The influence of baseline capability on intervention effects in strength and conditioning: A review of concepts and methods with meta-analysis.

Paul, A. Swinton

Doi: 10.51224/SRXIV.285

SportRχiv hosted preprint version 1

19/04/2023

PREPRINT - NOT PEER REVIEWED

Contact details

Dr. Paul Swinton

School of Health Sciences, Robert Gordon University

Garthdee Road

Aberdeen, UK,

AB10 7QG

p.swinton@rgu.ac.uk, +44 (0) 1224 262 3361

Twitter Handle:
@PaulSwinton9

Please cite as: Swinton, PA. The influence of baseline capability on intervention effects in strength and conditioning: A review of concepts and methods with meta-analysis. Pre-print available from SportRχiv. https://doi.org/10.51224/SRXIV.285

Abstract

Background: In strength and conditioning (S&C) it is commonly believed that baseline capability influences response to an intervention, such that in general, those with the higher baseline values experience reduced change. Such differences are referred to as intervention differential effects (IDE) and are important in the study of tailoring training programs in S&C. There are, however, several conceptual and technical issues that present a challenge when investigating whether baseline capability causes IDE. The present review provides an overview of these conceptual and technical issues, highlighting important differences between changes within and between populations, and the role of measurement error and subsequent regression to the mean when performing standard analyses.

Methods: The present review also includes a meta-analysis to explore more generally, whether those with higher baseline values experience reduced change. Baseline and post-intervention standard deviations were extracted from 421 S&C training studies including the 1RM squat (121 studies; 329 outcomes), 1RM bench press (103 studies; 307 outcomes), vertical jump (312 studies; 896 outcomes), 10 m sprint time (95 studies; 194 outcomes), 20 m sprint time (97 studies; 193 outcomes), and 30 m sprint time (58 studies; 118 outcomes). For each outcome, a Bayesian three-level hierarchical meta-analysis model was conducted to estimate the pooled mean difference of the standard deviations. Where results indicated that the post-intervention standard deviation was equal to, or less than the baseline standard deviation, this was interpreted as evidence of a negative relationship between baseline and change values. Where results indicated a greater post-intervention standard deviation, this was judged as indeterminate due to the potential for random variation in intervention effects to increase the post-intervention standard deviation.

Results: Moderate evidence was obtained for a reduction in standard deviation post-intervention for the vertical jump ($\text{Difference}_{0.5}$ = -0.07 [95%CrI: -0.16 to 0.02 cm]; p(Difference <0)= 0.933); and strong evidence for the same with sprint time across all three distances (10 m: $\text{Difference}_{0.5}$ = -0.007 [95%CrI: -0.012 to -0.003 s]; p(Difference <0)>0.999; 20 m: $\text{Difference}_{0.5}$ = -0.020 [95%CrI: -0.034 to -

0.009 s]; *p*(Difference <0)>0.999; 30 m: $Difference_{0.5}$ = -0.011 [95%CrI: -0.020 to -0.002 s]; *p*(Difference <0)=0.992). In contrast, strong evidence was obtained for an increase in post-intervention standard deviation for the 1RM squat ($Difference_{0.5}$ = 0.93 [95%CrI: 0.52 to 1.34 kg]; *p*(Difference <0)<0.001) and bench press ($Difference_{0.5}$ = 0.69 [95%CrI: 0.40 to 0.98 kg]; *p*(Difference <0)<0.001).

Conclusion: Collectively, the results present evidence for a negative IDE of baseline capability for sprint and vertical jump performance, but not maximum strength. Further research that is cognisant of the conceptual and analytical challenges in determining if baseline capability causes IDE is required, including the contexts and populations in S&C which may alter the interactions.

1.0 Introduction

Strength and conditioning (S&C) is frequently used in the development of a range of outcomes important for sporting performance and the ability to carry out demanding physical tasks. Many studies have been conducted to identify the types of interventions that best improve performance of a population on average, across a range of outcomes and domains. As knowledge has been gained and general principles developed, interest has grown in the study of maximising improvements through processes such as individualisation of interventions. To achieve individualisation, the factors that tend to influence response to an intervention must be identified and modifications then mapped to the individual's level. Non-random variation in response to an intervention is referred to as intervention differential effects (IDE), with an individual's baseline capability believed to be an important cause of IDE in S&C. It is generally posited that baseline capability is inversely related to improvements such that those with higher baseline performances experience reduced improvements (Appleby et al, 2012; Wetmore et al, 2020).

Despite the general belief that baseline capability causes IDE, there are several conceptual issues and analytical technicalities that create challenges when investigating any influence. Conceptual issues include the important difference of comparisons between and within populations, and specifics such as intervention duration. Analytical technicalities include phenomena such as mathematical coupling and regression to the mean that can interfere with statistical approaches frequently used in S&C, creating bias and, in some cases, spurious results. The purpose of this review is to highlight and discuss these conceptual issues and analytical technicalities. A more general assessment of baseline capability and IDE in S&C is also explored using meta-analyses with a large aggregate data set extracted from S&C interventions across a range of popular outcomes.

1.1 Conceptual Issues

Conceptually it is important to distinguish between the relative improvements of different populations and the relative improvements of individuals within a population that start with different capabilities. Across different populations, the influence of baseline capability and IDE is relatively simple, and research has consistently shown that more advanced populations experience smaller improvements than less advanced populations. Across a twenty-one-week resistance training intervention comparing experienced strength athletes and untrained males, Ahtiainen et al, (2003) reported increases of 20.9% in maximum force and 5.6% in muscle cross-sectional area for the untrained group, compared with changes of 3.9 and -1.8% for the trained group. Additionally, multiple long-term (≥1 year) studies have identified relatively small magnitude (Appleby et al, 2012) or non-significant changes (Häkkinen et al, 1987) in the strength of elite athletes clearly indicating the existence of ceiling effects. What is less clear, however, is whether those with higher baseline values experience different magnitudes of improvement compared to those with lower baseline values within the same population. For example, whether the strongest individuals within an untrained population (e.g. have never performed structured resistance exercise) will increase strength more or less than those with lower baseline values. Additionally, it is not clear whether any IDE caused by baseline capability would be similar within untrained, intermediate, and elite populations. Whilst often not stated explicitly in these terms, this appears to be one of the main questions of interest for individualising S&C interventions.

Another important conceptual issue is the duration of the intervention. Most research in S&C is conducted over a single intervention with durations between 6 and 12 weeks, and 95% of interventions lasting less than 25 weeks (Swinton et al, 2022). The change across such short time periods is likely to be relatively simple and constrain the form of any relationship with baseline capability. In contrast, long-term interventions are likely to create more complex and non-linear changes (Steele et al, 2022), such that any relationship may be different to those obtained with short interventions. Future research should be cognisant of these potential differences and seek to explore IDE and baseline capability relationships across a range of durations employing suitable research designs and analyses.

1.2 Analytical

Previous research in S&C has generally adopted one of two different approaches to investigate baseline capability and IDE within a population. The first is to perform the intervention and then include basic correlation or regression analyses of change and baseline values (Appleby et al, 2012; Latella 2020). The second is to perform the intervention and then group individuals using baseline values and a threshold, comparing differences in mean change between the created groups (James et al, 2018; Wetmore et al, 2020). Both approaches can be severely impaired and result in spurious relationships or differences where none exist, or bias results where there is an underlying effect. These limitations are due to mathematical coupling and regression to the mean. In the following sections these processes are explained with a plausible data generating model used to provide interpretable formulae.

1.2.1 Mathematical coupling

The correlation between baseline (Pre) and change ($Post - Pre$) values results in part from the mathematical coupling between the two terms. Mathematical coupling occurs when one variable is part of another (Chiolero et al, 2013) and occurs independently of measurement error. Previous discussions of mathematical coupling and IDE often begin by highlighting the spurious relationship that occurs when correlating baseline and change values when Pre and $Post$ values are unrelated (see Supplementary B1). Baseline and post-intervention values are, however, always correlated in S&C, such that mathematical coupling does not produce spurious results, but is best understood as a process that constrains the variances and relationships between baseline, post-intervention, and change values. To highlight these constraints, it is best to introduce a plausible data generating model that describes IDE due to baseline values. A simple data generating model includes the following:

$$Post = Pre + \beta_0 + \beta_1 Pre + \xi. \qquad eq. 1$$

$Post$ and Pre are the true baseline and post-intervention values without measurement error and β_0 is a constant describing systematic change. Where IDE due to baseline capability exists, β_1 will be non-zero, and in general $-1 < \beta_1 < 0$. True variation in intervention effects beyond baseline capability (e.g. variation due to the participant, training, and external factors such as nutrition and sleep) is described by the random error term $\xi \sim N(0, \nu^2)$. For the purposes of regression and correlation, we can also express the data generating model as:

$$Post = \beta_0 + (1 + \beta_1)Pre + \xi, \qquad eq. 2$$

where we regress post-intervention values on baseline values and obtain an estimate of β_1 by subtracting 1. Additionally, we can express the model as:

$$Post - Pre = \beta_0 + \beta_1 Pre + \xi, \qquad eq. 3$$

where we regress change on baseline values to obtain an estimate of β_1. To identify the constraints induced by mathematical coupling (Supplementary B2), we first note that the correlation between true baseline and post-intervention values (ρ) is predominantly determined by the variance ratio of baseline values (σ^2) and the random intervention effect (ν^2):

$$\text{Cor}(Pre, Post) = \frac{1+\beta_1}{\sqrt{(1+\beta_1)^2 + \frac{\nu^2}{\sigma^2}}}. \qquad eq. 4$$

The correlation between true baseline and change values (Supplementary B2) is then given by:

$$\text{Cor}(Pre, Post - Pre) = \frac{\rho\sqrt{\text{Var}(Post)} - \sqrt{\text{Var}(Pre)}}{\sqrt{\left(\text{Var}(Pre) + \text{Var}(Post) - 2\rho\sqrt{\text{Var}(Pre)\text{Var}(Post)}\right)}}. \qquad eq. 5$$

Here the constraints from mathematical coupling show that as $0 < \rho < 1$, the correlation between true baseline and change values will always be negative if the variance of post-intervention values is equal to or less than the variance of baseline values.

1.2.2 Regression to the mean

In the previous section it was shown that mathematical coupling imposes several constraints that are well understood when relating baseline and change values from a simple data generating model. Importantly, if outcomes were measured without error, simple correlation or regression could be used to obtain appropriate parameter estimates including any IDE due to baseline capability. In many measurements routinely used in S&C, however, large measurement errors occur, and it is these errors that create regression to the mean and can create biased estimates and spurious results. Regression to the mean is defined as a statistical phenomenon occurring due to errors in repeated measurements made on the same individual that can be the cause of observed change (Barnett et al, 2005). In general, we model measurement errors as being normally distributed around a hypothetical true value which remains unknown, but can be considered the average of a very large number of independent trials:

$$pre = Pre + \epsilon; post = Post + \epsilon, \qquad eq. 6$$

where lower case pre and $post$ are the observed baseline and post-intervention values, and measurement errors $\epsilon \sim N(0, \delta^2)$ are independent for each individual and each time point.

Measurement error is any difference between the observed and true value and comprises instrumentation noise and biological noise. The former includes error due to the measurement apparatus and variation in the instantiation of a test. For example, in a one-repetition maximum (1RM) test, instrumentation noise can comprise non-calibrated barbells and weights, and the test administrator failing to notice a participant shortening the required range of motion. Similarly, in a vertical jump test, instrumentation noise can comprise non-calibrated jump-mat timers, and participants extending time in the air by flexing lower-body joints. In contrast, biological noise comprises error due to biological processes such as circadian rhythm, nutritional intake, previous sleep, and motivation. It is important to note that measurement error definitions and associated gaussian models are appropriate for maximum tests such as the 1RM. That is, we conceive of an individual's true 1RM value not as the maximum performance that could be achieved under any permissible state (e.g. ruling out stimulants and any other acute enhancement), but as the theoretical average of a very large number of independent trials that will exhibit variation due to a range of instrumentation and biological factors.

