Electronic Supplementary Material (ESI) for Lab on a Chip.

Supplementary Information

Translating Diagnostic Assays from the Laboratory to the Clinic: Analytical and Clinical Metrics for Device Development and Evaluation

Mark D. Borysiak[1], Matthew J. Thompson[2], and Jonathan D. Posner[1,2,3]

[1]Chemical Engineering, University of Washington, Seattle, WA 98195

[2]Department of Family Medicine, University of Washington, Seattle, WA 98195

[3]Mechanical Engineering, University of Washington, Seattle, WA 98195

[*]Corresponding Author, Email: jposner@uw.edu

Abstract

The supporting information shows:

1. An analytical sensitivity vs analyte concentration for a sigmoidal calibration functions,
2. Analytical sensitivity plots of representative quantitative polymerase chain reaction (qPCR) data,
3. The non-parametric procedure for determining the limit of detection (LoD) when blank data are truncated at zero and how to calculate pooled sample standard devations,
4. Natural frequencies diagrams for the QuickVue Chlamydia test and Chlamydia Rapid Test,
5. The non-linear dependence of LR+ on diagnostic specificity.

1. Analytical sensitivity vs. concentration plot

As discussed in Section 1.2, analytical sensitivity is the ability of an analytical procedure to differentiate between different amounts of analyte, dR/dx, which is typically shown as the slope of the calibration curve.[1] Below we show two different response versus concentrations plots for hypothetical tests with higher reportable range (A) and lower reportable range (B). Figure S1A shows the data from Figure 2 in the manuscript along with a best-fit calibration function, $y = \frac{1}{1+e^{-a(x-c)}}$, where *a* and *c* are fitting constants equal to 0.06 and 75 respectively. Figure S1B shows a plot of a test with very limited reportable range, with the same sigmoidal fit as (A), except *a* and *c* are are equal to 0.90 and 45 respectively.

By taking the derivative of the calibration function with respect to x, the analyte concentration, we are able to obtain a plot of analytical sensitivity versus concentration for the entire calibration function range, as shown in Figure S1C and S1D, for Figure S1A and S1B respectively. The specific derivative that we plot is: $\frac{dy}{dx} = ae^{-a(x-c)}/(1 + e^{-a(x-c)})^2$, with the a=0.06 or 0.90 and c=75 or 45. The difference between reportable ranges in Figure S1A and S1B is distinctly shown by the width of the Gaussian-type curve in the respective analytical sensitivity plots. Figure S1C has moderate analytical sensitivity over a larger analyte concentration, which means that it could provide quantitative results with greater certainty for more analyte values. Figure S1D, on the other hand, has very high analytical sensitivity at its peak, but it only covers a small analyte concentration range. In this small analyte concentration range the test from S1D could provide extremely high analyte concentration certainty, but any analyte concentrations outside of this range effectively could not be quantified at all. These figures show the trade-off that can occur between high analytical sensitivity and reportable range, which is explored more in the next section.

