Sources of variation, sample size calculations, and quality control

Matthijs Vynck[1]
Jo Vandesompele[2], Olivier Thas[1,3]

[1] Department of Data Analaysis and Mathematical Modelling
Faculty of Bioscience Engineering, Ghent University, Belgium
[2] Center for Medical Genetics, Ghent University, Belgium
[3] National Institute for Applied Statistics Research Australia (NIASRA), School of Mathematics and Applied Statistics, University of Wollongong, Australia

Contents

Sources of variation and bias

Jacobs *et al. BMC Bioinformatics* 2014, **15**:283
http://www.biomedcentral.com/1471-2105/15/283

RESEARCH ARTICLE Open Access

Impact of variance components on reliability of absolute quantification using digital PCR

Bart KM Jacobs*, Els Goetghebeur and Lieven Clement*

Abstract

Background: Digital polymerase chain reaction (dPCR) is an increasingly popular technology for detecting and quantifying target nucleic acids. Its advertised strength is high precision absolute quantification without needing reference curves. The standard data analytic approach follows a seemingly straightforward theoretical framework but ignores sources of variation in the data generating process. These stem from both technical and biological factors, where we distinguish features that are 1) hard-wired in the equipment, 2) user-dependent and 3) provided by manufacturers but may be adapted by the user. The impact of the corresponding variance components on the accuracy and precision of target concentration estimators presented in the literature is studied through simulation.