The process by which measurement errors create regression to the mean is straightforward when we consider errors to be independent. Where for example, an individual experiences a large positive error, the subsequent measurement error is likely to be lower such that the observed difference ($post - pre$) will be smaller than the true difference $(Post - Pre)$. The opposite effect occurs for individuals that experience an initial large negative error. This phenomenon can be easily illustrated by correlating the initial measurement error with the change value which creates a negative correlation with greater absolute value for tests with higher measurement errors (Supplementary C1).

To quantify the effects of regression to the mean when estimating the influence of baseline capability on change, we consider separately the two standard analytical approaches used in S&C research. Firstly, we consider the continuous case where an intervention is conducted, and change values are regressed on baseline values. Given our data generating model presented in $eq. 1$, we wish to estimate the parameter β_1. Where there is no measurement error, we obtain an unbiased estimate. We can show, however (Supplementary C2), where we have measurement error simply regressing observed change and baseline values gives the estimate $\hat{\beta}_1$, where:

$$\hat{\beta}_1 = \frac{\beta_1\sigma^2 - \delta^2}{\sigma^2 + \delta^2}. \qquad eq.7$$

This equation shows that where there is no IDE from baseline values ($\beta_1 = 0$), we will estimate a negative regression coefficient that increases in absolute magnitude as measurement errors increase and the group studied is more homogenous. Where there is IDE from baseline values ($\beta_1 \neq 0$) the equation shows that bias will increase with greater measurement error (approaching -1) and will tend to decrease towards the true population value as the baseline variation increases.

The second analytical approach commonly used in S&C to investigate potential IDE from baseline values is to conduct an intervention and post-hoc, split the sample into groups using baseline values and a threshold. The threshold may be determined a priori (Wetmore et al, 2020) or with an arbitrary midpoint generating what can be labelled for example 'stronger' and 'weaker' groups (James et al, 2018). Conceptually, the limitation of this method is that some of the individuals in the stronger group will have exceeded the threshold based on a positive measurement error, and conversely, some of those in the weaker group due to a negative measurement error. Regression to the mean tends to result in those mislabelled as stronger observing lower change and those mislabelled as weaker observing greater change than is true. When the mean of the weaker group is subtracted from the stronger group regression to the mean will bias results and where there is no IDE from baseline values, we will tend to observe a negative value indicating greater improvement of the weaker group. More formally, it can be shown (Supplementary C3) where there is no IDE from baseline values the regression to the mean when selecting a threshold c is equal to:

$$-\frac{\delta^2}{\sqrt{(\sigma^2+\delta^2)}}\left(C\left(z_{pre}^{-}\right) + C\left(z_{pre}^{+}\right)\right), \qquad eq.8$$

where $z_{pre}^{+} = \left(\frac{c-\mu_{pre}}{\sqrt{\sigma^2+\delta^2}}\right)$, $z_{pre}^{-} = \left(\frac{\mu_{pre-c}}{\sqrt{\sigma^2+\delta^2}}\right)$, and $C(z)$ is the ratio of the standard normal density function ($\phi(z)$) and one minus the cumulative distribution function ($\Phi(z)$). As with the continuous case, the analysis shows that the magnitude of the regression to the mean effect is increased as measurement error increases and the group studied is more homogenous. Additionally, eq.8 shows that the regression to the mean effect is increased as the threshold value moves further from the population mean.

1.2.2 Accounting for regression to the mean

Multiple methods have been proposed to account for regression to the mean and obtain unbiased estimates of the relationship between baseline capability and change. One popular method includes Oldham's method (1962) which regresses change values on the average of the baseline and post-intervention values. This method is reflective of the Bland-Altman method and associated plot frequently used in criterion validity and reliability analyses (Bland and Altman 1986). In some cases, Oldham's method (1962) obtains non-biased estimates of the relationship between baseline capability and change. The biological significance of a correlation between change values and the middle value of an intervention has, however, been criticised (MacGregor et al, 1985). In addition, the method does not perform well when there is variation in intervention effect reflected by ν^2 (Hayes 1988). It is important to note the difference between variation in intervention effect and the concept of trainability that has been discussed more recently in S&C (Hecksteden et al, 2015). Here, trainability refers to individual participant factors (e.g. genotype, previous training history) that interact with the training stimulus to cause systematic variation in the intervention effect (Hecksteden et al, 2015). These factors could be added to the data generating model and reduce the random variation ν^2. The extent to which these factors exist, however, is debated, with many recent reviews failing to provide quality evidence (Williamson et al, 2017;Bonafiglia et al, 2022). Approaches to test for trainability have typically compared variation in change values between intervention and control groups, and concluded there is evidence of trainability when the variance is greater in those performing the intervention (Atkinson et al, 2019). Whilst this approach may appropriately identify the existence of trainability when systematic variation is large and positive, there are several potential subtleties that may conceal any relationship and cause observed variation to be similar between intervention and control groups (

An alternative method that can be used to estimate the relationship between baseline capability and change whilst accounting for regression to the mean is Blomqvist's method (1977). In $eq. 7$ we presented the estimate of the IDE from baseline values ($\hat{\beta}_1$) in terms of the true population value (β) and showed that the two were not equal. In Blomqvist's method (1977), we rewrite the equation and identify the adjustment required so that our new estimate ($\tilde{\beta}_1$) provides an unbiased estimate of the true population value using the original biased estimate:

$$\tilde{\beta}_1 = \frac{\hat{\beta}_1(\sigma^2+\delta^2)+\delta^2}{\sigma^2}. \qquad eq. 9$$

Blomqvist also provided an approximate standard error (Blomqvist 1977), and we note that the expression does not feature ν^2, such that the relationship holds regardless of whether there is variation in the intervention effect. To use Blomqvist's method we must include an estimate of the measurement error δ^2 which can be obtained from reliability studies reporting typical error. Additionally, the baseline value standard deviation σ should represent the population standard deviation, and not just those included in the study (Hayes 1988). Multiple simulations have confirmed that Blomqvist's method can be used to obtain suitable estimates of the relationship between change and baseline capability (Hayes 1988; Chiolero et al, 2013), however, additional methods including the use of multi-level models can also be used and may be more effective where measurements are collected multiple times during the intervention (Chiolero et al, 2013).

2.0 Meta-analysis

The meta-analysis was conducted on a database of S&C training studies obtained from a search of the literature comprising studies from 1962 to 2018. The database included information describing outcome variables along with baseline and follow-up means and standard deviations and has been described elsewhere (Swinton et al, 2022). For the current meta-analysis, baseline and follow-up standard deviations were extracted from the following outcomes: 1) 1RM squat (kg); 2) 1RM bench press (kg); 3) maximum vertical jump (cm); 4) 10 m sprint time (s); 5) 20 m sprint time (s); and 6) 30 m sprint time (s). Data were extracted from a total of 421 studies comprising 819 groups, 10,267 participants and 2,037 outcomes (Table 1).

Table 1: Description of data extracted for different outcomes.

	1RM squat	**1RM bench press**	**Vertical jump**	**10 m sprint**	**20 m sprint**	**30 m sprint**
Number of studies	121	103	312	95	97	58
Number of outcomes	329	307	896	194	193	118
Intervention duration (median [IQR]) Weeks	8 [6-10]	8 [6-12]	8 [6-10]	8 [6-10]	8 [6-10]	8 [6-10]
Number of groups	254	219	595	170	174	102
Number of participants	3137	2959	7374	2181	2372	1346

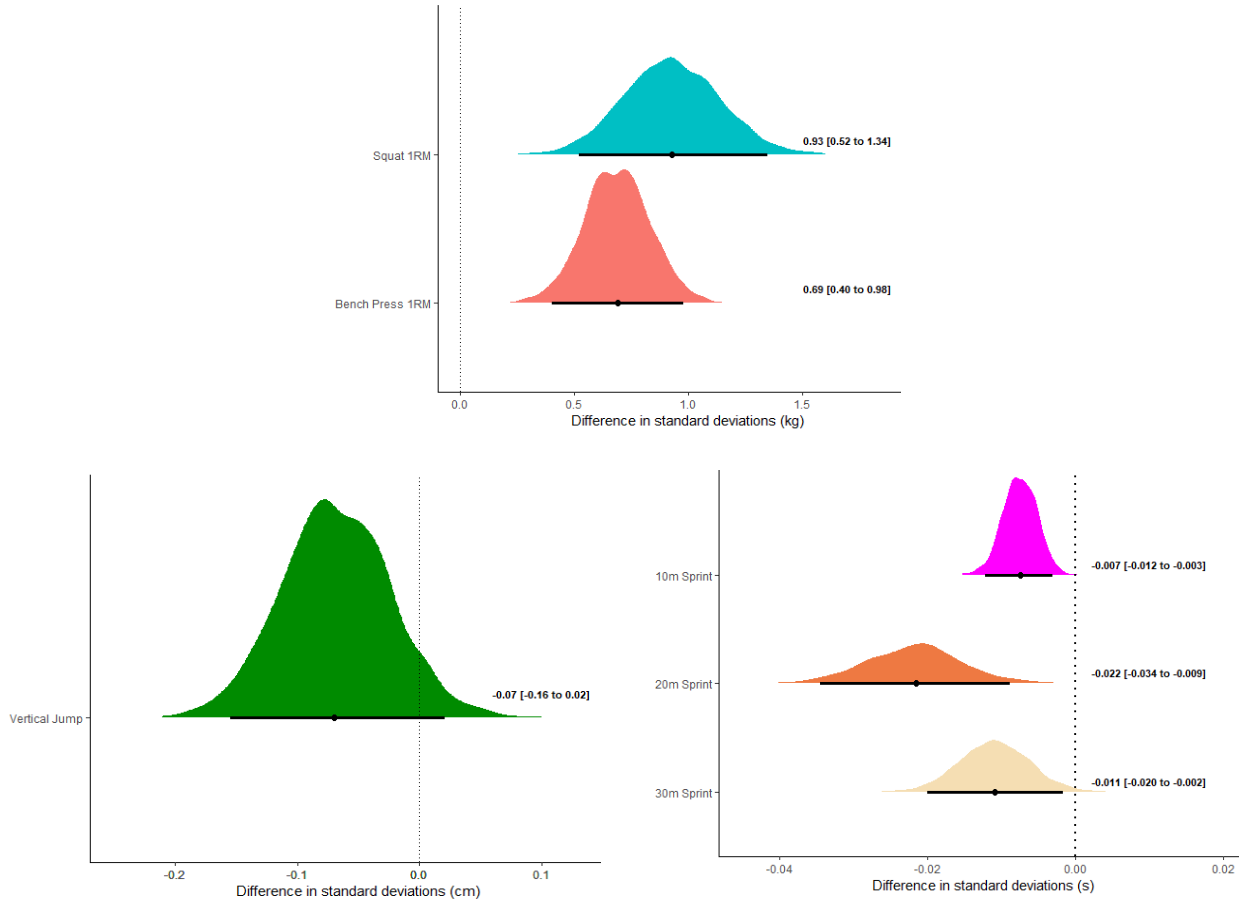
Meta-analyses were conducted for each outcome to quantify and pool the difference in sample standard deviation s, at each time point with baseline. Within-study errors $\sqrt{\text{Var}(s_{post} - s_{pre})}$ were and accounted for the correlation between time points (Supplementary D). Briefly, the standard error of the standard deviation at each time point was calculated from $\sqrt{\text{Var}(s)} = \sqrt{E(s^2) - E(s)^2}$. Given standard distributional assumptions and knowledge that the sample variance is an unbiased estimator we obtain $\sqrt{\text{Var}(s)} = s\sqrt{1 - \frac{2\lambda_n^2}{n-1}}$, where $\lambda_n = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{(n-1)}{2}\right)}$. Using the correlation between sample standard deviations $r_{s_{pre}s_{post}} = \frac{2\lambda_n^2(H(\rho^2)-1)}{n-1-2\lambda_n^2}$, where ρ is the correlation between baseline and other time points, and H is a hypergeometrical series (Pearson 1925 and Supplementary D), we have: $\sqrt{\text{Var}(s_{post} - s_{pre})} = \sqrt{\left(1 - \frac{2\lambda_n^2}{n-1}\right)\left(s_{pre}^2 + s_{post}^2 - 4\frac{\lambda_n^2(H(\rho^2)-1)}{n-1-2\lambda_n^2} s_{pre}s_{post}\right)}$.