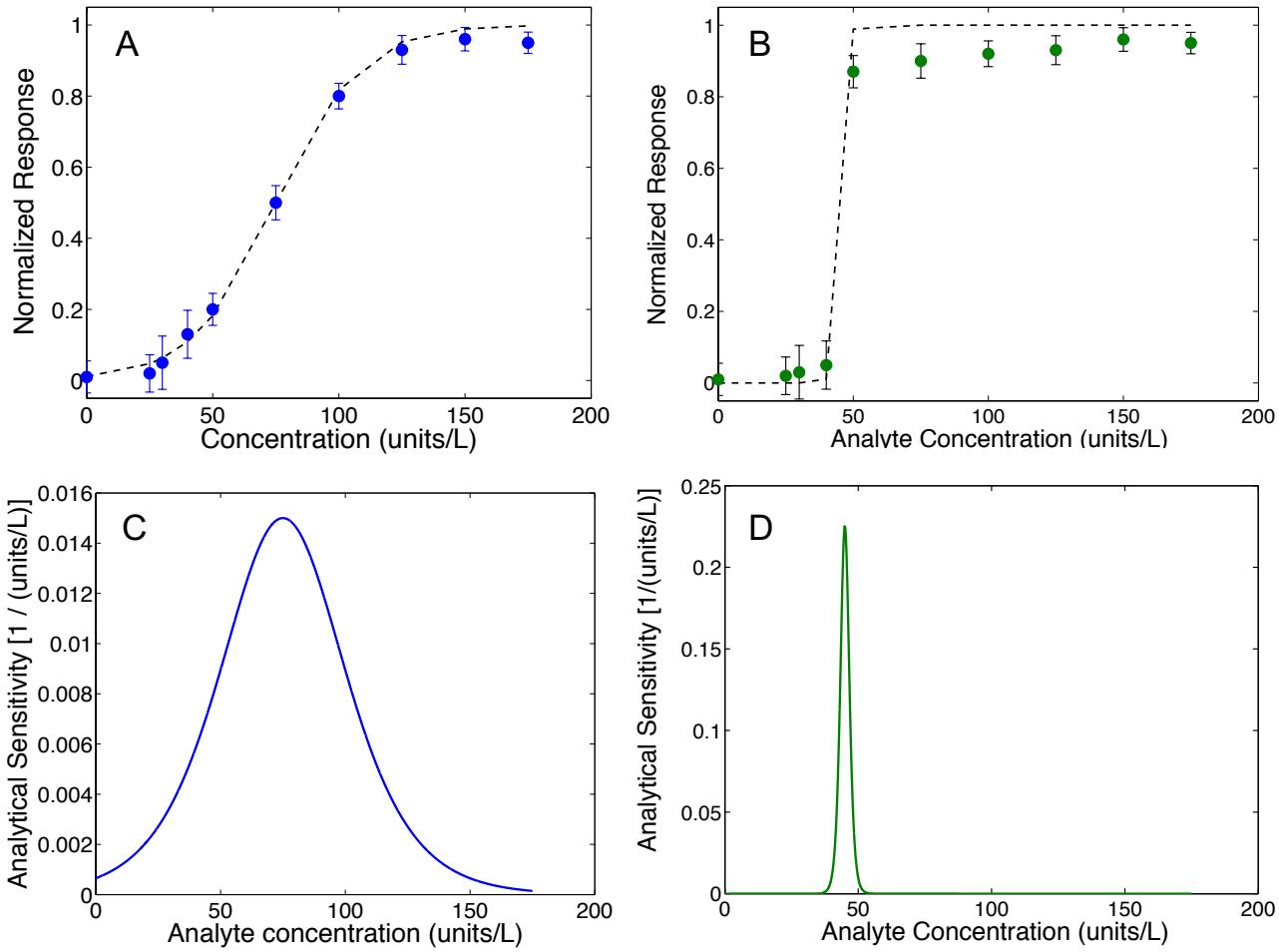

Figure 1. Theoretical response plots for a test that has higher reportable range (A) versus lower reportable (B) with sigmoidal curve fits. Plots (C) and (D) show the derivative of the sigmoidal calibration function as a function of analyte concentration, which gives the analytical sensitivity at each different concentration. The width and peaks of the analytical sensitivity plots show the broader reportable range with less analytical sensitivity for (A) versus the very narrow reportable range for (B). These analytical sensitivities can be used to calculate the quantitative response for a given measurement uncertainty at each concentration, as discussed in section 1.1 of the manuscript.