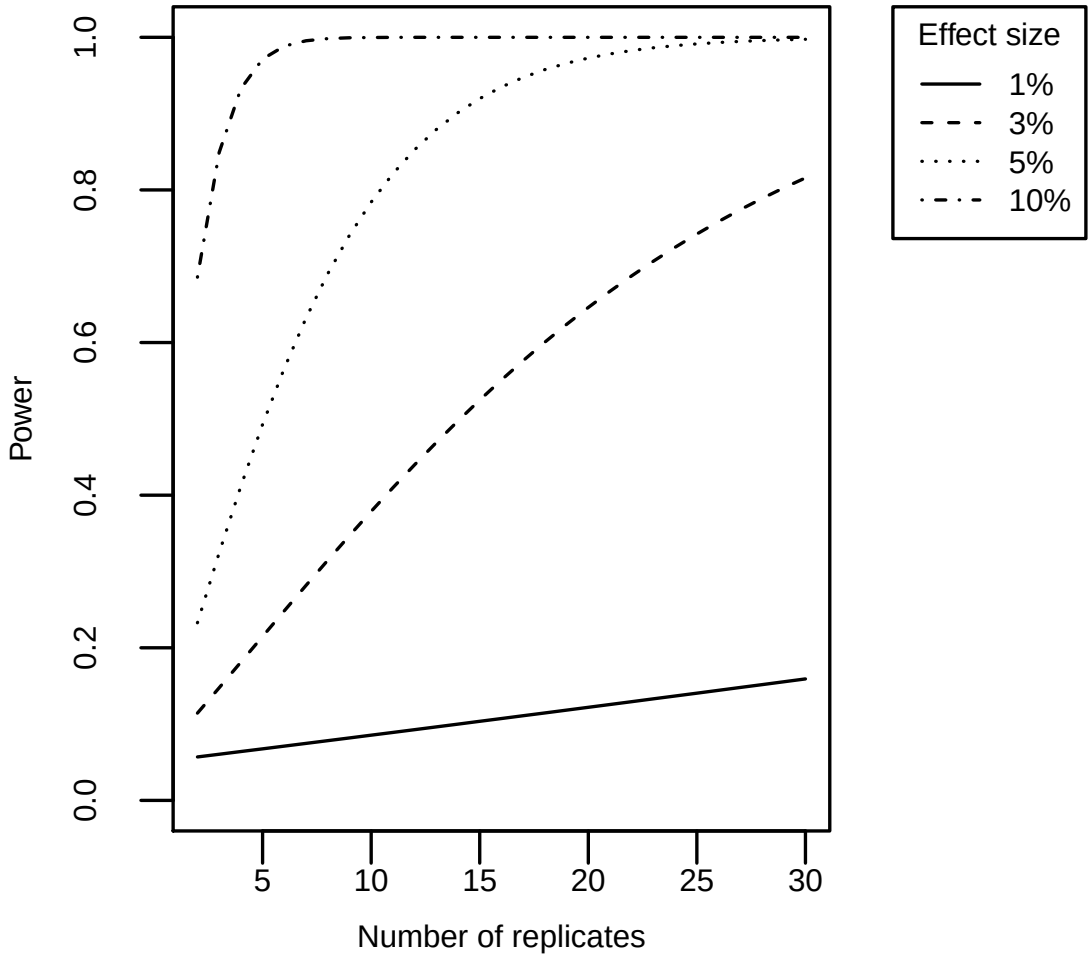
Sources of variation and bias

Why important? Determines reliability of measurements!

- In terms of trueness (bias)
  - Partition misclassification (solutions exist)
  - Wrong average partition volume (solution: measure your volumes!)
  - Partition volume variability (solutions exist)
- In terms of precision (uncertainty, variance)
  - Between-replicate variance handling, e.g. pipette errors (solutions exist)
  - Partition occupancy (neither too high, nor too low)

Sample size calculations: factors affecting power

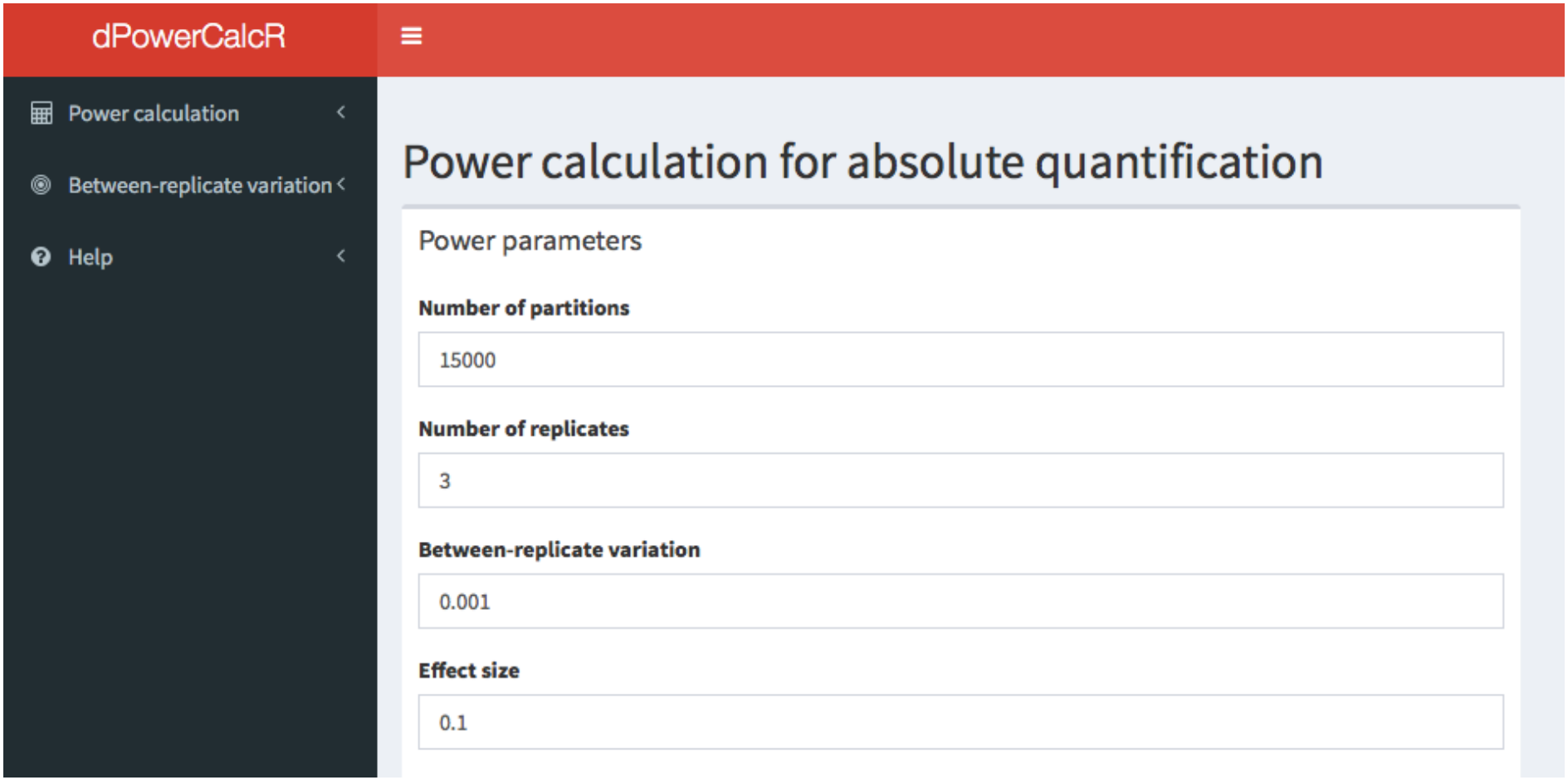
- Number of partitions
- Number of replicates
- Fraction of negative (positive) partitions
- Effect size
- Between-replicate variation
- Significance level



