doi: 10.1255/nirn.1349

The Replication Myth 1

Column Editors: Kim H. Esbensen,[a] Paul Geladi[b] and Anders Larsen[c]
[a]**Geological Survey of Denmark and Greenland (GEUS), Denmark**
[b]**Swedish University of Agricultural Sciences, Umeå, Sweden**
[c]**Q-Interline, Denmark**

This column, the first in a series of two, is an overview of an issue that presents a muddled appearance and which has not received proper attention for decades—the issue of "replication". What *exactly* is it that is *replicated*? We here describe what turns out to be a complex issue—and some of the inherent confusion involved. In a later column (number 2) we will suggest one or two ways out of this difficulty. We also invite readers to present their views on this issue (for a possible column number 3).

Setting the scene

Here are three examples of what is often stated in response to the fundamental question: "What exactly is replicated?"

i) replicate sampling, replicate samples
ii) replicate measurements
iii) replicate analysis, replicate analytical results

Upon even a little reflection it is clear that these three issues are not identical.

The obvious, but often only implied, understanding for all three is that a beneficial *averaging* is carried out by the process of "replication". It is also implied that important differential insight can be gained by "replication", i.e. by replicating the specific process behind replicated samples, measurements and results, respectively, some sort of measure of variability is obtained. This is undoubtedly true, but a measure of what? There are many vague pre-requisites and imprecise assumptions involved and these need careful analysis. Just for starters, i) addresses the **pre**-laboratory realm while ii) and iii) play out their role **in** the analytical laboratory—but even in this case, is replicate analysis the same as replicate measurements?

Background

From the discipline of experimental design (design of experiments, DOE) comes a well-organised, strict concept and terminology regarding "replicate measurement"; this is because of the controlled situation surrounding an experimental design. For example, in the situation of chemical synthesis influenced by several experimental factors—temperature, pressure, concentration of co-factors for example—it is easy to understand what a replicate measurement means. The operator must repeat the run(s) under **identical** conditions for all controllable factors, taking care to randomise all other potential factors, in which case the variance of the experimental outcome (replicated analytical results), be it small or large, is supposed to furnish a measure of the "total analytical uncertainty", the kind which in the analytical realm is known as **repeatability**. However, in routine operations in the analytical laboratory, variability also encompasses effects from other uncertainty contributions stemming, for example, from small-scale sampling of reactants involved which may not necessarily represent "homogeneous stocks". Added uncertainty contributions may also occur from resetting the experimental setup—with what precision can one "reset", e.g. temperature, pressure, concentration levels of co-factor chemical species after having turned the setup off and cleaned all the experimental equipment? Still, such uncertainty contributions are in the main usually considered negligible or, if not, at least controllable. Often all of the above turn out to be of small effect because of the regular situation surrounding DOE.

Taking one step back, it will be equally relevant to repeat the experiment using another technician, researcher and/or in another laboratory—here we enter the well-known analytical concept of **reproducibility**. There may be smaller or larger effects in this realm and careful, empirical total effect estimations must always be carried out in order to obtain a valid estimate of *TAE* (Total Analytical Error).

But we are here addressing more extensive issues, not always on the agenda regarding replication, in fact quite often left out or forgotten…

There are many scenarios that differ from a nicely bracketed DOE situation. Indeed, most data sets do not originate exclusively from within the well-controlled environment of an analytical laboratory. What shall be described below constitutes the opposite end of a full spectrum of possibilities in which the researcher/data analyst must also recognise *significant* sampling, handling and other errors in addition to the *TAE per se*. The Total Sampling Error (*TSE*) will include all sampling and mass-reduction error effects. It is self-evident that these errors must also be included; *TAE* alone will not give a relevant, valid estimate of the effective total effects associated with all analytical results. We are thus forced to be able to furnish a valid estimate of the total sampling + handling + analysis uncertainty estimate (termed the Global Estimation Error) ($GEE = TSE + TAE$).

The description below is supposed to deal fairly comprehensively with the different manifestations surrounding the replication issue, such that most realistic scenarios are covered.

At the heart-of-the-matter is one key question: what is meant by "replicate samples"? This issue will appear more complex than may seem the case at first sight and shall receive careful attention with respect to definitions and terminology. As will become clear, the issue is also intimately related to validation in statistics and data analysis (NIR multivariate calibration is no exception).