Interpretations of meta-analyses were based on the pooled mean difference between post-intervention and baseline standard deviations. From eq5, where the standard deviation was judged to decrease or remain constant across the intervention, this was interpreted as evidence of a negative relationship between baseline capability and change. For the case where standard deviation increased across the intervention this was deemed to be indeterminate, as the increase could be caused by random variation in intervention effects combined with either a positive, negative, or no relationship between baseline capability and change. All meta-analyses were conducted using Bayesian three-level hierarchical models to account for covariances between multiple outcomes reported in the same study due to inclusion of multiple groups and/or reporting across multiple time-points. Weakly informative Student's and half Student's t priors with 3 degrees of freedom were used for intercept and variance parameters, respectively. Inferences from all analyses were performed on posterior samples generated using the Hamiltonian Markov Chain Monte Carlo method with 4 chains for 20,000 iterations with a burn-in period of 10,000. Interpretations were based on the median value ($\text{Difference}_{0.5}$: 0.5-quantile), the range within the credible interval (CrI), and the probability that the pooled mean value was less than 0. Analyses were performed using the R wrapper package brms interfaced with Stan to perform sampling (Bürkner 2017). Convergence of parameter estimates was obtained for all models with Gelman-Rubin R-hat values below 1.1 (Gelman et al, 2014).

3.0 Results

Distributions of posterior estimates are presented in Figure 1, with results showing an increase in standard deviation from baseline to post-intervention for the 1RM squat (Difference$_{0.5}$ = 0.93 [95%CrI: 0.52 to 1.34 kg]; *p*(Difference <0)<0.001) and bench press (Difference$_{0.5}$ = 0.69 [95%CrI: 0.40 to 0.98 kg]; *p*(Difference <0)<0.001). Results provided some evidence for a decrease in standard deviation from baseline to post-intervention for the vertical jump (Difference$_{0.5}$ = -0.07 [95%CrI: -0.16 to 0.02 cm]; *p*(Difference <0)= 0.933) and strong evidence for sprint time across all three distances (10 m: Difference$_{0.5}$ = -0.007 [95%CrI: -0.012 to -0.003 s]; *p*(Difference <0)>0.999; 20 m: Difference$_{0.5}$ = -0.020 [95%CrI: -0.034 to -0.009 s]; *p*(Difference <0)>0.999; 30 m: Difference$_{0.5}$ = -0.011 [95%CrI: -0.020 to -0.002 s]; *p*(Difference <0)=0.992).

Figure 1: Meta-analysis results illustrating posterior distributions of difference in baseline and post-intervention standard deviations across outcomes.



4.0 Discussion

It is commonly believed in S&C that baseline capability causes IDE and may be one of the main factors influencing observed variation in response to an intervention. As highlighted by this review, there are several important conceptual issues and technical challenges that must be considered when investigating whether baseline capability causes IDE. This is especially relevant in S&C where there is interest in maximum performance and as a result, the potential for large measurement errors that can cause substantive regression to the mean effects. In addition to highlighting some of the important conceptual and technical issues, this review adopted a distinct approach to the analysis of IDE caused by baseline capability, seeking to explore the phenomena more generally. To achieve this, meta-analyses were conducted comparing baseline and post-intervention standard deviations and pooling across a large number of studies. As identified in the introduction, post-intervention standard deviations should be expected to increase relative to baseline due to variability in response to the intervention. Based on the data generating mechanism presented, however, a negative relationship between baseline capability and change would counteract this increase. Where there is evidence that the post-intervention standard deviation is equal to, or certainly less than baseline standard deviation, this can be interpreted as providing evidence for a negative relationship between baseline capability and change. In the meta-analyses presented, moderate evidence was obtained for a reduction in standard deviation for vertical jump, and strong evidence for a reduction in standard deviation for sprint performance across all three distances. In contrast, strong evidence was obtained for an increase in standard deviation for maximum strength as measured during the squat and bench press. At present it is unknown why results were distinct across the different outcomes. It remains possible that in general a negative relationship exists between baseline capability and change for maximum strength, but that this relationship is counteracted by a greater variation in treatment response. Most interventions investigated in S&C are primarily aimed at improving maximum strength and this may explain greater variation in treatment response compared to other outcomes. Further research is required to investigate further the results presented here and the different contexts which influence the magnitude of variation in treatment response and the causes of

IDE. To conduct this research, future studies should be cognisant of the potential data generating mechanisms and include appropriate research designs and statistical approaches to account for the conceptual and technical challenges present in this area of investigation.

5.0 References

Appleby B, Newton RU, Cormie P. Changes in strength over a 2-year period in professional rugby union players. The Journal of Strength & Conditioning Research. 2012;26(9):2538-46. Doi:10.1519/JSC.0b013e31823f8b86.

Ahtiainen JP, Pakarinen A, Alen M, Kraemer WJ, Häkkinen K. Muscle hypertrophy, hormonal adaptations and strength development during strength training in strength-trained and untrained men. European journal of applied physiology. 2003l89(6):555-63. Doi:10.1007/s00421-003-0833-3.

Atkinson G, Williamson P, Batterham AM. Issues in the determination of 'responders' and 'non-responders' in physiological research. Experimental Physiology. 2019;104(8):1215-25. Doi:10.1113/EP087712.

Barnett AG, Van Der Pols JC, Dobson AJ. Regression to the mean: what it is and how to deal with it. International journal of epidemiology. 2005;34(1):215-20. Doi:10.1093/ije/dyh299.

Bland JM, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. The lancet. 1986;327(8476):307-10. Doi:10.1016/S0140-6736(86)90837-8.

Blomqvist N. On the relation between change and initial value. Journal of the American Statistical Association. 1977;72(360a):746-9. Doi:10.1080/01621459.1977.10479950.

Bonafiglia JT, Swinton PA, Ross R, Johannsen NM, Martin CK, Church TS, Slentz CA, Ross LM, Kraus WE, Walsh JJ, Kenny GP. Interindividual Differences in Trainability and Moderators of Cardiorespiratory Fitness, Waist Circumference, and Body Mass Responses: A Large-Scale Individual Participant Data Meta-analysis. Sports Medicine. 2022;52(12):2837-51. Doi:10.1007/s40279-022-01725-9.

Bürkner PC. brms: An R package for Bayesian multilevel models using Stan. Journal of statistical software. 2017;80:1-28. Doi:10.18637/jss.v080.i01.

Chiolero A, Paradis G, Rich B, Hanley JA. Assessing the relationship between the baseline value of a continuous variable and subsequent change over time. Frontiers in public health. 2013;1:29. Doi:10.3389/fpubh.2013.00029.

Cochran WG. The distribution of quadratic forms in a normal system, with applications to the analysis of covariance. In Mathematical Proceedings of the Cambridge Philosophical Society 1934 Apr (Vol. 30, No. 2, pp. 178-191). Cambridge University Press. Doi: 10.1017/S0305004100016595.

Davis CE. The effect of regression to the mean in epidemiologic and clinical studies. American journal of epidemiology. 1976 Nov 1;104(5):493-8. Doi: 10.1093/oxfordjournals.aje.a112321.

Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian data analysis (Vol. 2). 2014. Taylor & Francis

Häkkinen K, Komi PV, Alén M, Kauhanen H. EMG, muscle fibre and force production characteristics during a 1 year training period in elite weight-lifters. European journal of applied physiology and occupational physiology. 1987;56:419-27. Doi:10.1007/BF00417769.

Hayes RJ. Methods for assessing whether change depends on initial value. Statistics in medicine. 1988;7(9):915-27. Doi:10.1002/sim.4780070903.

Hecksteden A, Kraushaar J, Scharhag-Rosenberger F, Theisen D, Senn S, Meyer T. Individual response to exercise training-a statistical perspective. Journal of applied physiology. 2015;118(12):1450-9. Doi:10.1152/japplphysiol.00714.2014.

James LP, Gregory Haff G, Kelly VG, Connick MJ, Hoffman BW, Beckman EM. The impact of strength level on adaptations to combined weightlifting, plyometric, and ballistic training. Scandinavian journal of medicine & science in sports. 2018;28(5):1494-505. Doi:10.1111/sms.13045.

Latella C, Teo WP, Spathis J, van den Hoek D. Long-term strength adaptation: A 15-year analysis of powerlifting athletes. Journal of strength and conditioning research. 2020;34(9):2412. Doi:10.1519/JSC.0000000000003657.

MacGregor GA, Sagnella GA, Macrae KD, Altman DG, Hill ID. Misleading paper about misleading statistics. The Lancet. 1985;325(8434):926-7. Doi:10.1016/S)140-6736(85)9169-4.

Pearson K. Further contributions to the theory of small samples. Biometrika. 1925:176-200. Doi:10.2307/2332035.

Ross R, Goodpaster BH, Koch LG, Sarzynski MA, Kohrt WM, Johannsen NM, Skinner JS, Castro A, Irving BA, Noland RC, Sparks LM. Precision exercise medicine: understanding exercise response variability. British journal of sports medicine. 2019;53(18):1141-53. Doi:10.1136/bjsports-2018-100328.

Oldham PD. A note on the analysis of repeated measurements of the same subjects. Journal of chronic diseases. 1962;15(10):969-77. Doi:10.1016/0021-9681(62)90116-9.

Steele J, Fisher JP, Giessing J, Androulakis-Korakakis P, Wolf M, Kroeske B, Reuters R. Long-term time-course of strength adaptation to minimal dose resistance training through retrospective longitudinal growth modeling. Research Quarterly for Exercise and Sport. 2022:1-8. Doi:10.1080/02701367.2022.2070592.

Swinton PA, Burgess K, Hall A, Greig L, Psyllas J, Aspe R, Maughan P, Murphy A. Interpreting magnitude of change in strength and conditioning: Effect size selection, threshold values and Bayesian updating. Journal of sports sciences. 2022;40(18):2047-54. Doi:10.1080/02640414.2022.2128548.

Wetmore AB, Moquin PA, Carroll KM, Fry AC, Hornsby WG, Stone MH. The effect of training status on adaptations to 11 weeks of block periodization training. Sports. 2020;8(11):145. Doi:10.3390/sports8110145.

Williamson PJ, Atkinson G, Batterham AM. Inter-individual responses of maximal oxygen uptake to exercise training: a critical review. Sports Medicine. 2017;47:1501-13. Doi:10.1007/s40279-017-0680-8.

Doi:10.51224/SRXIV.285 SportRχiv Preprint version 1

The influence of baseline capability on intervention effects in strength and conditioning: A review of concepts and methods with meta-analysis.

Swinton, P.A.

Supplementary files

The following supplementary files derive the required results presented in the main paper and provides R code to illustrate and provide checks.

