Biological variability studies: design, analysis and reporting

Alice Sitch, Sue Mallett and Jon Deeks
With thanks to Jac Dinnes, Ed Lamb and Ceri Rowe
Monitoring process

- Patient has potential disease progression or recurrence

  Monitoring test performed

  - Test considered positive
    - Change in patient management
  - Test considered negative
Biological variability studies importance

- Appropriate and optimal use of tests requires accurate information regarding the variability of a test
- General threshold value (‘snap-shot’ rule) or differences from previous values for each individual should be considered (‘track-shot’ rule)
- Results of variability studies will also guide the threshold values used to define a positive test result
Biological variability studies aims

- To quantify the inherent variability of test results
- To understand the amount of variability at each level

Knowledge of the variability components allows tests to be used in the most efficient way which is especially important when using tests to monitor patient health.
Biological variability studies design

Patient 1

Time 1

Assessment 1

Assessment n3

Time n2

Assessment 1

Assessment n3

Patient n1

Time 1

Assessment 1

Assessment n3

Time n2

Assessment 1

Assessment n3
Biological variability studies design
Between-individual variability
Within-individual variability
Analytical variability
Pre-analytical variability
Variability

- Pre-analytical variation
- Analytical variation
- Within-individual variation
- Between-individual variation
eGFR-C study

- The eGFR-C study is a large multicentre HTA funded study being conducted in the UK.
- eGFR-C sub-study to assess the biological variability of the reference test for the measurement of GFR, iohexol clearance, and measures required for the calculation of estimated GFR, creatinine and Cystatin C.
Standard design
Standard analysis

- Use of ANOVA or GLM
- Normality assessed using Shapiro-Wilk tests and ln transformation if non-normal
- Test for outliers using Cochran’s C test and Reed’s criterion
- Fraser & Harris 1989
Standard measures calculated and reported

<table>
<thead>
<tr>
<th></th>
<th>lohexol</th>
</tr>
</thead>
<tbody>
<tr>
<td>$CV_A$</td>
<td>2.22</td>
</tr>
<tr>
<td>$CV_I$</td>
<td>6.67</td>
</tr>
<tr>
<td>$CV_G$</td>
<td>16.17</td>
</tr>
<tr>
<td>$CV_T$</td>
<td>17.63</td>
</tr>
<tr>
<td>RCV</td>
<td>19.49</td>
</tr>
<tr>
<td>$II$</td>
<td>0.41</td>
</tr>
</tbody>
</table>

\[
CV = \frac{SD}{\text{mean}}
\]

\[
RCV = \sqrt{2Z\sqrt{CV_A^2 + CV_I^2}}
\]

\[
II = \frac{CV_I}{CV_G}
\]
Key papers

**Generation and Application of Data on Biological Variation in Clinical Chemistry**

Authors: Callum G. Fraser
Department of Biochemical Medicine
Ninewells Hospital and Medical School
Dundee, Scotland

Eugene K. Harris
Department of Pathology
University of Virginia Health Sciences Center
Charlottesville, Virginia

**I. INTRODUCTION**

Numerical data on the biological variation of analytes assayed in the clinical chemistry laboratory have important applications. Desirable performance standards, or analytical goals, for the imprecision of analytical methods are best derived from data on biological variation.

The critical evaluation of the significance of changes in results obtained on analysis of serial specimens can be performed only by consideration of biological in addition to analytical variation. The likely utility of conventional population based reference values can be best assessed from consideration of within and between subject biological variation.

The data have a number of other useful applications, including assessment of the number of specimens to be collected to assess the homeostatic set point of an individual, determination of the optimum mode to report numerical results, selection of the best specimen to collect, comparison of available tests, and assessment of the clinical utility of tests.

In view of these applications, and the increasing interest in this aspect of clinical chemistry, this review has been prepared to propose a standard approach to the definition and analysis of biological variation, and to consider ways in which the data generated can be applied usefully.
Opinion Paper

William A. Bartlett*, Federica Braga, Anna Carobene, Abdurrahman Coşkun, Richard Prusa, Pilar Fernandez-Calle, Thomas Røraas, Neils Jonker and Sverre Sandberg, on behalf of the Biological Variation Working Group, European Federation of Clinical Chemistry and Laboratory Medicine (EFLM)

A checklist for critical appraisal of studies of biological variation
Confidence Intervals and Power Calculations for Within-Person Biological Variation: Effect of Analytical Imprecision, Number of Replicates, Number of Samples, and Number of Individuals