Sample size calculations: the easy way

Web tool: easy power calculation / report generation

`http://statapps.ugent.be/dPCR/dPowerCalcR/`



Sample size calculations: example

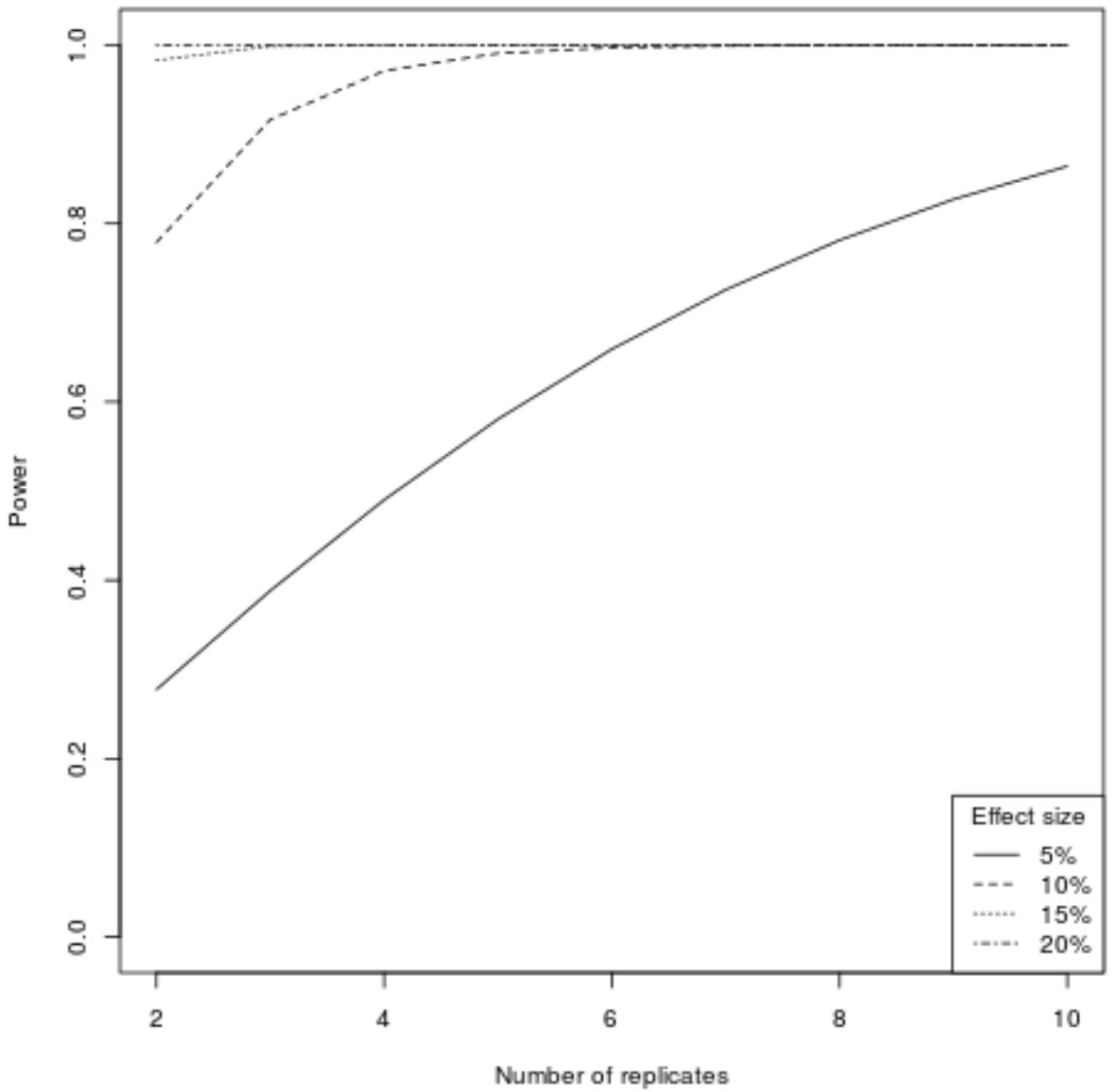
Example:

