

A Spreadsheet for Monitoring an Individual's Changes and Trend

Will G Hopkins

Sportscience 21, 5-9, 2017 (sportsci.org/2017/wghtrend.htm)

Institute of Sport Exercise & Active Living, Victoria University, Melbourne, Australia. [Email](#). Reviewer: Martin Buchheit, Paris Saint-Germain, 78100 Saint-Germain-en-Laye, Paris, France.

A spreadsheet was available previously at the Sportscience site to apply magnitude-based inference to an individual's change scores between two tests. This spreadsheet has been updated and included in a workbook that has a spreadsheet devoted to monitoring changes and a linear trend between more than two tests. The monitoring spreadsheet applies magnitude-based inference to the following outcomes: change of each test from a chosen reference test or from the average of several chosen tests; changes between consecutive tests; deviation of one test or the average of several tests from a linear trend fitted to chosen time points; and the magnitude of the trend. Along with the test scores, the user inputs appropriate values of the smallest important change between tests and a desired change over a longer period to define a target trend. Input of a short-term typical error of measurement is optional; the spreadsheet otherwise estimates typical error from the linear trend. A time-series graphic helps the user to choose tests for estimating the trend and to identify obvious changes and deviations. Numeric chances of substantial and trivial real changes and trend are displayed and used to make non-clinical magnitude-based inferences about the true changes and trend. The user can gain some idea of the inferential error rates with a panel of cells in another spreadsheet (included in the workbook) that estimates chances of seeing various outcomes for a chosen true change score. A further spreadsheet can be used for the same purpose: it simulates realistic data from given true values of starting score, trend, extra changes at given time points, and typical error; it then performs inferential analyses of the simulated data in the same manner as the first spreadsheet. **KEYWORDS:** athlete, evaluation, magnitude-based inference, patient, performance indicator, probability, precision of estimation, simulation, test.

[Reprint pdf](#) · [Reprint docx](#) · [Workbook](#) · [Slideshow](#) · [Old spreadsheet](#)

Introduction	6
Change between Measurements	6
Deviation from the Trend	7
Magnitude of the Trend	7
Simulating Test Scores.....	7
References.....	8
Appendix 1: Smallest Important Changes	8
Performance tests for athletes competing for a best competition score	8
Tests and performance indicators	
with uncertain relationships to performance.....	8
Measures unrelated to performance	8
Measures where standardization is not appropriate	9
Appendix 2: A Target Trend.....	9
Appendix 3: Typical Error	10

Update Feb 2023. A short [slideshow](#) (5 slides) is now available. The spreadsheet for assessing a trend now includes panels for assessing changes from a **chosen reference test** or the mean of several tests (since changes between

consecutive tests several months apart could all be trivial and/or unclear, yet changes between an early test and later tests could be more decisive and encouraging for the athlete and the practitioner.) **Smallest important for psy-**

chometric measures are now included in Appendix 1, and advice on **averaging repeats to reduce typical error** is included in Appendix 3.

Update June 2018. It's a bit hard to find at this site, so here is the link to a [slideshow on medal-winning enhancements of performance](#) presented at the [performance-analysis conference](#) and [Olympic conference](#) in 2016.

Update May 2018. The original spreadsheet for assessing a simple change score has been updated, simplified, and included as the first tab in the workbook of spreadsheets for monitoring an individual.

Updates Feb 2018. [Martin Buchheit](#), the reviewer of this article, has published a very useful [summary of methods](#) (Buchheit, 2018) for defining **smallest important changes** in the new post-publication peer-reviewed journal [Sport Performance & Science Reports](#). In response to his article, I have now included instructions in the spreadsheet and [below](#) on monitoring for changes that are "unusual" (very unlikely) for an individual, when you don't have a well-defined smallest important change. I have now also provided a panel of cells to estimate the chances of seeing the various outcomes (*unclear, possible increase, very likely increase, possible decrease, and very likely decrease*) for a given true change. (See [these slides](#) for a partial explanation of the derivation of the chances—for experts only.) These cells offer a more direct way than the simulation spreadsheet to see how often you will make the wrong conclusion about changes. See the comments associated with these cells and with the cell for setting the meaning of *very unlikely*.