Thomas Røraas, Per H. Petersen, and Sverre Sandberg
Biological variability studies review: questions

- What is the current state of the field?
- How are studies assessing biological variability of tests designed?
- How are studies assessing biological variability of tests analysed?
- How are studies assessing biological variability of tests reported?
- What are the differences between studies assessing biological variability of laboratory, imaging and physiological tests?
Biological variability studies review: searches

- Key word search (bio* AND vari*)
- Clinical Biochemistry
- Radiology
- Clinical Chemistry
- Westgard QC database
- Detailed searches for three different test types in specific clinical areas
- Additional papers identified to enrich the sample
Current state of the field: frequency and areas

- Physiological (n=6, 6%) and imaging tests (n=20, 20%)—majority of studies were laboratory tests (n=75, 74%)

- Examples:
  - FEV
  - bladder thickness, tumour size
  - Vitamin uptake, hepatic enzymes, Hba1c, GFR, cardiac troponin
Current state of the field: aims

- 37 (37%) variability of just one test
- 27 (27%) also evaluated tests in multiple populations (or subpopulations) and 11 (11%) over different time ranges
- 25 (25%) primarily evaluating other aspect of test performance
- Median (Q1, Q3)= 4 (2, 7) number of testing situations
- 16 (16%) inter-intra reader studies
Design—population

- 48 (48%) used healthy populations
- 21 (21%) studies use a mixed population
- 22 (22%) studies only testing patients with disease
Design—sample size

- Rarely performed (n=1, 1%)
- Range of sample sizes- smallest 4 patients and the largest having 7101 patients.
- Many studies had very few participants with the median number of patients being 24 (Q1, Q3: 15, 40).
- Studies with larger sample sizes were utilising routinely collected data
Design—sample size

*excluding 8 studies with sample size greater than 100
Analysis—methods

- ANOVA and/or random effects modelling, which was used by 88 (87%) studies
- Assume ANOVA or random effects modelling must have been performed but this is not explicitly expressed (n=40, 40%).
- Fraser & Harris methods referenced by 52 (51%) studies and referred to in the methods section by 32 (32%) studies
- Bland-Altman (n=7, 7%), Kappa (n=1, 1%), other modelling (n=1, 1%) and other methods (n=10, 10%).
Analysis—outliers and normality

- Outliers were reported to have been tested for in 25 (25%) studies
- Outliers had been excluded in 27 (27%) studies
- The normality of obtained test data was tested in 38 (38%) studies
- 22 (22%) studies reported log transforming the data
Reporting—clarity

- not adequately defining/justifying:
  - population (n=10, 10%)
  - sample size (n=100, 99%)
  - number of repeats (n=39, 39%)
  - timing of repeats (n=42, 42%)
  - number of assessments duplicated (n=4, 4%)
  - duration of study (n=17, 17%)
  - variability of measure of assessment external (n=14, 14%)
  - method for analyses (n=40, 40%)
  - outlier procedure (n=10, 10%)
  - normality procedure (n=31, 31%).
Reporting—estimates

- CV estimates of assessment (analytical) variability (n=35, 35%), within-individual variability (n=72, 71%) and between individual variability (n=60, 59%)
- The RCV (or repeatability coefficient) was reported for 48 (48%) studies
- Forty-four (44%) studies reported the index of individuality (II) and 18 (18%) reported an ICC/reliability parameter
- Few studies (n=17, 17%) provided confidence intervals
Differences between test types

- Physiological (n=6, 6%) and imaging test (n=20, 20%)—majority of studies were laboratory tests (n=75, 74%)
- 52 (51%) of studies assessing laboratory tests reference the framework of design and analysis dictated by Fraser and Harris
- The majority of laboratory test studies use similar design and methods for analyses, others more ad hoc
Summary

- Studies usually use healthy population
- Sample size justification is often not considered in biological variability studies and these studies are often small
- Mainly ANOVA/GLM is used for primary analysis
- Outlier detection and exclusion is often used
- Normality assessment and log transformation is often performed
- CV often reported (especially for laboratory based tests)
- Clarity of reporting is weak in some areas
Conclusions

- Scientist and researchers need to be aware of the need for variability estimates
- The design, methods of analysis and clarity of reporting for biological variability studies can be improved
- Methods for sample size calculation are required
- The impact of outlier detection and data transformation requires investigation
- The reporting of biological variability studies needs to be detailed and transparent