2. Analytical sensitivity qPCR example

Higher slope magnitudes (high analytical sensitivity) mean that a test is more sensitive to changes in analyte amount and the analyte can be determined with higher certainty, while lower slope magnitudes (low analytical sensitivity) means that a test is less sensitive to changes in analyte amount. High analytical sensitivity is preferred for quantitative tests where small changes in analyte concentration should be monitored. For example, it is recommended that 3-fold changes in viral load should be able to be detected by a test monitoring the viral load of HIV patients who are receiving antiretroviral treatment (ART).[2] However, a trade-off between high analytical

sensitivity and the reportable range of analytes that can be quantified must often be considered because higher analytical sensitivity often leads to lower reportable range due to signal saturation at high analyte amounts or noise from blank samples at low analyte amounts.[3] Ideally, the high analytical sensitivity region covers the concentration range of interest for the clinical application, where the noise and saturated ends of the curve are clinically irrelevant. Figure S2 below demonstrates these ideas.

Figures S2A and S2C show representative qPCR data for a serial dilution of target nucleic acid copies. For detailed descriptions of qPCR data and analysis, see Bustin *et al* and Pabinger *et al*.[4,5] Normalized fluorescence intensity is plotted versus the number of cycles required to amplify the nucleic acid for detection. The copy numbers are labeled for each curve and the data shows that as copy number decreases, more thermal cycles are required in order to amplify the nucleic acid. The typical data analysis for qPCR is to set a threshold near the intensity where the fluorescence curves begin to increase exponentially. The point where the fluorescence crosses the threshold is known as the threshold cycle, which is the number of cycles required to create amplification equal to the threshold. The threshold cycles for each number of copies can be plotted versus the logarithm of copy number to create a linear calibration plot, as shown in Figure S2B and S2D for the data in Figure S2A and S2C respectively.

The exponential amplification curves in Figure S2A are more closely spaced compared to the curves from Figure S2C. The closer spacing in S2A leads to less differentiation between different concentrations than if the spacing is further apart, such as S2C. As a result, the corresponding analytical sensitivity slope magnitude in Figure S2B (–3.1) is not as steep as the slope in Figure S2D (–5.4). The lower analytical sensitivity for the test shown by Figures S2A and S2B would result in a qPCR procedure that is less able to differentiate between number of nucleic acid copies. It also means that assuming equal measurement variance between the two procedures, that the (A,B) procedure will give less certainty in the determined copy number due to worse quantitative resolution.[6] The test shown by Figures S2C and S2D has higher analytical sensitivity that would give better differentiation and improved quantitative resolution. However, the qPCR procedure in (C,D) reaches a saturation point and cannot quantify samples containing less than 10 copies because the dynamic range has decreased (35 cycles is a common end-point due to reagent degradation and primer-dimer false positives). If the test requires high diagnostic sensitivity, the inability to detect and quantify 10 copies would hinder diagnosis. Overall, the test that would be preferred for a given application would likely depend on the relevant clinical concentrations, as

well as the confidence in analyte concentration that would be needed to be detected for a given application.