Clarification

Upon reflection it will be appreciated that "replication" can concern (at least) the following alternatives in the lot-to-aliquot pathway from primary sampling to analytical result:

1) Replication of the primary sampling process
2) Replication starting at the secondary sampling stage (i.e. first mass reduction)
3) Replication starting with the tertiary sampling process (i.e. laboratory mass reduction)
4) Replication starting with repeated aliquot preparation (e.g. powder compaction)
5) Replication starting with aliquot instrument presentation (e.g. surface conditioning)

6) Replication of the analysis (measurement operation) only (*TAE*)

For the present discussion, we may assume that measurement and analysis are synonymous activities.

Option 6 is the situation corresponding to "replicate measurement" in the most restricted case. But does this mean that the analytical aliquot (the vial) stays in the analytical instrument all the time while the analyst "presses the button" repeatedly, say 10 times? Possibly—in which case this procedure would further a strict estimate of *TAE* only, but it may indeed seem equally relevant to extract the vial and insert it in the instrument repeatedly allowing a realistic temperature variation to influence *TAE* because this is a more *realistic* repetition of the general measurement process in any laboratory than simply leaving the test portion in the instrument. This is a first foray into what is known as "Taguchi thinking",[1] which opens up the possibility to focus on possible influencing factors which are not embedded in the experimental design; this could well be of interest in some cases. One important dictum of Taguchi was: do not necessarily look only for optimal results (which *may* have large variability) but to results where the response variability is low over a large span of the experimental domain (even if less optimal). This is what Toyota has been practicing for years. Certain scepticism regarding the merits of this approach has been voiced but here we will let the reader decide.

Opening up the relevance of this type of perturbation of the analytical process, to another analyst it may appear equally reasonable to include some, or all, of the "sample preparation" procedures in the replication as well, which should then also be repeated 10 times (point 4 and/or 5 above). But having broadened the horizon this far, it is an unavoidable logical step to follow up with still further realistic perturbations, which means to also include the tertiary, secondary and in the full measure of things, even also primary sampling errors in the replication concept. Why? Because these are potential uncertainty contributions that of necessity will be in play for any-and-all analytical aliquots ever subjected to measurement! Following the full impact of the *Theory of Sampling (TOS)* and its detailed treatment of the phenomenon of *heterogeneity*, it is in fact clear that the only **complete** "sampling-and-analysis" scenario that is guaranteed to include **all** uncertainty contributions must start with replication of the primary sampling ("replication from the top"). Any less comprehensive replication scenario is bound to be incomplete.

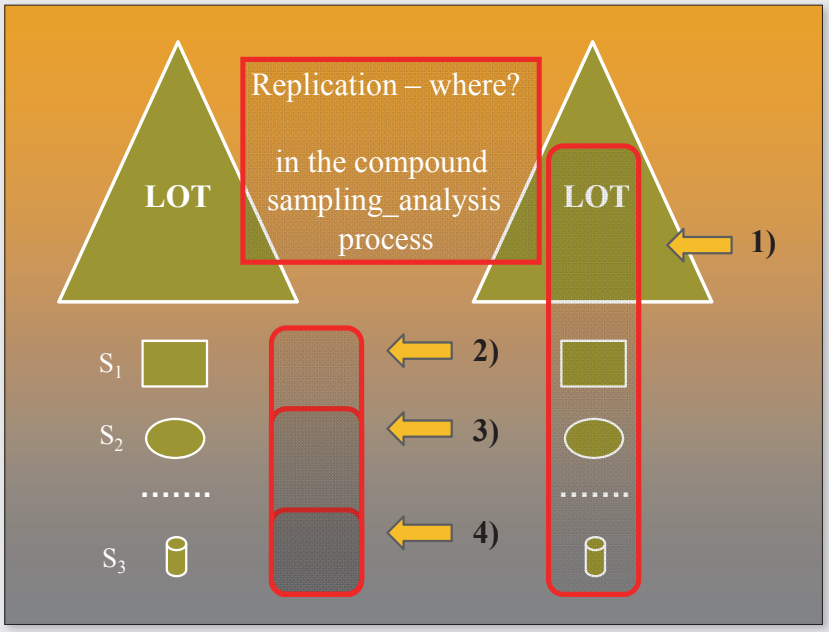

Figure 1. Replication can be performed at many stages in the full lot-to-aliquot pathway, but which is the most realistic situation pertaining to the general operations not only in the analytical laboratory? It turns out that all replication must meaningfully start "from the top".