Supplementary A: Properties of statistical models

In this section basic properties of statistical models are outlined that will be used to derive subsequent results.

Property 1 (P1): Jointly Normal random variables: Two random variables X, Y are said to be jointly normal if they can be expressed in the form $X = aU + bV$; $Y = cU + dV$ where U and V are independent normal random variables.

Property 2 (P2): Population mean $E(X) = \mu$ and the linearity of expectation: $E(aX + bY) = aE(X) + bE(Y)$, where a and b are constants.

Property 3 (P3): Expectation of an independent product: if X and Y are independent then $E(XY) = E(X)E(Y)$.

Property 4 (P4): Population variance and expectation: $\text{Var}(X) = E(X^2) - \mu^2$.

Property 5 (P5): Variance of a linear combination: $\text{Var}(aX + by) = a^2\text{Var}(X) + 2ab\text{Cov}(X, Y) + b^2\text{Var}(Y)$.

Property 6 (P6): Covariance and expectation: $\text{Cov}(X, Y) = E(XY) - \mu_X\mu_Y$.

Property 7 (P7): Covariance and correlation: $\text{Corr}(X, Y) = \rho_{XY} = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$.

Property 8 (P8): Bivariate normal distribution:

$$\begin{matrix} X \\ Y \end{matrix} \sim \mathbb{N}\left(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \text{Var}(X) & \rho_{XY}\sqrt{\text{Var}(X)\text{Var}(Y)} \\ \rho_{XY}\sqrt{\text{Var}(X)\text{Var}(Y)} & \text{Var}(Y) \end{bmatrix}\right).$$

Property 9 (P9): Conditional expectations

$$E(XY) = E[E(XY|Y)] = E[YE(X|Y)]$$

Property 10 (P10): Conditional expectation in the bivariate general normal distribution:

$$E(Y|X = x) = \mu_Y + \rho\left(\frac{\sigma_Y}{\sigma_X}\right)(x - \mu_X).$$

Supplementary B: Mathematical Coupling

We consider two situations when calculating the correlation between true pre-intervention values (Pre) and true change values ($Post - Pre$). The first situation is where there is no correlation between Pre and $Post$, the second situation is where there is a positive correlation. Across both conditions we assume no measurement error.

Supplementary B1 - Situation 1: No correlation between Pre and Post

$$\text{Corr}(Pre, Post - Pre) = \frac{\text{Cov}(Pre, Post - Pre)}{\sqrt{\text{Var}(Pre) + \text{Var}(Post - Pre)}}$$

$$= \frac{E\big(Pre(Post - Pre)\big) - \mu_{Pre}(\mu_{Post} - \mu_{Pre})}{\sqrt{\text{Var}(Pre)\big(\text{Var}(Pre) + \text{Var}(Post)\big)}}$$

$$= \frac{E(PrePost) - E(Pre^2) - \mu_{Pre}\mu_{Post} + \mu_{Pre}^2}{\sqrt{\text{Var}(Pre)\big(\text{Var}(Pre) + \text{Var}(Post)\big)}}$$

$$= \frac{\mu_{Pre}\mu_{Post} - (\text{Var}(Pre) + \mu_{Pre}^2) - \mu_{Pre}\mu_{Post} + \mu_{Pre}^2}{\sqrt{\big(\text{Var}(Pre) + \text{Var}(Post)\big)\text{Var}(Pre)}}$$

$$= \frac{-\text{Var}(Pre)}{\sqrt{\big(\text{Var}(Pre) + \text{Var}(Post)\big)\text{Var}(Pre)}}. \qquad \text{Result 1}$$

For result 1, where $\text{Var}(Pre) = \text{Var}(Post)$ we have $\text{Corr}(Post - Pre, Pre) = \frac{-\text{Var}(Pre)}{\sqrt{2\text{Var}(Pre)^2}} = -\frac{1}{\sqrt{2}}$.

Supplementary B2 - Situation 2: Positive correlation between Pre and Post

To obtain an expression for $\text{Corr}(Pre, Post - Pre)$ under the usual circumstance where baseline and post-intervention values are correlated, we first introduce our data generating model and some basic results. The data generating model is:

$Post = Pre + \beta_0 + \beta_1 Pre + \xi$, Where $Pre \sim N(\mu, \sigma^2)$.

The model states that $Post$ values are a function of Pre values $\big(Pre \sim N(\mu_{Pre}, \sigma^2)\big)$, plus an average intervention effect β_0, some intervention differential effect (IDE) based on Pre values where $\beta_1 \neq 0$, and an independent term describing random offset of intervention effects $\xi \sim N(0, \nu^2)$.

We can see that $\mu_{Post} = E(Pre + \beta_0 + \beta_1 Pre + \xi) = (1 + \beta_1)\mu_{Pre} + \beta_0$, and

$\text{Var}(Post) = \text{Var}(Pre + \beta_0 + \beta_1 Pre + \xi) = \sigma^2(1 + \beta_1)^2 + \nu^2$.

We now consider the correlation between baseline and post-intervention values given the data generating model.

$$\text{Corr}(Pre, Post) = \frac{\text{Cov}(Pre, Post)}{\sqrt{\text{Var}(Pre)\text{Var}(Post)}}$$

$$= \frac{E(PrePost) - \mu_{Pre}\mu_{Post}}{\sqrt{\sigma^2(\sigma^2(1+\beta_1)^2 + \nu^2)}}$$

$$= \frac{E\big(Pre(Pre + \beta_0 + \beta_1 Pre + \xi)\big) - \mu_{Pre}\big(\mu_{Pre} + \beta_0 + \beta_1\mu_{Pre}\big)}{\sqrt{\sigma^2(\sigma^2(1+\beta_1)^2 + \nu^2)}}$$

$$= \frac{E\big(Pre^2\big) + \beta_0 E(Pre) + \beta_1 E\big(Pre^2\big) + E(\xi Pre) - \mu_{Pre}^2 - \beta_0\mu_{Pre} - \beta_1\mu_{Pre}^2}{\sqrt{\sigma^2(\sigma^2(1+\beta_1)^2 + \nu^2)}}$$

$$= \frac{\text{Var}(Pre) + \mu_{Pre}^2 + \beta_0\mu_{Pre} + \beta_1\big(\text{Var}(Pre) + \mu_{Pre}^2\big) + E(Pre)E(\xi) - \mu_{Pre}^2 - \beta_0\mu_{Pre} - \beta_1\mu_{Pre}^2}{\sqrt{\sigma^2(\sigma^2(1+\beta_1)^2 + \nu^2)}}$$

$$= \frac{(1+\beta_1)\sigma^2}{\sqrt{\sigma^2(\sigma^2(1+\beta_1)^2 + \nu^2)}}$$

$$= \frac{(1+\beta_1)\sigma}{\sqrt{\sigma^2(1+\beta_1)^2+\nu^2}}$$

$$= \frac{1+\beta_1}{\sqrt{(1+\beta_1)^2+\frac{\nu^2}{\sigma^2}}}. \qquad \text{Result 2}$$

Now that we have the correlation ρ between Pre and $Post$, we know that the variables follow a bivariate normal distribution and from properties 10 and 11 have $E(PrePost) = E[Post\ E(Pre|Post)]$ and $E(Pre|Post) = \mu_{Pre} + \rho\sqrt{\frac{\text{Var}(Pre)}{\text{Var}(Post)}}(Post - \mu_{Post})$.

Therefore,

$$\text{Corr}(Pre, Post - Pre) = \frac{\text{Cov}(Pre, Post-Pre)}{\sqrt{\text{Var}(Pre)\text{Var}(Post-Pre)}}$$

$$= \frac{E\big(Pre(Post-Pre)\big)-\mu_{Pre}(\mu_{Post}-\mu_{Pre})}{\sqrt{\text{Var}(Pre)\Big(\text{Var}(Pre)+\text{Var}(Post)-2\rho\sqrt{\text{Var}(Pre)\text{Var}(Post)}\Big)}}$$

$$= \frac{E(PrePost)-E(Pre^2)-\mu_{Pre}\mu_{Post}+\mu_{Pre}^2}{\sqrt{\text{Var}(Pre)\Big(\text{Var}(Pre)+\text{Var}(Post)-2\rho\sqrt{\text{Var}(Pre)\text{Var}(Post)}\Big)}}$$

$$= \frac{E[Post\ E(Pre|Post)]-\big(\text{Var}(Pre)+\mu_{Pre}^2\big)-\mu_{Pre}\mu_{Post}+\mu_{Pre}^2}{\sqrt{\text{Var}(Pre)\Big(\text{Var}(Pre)+\text{Var}(Post)-2\rho\sqrt{\text{Var}(Pre)\text{Var}(Post)}\Big)}}$$

$$= \frac{E\left[Post\left(\mu_{Pre}+\rho\sqrt{\frac{\text{Var}(Pre)}{\text{Var}(Post)}}(Post-\mu_{Post})\right)\right]-(\text{Var}(Pre)+\mu_{Pre}\mu_{Post})}{\sqrt{\text{Var}(Pre)\Big(\text{Var}(Pre)+\text{Var}(Post)-2\rho\sqrt{\text{Var}(Pre)\text{Var}(Post)}\Big)}}$$

$$= \frac{\mu_{Pre}E(Post)+\rho\sqrt{\frac{\text{Var}(Pre)}{\text{Var}(Post)}}\big[E(Post^2)-\mu_{Post}E(Post)\big]-(\text{Var}(Pre)+\mu_{Pre}\mu_{Post})}{\sqrt{\text{Var}(Pre)\Big(\text{Var}(Pre)+\text{Var}(Post)-2\rho\sqrt{\text{Var}(Pre)\text{Var}(Post)}\Big)}}$$

$$= \frac{\mu_{Pre}\mu_{Post}+\rho\sqrt{\frac{\text{Var}(Pre)}{\text{Var}(Post)}}\text{Var}(Post)-(\text{Var}(Pre)+\mu_{Pre}\mu_{Post})}{\sqrt{\text{Var}(Pre)\Big(\text{Var}(Pre)+\text{Var}(Post)-2\rho\sqrt{\text{Var}(Pre)\text{Var}(Post)}\Big)}}$$

$$= \frac{\rho\sqrt{\text{Var}(Pre)\text{Var}(Post)}-\text{Var}(Pre)}{\sqrt{\text{Var}(Pre)\Big(\text{Var}(Pre)+\text{Var}(Post)-2\rho\sqrt{\text{Var}(Pre)\text{Var}(Post)}\Big)}}$$

$$= \frac{\rho\sqrt{\text{Var}(Post)}-\sqrt{\text{Var}(Pre)}}{\sqrt{\Big(\text{Var}(Pre)+\text{Var}(Post)-2\rho\sqrt{\text{Var}(Pre)\text{Var}(Post)}\Big)}}. \qquad \text{Result 3}$$

From result 3, we can see that as ρ is positive, then the correlation between pre-intervention and change values will be negative when $\text{Var}(Pre) = \text{Var}(Post)$.

