CellPress

Quantitative metabolomics: a phantom?

Stephan Noack and Wolfgang Wiechert

Institute of Bio- and Geosciences, IBG-1: Biotechnology, Forschungszentrum Jülich, D-52425 Jülich, Germany

'Mass specs are precise but biology is not!' is a frequently heard argument when quantitative experimental data do not fit into the overall picture. The problem with this opinion is that the significance of measured biological data becomes a matter of gut feeling. Doubtlessly, the measurement precision of modern mass spectrometers is far better than the reproducibility of biological experiments. However, precisely for this reason, technical reproduction of mass spectrometric measurements neither characterizes the whole experiment from cell cultivation to producing biological data nor says anything about systematic errors in the overall measurement procedure. Taking quantitative metabolomics as a fruitful example, we deal with the question of why it is so difficult to say something precise about imprecision in biology.

Quantifying biology

As a result of its historical development, biology is not really a quantitative discipline. 'Yes or no' questions can still be answered on the basis of rather crude or comparative data. But if the mission of systems biology (that is, obtaining a quantitative understanding of cellular behavior through integration of experimentation and mathematical modelling) is to be taken seriously, there is no way around quantitation. Particularly, mechanistic mathematical models are highly sensitive to systematic errors in the underlying data, and predictions can easily become worthless [1]. If economic considerations play a part, as in the case of applied biotechnology, the challenge is even greater. We must be able to compare quantitatively the performance of genetically modified organisms or cultivation regimes in which differences can be in the range of a few percent [2]. Thus, quantitative biology becomes a cornerstone of systems and synthetic biology [3], as well as of all technically oriented biotechnological disciplines.

Interestingly, when discussing biological experiments with physicists, their first question always concerns data quality. In physics, the quantitative analysis of measurement protocols and instruments is absolutely fundamental and highly developed [4]. By contrast, the general theory of experimental error analysis (see Glossary) is often not part of basic biological training. Can we really establish quantitative biology without such a sound methodological platform? And what precisely distinguishes biology from physics?

The accuracy and precision of multi-step bioanalytical measurement procedures are always the sum of many different elementary errors with their systematic and stochastic influences. Particularly for the complex measurement procedures of modern metabolomics, we need a scientific basis for assessing the degree of confidence we have in quantitative biological data.

In the following sections, we analyze the known sources of errors in common metabolomics measurement protocols by addressing the problems of precision, accuracy and error propagation (Box 1). It will become clear that it is not

Corresponding author: Wiechert, W. (w.wiechert@fz-juelich.de).
Keywords: metabolomics; mass spectrometry; intracellular metabolites; quantification.

0167-7799/
 http://dx.doi.org/10.1016/j.tibtech.2014.03.006

Glossary

Absolute quantification: aims at the absolute comparability of two measurements taken at different times and/or places.
Bioanalysis: deals with the development of standardized measurement protocols, which are sequences of cell cultivation, sampling, cell disruption, extraction, chromatography, detection, or spectral evaluation steps. In general, these protocols are lengthy and complicated. Consequently, the measurement is subject to many disturbances and the elimination of systematic errors is of paramount importance for bioanalysis and constitutes an ongoing process.
Error analysis: a methodology that deals with the whole complexity of a measurement protocol by separately characterizing every single step of it (cf. Box 1). By contrast, for complex protocols, the experimental multiple execution of a complete experiment including measurement is only a theoretical option.
Gold standard: a measurement protocol that may be extremely complicated and expensive but reproduces the true physical quantity very well. Unfortunately, the continuing improvement of instrumentation and procedures in metabolomics is a long-term endeavor that is still far from converging to a quasi-standard. Moreover, there are some conceptual issues about 'true' quantities (see 'measured quantity').
Measured quantity: a physical entity that can be quantitatively described by a number and exists independently of the measurement protocol used. Defining the measured quantity in biology is part of the measurement problem. Taking, for example, the ATP concentration in a cell, it quickly becomes clear that this is not really what is measured. Apart from the notorious cellular compartmentation problem, the chemical name disguises the fact that ATP simultaneously occurs in different forms as ATP^{4-}, $HATP^{3-}$, H_2ATP^{2-}, $MgHATP^{-}$, and Mg_2ATP. They all have different biological activity but can hardly be separated analytically.
Relative quantification: aims at the ratio between a reference sample and a measured sample. It can be shown (Box 1) that relative quantification is less problematic than the ultimate goal of absolute quantification.
Reproducibility: means that a result is exactly reproduced in repeated experiments, but it may, nevertheless, be completely wrong. In chemical analysis, data reproduction is described by the term precision, whereas a statistician would use the expression (standard) deviation. In contrast to accuracy (needing a gold standard), precision can be experimentally assessed. As a rule of thumb, lengthy protocols tend to be more accurate but imprecise, whereas short protocols are rather inaccurate but precise.
Systematic error: has different synonyms in different disciplines. In chemical analysis it is called accuracy, a statistician would say bias. Keeping the discrepancy between the average measured quantity and the corresponding 'true' biological quantity as small as possible is the great challenge for bioanalysis. Accuracy should not be confused with reproducibility of a measurement procedure.