- My samples have a negative fraction of 20%
- My machine generates 20 000 partitions
- My between-replicate variation is 0.01
- My significance level is 5%
- I want to detect an increase in copy number of 10%, which I consider of biological (clinical) relevance
- If such a relevant difference is indeed present, I want to be 90% sure to detect it (required power)

How many replicates should I run?

Sample size calculations: example

Plug into web application, several figures result:



Quality control

- Linearity (dynamic range)
- Trueness (not discussed)
- Precision (repeatability, reproducibility)

Anal Bioanal Chem (2017) 409:5919–5931
DOI 10.1007/s00216-017-0538-9

PAPER IN FOREFRONT

Quality control of digital PCR assays and platforms

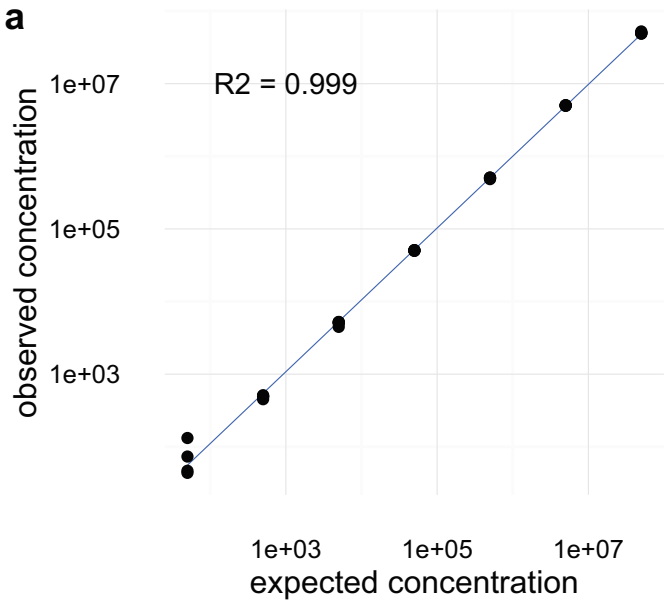
Matthijs Vynck[1] · Jo Vandesompele[2,3,4] · Olivier Thas[1,5,6]

Linearity

Current assessment:

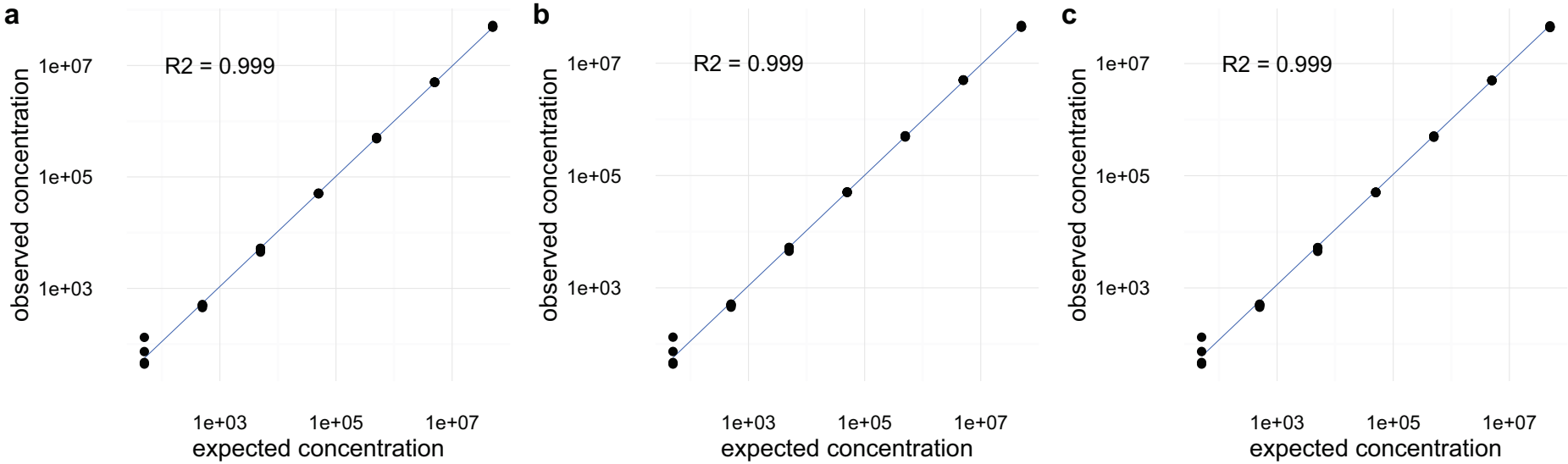
- Graphical: log-log plots
- Numerical: R^2 values