Plugging in our previously determined variances and correlation gives

$$\text{Corr}(Pre, Post - Pre) = \frac{\rho\sqrt{\text{Var}(Post)}-\sqrt{\text{Var}(Pre)}}{\sqrt{\Big(\text{Var}(Pre)+\text{Var}(Post)-2\rho\sqrt{\text{Var}(Pre)\text{Var}(Post)}\Big)}}$$

$$= \frac{\frac{(1+\beta_1)\sigma}{\sqrt{\sigma^2(1+\beta_1)^2+\nu^2}}\sqrt{\sigma^2(1+\beta_1)^2+\nu^2}-\sigma}{\sqrt{\sigma^2+\sigma^2(1+\beta_1)^2+\nu^2-2\frac{(1+\beta_1)\sigma}{\sqrt{(\sigma^2(1+\beta_1)^2+\nu^2)}}\sqrt{\sigma^2(\sigma^2(1+\beta_1)^2+\nu^2)}}}$$

$$= \frac{\beta_1\sigma}{\sqrt{\sigma^2+\sigma^2(1+\beta_1)^2+\nu^2-2(1+\beta_1)\sigma^2}}$$

$$= \frac{\beta_1\sigma}{\sqrt{\sigma^2(1+(1+\beta_1)^2-2(1+\beta_1))+\nu^2}}$$

$$= \frac{\beta_1\sigma}{\sqrt{\sigma^2\left(\beta_1^2+\frac{\nu^2}{\sigma^2}\right)}}$$

$$= \frac{\beta_1}{\sqrt{\beta_1^2+\frac{\nu^2}{\sigma^2}}}. \qquad \text{Result 4}$$

Note, we can see that from this data generating model, the example discussed in Supplementary B1 (Result 1) is simply the case where $\beta_1 = -1$, here we have $Post = \beta_0 + \xi$, $\text{Cor}(Pre, Post) = 0, \text{Var}(Post) = \nu^2$, and if $\nu^2 = \sigma^2$, then $\text{Corr}(Pre, Post - Pre) = -\frac{1}{\sqrt{2}}$.

Supplementary C: Regression to the mean

Supplementary C1 – Correlating errors and repeated measurements

Regression to the mean phenomena occur in the presence of measurement error. Up to this point we have considered measurements with no error, that is measurements that return the true values Pre and $Post$. In practice we do not have access to true values and instead our observed values pre and $post$ are assumed to comprise measurement error ϵ that are normally distributed with the same standard deviation, independent of each other, and independent of the true score such that:

$$pre = Pre + \epsilon; post = Post + \epsilon; \epsilon \sim N(0, \delta^2).$$

Derivations in subsequent sections based on our data generating model will require knowledge of the following correlations:

$$\text{Cor}(Pre, pre) = \frac{\text{Cov}(Pre,pre)}{\sqrt{\text{Var}(Pre)\text{Var}(pre)}}$$

$$= \frac{E(Pre\ pre) - \mu_{pre}\mu_{Pre}}{\sqrt{\sigma^2(\sigma^2+\delta^2)}}$$

$$= \frac{E\big((Pre+\epsilon_{pre})(Pre)\big) - \mu_{pre}^2}{\sqrt{\sigma^2(\sigma^2+\delta^2)}}$$

$$= \frac{E(Pre^2) + E(\epsilon_{pre}Pre) - \mu_{pre}^2}{\sqrt{\sigma^2(\sigma^2+\delta^2)}}$$

$$= \frac{\text{Var}(Pre) + \mu_{Pre}^2 - \mu_{pre}^2}{\sqrt{(\sigma^2+\delta^2)(\sigma^2+\nu^2+\delta^2)}}$$

$$= \frac{\sigma^2}{\sqrt{\sigma^2(\sigma^2+\delta^2)}}$$

$$= \frac{\sigma}{\sqrt{\sigma^2+\delta^2}} \qquad \text{Result 5}$$

$$\text{Cor}(Pre, post) = \frac{\text{Cov}(Pre,post)}{\sqrt{\text{Var}(Pre)\text{Var}(post)}}$$

$$= \frac{E(Pre\ post) - \mu_{Pre}\mu_{post}}{\sqrt{\sigma^2(\sigma^2(1+\beta_1)^2+\nu^2+\delta^2)}}$$

$$= \frac{E\big(Pre(Pre+\beta_0+\beta_1 Pre+\xi+\epsilon)\big) - \mu_{Pre}(\mu_{Pre}+\beta_0+\beta_1\mu_{Pre})}{\sqrt{\sigma^2(\sigma^2(1+\beta_1)^2+\nu^2+\delta^2)}}$$

$$= \frac{E(Pre^2)+\beta_0 E(Pre)+\beta_1 E(Pre^2)+E(\xi Pre)+E(Pre\epsilon)-\mu_{Pre}^2-\beta_0\mu_{Pre}-\beta_1\mu_{Pre}^2}{\sqrt{\sigma^2(\sigma^2(1+\beta_1)^2+\nu^2+\delta^2)}}$$

$$= \frac{\text{Var}(Pre)+\mu_{Pre}^2+\beta_0\mu_{Pre}+\beta_1(\text{Var}(Pre)+\mu_{Pre}^2)+E(Pre)E(\xi)+E(Pre)E(\epsilon)-\mu_{Pre}^2-\beta_0\mu_{Pre}-\beta_1\mu_{Pre}^2}{\sqrt{\sigma^2(\sigma^2(1+\beta_1)^2+\nu^2+\delta^2)}}$$

$$= \frac{(1+\beta_1)\sigma^2}{\sqrt{\sigma^2(\sigma^2(1+\beta_1)^2+\nu^2+\delta^2)}}$$

$$= \frac{(1+\beta_1)\sigma}{\sqrt{\sigma^2(1+\beta_1)^2+\nu^2+\delta^2}}$$

$$= \frac{1+\beta_1}{\sqrt{(1+\beta_1)^2+\frac{\nu^2+\delta^2}{\sigma^2}}}. \qquad \text{Result 6}$$

$$\text{Corr}(pre, post) = \frac{E(prepost) - \mu_{pre}\mu_{post}}{\sqrt{\text{Var}(pre)\text{Var}(post)}}$$

$$= \frac{E\left((Pre+\epsilon_{pre})(Pre+\beta_0+\beta_1 Pre+\xi+\epsilon_{post})\right)-\mu_{pre}(\mu_{pre}+\beta_0+\beta_1\mu_{pre})}{\sqrt{(\sigma^2+\delta^2)(\sigma^2(1+\beta_1)^2+\nu^2+\delta^2)}}$$

$$= \frac{E(Pre^2)+\beta_0 E(Pre)+\beta_1 E(Pre^2)+E(\xi Pre)+E(\epsilon_{post}Pre)+E\left(\epsilon_{pre}(Pre+\beta_0+\beta_1 Pre+\xi+\epsilon_{post})\right)-\mu_{pre}^2-\beta_0\mu_{pre}-\beta_1\mu_{Pre}^2}{\sqrt{(\sigma^2+\delta^2)(\sigma^2(1+\beta_1)^2+\nu^2+\delta^2)}}$$

$$= \frac{(1+\beta_1)(\text{Var}(Pre)+\mu_{Pre}^2)+\beta_0\mu_{Pre}-\mu_{pre}^2-\beta_0\mu_{pre}-\beta_1\mu_{Pre}^2}{\sqrt{(\sigma^2+\delta^2)(\sigma^2(1+\beta_1)^2+\nu^2+\delta^2)}}$$

$$= \frac{(1+\beta_1)\sigma^2}{\sqrt{\sigma^2\left(1+\frac{\delta^2}{\sigma^2}\right)\sigma^2\left((1+\beta_1)^2+\frac{\nu^2+\delta^2}{\sigma^2}\right)}}$$

$$= \frac{1+\beta_1}{\sqrt{\left(1+\frac{\delta^2}{\sigma^2}\right)\left((1+\beta_1)^2+\frac{\nu^2+\delta^2}{\sigma^2}\right)}}. \qquad \text{Result 7}$$

We now illustrate the general regression to the mean phenomena and start with repeated measurements and quantify the correlation between the error on the first measurement and the difference between measurements one and two.

We have $pre_i = Pre + \epsilon_i, Pre \sim N(\mu, \sigma^2), \epsilon_i \sim N(0, \delta^2), i = 1,2.$

$$\text{Corr}(\epsilon_1, pre_2 - pre_1) = \frac{E(\epsilon_1(pre_2-pre_1))}{\sqrt{\text{Var}(\epsilon_1)\text{Var}(pre_2-pre_1)}}$$

$$= \frac{E(\epsilon_1(Pre+\epsilon_2-Pre+\epsilon_1))}{\sqrt{\delta^2(\text{Var}(pre_1)+\text{Var}(pre_2)-2\text{Cov}(pre_1,pre_2))}}$$

$$= \frac{E(\epsilon_1)E(Pre)+E(\epsilon_1)E(\epsilon_2)-E(\epsilon_1)E(Pre)-E(\epsilon_1^2)}{\sqrt{\delta^2(\text{Var}(Pre+\epsilon_1)+\text{Var}(Pre+\epsilon_2)-2\text{Cov}(pre_1,pre_2))}}$$

$$= \frac{-\delta^2}{\sqrt{\delta^2\left(2(\sigma^2+2\delta^2-\text{Cov}(pre_1,pre_2))\right)}}.$$

We also have

$$\text{Cov}(pre_1, pre_2) = E(pre_1 pre_2) - \mu_{pre}^2$$

$$= E\left((Pre + \epsilon_1)(Pre + \epsilon_2)\right) - \mu_{pre}^2$$

$$= E(Pre^2) + E(Pre)E(\epsilon_2) + E(\epsilon)E(Pre) + E(\epsilon_1)E(\epsilon_2) - \mu_{pre}^2$$

$$= \sigma^2.$$

Hence

$$\text{Corr}(\epsilon_1, pre_2 - pre_1) = \frac{-\delta^2}{\sqrt{\delta^2(2(\sigma^2+2\delta^2-\sigma^2))}}$$

$$= \frac{-1}{\sqrt{2}}. \qquad \text{Result 8}$$

Supplementary C2 – Regression to the mean and continuous analyses.

For simple linear regression of y on x we have $\beta_1 = \frac{\text{Cov}(x,y)}{\text{Var}(x)}$. Therefore, given our data generating model we would estimate β_1 from

$$\hat{\beta}_1 = \frac{\text{Cov}(post-pre, pre)}{\text{Var}(pre)}$$

$$= \frac{E(pre(post-pre)) - E(pre-post)E(pre)}{\text{Var}(pre)}$$

$$= \frac{-E(pre^2) + E(pre\ post) - E(post)E(pre) + E(pre)^2}{\text{Var}(pre)}$$

$$= \frac{-(\text{Var}(pre) + \mu_{pre}^2) + \text{Cov}(pre, post) + \mu_{pre}\mu_{post} - \mu_{pre}\mu_{post} + \mu_{pre}^2}{\text{Var}(pre)}$$

$$= \frac{\text{Cov}(pre, post) - \text{Var}(pre)}{\text{Var}(pre)}.$$

From $Post = Pre + (\beta_0 + \beta_1 Pre) + \xi$, we have $Post = \beta_0 + Pre(1 + \beta_1) + \xi$.

Hence $\text{Cov}(Post, Pre) = (1 + \beta_1)\text{Var}(Pre)$.