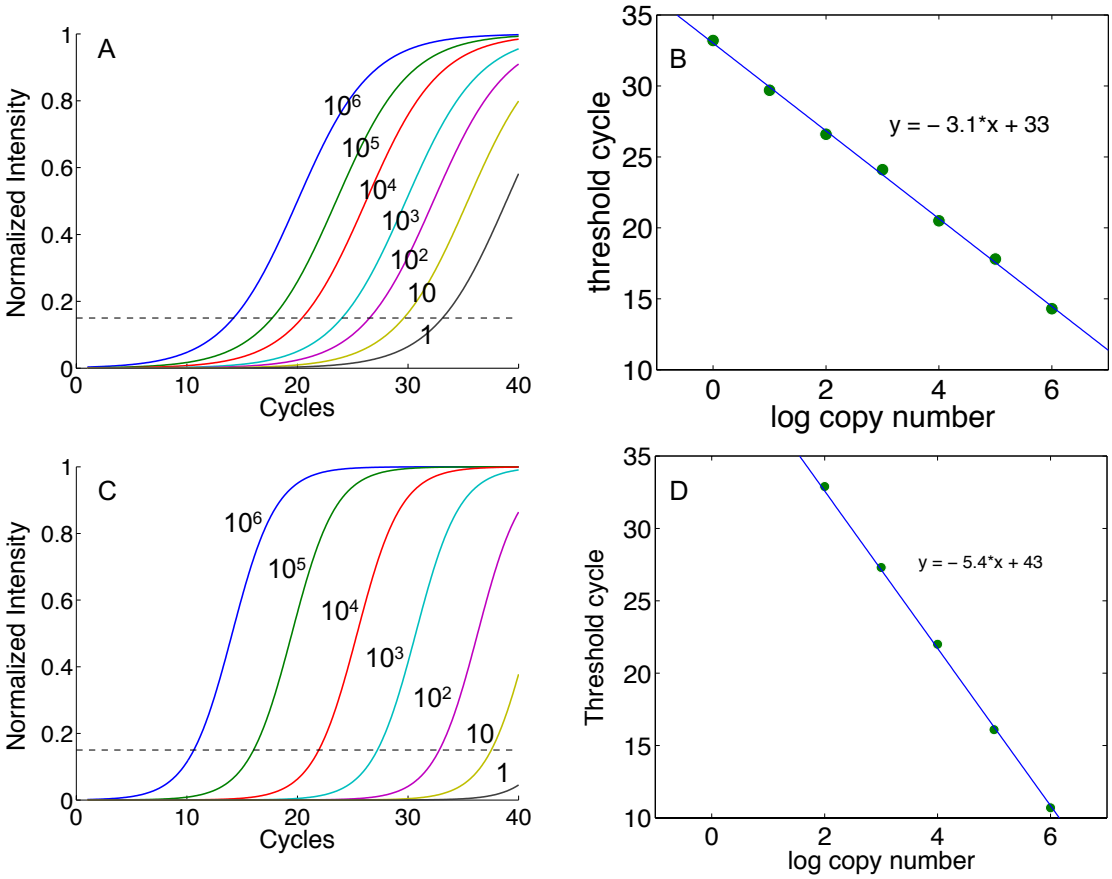

Figure S2. Example demonstrating the trade-off between high analytical sensitivity and dynamic range. Plots A and C show representative qPCR data for a test with (A) lower analytical sensitivity and (C) higher analytical sensitivity. Plots B and D show the calibration curves that can be generated from the data in A and C respectively. Test (A) has more closely spaced amplification curves than test (C) which leads to less analyte differentiation and the lower analytical sensitivity shown in the corresponding calibration curve. The lower analytical sensitivity for test (A,B) would result in worse quantitative resolution than the test for (C,D). However, the qPCR procedure for (C,D) is only able to quantify 5 out of the 7 concentrations, assuming a cut-off at 35 cycles. The test that would be preferred for a given application would likely depend on the relevant clinical concentrations, as well as the change in analyte concentration that would be needed to be detected for a given application.

3. Non-parametric determination of LoB pooled standard deviation for LoD

If the test developer is unable to determine or assume that the blank distribution is Gaussian and comes from a normal distribution, which may occur when blank values are truncated at zero, a non-parametric procedure is required for determining the LoB. In this case, the LoB is determined using,[7]

$$LoB = Result\ at\ position\ [N_B\left(\frac{p}{100}\right) + 0.5], \qquad (1)$$

where N_B is the number of blank measurements, and p is the percentile of interest such as the 95th percentile. In this non-parametric case, the blank measurements are ordered according to their values, *i.e.* lowest to highest, and the 95th percentile is estimated by choosing the blank values that occur at approximately the 95% position of the order. For 200 samples, $LoB = Result\ at\ position\ [200 \cdot 0.95 + 0.5] = 190.5$, which means the LoB is the value of a linear interpolation between the 190th and 191th ranked LoB measurements. For ten ordered blank samples such as: [0 0 0 1 1 2 2 4 5 6], the LoB would be the average between the 9th and 10th measurements or (5+6)/2 = 5.5. Advantages of this method are that it does not rely on the data being normally distributed and that it is not as sensitive to outliers as a parametric model where the standard deviation can be skewed due to a small number of samples that vary greatly from the mean. A disadvantage however, is that it only takes into account a very small subset of the data at the tail of the distribution when determining the LoB instead of using the entire set to calculate the estimated mean and variance of the population.