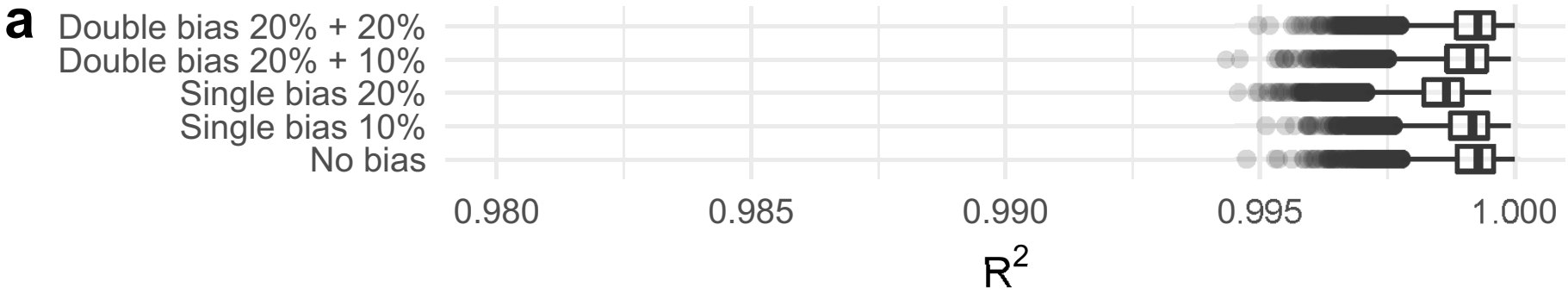
Typical linearity plots:



Linearity

Small to medium deviations not detected on log-log plots / R^2 values





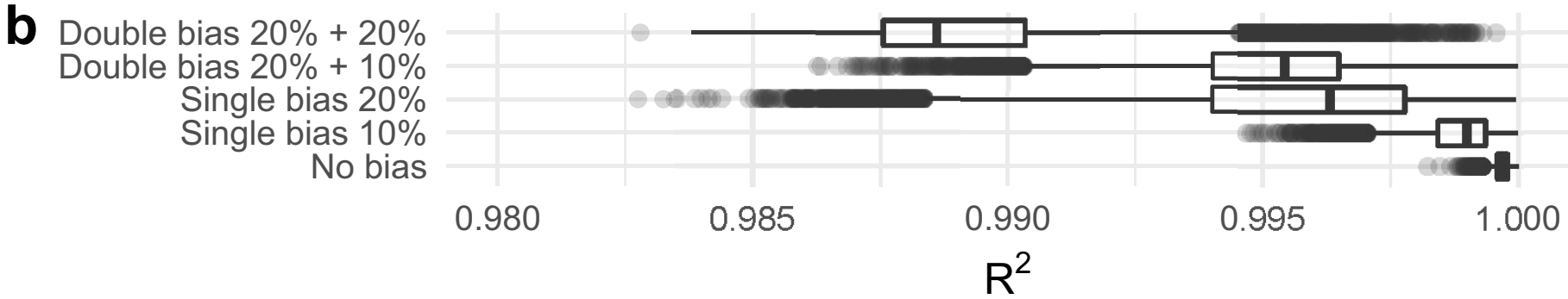
Linearity

Why?