Introduction

This article and the accompanying spreadsheet represent a long overdue resource for practitioners in the disciplines of exercise and sport science who want to monitor the progress of their individual athletes, patients, clients or teams in a reasonably rigorous quantitative fashion. In a previous [article/slideshow](#) and [spreadsheet](#) I focused on quantitative assessment of change since the previous test or measurement, but there has always been a need for quantitative assessment of trend over multiple tests to determine the extent to which the individual is on track for improvement or recovery.

Equally, there is a need for assessment of acute deviations from the trend to determine the likelihood of benefit or harm from short-term changes in training or other strategies. This new spreadsheet meets all these needs.

The [previous spreadsheet](#) included a panel for inserting a single measurement to estimate the chances that the individual's true value falls either side of a reference value or threshold, such as a critical blood concentration of vitamin D. The present spreadsheet cannot be used for such calculations, but I will produce a similar upgraded resource, if there is enough demand for it.

The new spreadsheet can be used to track an individual team's scores, but if the data for the team come from the sum of the scores of individual athletes, you should consider analyzing the athletes' data as a time series with either the [post-only crossover spreadsheet](#) or [mixed modeling](#). You can then perform the usual sample-based inferential analyses, which are potentially publishable, if you have got some interesting results with the team.

The new spreadsheet comes as a workbook with a second spreadsheet for simulating realistic test scores and thereby investigating error rates with known true values for starting score, trend, changes on chosen time points, and typical error of measurement. Instructions on their use are included at the top of both spreadsheets and are supplemented with comments in some cells.

Both spreadsheets require input of values for the smallest important change and for a target trend. Input of a value for the typical (standard) error of measurement is also desirable. There are numerous articles at this site and elsewhere about the meaning and estimation of these values, which at the request of the reviewer I have summarized in appendices. What follows is a brief description of how the inferences are derived for the three kinds of monitoring: change between measurements, deviation of one measurement or the average of several measurements from the trend, and magnitude of the trend. The operation of the spreadsheet for simulating data is also described.

Change between Measurements

The probability that the true change is a substantial increase is calculated by combining the observed change, the smallest important change, and the short-term typical error (with

its degrees of freedom) into the relevant value of the t statistic. The probability of a substantial decrease is calculated in a similar manner. The probability of a trivial change is the difference between 100% and the sum of the percent probabilities of a substantial increase and decrease.

To the extent that measurements on an individual represent a sample of the individual's true values, it seems reasonable to apply the rules of [magnitude-based inference](#) to make conclusions about the individual's true changes. I have chosen the non-clinical version of magnitude-based inference, according to which a change is unclear if the true change *could* be both a substantial increase and a substantial decrease. I have set the default value for *could* to 10%, rather than the usual 5%, which in simulations (described below) does not produce enough clear outcomes for real changes with noisy measures (typical error \geq smallest important change). Unclear changes are indicated with a "?". Clear changes are indicated with \leftrightarrow , \uparrow , and/or \downarrow for trivial changes, substantial increases and/or substantial decreases respectively, and very likely changes (those with chances $>90\%$) are indicated with an asterisk (*).

Inserting a value for the short-term typical error is optional. If you do not have a value, the spreadsheet uses the typical error of the estimate from the trend analysis, which will generally overestimate the error of measurement (if the individual made sporadic real changes off the trend) and thereby make it harder to get clear outcomes. Also, you need enough test scores (at least 10) on the individual to get even modest precision for the typical error of the estimate.

Deviation from the Trend

A scatterplot of the data allows for easy identification of the scores to include for calculating the trend line. The plot shows the regression line flanked by dashed lines defining a zone of trivial changes plus uncertainty in predicted scores. When a score falls outside the zone defined in this manner, the true difference from the predicted score is approximately likely ($>75\%$ chance) to be substantial.

A trend line is fitted to chosen assessments using the Excel functions for simple linear regression, then the difference between the score predicted by the linear trend and the other test scores is analyzed in the same manner as for changes between measurements. (Formulae for

the predicted scores and their uncertainty were copied from the [spreadsheet for validity](#).)