Box 1. Error propagation

Error propagation analysis is based on the laws of probability. It serves to calculate the overall standard deviation of a complex measurement protocol by accumulating all single stochastic influences along its processing steps. On the one hand, it is the method of choice for assessing the precision of a protocol, but on the other hand, it cannot say anything about its accuracy. Without going into the details of the methodology, a simple example illustrates how it works.

In the following, all errors are written as relative errors (percentages) ε, δ, etc. For example, a $\pm 10\%$ error is indicated by $\varepsilon = \pm 0.1$. If a true quantity, x, is measured with some error, ε, the resulting measurement is $y = (1 + \varepsilon) \cdot x$. Note that ε is a stochastic term and can have both signs. In the following calculations, higher-order products such as $\varepsilon \cdot \delta$, ε^2, $\varepsilon \cdot \delta \cdot \gamma$, etc. are very small and, thus, can be neglected. This simple approximation is the centerpiece of classical error propagation methods based on the calculation of partial derivatives.

As a highly simplified example, consider a measurement protocol covering three steps: (i) biological variability, cultivation, and sampling; (ii) sample processing; and (iii) chromatography and MS measurement (Figure I, top). This corresponds to three relative errors ε, δ, γ. The relation between the true quantity, x, and the measured quantity, y, is then described by (neglect higher-order terms):

$$y = (1 + \varepsilon) \cdot (1 + \delta) \cdot (1 + \gamma) \cdot x \approx (1 + \varepsilon + \delta + \gamma) \cdot x$$

It turns out that relative errors add up in a simply linear error propagation pipeline.

Consider now the case that the ratio of two independently processed measurements (indexed with 1 and 2) is wanted. By expanding the fraction it holds:

$$\begin{aligned}\frac{y_1}{y_2} &= \frac{(1+\varepsilon_1)(1+\delta_1)(1+\gamma_1)}{(1+\varepsilon_2)(1+\delta_2)(1+\gamma_2)} \cdot \frac{x_1}{x_2} \\ &= \frac{(1+\varepsilon_1)(1-\varepsilon_2)(1+\delta_1)(1-\delta_2)(1+\gamma_1)(1-\gamma_2)}{(1+\varepsilon_2)(1-\varepsilon_2)(1+\delta_2)(1-\delta_2)(1+\gamma_2)(1-\gamma_2)} \cdot \frac{x_1}{x_2} \\ &\approx (1 + \varepsilon_1 - \varepsilon_2 + \delta_1 - \delta_2 + \gamma_1 - \gamma_2) \cdot \frac{x_1}{x_2}\end{aligned}$$

Keep in mind that all terms ε_i, δ_i, γ_i can have both signs. Again, it turns out that all relative errors contribute additively to the total error, although some of the errors might cancel out. Consequently, the average error of a quotient of two absolute measurements is even higher than the relative error of each single measurement.

