Regression Accuracy and Power Simulations

Version 4.0

(for Windows 2000 and later)

January 2013

W.A. Sadler
71B Middleton Road
Christchurch 8041
New Zealand

Ph: +64 3 343 3808
e-mail: bill.sadler@xtra.co.nz

(formerly at Nuclear Medicine Department, Christchurch Hospital)
Regression Accuracy and Power Simulations (RAPS.exe, version 4.0)

This program assesses the accuracy and statistical power of parametric linear regression using sets of simulated X, Y pairs that are randomly generated by the program itself (although users do have the option of importing their own simulated data). The program incorporates ordinary (unweighted) least squares regression (OLS), iteratively reweighted least squares regression (WLS) and iteratively reweighted Deming (I) regression (errors in X and Y; also known as errors-in-variables regression). Significance testing is based on the usual confidence intervals for slope and intercept, but also uses the elliptically shaped joint slope, intercept confidence region. A major program aim is to contrast these two approaches; in particular the improved statistical power that can be realised in most situations by using the joint slope, intercept confidence region rather than confidence intervals.

This document gives a broad description of the program. The data simulation and numerical methods are fully described in Ref. 2. The fine detail of using the program is given in program on-line help. A brief popup description of all controls on all dialogs can be obtained by pressing F1 when the control has focus, or by clicking the right mouse button over the control (provided it is enabled), or by using the ? icon at the top right of the dialog. All dialogs have a Help button which provides detailed information about the dialog.

Data

It must be emphasized that meaningful use of RAPS.exe requires computer simulated data. The program is designed to produce such data. The aim is to investigate accuracy and statistical power by randomly drawing many sets of X, Y pairs, each of a specified size (N), which have a precisely known underlying relationship (eg. Y = X) and precisely known error characteristics. In the program Accuracy dialog we are attempting to verify that the mean of the sampling slopes and the mean of the sampling intercepts equate to the values that were used to generate the data. The program performs a significance test on the difference between true and recovered values. Equally importantly, we are also seeking to verify that the expected proportion of slope and intercept confidence intervals and joint slope, intercept confidence regions enclose the true underlying slope and intercept values. A set of real method comparison X, Y pairs (even with N as large as several hundred) is unsuitable for this type of sampling experiment because the true underlying slope and intercept are unknown. The data could of course provide an estimate of slope and intercept, but such values are not necessarily the true underlying values.

In practice, investigators will invariably have a reasonably precise knowledge of the X and Y ranges of values that will occur in a method comparison. They can also control how specimen results are distributed over those ranges. Common patterns are roughly uniform, concentrated in the central part of the range and concentrated at the lower end of the range (right skewed). Knowledge of the error (precision) characteristics of the X and Y results should be readily available (at least in a medical laboratory context) from analysis of internal QC data or precision data from a method evaluation (Ref. 3 provides one way of summarising replicated precision data). Use the program Accuracy dialog to generate simulated data that mimics as closely as possible the data that will be collected in a real method comparison, by specifying the likely range, distribution and error characteristics of the data. Assuming that accuracy is verified, the next step is to transfer the data characteristics to the program Power dialog to determine the minimum number (N) of X, Y pairs that are necessary to give a high probability of detecting the maximum deviation from the line of identity that is considered tolerable. The evaluation of statistical power (ie. necessary sample size) should precede the actual experiment.

The option of importing externally simulated data (eg. using Ref. 2 as a guideline) is mainly to provide for data distributions that differ from the uniform, centrally located and right skewed distributions offered by the program. It should be noted, however, that data distribution is a less important factor than range and error characteristics.

Version 3.0 to Version 4.0 Changes

A history of previous program changes and bug fixes is given on page 7. The upgrade from version 3.0 to 4.0 was made using Embarcadero’s Delphi XE3. The program requires Windows 2000 or later, a screen resolution of 1024 x 768 pixels or better and 16-bit colors or higher. The following changes were made in the upgrade to version 4.0:

- A button has been added to automate the transfer of regression type and weighting settings between the Accuracy and Power dialogs (either direction). Since a major program aim is to verify accuracy then assess statistical power this can save a little time and avoid data entry errors.

