-1|}\right)+\mathrm{CV_Z}\left(\frac{1}{|h-1|}\right)=\frac{1-Z'}{3}$$

If HZ = 1/*h*

$$\mathrm{CV_H}\left(\frac{\frac{1}{h}}{\left|\frac{1}{h}-1\right|}\right)+\mathrm{CV_Z}\left(\frac{1}{\left|\frac{1}{h}-1\right|}\right)=\frac{1-Z'}{3}$$

$$\mathrm{CV_H}\left(\frac{\frac{1}{h}}{\left|\frac{1-h}{h}\right|}\right)+\mathrm{CV_Z}\left(\frac{1}{\left|\frac{1-h}{h}\right|}\right)=\frac{1-Z'}{3}$$

$$\mathrm{CV_H}\left(\frac{1}{|h-1|}\right)+\mathrm{CV_Z}\left(\frac{h}{|h-1|}\right)=\frac{1-Z'}{3}\ [\text{NB: } |1-h|=|h-1|]$$

Note that both equations are not similar, with the coefficient of $\mathrm{CV_H}$ and $\mathrm{CV_Z}$ swapped in different assay directionality. This implies that equation 2 is affected by the assay 'directionality'. As equation 3 is derived from equation 2, the same conclusion can be drawn.

For equation 4, if HZ = *h*

$$\mathrm{SSMD}=\frac{h-1}{\sqrt{(\mathrm{CV_H}h)^2+\mathrm{CV_Z}^2}}$$

If HZ = 1/*h*

$$\mathrm{SSMD}=\frac{\frac{1}{h}-1}{\sqrt{(\mathrm{CV_H}\frac{1}{h})^2+\mathrm{CV_Z}^2}}$$

$$\text{SSMD} = \frac{1-h}{h\sqrt{(\text{CV}_\text{H}\frac{1}{h})^2 + \text{CV}_\text{Z}{}^2}} \Rightarrow \frac{1-h}{\sqrt{\text{CV}_\text{H}{}^2 + (\text{CV}_\text{Z}h)^2}}$$

Note that both equations are not similar, with the coefficient of CV_H and CV_Z swapped in different assay directionality. This implies that equation 4 is affected by the assay 'directionality'.

Acknowledgement

This work is supported by Alzheimer's Research UK (ARUK-2022DDI-CAM) and the Rainwater Charitable Foundation (Tau Consortium).

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

References

1. Sittampalam, G. S.; Iversen, P. W.; Boadt, J. A.; et al. Design of Signal Windows in High Throughput Screening Assays for Drug Discovery. *Journal of Biomolecular Screening* **1997,** *2*, 159-169.
2. Zhang, J. H.; Chung, T. D. Y.; Oldenburg, K. R. A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *Journal of Biomolecular Screening* **1999,** *4*, 67-73.

3. Iversen, P. W.; Eastwood, B. J.; Sittampalam, G. S.; et al. A Comparison of Assay Performance Measures in Screening Assays: Signal Window, Z' Factor, and Assay Variability Ratio. *Journal of Biomolecular Screening* **2006,** *11*, 247-252.
4. Bar, H.; Zweifach, A. Z' Does Not Need to Be > 0.5. *SLAS DISCOVERY: Advancing the Science of Drug Discovery* **2020,** *25*, 1000-1008.
5. Zhang, X. D.; Wang, D.; Sun, S.; et al. Issues of Z-factor and an approach to avoid them for quality control in high-throughput screening studies. *Bioinformatics* **2020,** *36*, 5299-5303.
6. Sui, Y.; Wu, Z. Alternative Statistical Parameter for High-Throughput Screening Assay Quality Assessment. *Journal of Biomolecular Screening* **2007,** *12*, 229-234.
7. Zhang, X. D. A pair of new statistical parameters for quality control in RNA interference high-throughput screening assays. *Genomics* **2007,** *89*, 552-561.