Repeating the primary sampling, say 10 times, means that each primary sample is being subjected to the exact same protocol governing **all** the ensuing sub-sampling (mass-reduction), sample handling and preparation procedures in the laboratory. From the logic of the full representativity pathway "from lot-to-analytical aliquot", this is the only procedure incorporating the complete ensemble of uncertainties and errors encountered. For each primary sample subjected to this pathway, all errors (sampling, handling, splitting, preparation, analytical …) will be manifested differently ten individual times giving rise to an accumulated total variance which would be the most realistic estimate of the total sampling plus analysis error, indeed the **total** measurement uncertainty (*MU*).[2] In particular, this estimate is bound to include the full sampling error effects (*TSE*). In clear contrast, starting at any other of the levels in the list above, e.g. 2–6, will guarantee an incomplete, inferior *TSE*+*TAE* estimations, which of necessity is structurally destined to be too low.

Should one nevertheless feel compelled to cut short the full replication procedure starting from the top, one is mandated to describe the rationale behind such a choice and to report fully what was, in fact, done, otherwise the user of the analytical data has absolutely no way of knowing in full detail what was meant by the umbrella term "replication", i.e. users and decision makers, acting on the analytical data, are kept in the dark.

Undocumented, or unexplained, application of the term "replicate experiments" (or "repeated experiments") has been the source of a significant amount of unnecessary confusion in the past. Many times $s^2(TAE)$ has simply been *misconstrued* to imply $s^2(TSE+TAE)$, a grave error and one for which someone or some ill-considered, incomplete protocol is responsible. But nobody is interested in pointing fingers at any person—it is sufficient to stop such practice.

The above scenario illustrates an unfortunate compartmentalisation of responsibility which is, however, sometimes found in scientific and industrial publishing or regulatory contexts, e.g.

> "The analyst is not supposed to deal with sampling outside the laboratory"
> "This department is only charged with the task of reducing the primary sample to manageable proportions, as per codified laboratory's instructions"

continued on page 20

continued from page 18

> "Sampling is automated and carried out by Process Analytical Technology (PAT) sensors; there is no sampling issue involved here"
>
> "I am not responsible for sampling, I only analyse/model the *data*"…

and similar excuses for not seeing the complete measurement uncertainty issue. All too often the problem belongs to "somebody else", with the unavoidable result that the problem does not receive further attention. For such reasons, this attitude ("not our responsibility") is in danger of being perpetuated: "replicate analysis" will then still take as its starting point stage 3 or maybe stage 2 but almost never stage 1, the primary sampling stage. This is not an acceptable situation.

There have also been occasions on which authors, reviewers or even editors have failed to crack down with the necessary firmness on demonstrated ambiguities regarding "replication". The result is, then, that the reader is not able to fully understand what was intended or what indeed was actually carried out because of incomplete descriptions in the "Method" sections of scientific publications and technical reports. The issue is therefore far from trivial, indeed grave errors are still sometimes committed, but rather than address the obvious first question (who is responsible?), the way forward shall be constructive. The next column will suggest and illustrate ways and means to put an effective end to the replication issue where confusion still exists.

The replication myth: Something needs to be done!

Notes

1. Taguchi approach: http://en.wikipedia.org/wiki/Taguchi_methods
2. Issues related to the concept of Measurement Uncertainty (*MU*), which too often in practice only covers the parts of the analysis process that can be brought under direct laboratory control, while in its full definition purports to cover the entire sampling–handling–analysis pathway, shall be treated in later columns.

DIARY

2013

17–21 March, Philadelphia, PA, USA. **Pittcon 2013.** www.pittcon.org

27 March, Gembloux, Belgium. **NIR Platform: Quality Control and Contaminant Detection.** fernandez@cra.wallonie.be

1–7 June, Montpellier, France. **NIR-2013.** icnirs2013@cermagref.fr, www.icnirs2013.com/

30 June–4 July, Bevagna, Italy. **8th Colloquium Chemiometricum Mediterraneum (CCM VIII 2013).** www.gruppochemiometria.it/ccm2013/index.htm

2014

23–26 June, Kyungpook National University, Daegu, South Korea. **4th Asian NIR Symposium**. http://nir.ac.affrc.go.jp/Web-ANC/ANC-index.html

2–8 August, Chambersburg, PA, USA. **IDRC 2014.** www.idrc-chambersburg.org, rromanac@yahoo.com

2015

September or October, Brazil. **NIR-2015.**

CONFERENCE ORGANISERS

Remember to let us know of any conferences or other events that you would like listed in the *NIR news* Diary. Just e-mail the details to ian@impublications.co.uk.