- In previous program versions externally generated X, Y pairs could only be imported from a text file. External data can now be pasted from the Windows clipboard.

- The program produces plots of variance (weighting) functions in terms of either Variance, SD or CV(%) versus the mean. Previously Y-axes were restricted to a linear scale but a logarithmic scale can now be selected for the Y-axis of Variance and SD plots.
Confidence Intervals and Regions

Figure 1 illustrates a sampling experiment in which OLS was applied to 50 sets of N = 20 randomly drawn X, Y pairs from a hypothetical Y = X relationship. Note the characteristic ‘bow-tie’ appearance (the same pattern is seen with WLS and Deming regression). Highest values of slope are associated with lowest values of intercept and vice versa. High values of slope and intercept are not jointly observed. Likewise, low values of slope and intercept are not jointly observed. Figure 1 illustrates the correlation of slope and intercept. For any given error characteristics, the smaller the range of X-values the higher the slope, intercept correlation (increasingly pronounced bow-tie shape). Conversely, increasing the range of X-values reduces slope, intercept correlation.

Figure 2 illustrates a typical slope and intercept parameter space for a straight line model in the case of a narrow range of X-values (maximum:minimum ratio ≤ 2:1). The data point in the centre of the graph represents the estimated [slope, intercept] point. The elliptical joint confidence region encompasses a region of slope, intercept values that are jointly reasonable for the data. Note that it does not extend into the area of large slope and large intercept values (top right) or low slope and low intercept values (bottom left). The joint confidence region takes account of the correlation of slope and intercept values whereas confidence intervals do not.

Confidence intervals and regions are identical in the sense that in sampling experiments, such as illustrated in Fig. 1, a specified proportion of intervals and regions (depending in the significance level) are expected to enclose the true underlying [slope, intercept] point (eg. the point [1, 0] in the case of Fig. 1). Either could be used as the basis of significance tests. Intuition suggests that the confidence region is preferable because it takes account of parameter correlation, but there is a further consideration. The usual purpose of significance tests is to detect differences from a hypothetical relationship (such as the line of identity, Y = X) and this is determined by the failure of confidence intervals or confidence regions to enclose the hypothetical [slope, intercept] point. The relative shapes in Fig 2 clearly indicate that confidence ellipses will detect smaller horizontal or vertical shifts from a target point (pure proportional bias and pure constant bias, respectively). This implies the ability to detect smaller differences with a given sample size (N) or, alternatively, require smaller sample sizes to detect a specified difference.

The idea of using the joint parameter confidence region is far from new. Munson and Rodbard wrote an excellent paper (4) advocating the use of joint slope, intercept confidence regions in an immunoassay context. Text book authors make the same general recommendation (eg. Refs. 5, 6). However, for one reason or another, confidence intervals appear to have become entrenched as the standard mode of significance testing in medical laboratory regression analysis. This computer program is a tool for acquiring a feel for the properties of confidence regions and can also be used to formally estimate necessary sample size by simulating data whose properties mimic as closely as possible the range and error characteristics of real method comparison data. A separate computer program (3) estimates variance (weighting) functions and offers OLS, WLS and Deming analysis of real data, including plots of joint parameter confidence regions at user defined significance levels.

To give a more complete picture, Fig. 3 illustrates a typical parameter space when the maximum:minimum ratio of X-values is several thousand-fold (very low slope, intercept correlation). In these cases the more circular joint confidence region confers no statistical power advantage. However, Fig. 3 illustrates an extreme case that is only likely to be relevant for a small number of medical laboratory analytes measured by immunoassay.
Accuracy

Accuracy in this context refers to the reliability of recovery of the slope and intercept values that were used to generate sets of paired X, Y values and also the concordance between expected and observed frequencies of confidence intervals and confidence regions enclosing target values. Figure 4 illustrates the program Accuracy dialog. Use the Data box to specify properties of the X, Y data that the program will randomly generate, or to import X, Y pairs generated externally. Use the Regression Type and Weighting box to specify error properties of the Y-variate (OLS or WLS) or both the X and Y-variates (Deming). A detailed explanation of how to specify error properties is given in program on-line help. The Results box includes significance tests on the differences between true and recovered slope and intercept values (see on-line help for more information). The lower panel displays the current date and time. Use the adjacent buttons to produce a permanent copy of the design and outcome of a simulation run by copying an image of the dialog to the Windows clipboard for subsequent pasting into another application, such as Word, or directly printing an image of the dialog.