Now we show that $\text{Cov}(pre, post) = \text{Cov}(Post, Pre)$.

$$\text{Cov}(pre, post) = \text{Cov}(Post + \epsilon_{post}, Pre + \epsilon_{pre})$$

$$= E\left((Post + \epsilon_{post})(Pre + \epsilon_{pre})\right) - \mu_{Pre}\mu_{Post}$$

$$= E(PostPre + Post\epsilon_{pre} + Pre\epsilon_{post} + \epsilon_{post}\epsilon_{pre}) - \mu_{Pre}\mu_{Post}$$

$$= \text{Cov}(Pre, Post).$$

Inserting this expression to that above gives:

$$\hat{\beta}_1 = \frac{\text{Cov}(pre, post) - \text{Var}(pre)}{\text{Var}(pre)}$$

$$= \frac{(1+\beta_1)\text{Var}(Pre) - \text{Var}(pre)}{\text{Var}(pre)}$$

$$= \frac{(1+\beta_1)\sigma^2 - (\sigma^2 + \delta^2)}{\sigma^2 + \delta^2}$$

$$= \frac{\beta_1\sigma^2 - \delta^2}{\sigma^2 + \delta^2}. \qquad \text{Result 9}$$

Here we have shown that if $\beta_1 = 0$ we obtain a spurious negative relationship. If $\beta_1 \neq 0$ there is no bias if $\delta^2 = 0$, and a bias which increases as δ^2 increases resulting in a more negative relationship. To address this bias, we can use Blomqvist's method (1977) and rewrite the equation and identify the adjustment required so that our new estimate ($\tilde{\beta}_1$) provides an unbiased estimate of the true population value using the original biased estimate:

$$\hat{\beta}_1 = \frac{\tilde{\beta}_1\sigma^2 - \delta^2}{\sigma^2 + \delta^2} \rightarrow \tilde{\beta}_1 = \frac{\hat{\beta}_1(\sigma^2 + \delta^2) + \delta^2}{\sigma^2}. \qquad \text{Result 10}$$

Supplementary C3 – Regression to the mean based on groups and threshold values.

To quantify regression to the mean in the scenario where a sample is split into two groups using baseline values relative to a threshold and the subsequent change values compared, we need to quantify conditional expectations based on true and observed values. If we have a variable $X \sim N(\mu, \sigma^2)$ and we truncate so we have two groups $X > c$ and $X < c$, then from Davis (1976) we have:

$$E(X|X > c) = \mu + C(z_X^+)\sqrt{\mathrm{Var}(X)}$$

$$E(X|X < c) = \mu - C(z_X^-)\sqrt{\mathrm{Var}(X)}$$

where $z_X^+ = \left(\frac{c-\mu}{\sqrt{\mathrm{Var}(X)}}\right)$, $z_X^- = \left(\frac{\mu-c}{\sqrt{\mathrm{Var}(X)}}\right)$, and $C(z) = \frac{\phi(z)}{1-\Phi(z)}$.

Where we don't have measurement error this would give:

$$E(X|X > c) = \mu + \left(\phi\left(\frac{c-\mu}{\sigma}\right) / \left(1 - \phi\left(\frac{c-\mu}{\sigma}\right)\right)\right)\sigma$$

$$E(X|X < c) = \mu - \left(\phi\left(\frac{\mu-c}{\sigma}\right) / \left(1 - \phi\left(\frac{c-\mu}{\sigma}\right)\right)\right)\sigma. \qquad \text{Result 11}$$

Where we introduce measurement error $x = X + \epsilon, \epsilon \sim N(0, \delta^2)$, we have:

$$E(x|x > c) = \mu + \left(\phi\left(\frac{c-\mu}{\sqrt{\sigma^2+\delta^2}}\right) / \left(1 - \phi\left(\frac{c-\mu}{\sqrt{\sigma^2+\delta^2}}\right)\right)\right)\sqrt{\sigma^2 + \delta^2}$$

$$E(x|x < c) = \mu - \left(\phi\left(\frac{\mu-c}{\sqrt{\sigma^2+\delta^2}}\right) / \left(1 - \phi\left(\frac{\mu-c}{\sqrt{\sigma^2+\delta^2}}\right)\right)\right)\sqrt{\sigma^2 + \delta^2}. \qquad \text{Result 12}$$

Also from Davis (1976), where x, y follow a bivariate normal distribution with correlation ρ, we have:

$$E(y|x > c) = \mu_y + \rho C(z_x^+)\sqrt{\mathrm{Var}(y)}$$

$$E(y|x < c) = \mu_y - \rho C(z_x^-)\sqrt{\mathrm{Var}(y)}. \qquad \text{Result 13}$$

To quantify the effects of regression to the mean we first examine the case where there is no IDE due to baseline values $(\beta_1 = 0)$ such that $Post = Pre + \beta_0 + \xi$. We show that if we were able to split the sample based on true baseline values, we obtain a non-biased estimate of the change. Using Result 2, setting $\beta_1 = 0$ and subsequently $\sqrt{\mathrm{Var}(Post)} = \sqrt{\sigma^2 + \nu^2}$, gives:

$$E(Post|Pre > c) = \mu_{Post} + \frac{1}{\sqrt{1+\frac{\nu^2}{\sigma^2}}} C(z_{Pre}^+)\sqrt{\sigma^2 + \nu^2} = \mu_{Post} + C(z_{Pre}^+)\sigma$$

$$E(Post|Pre < c) = \mu_{Post} - \frac{1}{\sqrt{1+\frac{\nu^2}{\sigma^2}}} C(z_{Pre}^-)\sqrt{\sigma^2 + \nu^2} = \mu_{Post} - C(z_{Pre}^-)\sigma$$

Note using Result 6, we obtain the same expression if we use the observed post-intervention values:

$$E(post|Pre > c) = \mu_{Post} + \frac{1}{\sqrt{1+\frac{\nu^2+\delta^2}{\sigma^2}}} C(z_{Pre}^+)\sqrt{\sigma^2 + \nu^2 + \delta^2} = \mu_{Post} + C(z_{Pre}^+)\sigma$$

$$E(post|Pre < c) = \mu_{Post} - \frac{1}{\sqrt{1+\frac{\nu^2+\delta^2}{\sigma^2}}} C(z_{Pre}^{-})\sqrt{\sigma^2+\nu^2+\delta^2} = \mu_{Post} - C(z_{Pre}^{-})\sigma.$$

We can see there is no regression to the mean when using true baseline values and the expected change across the intervention is correct:

$$E(post|Pre > c) - E(Pre|Pre > c) = \mu + \beta_0 + C(z_{Pre}^{+})\sigma - \mu + C(z_{Pre}^{+})\sigma = \beta_0$$

$$E(post|Pre < c) - E(Pre|Pre < c) = \mu + \beta_0 - C(z_{Pre}^{-})\sigma - (\mu - C(z_{Pre}^{-})\sigma) = \beta_0.$$

Where regression to the mean occurs, is when we split groups based on ***observed*** values beyond and below the threshold. We find that the expected values are biased at both baseline and post-intervention, creating the regression to the mean effect. To quantify this bias, we start with the group comprising those with observed values above the threshold and use Results 5-7:

$$[E(post|pre > c) - E(post|Pre > c)] - [E(pre|pre > c) - E(pre|Pre > c)] =$$

$$\left[\mu + \beta_0 + \frac{\sqrt{\sigma^2+\nu^2+\delta^2}}{\sqrt{\left(1+\frac{\delta^2}{\sigma^2}\right)\left(1+\frac{\nu^2+\delta^2}{\sigma^2}\right)}} C(z_{pre}^{+}) - (\mu + \beta_0 + C(z_{Pre}^{+})\sigma)\right] - \left[\mu + C(z_{pre}^{+})\sqrt{\sigma^2+\delta^2} - \left(\mu + \frac{\sigma}{\sqrt{\sigma^2+\delta^2}} C(z_{Pre}^{+})\sqrt{\sigma^2+\delta^2}\right)\right]$$

$$= \left[\sigma\left(\frac{C(z_{pre}^{+})}{\sqrt{\left(1+\frac{\delta^2}{\sigma^2}\right)}} - C(z_{Pre}^{+})\right)\right] - \left[\sigma\left(C(z_{pre}^{+})\sqrt{1+\frac{\delta^2}{\sigma^2}} - C(z_{Pre}^{+})\right)\right]$$

$$= \sigma C(z_{pre}^{+})\left(-\frac{\delta^2/\sigma^2}{\sqrt{\left(1+\frac{\delta^2}{\sigma^2}\right)}}\right)$$

$$= -C(z_{pre}^{+})\left(\frac{\delta^2}{\sqrt{(\sigma^2+\delta^2)}}\right).$$

For the group comprising those with observed values below the threshold we have:

$$[E(post|pre < c) - E(post|Pre < c)] - [E(pre|pre < c) - E(pre|Pre < c)] =$$

$$\left[\mu + \beta_0 - \frac{\sqrt{\sigma^2+\nu^2+\delta^2}}{\sqrt{\left(1+\frac{\delta^2}{\sigma^2}\right)\left(1+\frac{\nu^2+\delta^2}{\sigma^2}\right)}} C(z_{pre}^{-}) - (\mu + \beta_0 - C(z_{Pre}^{-})\sigma)\right] - \left[\mu - C(z_{pre}^{-})\sqrt{\sigma^2+\delta^2} - \left(\mu - \frac{\sigma}{\sqrt{\sigma^2+\delta^2}} C(z_{Pre}^{-})\sqrt{\sigma^2+\delta^2}\right)\right]$$

$$= \left[\sigma\left(C(z_{Pre}^{-}) - \frac{C(z_{pre}^{-})}{\sqrt{\left(1+\frac{\delta^2}{\sigma^2}\right)}}\right)\right] - \left[\sigma\left(C(z_{Pre}^{-}) - C(z_{pre}^{-})\sqrt{1+\frac{\delta^2}{\sigma^2}}\right)\right]$$

$$= \sigma C(z_{pre}^{-})\left(\frac{\delta^2/\sigma^2}{\sqrt{\left(1+\frac{\delta^2}{\sigma^2}\right)}}\right)$$

$$= C(z_{pre}^{-})\left(\frac{\delta^2}{\sqrt{(\sigma^2+\delta^2)}}\right).$$

The total regression to the mean accounting for the underestimation of the group above the threshold and the overestimation of the group below the threshold is then

$$-\frac{\delta^2}{\sqrt{(\sigma^2+\delta^2)}}\left(C(z_{pre}^{-}) + C(z_{pre}^{+})\right).$$

Supplementary D: Meta-analysis of standard deviations

To conduct a meta-analysis with standard deviations, we have the random variable for the sample standard deviation $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$, we have from Cochran's theorem (1934) that $\frac{(n-1)S^2}{\sigma^2} = U,$ where U is a random variable with chi-squared distribution χ_ν^2 with degrees of freedom $\nu = n - 1.$

To perform meta-analyses with the sample standard deviation, we require the standard error $\sqrt{\text{Var}(S)} = \sqrt{E(S^2) - E(S)^2}.$

The sample variance S^2 is an unbiased estimator hence $E(S^2) = \sigma^2$. To calculate $E(S),$ we note that $S = \sqrt{\frac{(n-1)S^2}{\sigma^2}\frac{\sigma^2}{n-1}}$ and the probability density function for χ_n^2 is $p(x) = \frac{(1/2)^{n/2}}{\Gamma\left(\frac{n-1}{2}\right)} x^{(n/2)-1}e^{-\frac{x}{2}}$. Hence

$$E(S) = \sqrt{\frac{\sigma^2}{n-1}}E\left(\sqrt{\frac{(n-1)S^2}{\sigma^2}}\right).$$

The expectation of function of a random variable $E\big(g(X)\big) = \int g(x)f(x)dx$, hence

$$E(S) = \sqrt{\frac{\sigma^2}{n-1}}\int_0^\infty \sqrt{s}\frac{(1/2)^{(n-1)/2}}{\Gamma\left(\frac{n-1}{2}\right)} s^{((n-1)/2)-1}e^{-\frac{s}{2}}\,ds$$

$$= \sqrt{\frac{\sigma^2}{n-1}}\int_0^\infty \frac{(1/2)^{(n-1)/2}}{\Gamma\left(\frac{n-1}{2}\right)} s^{(n/2)-1}e^{-\frac{s}{2}}\,ds.$$

To turn the latter expression back into a chi-squared distribution to integrate to 1, we perform the following manipulation

$$E(S) = \sqrt{\frac{2\sigma^2}{n-1}}\frac{\Gamma(n/2)}{\Gamma\left(\frac{n-1}{2}\right)}\int_0^\infty \frac{(1/2)^{n/2}}{\Gamma(n/2)} s^{(n/2)-1}e^{-\frac{s}{2}}ds.$$

Hence $E(S) = \sigma\sqrt{\frac{2}{n-1}}\frac{\Gamma(n/2)}{\Gamma\left(\frac{n-1}{2}\right)}$.