- Large range of values: 1 - 10 000 000 copies: high leverage
- Substantial heterogeneity in variance: heteroscedasticity

Standard linear regression, a.k.a. (ordinary) least squares (from which R^2 can be calculated) does not deal well with high leverage and heteroscedasticity.

Solution: R^2 values derived from a robust weighted least squares?



Linearity

Improved ways to detect non-linearity? Quadratic regression, lack-of-fit, runs test, frequency within a block test, more detailed plots ...

Additional factors influencing ease to detect non-linearity:

- Number of replicates: more is better
- Dynamic range (cfr. leverage): less is better (in terms of detection of deviation from linearity)
- Concentration at which deviation happens: easier detection for extremal (lowest and highest) concentrations

Precision

- Assessment of precision in terms of the coefficient of variation (CV)
- Often comparison of CV between dPCR and qPCR
- Often point estimates: how reliable?

Example (Morriset et al., 2013):

Five replicates of the dilution series [...] were measured by ddPCR. For qPCR, measurements were made in duplicate.

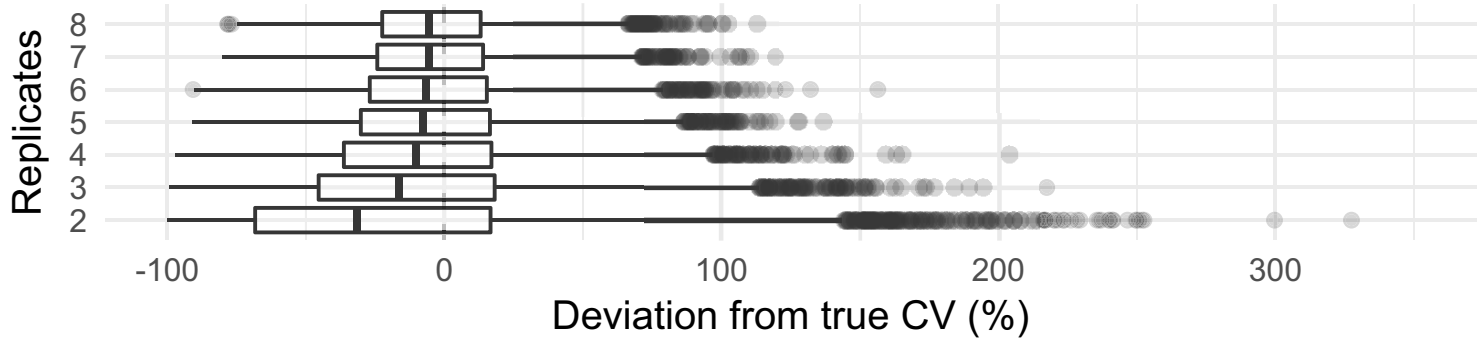
Conclusion (partly):

All along the dynamic range, the CV of the determined hmg copies, MON 810 copies, and MON810 content remained below the threshold for acceptance of quantitative methods (CV <25%).

How uncertain are these estimates?

Precision

How uncertain are these estimates?



It is recommended to also report confidence intervals of the CV!

For the Morisset et al. (2013) data (MON810):

Dilution	CV	CV 95% CI
3	1.8%	[1.1%, 5.9%]
4	2.1%	[1.3%, 6.0%]
16	4.8%	[3.0%, 13.9%]
81	8.3%	[5.1%, 24.2%]
243	16.5%	[9.8%, 50.2%]
729	19.7%	[11.7%, 61.5%]