When measures are noisy, averaging repeated measurements is the best strategy to reducing the noise. I have therefore included an option for averaging a cluster of values to compare with the values predicted by the trend line.

Magnitude of the Trend

If you insert a value for a change that the individual wants or needs to show over an extended period (which you also have to insert; e.g., 10 weeks, a season, or a year), the spreadsheet divides the change by the period to give a target trend. This trend is then treated as a smallest important trend, and the spreadsheet combines this value with the slope of the regression line and its uncertainty to give probabilities that the true trend is greater than the positive trend, less than the negative trend, and trivial (falls between \pm this trend). Magnitude-based inferences for the trend are provided in the same manner as for changes. (The formula for the uncertainty in the slope was copied from the [spreadsheet for validity](#).)

Simulating Test Scores

You will be a better practitioner if you have some idea of the errors you can make when assessing changes and trend in an individual. I have therefore gone to considerable trouble to make a version of the spreadsheet that simulates and analyses test scores similar to yours, so you can see directly how often errors will occur. To generate the score at a given time point, the spreadsheet combines a true starting score with a true trend multiplied by the elapsed time, adds any extra true change at one or more time points, then adds a value of the true typical error multiplied by $\text{NORMSINV}(\text{RAND}())$, which simulates the effect of normally distributed random measurement error. In this way the spreadsheet generates a sample of the kind of scores you will get with tests on your individual, if you have inserted realistic values of true typical error, true trend, true extra changes, and of course a true starting score. The spreadsheet then performs inferences on changes and trends in the same manner as for the other spreadsheet, using values for the smallest important change and target trend.

In the absence of an estimate of the typical error from a reliability study, a value can come from an initial analysis of the individual with the first spreadsheet. Find the cell showing the

typical error of the estimate from the fitting of the trend line (Cell E81), and use that as the typical error in the simulation sheet. You can also use a typical error obtained by combining the typical errors of the estimate from several individuals, by adding their squares and taking the square root.

To use the simulation spreadsheet, start with a zero true trend and zero extra changes, so you can see how often you get Type-1 (false-positive) errors: that is, how often the inference shows a non-existent substantial trend and non-existent substantial changes. You should play with the number of time points included in the trend analysis, to get some idea of how many you will need with your real data. Then insert a realistic true trend to see how often you get false-negative (Type-2) errors: how often the inference fails to show a substantial trend. Similarly insert substantial changes at individual time points to see how often the inferences fail to show changes from the linear trend and changes from the previous test.

The Type-1 error rate for changes from the previous test depends on the relative magnitudes of typical error and smallest important change. The Type-1 error rate for change from the trend depends additionally on the number of time points included in the trend, which along with the smallest important trend also contributes to the Type-1 error rate for the trend itself. All these factors contribute to the Type-2 error rates, along with the magnitude of the true changes and trend. For noisy measures you will have to learn to live with high error rates.

It is inevitable that the two new spreadsheets have bugs, hopefully only minor formatting errors or ambiguous instructions. Please get back to me by [email](#) if you encounter problems or especially if you have any suggestions for improvement.

References

- Buchheit M (2018). Magnitudes matter more than beetroot Juice. Sport Performance & Science Reports 1, <https://sportperfsci.com/magnitudes-matter-more-than-beetroot-juice/>
- Hopkins WG (2010). Linear models and effect magnitudes for research, clinical and practical applications. Sports Science 14, 49-57
- Malcata RM, Hopkins WG (2014). Variability of competitive performance of elite athletes: a systematic review. Sports Medicine 44, 1763-1774

Published June 2017

©2017

Appendix 1: Smallest Important Changes

For a description of smallest important effects for all types of outcomes, see [Linear Models and Effect Magnitudes](#) (Hopkins, 2010). See also [a slideshow](#) focusing on medal-winning enhancements presented at the [performance analysis conference](#) in 2016. For the following summary I have also drawn on Martin Buchheit's [recent article](#) on monitoring athletes in the [Aspetar Journal](#) (now published [here](#)).