Let $\lambda_n = \frac{\Gamma(n/2)}{\Gamma\left(\frac{n-1}{2}\right)}$, then

$$\sqrt{\text{Var}(S)} = \sqrt{E(S^2) - E(S)^2}$$

$$= \sqrt{\sigma^2 - \sigma^2\frac{2\lambda_n^2}{n-1}}$$

$$= \sigma\sqrt{1 - \frac{2\lambda_n^2}{n-1}}.$$

As an estimate, we then replace σ with our observed sample standard deviation s.

Note, frequently the standard error term for s is reported as $\frac{s}{\sqrt{2(n-1)}}$, we can see this from

$$\sqrt{\text{Var}(S)} = \sigma\sqrt{1 - \frac{2}{n-1}\left(\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\right)^2}$$

$$= \sigma\sqrt{\frac{2}{n-1}}\sqrt{\frac{n-1}{2} - \left(\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\right)^2}.$$

From general properties of the Gamma function $\Gamma(a+1) = a\Gamma(a)$ we have, and $\Gamma\left(\frac{n+1}{2}\right) = \frac{n-1}{2}\Gamma\left(\frac{n-1}{2}\right)$ and $\frac{n-1}{2} = \Gamma\left(\frac{n+1}{2}\right)/\Gamma\left(\frac{n-1}{2}\right)$. Hence we can express $\sqrt{\mathrm{Var}(S)}$ as:

$$\sqrt{\mathrm{Var}(S)} = \sigma\sqrt{\frac{2}{n-1}}\sqrt{\frac{n-1}{2} - \left(\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\right)^2}$$

$$= \sigma\sqrt{\frac{2}{n-1}}\sqrt{\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} - \left(\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\right)^2}.$$

The term $\sqrt{\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} - \left(\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\right)^2}$ rapidly approaches $\frac{1}{2}$, hence

$$\sqrt{\mathrm{Var}(S)} \approx \frac{1}{2}\sigma\sqrt{\frac{2}{n-1}} = \frac{\sigma}{\sqrt{2(n-1)}}$$

Where again as an estimate, we then replace σ with s.

Now that we have the within study error for S we can calculate the within study error for the actual statistic used in the study which is $S_{Post} - S_{Pre}$. We have

$$\sqrt{\mathrm{Var}(S_{Post} - S_{Pre})} = \sqrt{\mathrm{Var}(S_{Post}) + \mathrm{Var}(S_{Pre}) - 2\mathrm{Cov}(S_{Pre}, S_{Post})}$$

$$= \sqrt{\mathrm{Var}(S_{Post}) + \mathrm{Var}(S_{Pre}) - 2\mathrm{Corr}(S_{Pre}, S_{Post})\sqrt{\mathrm{Var}(S_{Post})}\sqrt{\mathrm{Var}(S_{Pre})}}.$$

From Pearson (1925), If we have a population with σ_1, σ_2 and correlation between $X_1, X_2 = \rho$, then we have $\mathrm{Corr}(S_{X1}, S_{X2}) = \frac{2\lambda n^2(H(\rho)-1)}{n-1-2\lambda n^2}$, where H is a hypergeometrical series

$H(\rho^2) = 1 + \frac{\rho^2}{1!}\frac{1}{2n-2} + \frac{\rho^4}{2!}\frac{1}{2(n-1)(2n+2)} + \cdots \frac{\rho^{2p}}{p!}\frac{1}{(2n-2)(2n+2)\ldots(2n+4p-6)} + \cdots$, where we use up to $p = 3$ here.

Therefore, we estimate $\sqrt{\mathrm{Var}(S_{Post} - S_{Pre})}$ with

$$\sqrt{s_{Pre}^2\left(1 - \frac{2\lambda_n^2}{n-1}\right) + s_{Post}^2\left(1 - \frac{2\lambda_n^2}{n-1}\right) - 2\frac{2\lambda n^2(H(\rho)-1)}{n-1-2\lambda n^2}\sqrt{s_{Pre}^2\left(1 - \frac{2\lambda_n^2}{n-1}\right)}\sqrt{s_{Post}^2\left(1 - \frac{2\lambda_n^2}{n-1}\right)}}$$

$$= \sqrt{\left(1 - \frac{2\lambda_n^2}{n-1}\right)\left(s_{Pre}^2 + s_{Post}^2 - 4\frac{\lambda_n^2(H(\rho^2)-1)}{n-1-2\lambda_n^2}s_{Pre}s_{Post}\right)}.$$