The equation for pooled variance for low-level samples to determine the LoD is:

$${SD_S}^2 = (n_1\{Citation\}SD_1^2 + n_2SD_2^2 + \cdots)/(n_1 + n_2 + \cdots) \qquad (2)$$

where n_i is the number of replicates tested and SD_i^2 is the variance of sample concentration i.[7] The pooled standard deviation is simply the square root of this pooled variance. The CLSI EP17 document recommends testing and pooling 4–6 low level samples for LoD determination, but F-tests or Cochran's test must be run prior to pooling to ensure that all variances are equal. The CLSI EP17 document provides further details.[7]

4. Frequency diagram for QuickVue and Chlamydia Rapid Test

Section 4.3 in the manuscript discusses the accuracy of a number of lateral flow Chlamydia tests including the QuickVue Chlamdyia test and Chlamydia Rapid Test (CRT). van Dommelen *et al* showed the the QuickVue test had a diagnostic sensitivity of 25% and a diagnostic specificity of 99.7%, with a high positive likelihood ratio (LR+) of 83.3.[8] The high specificity and LR+ suggest that the QuickVue test may have utility has a rule-in test. Natural frequencies diagrams are useful for visualizing testing results for a theoretical patient population.[9,10] Figure S3A shows the frequency diagram for 10,000 patients and 11% prevalence being tested with the QuickVue test. The positive predictive value (PPV) for this test is high (11% pre-test to 91.2% post-test), meaning that a positive result would likely lead to conclusive diagnosis and treatment. However, detecting 275 patients (25% of those infected) would require 10,000 tests to be run, as well as 9,725 backup

tests. Section 4.3 in the manuscript and Gift et al. discuss the implications of using a test with poor sensitivity, yet high specificity in more detail.[11]

The Chlamydia Rapid Test (CRT) has a higher diagnostic sensitivity (42%), but lower diagnostic specificity (96.8%) compared to the QuickVue test. This lower diagnostic specificity compared to the QuickVue may not appear to be largely signficant, but results in a much higher false positive rate and lower PPV (91% for QuickVue and 57% for CRT). Figure S3B shows a natural frequencies diagram for the CRT that can be compared to the QuickVue diagram, Figure S3.