Performance tests for athletes competing for a best competition score

For an athlete who is already winning medals regularly, the smallest improvement in performance results in an extra medal every 10 competitions. Assuming little interaction between the athletes in the competition, the change in performance score (time, distance, weight lifted, or judges' ratings) producing the extra medal is 0.3 of the variability that top athletes show from competition to competition. Expressed as a coefficient of variation, the variability ranges from ~0.5% to ~3.0% for top athletes, depending on the sport. You have to take into account a kinetic factor to convert 0.3 of this variability in competition time or distance into the smallest important change in test performance, if the test measure is power output. The test protocol also has to be taken into account, if the test measure is time in an incremental or constant-power test to exhaustion and/or there is a fatiguing pre-load. See Malcata and Hopkins (2014) for a comprehensive review.

Sub-elite athletes generally show more variability in competition than elites do (perhaps by a factor of 1.1-1.4), so the smallest important change for winning medals in local or age-group competitions is correspondingly larger. But a target trend may be more important than winning medals for these athletes. See Appendix 2.

Tests and performance indicators with uncertain relationships to performance

When you do not know exactly how much change in a test is associated with winning an extra match or medal every 10 competitions, it may still be worth monitoring with the test. Measures of psychological state (fatigue, pain, motivation, and so on) are in this category. If these are assessed with a Likert scale or visual-analog scale, the value can be rescaled to range

from a minimum possible of 0 to a maximum possible of 100; a change of 10 units then corresponds to half a step on a 6-point scale or one step on an 11-point scale, which is something like the smallest perceptible change and is a good default for the smallest important.

For other measures, including fitness tests and match performance indicators, standardization is the only way to define the smallest important. You will need an estimate of the between-subject standard deviation for athletes on a similar level to your athlete, and in the same playing position for team-sport athletes. (You may pool playing positions to get enough players for a reasonably accurate estimate, if you are satisfied that mean test scores are unlikely to differ substantially between the playing positions; if the means differ, you can pool playing-position SDs, if these appear to be similar.) One-fifth (0.2) of this standard deviation is then the smallest important change.

Measures unrelated to performance

A test unrelated to performance might still be worth administering if it is related to health. In particular, the smallest important change in a test related to risk of injury or illness can be defined as the change or difference associated with the smallest important reduction in risk in a relevant epidemiological study: a risk (hazard) ratio of 0.90.

In the absence of information on risk, standardization is the only approach. Strictly speaking, the between-subject standard deviation for a health indicator should be "pure" or free of short-term typical error of measurement (in contrast to that for measures of performance, which for team-sport athletes should be the observed between-subject standard deviation). The pure standard deviation is $\sqrt{(\text{observed SD}^2 - \text{typical error}^2)}$ and requires a short-term reliability study on subjects similar to your individual for estimation of the SD and error.

Measures where standardization is not appropriate

Unfortunately, some physiological measures have real differences between individuals that are only a fraction of the typical error (within-subject variation); that is, any observed differences are due almost entirely to day-to-day variation. The variation itself may differ substantially between individuals. Standardization fails for such measures. Martin Buchheit (2018) has suggested using a fraction or multiple of the

typical error as the smallest important change for such measures, in the sense that any observed change greater than this amount can be regarded as a change not due simply to error of measurement. This approach amounts to identifying when a change is "unusual" for the individual. With the current spreadsheet, you can identify such changes by specifying what you mean by *unusual* with the cell that defines the meaning of *very unlikely*. The default of 10% is probably about right for most situations; it corresponds to accepting a single change score as a real change when the change is greater than ~ 2.0 typical errors ($t \times \sqrt{2} \times e$, where t is the value of the t statistic for a cumulative probability of 0.9), as shown in one of the cells. You then set the smallest important change to zero. All unusual changes are now marked with either \uparrow^* or \downarrow^* , and all other changes are unclear (marked with $?$). Keep in mind that, if the real change is zero, you will see unusual increases 10% of the time and unusual decreases 10% of the time.