Supplementary E: su**R code**

```
# In the following file we demonstrate the supplementary results derived
# Load packages
library(ggplot2)
library(MASS)

# Supplementary B: Mathematical Coupling
# Supplementary B1 - Situation 1: No correlation between Pre and Post
# We simulate non-correlated variables and then correlate the first with
# the difference. We included two cases, one where the variances are not
# equal and the second where they are.
set.seed(1)
n_B1 = 100000
mu_Pre_B1 = 100
mu_Post_B1 = 100
sd_Pre_B1 = 10
sd_Post1_B1 = 12
sd_Post2_B1 = 10
Pre_B1 = rnorm(n_B1,mu_Pre_B1,sd_Pre_B1)
Post1_B1 = rnorm(n_B1,mu_Post_B1,sd_Post1_B1)
Post2_B1 = rnorm(n_B1,mu_Post_B1,sd_Post2_B1)

# Calculate correlation non equal variance
round(cor(Pre_B1,Post1_B1-Pre_B1),2)
# check with derivation
round(-sd_Pre_B1^2/
  sqrt((sd_Pre_B1^2+sd_Post1_B1^2)*sd_Pre_B1^2),2)

# Calculate correlation  equal variance
round(cor(Pre_B1,Post2_B1-Pre_B1),2)
# check with derivation
round(-1/sqrt(2),2)
############

# Supplementary B2 - Situation 2 Positive correlation between Pre and Post
# We simulate data using the model
# Post = Pre + \beta_0 + \beta_1*Pre + \xi

# We simulate two conditions, the first where there is no IDE of baseline values
# such that \beta_1 = 0, and the second where \beta_1 \neq 0.
set.seed(1)
n_B2 = 1000000
mu_Pre_B2 = 100
sd_Pre_B2 = 10
beta_B2_0 = 40
beta_B2_1a = 0
beta_B2_1b = -0.3
sd_random_B2 = 5

Pre_B2 = rnorm(n_B2,mu_Pre_B2,sd_Pre_B2)
random_term_B2 = rnorm(n_B2,0,sd_random_B2)
Post_B2_a = Pre_B2 + beta_B2_0 + beta_B2_1a*Pre_B2 + random_term_B2
Post_B2_b = Pre_B2 + beta_B2_0 + beta_B2_1b*Pre_B2 + random_term_B2

# check means and variance
round(mean(Post_B2_a),1)
round(mu_Pre_B2+beta_B2_0+beta_B2_1a*mu_Pre_B2,1)

round(mean(Post_B2_b),1)
round(mu_Pre_B2+beta_B2_0+beta_B2_1b*mu_Pre_B2,1)
```

```
round(var(Post_B2_a),0)
var_Post_B2_a = (sd_Pre_B2^2)*((1+beta_B2_1a)^2)+sd_random_B2^2
round(var_Post_B2_a,0)

round(var(Post_B2_b),0)
var_Post_B2_b = (sd_Pre_B2^2)*((1+beta_B2_1b)^2)+sd_random_B2^2
round(var_Post_B2_b,0)

# check correlations pre and post (result 2)
round(cor(Pre_B2,Post_B2_a),2)
corPre_B2Post_B2_a
=(1+beta_B2_1a)/sqrt(((1+beta_B2_1a)^2)+sd_random_B2^2/sd_Pre_B2^2)
round(corPre_B2Post_B2_a,2)

round(cor(Pre_B2,Post_B2_b),2)
corPre_B2Post_B2_b
=(1+beta_B2_1b)/sqrt(((1+beta_B2_1b)^2)+sd_random_B2^2/sd_Pre_B2^2)
round(corPre_B2Post_B2_b,2)

# check Correlation between pre and change (Result 3 and 4)
round(cor(Pre_B2,Post_B2_a-Pre_B2),2)

# Result 3
(corPre_B2Post_B2_a*sqrt(var_Post_B2_a)-sd_Pre_B2)/
  sqrt(sd_Pre_B2^2+var_Post_B2_a-
(2*corPre_B2Post_B2_a*sqrt((sd_Pre_B2^2)*var_Post_B2_a)))

# Result 4
round(beta_B2_1a/(sqrt(beta_B2_1a^2+(sd_random_B2^2)/sd_Pre_B2^2)),2)


round(cor(Pre_B2,Post_B2_b-Pre_B2),2)

# Result 3
round((corPre_B2Post_B2_b*sqrt(var_Post_B2_b)-sd_Pre_B2)/
        sqrt(sd_Pre_B2^2+var_Post_B2_b-
(2*corPre_B2Post_B2_b*sqrt((sd_Pre_B2^2)*var_Post_B2_b))),2)

# Result 4
round(beta_B2_1b/(sqrt(beta_B2_1b^2+(sd_random_B2^2)/sd_Pre_B2^2)),2)
############

# Supplementary C1 - Correlating errors and repeated measurements
# We start with using our data generating model and generating true and observed
# values and check the expressions for the various correlations derived.

# We simulate two sets of pre value and two sets of post values. For the latter
# we simulate with and without IDE of baseline values.
set.seed(1)
n_C = 1000000
mu_Pre_C = 100
sd_Pre_C = 10
beta_C_0 = 40
beta_C_1a = 0
beta_C_1b = -0.3
sd_random_C = 8
sd_epsilon_C = 5

Pre_C = rnorm(n_C,mu_Pre_C,sd_Pre_C)
random_term_C = rnorm(n_C,0,sd_random_C)
Post_C_a = Pre_C + beta_C_0 + beta_C_1a*Pre_C + random_term_C
```

```
Post_C_b = Pre_C + beta_C_0 + beta_C_1b*Pre_C + random_term_C

error_C_pre_a = rnorm(n_C,0,sd_epsilon_C)
error_C_pre_b = rnorm(n_C,0,sd_epsilon_C)
error_C_post_a = rnorm(n_C,0,sd_epsilon_C)
error_C_post_b = rnorm(n_C,0,sd_epsilon_C)

pre_C_a = Pre_C + error_C_pre_a
pre_C_b = Pre_C + error_C_pre_b
post_C_a = Post_C_a + error_C_post_a
post_C_b = Post_C_b + error_C_post_b

# Check Result 5
round(cor(pre_C_a,Pre_C),2)
Cor_Pre_pre_C = sd_Pre_C/sqrt(sd_Pre_C^2+sd_epsilon_C^2)
round(Cor_Pre_pre_C,2)

# Check Result 6
round(cor(post_C_a,Pre_C),2)
Cor_Pre_post_C_a = (1+beta_C_1a)/
  sqrt((1+beta_C_1a)^2 + (sd_random_C^2+sd_epsilon_C^2)/sd_Pre_C^2)
round(Cor_Pre_post_C_a,2)

round(cor(post_C_b,Pre_C),2)
Cor_Pre_post_C_b = (1+beta_C_1b)/
  sqrt((1+beta_C_1b)^2 + (sd_random_C^2+sd_epsilon_C^2)/sd_Pre_C^2)
round(Cor_Pre_post_C_b,2)

# Check Result 7
round(cor(pre_C_a,post_C_a),2)
Cor_pre_post_C_a = (1+beta_C_1a)/
  sqrt((1+((sd_epsilon_C^2)/(sd_Pre_C^2)))*
          ((1+beta_C_1a)^2 + (sd_random_C^2+sd_epsilon_C^2)/sd_Pre_C^2))
round(Cor_pre_post_C_a,2)

round(cor(pre_C_a,post_C_b),2)
Cor_pre_post_C_b = (1+beta_C_1b)/
  sqrt((1+((sd_epsilon_C^2)/(sd_Pre_C^2)))*
         ((1+beta_C_1b)^2 + (sd_random_C^2+sd_epsilon_C^2)/sd_Pre_C^2))
round(Cor_pre_post_C_b,2)

# To provide a general illustration of regression to the mean, we explore the
# correlation between measurement error and the difference between two observed
# values.

plotRTM1 = data.frame(x =error_C_pre_a, y =pre_C_b-pre_C_a)
ggplot(plotRTM1[1:10000,],aes(x=x,y=y)) + geom_point() +
  theme_classic() + labs(x="Measurement 1 error",
                         y = "Change value") +
  geom_vline(xintercept=0, linetype="dashed", color = "red") +
  geom_hline(yintercept=0, linetype="dashed", color = "red")

# Check Result 8
round(cor(error_C_pre_a,pre_C_b-pre_C_a),2)
round(-1/sqrt(2),2)
############

# Supplementary C2 - Regression to the mean and continuous analyses.

# We now investigate regression to the mean by estimating \beta_1 using
# both true and observed values
```

```
# Case where \beta_1 = 0
# True values
round(summary(lm(Post_C_a~Pre_C))$coefficients[2,1]-1,2)
round(summary(lm(Post_C_a-Pre_C~Pre_C))$coefficients[2,1],2)

# observed values
round(summary(lm(post_C_a~pre_C_a))$coefficients[2,1]-1,2)
round(summary(lm(post_C_a-pre_C_a~pre_C_a))$coefficients[2,1],2)

# Check Result 9
round((beta_C_1a*sd_Pre_C^2 - sd_epsilon_C^2)/
  (sd_Pre_C^2 + sd_epsilon_C^2),2)

# Case where \beta_1 \neq 0
# True values
round(summary(lm(Post_C_b~Pre_C))$coefficients[2,1]-1,2)
round(summary(lm(Post_C_b-Pre_C~Pre_C))$coefficients[2,1],2)

# observed values
round(summary(lm(post_C_b~pre_C_a))$coefficients[2,1]-1,2)
round(summary(lm(post_C_b-pre_C_a~pre_C_a))$coefficients[2,1],2)

# Check Result 9
round((beta_C_1b*sd_Pre_C^2 - sd_epsilon_C^2)/
        (sd_Pre_C^2 + sd_epsilon_C^2),2)

# use Blomqvist's method
# Case where \beta_1 = 0
# Check Result 10
((((beta_C_1a*sd_Pre_C^2 - sd_epsilon_C^2)/
     (sd_Pre_C^2                                            +
sd_epsilon_C^2))*(sd_Pre_C^2+sd_epsilon_C^2))+sd_epsilon_C^2)/
  sd_Pre_C^2


# Case where \beta_1 \neq 0
# Check Result 10
((((beta_C_1b*sd_Pre_C^2 - sd_epsilon_C^2)/
  (sd_Pre_C^2 + sd_epsilon_C^2))*(sd_Pre_C^2+sd_epsilon_C^2))+sd_epsilon_C^2)/
  sd_Pre_C^2
############

#  Supplementary  C3  –  Regression  to  the  mean  based  on  groups  and  threshold
values.
# To quantify regression to the mean in the scenario where a sample is split
# into two groups using baseline values relative to a threshold and
# the subsequent change values compared, we need to quantify
# conditional expectations based on true and observed values.

#  First  we  provide  checks  on  expectations  conditioned  on  true  and  observed
values
# generated previously and the threshold 110
C_thresh = 110

# Check Result 11
z_plus_Pre = (C_thresh-mu_Pre_C)/sd_Pre_C
z_minus_Pre = (mu_Pre_C-C_thresh)/sd_Pre_C
Cz_plus_Pre = dnorm(z_plus_Pre)/(1-pnorm(z_plus_Pre))
Cz_minus_Pre = dnorm(z_minus_Pre)/(1-pnorm(z_minus_Pre))
```

```
round(mean(Pre_C[Pre_C>C_thresh]),2)
round(mu_Pre_C + sd_Pre_C*Cz_plus_Pre,2)

round(mean(Pre_C[Pre_C<C_thresh]),2)
round(mu_Pre_C - sd_Pre_C*Cz_minus_Pre,2)

# Check Result 12
z_plus_pre = (C_thresh-mu_Pre_C)/sqrt(sd_Pre_C^2+sd_epsilon_C^2)
z_minus_pre = (mu_Pre_C-C_thresh)/sqrt(sd_Pre_C^2+sd_epsilon_C^2)
Cz_plus_pre = dnorm(z_plus_pre)/(1-pnorm(z_plus_pre))
Cz_minus_pre = dnorm(z_minus_pre)/(1-pnorm(z_minus_pre))

round(mean(pre_C_a[pre_C_a>C_thresh]),2)
round(mu_Pre_C + sqrt(sd_Pre_C^2+sd_epsilon_C^2)*Cz_plus_pre,2)

round(mean(pre_C_a[pre_C_a<C_thresh]),2)
round(mu_Pre_C - sqrt(sd_Pre_C^2+sd_epsilon_C^2)*Cz_minus_pre,2)

# We provide a check on the post intervention expectations conditioned on true
# baseline values. Here we assume that \beta_1 = 0.

# Check Result 13
round(mean(Post_C_a[Pre_C>C_thresh]),1)
round(mean(post_C_a[Pre_C>C_thresh]),1)
round(mu_Pre_C + beta_C_0 + Cz_plus_Pre*sd_Pre_C,1)

round(mean(Post_C_a[Pre_C<C_thresh]),1)
round(mean(post_C_a[Pre_C<C_thresh]),1)
round(mu_Pre_C + beta_C_0 - Cz_minus_Pre*sd_Pre_C,1)

# Still assuming that \beta_1 = 0, we quantify the regression to the mean
# that will occur when groups are determined based on observed baseline values

# We achieve this by calculating the difference between expectations when
# using observed versus true baseline values

round((mean(post_C_a[pre_C_a>C_thresh])-
    mean(post_C_a[Pre_C>C_thresh])) -
  (mean(pre_C_a[pre_C_a>C_thresh])-
     mean(pre_C_a[Pre_C>C_thresh])),1)

round(-Cz_plus_pre*(sd_epsilon_C^2/sqrt(sd_epsilon_C^2+sd_Pre_C^2)),1)

round((mean(post_C_a[pre_C_a<C_thresh])-
         mean(post_C_a[Pre_C<C_thresh])) -
        (mean(pre_C_a[pre_C_a<C_thresh])-
           mean(pre_C_a[Pre_C<C_thresh])),1)

round(Cz_minus_pre*(sd_epsilon_C^2/sqrt(sd_epsilon_C^2+sd_Pre_C^2)),1)

# Total regression to the mean
round(((mean(post_C_a[pre_C_a>C_thresh])-
         mean(post_C_a[Pre_C>C_thresh])) -
        (mean(pre_C_a[pre_C_a>C_thresh])-
           mean(pre_C_a[Pre_C>C_thresh])))-
  ((mean(post_C_a[pre_C_a<C_thresh])-
      mean(post_C_a[Pre_C<C_thresh])) -
     (mean(pre_C_a[pre_C_a<C_thresh])-
        mean(pre_C_a[Pre_C<C_thresh]))),1)
round(-
(Cz_minus_pre+Cz_plus_pre)*(sd_epsilon_C^2/sqrt(sd_epsilon_C^2+sd_Pre_C^2)),1)
```

```
############

# Supplementary D: Meta-analysis of standard deviations


# Standard error of standard deviation
# sd(s) = s*(\sqrt(1-2/n-1*\lambda_n)

# function for lambda
lambdan = function(n){
  gamma(n/2)/gamma((n-1)/2)
}

# Function for Se of S
SeS = function(s,n){
  s*sqrt(1-((2*lambdan(n)^2)/(n-1)))
  }

# We compare across group sizes of 10,25,50,100

# Collect results
set.seed(123)
sesCollectn = c(10,25,50,100)
  sesCollect = matrix(NA, nrow = 100000, ncol=4)
for(j in 1:4){
  for(i in 1:100000){
        sesCollect[i,j] = sd(rnorm(sesCollectn[j],100,20))
  }}

round(apply(sesCollect,2,sd),2)
# Check
round(SeS(20,sesCollectn),2)


# We now look at the correlation between standard deviations
Hrho2 = function(n,rho2){
  1 + rho2*(1/(2*n-2)) +
    (rho2^2/2)*(1/((2*n-2)*(2*n+2))) +
    (rho2^3/6)*(1/((2*n-2)*(2*n+2)*(2*n+6)))
}

cors12 = function(n,rho){
  (2*(lambdan(n)^2)*(Hrho2(n,rho^2)-1))/
    (n-1-(2*lambdan(n)^2))
}
# We simulate data from multivariate normal distribution with correlations equal
# to 0, 0.25, 0.5 and 0.75. across sample sizes of 10,25,50,100.
Sigma0 = matrix(c(20^2, 0*20*20, 0*20*20, 20^2),ncol=2)
Sigma025 = matrix(c(20^2, 0.25*20*20, 0.25*20*20, 20^2),ncol=2)
Sigma05 = matrix(c(20^2, 0.5*20*20, 0.5*20*20, 20^2),ncol=2)
Sigma075 = matrix(c(20^2, 0.75*20*20, 0.75*20*20, 20^2),ncol=2)

Sigmas = list(Sigma0,Sigma025,Sigma05,Sigma075)

set.seed(123)
sdcorCollect = array(NA, c(100000,2,4))
for(j in 1:4){
  for(i in 1:100000){
    Data = mvrnorm(sesCollectn[j],c(100,100),Sigmas[[j]])
    sdcorCollect[i,1,j] = sd(Data[,1])
    sdcorCollect[i,2,j] = sd(Data[,2])
```

```
  }}

round(cor(sdcorCollect[,1,1],sdcorCollect[,2,1]),2)
# 0
round(cors12(10,0),2)

round(cor(sdcorCollect[,1,2],sdcorCollect[,2,2]),2)
# 0.06
round(cors12(25,0.25),2)

round(cor(sdcorCollect[,1,3],sdcorCollect[,2,3]),2)
# 0.25
round(cors12(50,0.5),2)

round(cor(sdcorCollect[,1,4],sdcorCollect[,2,4]),2)
# 0.56
round(cors12(100,0.75),2)


# We now look at the standard error of the difference in standard deviations
SEsdiff = function(sd1,sd2,n,rho){

  sqrt((1-2*(lambdan(n)^2)/(n-1))*(sd1^2+sd2^2 - 2*cors12(n,rho)*sd1*sd2))
}


round(sd(sdcorCollect[,2,1]-sdcorCollect[,1,1]),2)
round(SEsdiff(20,20,10,0),2)

round(sd(sdcorCollect[,2,2]-sdcorCollect[,1,2]),2)
round(SEsdiff(20,20,25,0.25),2)

round(sd(sdcorCollect[,2,3]-sdcorCollect[,1,3]),2)
round(SEsdiff(20,20,50,0.5),2)

round(sd(sdcorCollect[,2,4]-sdcorCollect[,1,4]),2)
round(SEsdiff(20,20,100,0.75),2)
```