Precision

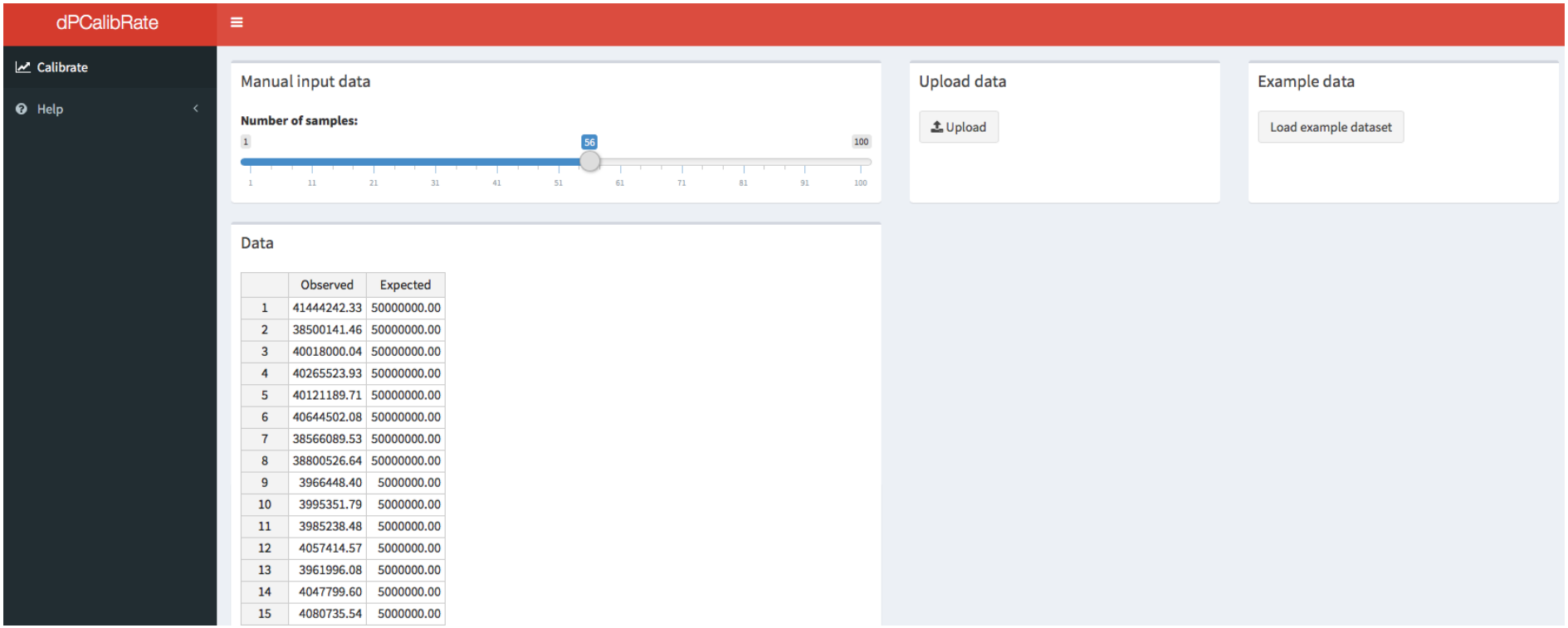
When studying precision:

- Consider the number of replicates
- Consider that the estimated CV is uncertain
  - When formulating conclusions on achieved precision
  - When comparing techniques, e.g. qPCR and dPCR
    (consider a two-sample statistical test, e.g. Feltz and Miller, 1996)
- Uncertainty decreases with increasing number of replicates
- Sample size calculations for desired width of the CI exist (e.g. Kelley, 2007)

Quality control: the easy way

Web tool: easy quality report generation

`http://statapps.ugent.be/dPCR/dPCalibRate/`



	Observed	Expected
1	41444242.33	50000000.00
2	38500141.46	50000000.00
3	40018000.04	50000000.00
4	40265523.93	50000000.00
5	40121189.71	50000000.00
6	40644502.08	50000000.00
7	38566089.53	50000000.00
8	38800526.64	50000000.00
9	3966448.40	5000000.00
10	3995351.79	5000000.00
11	3985238.48	5000000.00
12	4057414.57	5000000.00
13	3961996.08	5000000.00
14	4047799.60	5000000.00
15	4080735.54	5000000.00

Thank you!

References:

Feltz and Miller (1996) An asymptotic test for the equality of coefficients of variation from k populations. Statistics in Medicine

Jacobs et al. (2014) Impact of variance components on reliability of absolute quantification using digital PCR. BMC Bioinformatics

Kelley (2007) Sample size planning for the coefficient of variation from the accuracy in parameter estimation approach. Behaviour Research Methods

Morisset et al. (2013) Quantitative analysis of food and feed samples with droplet digital PCR. PLoS One

Vynck et al. (2017) Quality control of digital PCR assays and platforms. Analytical and Bioanalytical Chemistry