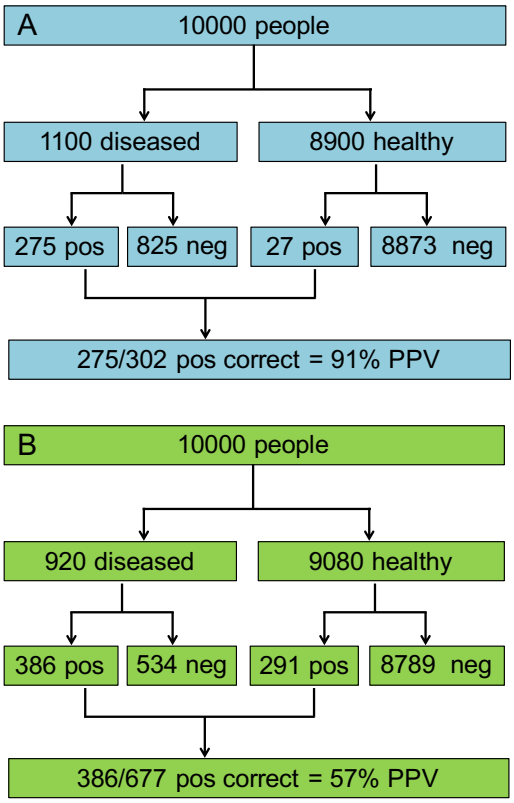

Figure S3. Natural frequencies diagrams for the (A) QuickVue and (B) Chlamydia Rapid Tests. (A) Natural frequency diagram of a 10,00 patient population with 11% prevalence being testing with the QuickVue Chlamydia test from van Dommelen et al. The high test diagnostic specificity results in a high PPV of 91.2%, but the overall economic and clinical benefit of the test requires consideration of additional factors outlined by Gift *et al*. in the "rapid test paradox". (B) The CRT has higher diagnostic sensitivity compared to the QuickVue test (42% vs. 25%), but lower diagnostic specificity (96.8% vs. 99.7%). The seemingly small decrease in diagnostic specificity leads to a much lower PPV (57%) for the CRT, although it should be noted that the prevalence is slightly lower in the CRT case.

5. Non-linearity of LR+ versus diagnostic specificity

The equations for positive (LR+) and negative (LR–) likelihood ratios are given below.

$$LR+ = \frac{\frac{TP}{TP+FN}}{\frac{FP}{FP+TN}} = \frac{sensitivity}{1-specificity} \quad (3)$$

$$LR- = \frac{\frac{FN}{TP+FN}}{\frac{TN}{TN+FP}} = \frac{1-sensitivity}{specificity} \quad (4)$$

Positive likelihood ratios are interesting due to the non-linear dependency on specificity. Figure S4 shows a plot of LR+ versus diagnostic specificity+ for various diagnostic sensitivities. As the diagnostic specificity approaches 1, the curve begins to show highly non-linear behavior due to the 1/(1-x) dependence. This plot demonstrates that when diagnostic specificities are over approximately 90%, that small increases can lead to drastically improved LR+ that give more confident positive diagnostic results. The specificity at which the non-linear behavior begins depends slightly on the test diagnostic sensitivity, as shown by plot A-D. Negative likelihood ratios do not exhibit the same non-linear dependence, and have linear dependency over the entire range of possible diagnostic sensitivities and specificities.