This changes when both samples are partly processed together. This is typical for many comparative studies with all kinds of 'omics' methods. In case of metabolomics, imagine two differently isotope-labeled samples. If the middle processing steps are identical (Figure I, bottom), it holds that $\delta_1 = \delta_2 = \delta$ and, thus:

$$\begin{aligned}\frac{y_1}{y_2} &= \frac{(1+\varepsilon_1)(1+\delta)(1+\gamma_1)}{(1+\varepsilon_2)(1+\delta)(1+\gamma_2)} \cdot \frac{x_1}{x_2} \\ &\approx (1 + \varepsilon_1 - \varepsilon_2 + \gamma_1 - \gamma_2) \cdot \frac{x_1}{x_2}\end{aligned}$$

Obviously, one error cancels out. If this processing error δ is the largest one in the pipeline, this will lead to a significant error reduction. However, what is frequently overlooked is that the biological stochasticity and cultivation errors ε_1 and ε_2 are still present and, thus, relative procedures do not necessarily perform much better than absolute ones.

Although this example is oversimplified, it demonstrates the basic principles of error propagation: a mathematical model is formulated explaining the influences of different error sources. Unlike the simple example, the model should contain a physical description of all protocol steps, including the necessary calibration steps. An approximate first-order calculation or, alternatively, a fully nonlinear Monte Carlo calculation (not shown here) then leads to the determination of the quantitative contribution of each individual error to the overall result. This generates a much greater understanding than just an experimental repetition of the whole multi-step procedure.

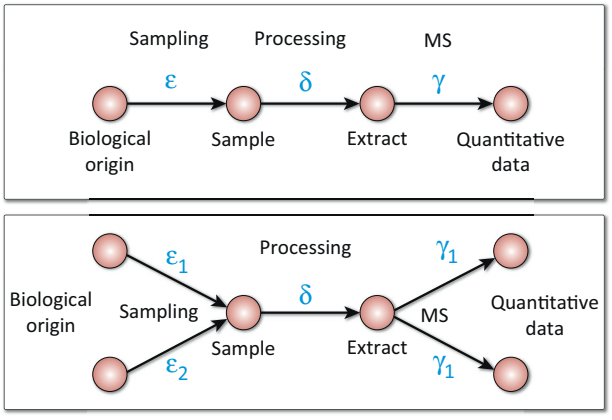

Figure I. Strongly simplified metabolomics processing pipelines for absolute (top) and relative (bottom) measurements.

sufficient to simply repeat experiments (biological replicates) or measurements (technical replicates) to obtain a proper judgment of data quality and error sources.

Biological variability: still a riddle

Metabolomics seems to be a simple task: we have to measure the total molar amount of metabolites found in a sample volume, then we have to relate it to the total biomass volume in the sample in order to obtain an intracellular concentration. Unfortunately, it is not that easy, and the very first question that arises is: 'What precisely is the biomass?' This question is as old as Monod's breakthrough papers on bacterial growth [5] and has still not been satisfactorily answered.

The major problems are to separate viable from nonviable cells and, subsequently, to measure the fraction of viable cells without bias. But which characteristics must a cell have in order to be considered 'viable'? Established criteria for narrowing the term 'cell viability' are manifold and can be based on, for example, cell growth, metabolic activity, membrane potential, or cell membrane integrity. Following the concept of Diaz *et al.* [6], a typical cell population can be roughly divided into viable cells, viable but nonculturable (VBNC) cells, damaged cells, and dead cells. By definition, both viable and VBNC cells possess intact membranes. Cell membrane integrity is essential because only intact cells are able to build gradients and potentials to enable energy production and cell proliferation. Clearly, only viable cells contribute to the metabolome and therefore represent the relevant biomass for intracellular metabolite quantification. Unfortunately, the simple equation 'all intact cells are viable' need not hold.

As a useful concept to relate cell viability to metabolomics, Harris *et al.* [7] introduced the term 'biovolume' to describe the liquid volume that is enclosed by an intact biological membrane. In general, the biovolume should be the preferred reference volume for intracellular metabolite quantification (Figure 1). Methods for direct biovolume measurement are already well established: for example, Coulter counter devices [8] or dielectric probes [7].

What makes the problem more complex is the fact that even the viable cells might divide into subpopulations