Finally, the above advice is based on the assumption that the test measure could have a useful relationship to performance or health. If you have a strong suspicion that the measure has no such relationship, stop measuring it and put your resources into measuring something that is more likely to have a relationship.

Appendix 2: A Target Trend

For an individual who wants or needs to improve over a period longer than simply the time to the next test, you can decide on a realistic change over a chosen period. The rate of change (the change divided the period) is then the target trend. For a development athlete aiming (eventually) for Olympic selection, work out the change needed over the next season or year to keep the athlete on target. For sub-elite team-sport athletes, the difference between the athlete's test or performance indicator and that of typical athletes in an elite team is a reasonable target; you then have to decide on a reasonable period over which you hope to gain that improvement. Similar reasoning would apply to measures of muscle mass, body fat, or health indices such as blood lipids or vitamin D.

The target trend evaluated over the period between consecutive tests is another approach to estimating the smallest important change. For example, if the target improvement for the coming year is 4%, then with three months between tests, the smallest important change of 1% will

keep the athlete on track for the target.

Appendix 3: Typical Error

The short-term typical or standard error of measurement represents the standard deviation of an individual's repeated measurements in the absence of any real change in the individual. The error can come from a short-term reliability study of individuals similar to your individual. Alternatively, if you have at least 10 or preferably more consecutive tests on the individual, and there is no reason why there should be substantial changes between the tests other than a gradual trend, the standard error of the estimate derived from fitting the trend line is a good estimate of the typical error. Hence, if you do not include a value for the typical error in the trend spreadsheet, the spreadsheet will use the standard error of the estimate.

Most tests with athletes are noisy (i.e., their errors of measurement are substantial—greater than half the smallest important), so some trivial changes will be unclear. With enough noise, all trivial changes and some small changes will be unclear. Hence, if you have a choice of tests, use the one with the smallest error. If possible, you can also reduce the error by getting the athlete to do the test several times on each testing occasion, then averaging the values and inserting the average values in the spreadsheet. Make sure you insert the appropriately smaller value for the typical error: if the test is performed n times on each occasion, the error of the mean is the usual typical error divided by \sqrt{n} . If you let the trend spreadsheet estimate the typical error, it will automatically provide an estimate of this smaller TE.

When starting tests with a new individual, you may use the standard error of the estimate from another individual. You may also average standard errors of the estimate from several individuals, but do it by averaging their squares and taking the square root. To be more accurate, weight the squares by their degrees of freedom (number of points in their trend line minus 2). The degrees of freedom of the mean is the sum the degrees of freedom.

If the performance test and score are similar to those of the competition (e.g., a time-trial time for a track athlete, a throwing or jumping distance for a field athlete, an ergometer time-trial time for a cyclist, rower or kayaker), and if you are satisfied that the athlete performs the test as reliably as a competition, then the typical error of the test is the published variability of top athletes in that sport, as stated in the Malcata and Hopkins (2014) review and in my recent [slideshow](#). Note that in this case the smallest important change is automatically $0.3\times$ the typical error of the test. You should consider using the smallest important change as calculated in this manner, but make the typical error a bit larger than the variability in competitions, by a factor of perhaps $\sim 1.3\times$, to allow for less reliability in tests than in competitions. Or you can use the value from a reliability study, or if that's not available, let the spreadsheet use the typical error of the estimate from the trend analysis.

If the test scores are counts and proportions of counts (e.g., most performance indicators from the analysis of players' actions in games), the typical error depends on the individual's mean count or proportion, but the estimate of typical error from a reliability study will be an average across all the subjects in the study. It is possible to estimate approximate typical errors for the counts and proportions of the individual by assuming they have Poisson and binomial sampling distributions respectively, but then there is a potential problem of non-randomness of the counts, which usually produces a larger typical error ("overdispersion"). I suggest using the estimate of typical error based on the sampling distribution^a until you have enough data to let the spreadsheet use the typical error of the estimate from the trend analysis.

^aCount: \sqrt{n} , where n is the number of events.

Proportion (%): $(100/m)\sqrt{[n(m-n)/m]}$, where n and m are the numbers of events and trials.