Figure 4. The program Accuracy dialog illustrating the options available, plus the results from an OLS simulation run using data generated by the program according to the Data box specifications.
Power

The program Power dialog, illustrated in Fig. 5, follows the same principles. Data can be either generated by the program or imported from an external source. The focus here is on the True slope and intercept values which are used by the program to generate sets of data. The resulting confidence intervals and regions are assessed for enclosure of the Target slope and intercept values. The issue is the sample size required to reliably detect a difference between True and Target (hypothetical) values when such a difference actually exists. The program offers automatic sample size determination (as illustrated).

It is important to appreciate that since the program is based on randomly generated data, there will always be some variation between results of simulation runs of the same design. Obviously the larger the number of samples and/or sample size (N), the smaller the run-to-run variation in outcomes, and vice versa. The manual sample size option on the Power dialog can be used repetitively to quickly get a feel for the variation associated with a particular design and sample size.

Figure 5. The program Power dialog illustrating a simulation run which automatically determined the smallest sample sizes (confidence intervals versus joint confidence region) necessary to achieve 90% detection of a statistically significant difference (p < 0.05) when the hypothetical slope is 1 and the true slope is 1.01 (ie. 1% proportional bias). In this OLS example, uniformly distributed (error-free) X-values cover a 10-fold range and Y-values were randomly drawn from a Gaussian distribution with variance = 1.

Imported Data

Externally generated data must be presented in a text file with a single X, Y pair on each line. The X, Y values must be separated by a single character (any character can used except 0 – 9, + (plus), - (minus) or period). The primary intention is allow distributions of X-values that are more extreme than those offered by the program and which may more accurately reflect a real data set. In the Accuracy dialog the option is given to either randomly draw X, Y pairs from the imported data or to draw sequential sets of X, Y pairs of a specified size. The latter could be used to compare results produced by this program with those produced elsewhere with the same data.

Program generated data represent an ideal situation because they are randomly drawn from Gaussian distributions with precise error properties as defined in the Regression Type and Weighting box. Imported data could be used to systematically assess the effect on both accuracy and power, of data drawn from non-Gaussian distributions, or the deliberate misspecification of errors, eg. generate and import data with certain error properties, then test the effect of deliberately specifying alternative error properties in the Regression Type and Weighting box (eg. to mimic the effect of over or underestimated method errors).
Graph Options

The program incorporates some graphical options. Variance functions can be viewed in both the Accuracy and Power dialogs by clicking the View Variance Functions button. The results of Accuracy simulation runs can be viewed by clicking the View Results button. The sampling lines can be viewed as either the lines themselves in the usual X, Y coordinate frame, or as [Slope, Intercept] points in a parameter space plot.

A Deming regression simulation, which was used to produce a set of example results in Ref. 2, is illustrated here to provide examples of program graphs. The Ref. 2 data were based around an automated immunoluminometric assay for the pituitary hormone thyrotropin (“Access” instrument, Beckman Coulter, Fullerton, CA, USA). The data range was set at 0.015 – 40 mIU/L. The Y-variance function was estimated from daily measurements of thyrotropin in five internal Quality Control specimens over a 7-month period. A hypothetical X-variance function was constructed such that predicted variance was smaller than predicted Y-variance by a factor of 4 at the lower end of the range, and larger than predicted Y-variance by a factor of 4 at the upper end of the range, with crossover at the mid-point of the range (20mIU/L). The variance functions are shown in Fig. 6, translated to CV(%) versus Mean. Variance function crossover appears shifted to the right because of the optional logarithmic scale. Although the main intention of RAPS.exe is to mimic real method comparisons, by specifying the range, distribution and error properties of real data, this particular setup was aimed at testing estimation accuracy under a “worst-case” example of non-parallel error properties. Despite the rather benign appearance of the CV profiles, the predicted relative increase in variance between 0.015 and 40 mIU/L is 650000-fold for the Y-variance function and 10.3 million-fold for the X-variance function (typical of some classes of immunoassays).