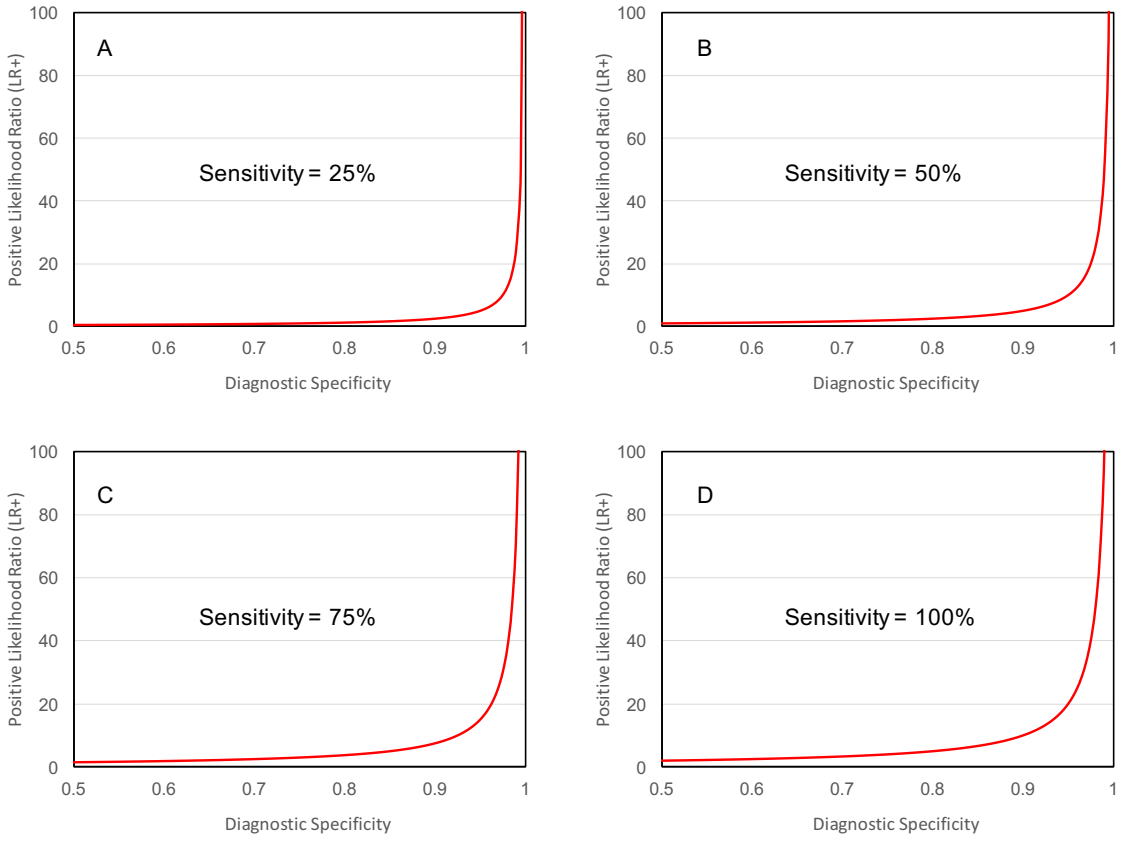

Figure S4. Non-linear dependence of LR+ on diagnostic specificity. As the diagnostic specificity approaches 1, the LR+ greatly increases, which means that much more confident positive diagnoses can be made with tests as the diagnostic specificity gets closer to 1. The point at which the non-linearity begins to drastically impact LR+ is slightly dependent on diagnostic sensitivity, as shown by plots A–D that increase in diagnostic sensitivity from 25% to 100%. The plots show diagnostic specificity from 0.5–1 and LR+ from 0 to 100 for clarity. The LR+ for a diagnostic specificity of 99.9% are 250, 500, 750, and 1000 for A–D respectively.

References

1 L. A. Currie, *Anal. Chim. Acta*, 1999, **391**, 105–126.
2 World Health Organization, World Health Organization and Department of HIV/AIDS, *Antiretroviral therapy for HIV infection in adults and adolescents: recommendations for a public health approach.*, 2010.
3 National Association of Testing Authorities, Australia, 2013.
4 S. A. Bustin, V. Benes, J. A. Garson, J. Hellemans, J. Huggett, M. Kubista, R. Mueller, T. Nolan, M. W. Pfaffl, G. L. Shipley, J. Vandesompele and C. T. Wittwer, *Clin. Chem.*, 2009, **55**, 611–622.
5 S. Pabinger, S. Rödiger, A. Kriegner, K. Vierlinger and A. Weinhäusel, *Biomol. Detect. Quantif.*, 2014, **1**, 23–33.
6 H. L. Pardue, *Clin. Chem.*, 1997, **43**, 1831–1837.
7 Clinical and Laboratory Standards Institute (CLSI), *Protocols for Determination of Limits of Detection and Limits of Quantitation; Approved Guidline*, EP17-A, 2004.
8 L. Van Dommelen, F. H. Van Tiel, S. Ouburg, E. E. Brouwers, P. H. Terporten, P. H. Savelkoul, S. A. Morré, C. A. Bruggeman and C. J. Hoebe, *Sex. Transm. Infect.*, 2010, **86**, 355–359.
9 G. Gigerenzer, *BMJ*, 2013, **347**, f5151–f5151.
10 G. Gigerenzer and U. Hoffrage, *Psychol. Rev.*, 1995, **102**, 684.
11 T. L. Gift, M. S. Pate, E. W. Hook and W. J. Kassler, *Sex. Transm. Dis.*, 1999, **26**, 232–240.