Figure 6. Illustration of the Y and X-variance functions used to generate data for the Deming regression simulation results in Ref. 2.
Figure 7. Parameter space for 10000 Deming regression sampling lines, each estimated from \( N = 100 \) X, Y pairs, with true underlying relationship \( Y = X \), and a narrow data range 14 – 28 mIU/L (maximum:minimum ratio 2:1). X and Y errors are illustrated in Fig. 6.

Figure 8. Parameter space for 10000 Deming regression sampling lines, each estimated from \( N = 100 \) X, Y pairs, with true underlying relationship \( Y = X \), and a wide data range 0.015 – 40 mIU/L (maximum:minimum ratio 2667:1). X and Y errors are illustrated in Fig. 6.
History

- Version 1.0 (December 2009; Ref. 2) was used to produce the simulation results in Ref. 2. Written with Borland Delphi version 5 and available to registered users at the website of the journal Accreditation and Quality Assurance.

- Version 2.0 (October 2010) was written with Embarcadero’s Delphi 2010. Visual themes were introduced, on-line help was upgraded from WinHelp to HTML help and graphical displays of X, Y variance functions and Accuracy sampling results were added (see Figs. 6 – 8).

- Version 3.0 (May 2011) was written with Embarcadero’s Delphi XE. A formal significance test on the accuracy of recovery of underlying slope and intercept values was added (see Fig. 4) and three version 2.0 bugs were corrected;
  1. Certain rare combinations of variance function parameter values produced inaccurate warning or error messages after clicking the View Variance Functions button; eg. the message might assert that the function equates to CV > 40% over part of the range when it actually equates to CV > 40% over the entire range (or vice versa). The functions always plotted correctly in these rare cases, but their appearance will not have matched the message.
  2. The specific sequence of starting the program, switching directly to the Power dialog, selecting an alternative parameterization value (Para spin edit) or variance function, then starting a simulation run without modifying any edit box values could result in incorrect power results; in particular, grossly low sample sizes associated with confidence intervals. These specific circumstances are hardly likely to occur in any meaningful simulation run and, moreover, the grossly low confidence interval sample sizes would almost certainly have been recognised as a program bug.
  3. Changing the parameterization value (Para spin edit) or switching variance functions in the program Power dialog, without changing any variance function parameter values, resulted in the program failing to update the Y-axis values in plots of the variance functions, ie. the revised functions were plotted with the Y-axis coordinates calculated for the previous variance functions (assuming that a plot of the previous variance functions had been invoked). This bug, in common with the previous one, was caused by a failure to reset some internal variables in the Power dialog when a variance function selection change occurred but no other change was made.

References


Program Delivery, Installation and Removal

The Australasian Association of Clinical Biochemists has kindly allowed the program installation file to be downloaded from its website (www.aacb.asn.au). Proceed to Resources > Useful Tools which contains a link to the Regression Accuracy and Power Simulations program. Click the installation file link and download file RAPS40Install.exe (3916 Kb) to a directory on your hard drive. Before proceeding further it is strongly recommended that you uninstall any previous version of the program. Double click RAPS40Install.exe to launch the installation process. C:\RAPS is suggested as the program home directory but you are free to specify an alternative location. RAPS.exe (2300 Kb) and on-line help file RAPSHelp.chm (53 Kb) are installed into the specified home directory and a shortcut icon is placed on the desktop.

Use the Add/Remove Programs icon (or equivalent) in the system Control Panel to remove the program icon and the installed files.