

This work was supported by funds from Ministero dell'Istruzione, dell'Università e della Ricerca and from the University of Bologna. The LightCycler instrument was made available by Centro Interdipartimentale Ricerche Biotecnologiche, University of Bologna.

References

- Zerbini M, Musiani M. Human parvoviruses. In: Murray PR, Baron EJ, Jorgensen JH, Pfaller MA, Tenover FC, Tenover FC, eds. *Manual of clinical microbiology*, 8th ed. Washington: ASM Press, 2003:1534–43.
- Gallinella G, Zuffi E, Gentilomi G, Manaresi E, Venturoli S, Bonvicini F, et al. Relevance of B19 markers in serum samples for a diagnosis of parvovirus B19-correlated diseases. *J Med Virol* 2003;71:135–9.
- Tabor E, Epstein JS. NAT screening of blood and plasma donations: evolution of technology and regulatory policy. *Transfusion* 2002;42:1230–7.
- Rasmussen R. Quantification on the LightCycler. In: Meuer S, Wittwer C, Nakagawara K, eds. *Rapid cycle real-time PCR: methods and application*. Berlin: Springer, 2001:21–34.
- Mackay IM, Arden KE, Nitsche A. Real-time PCR in virology. *Nucleic Acids Res* 2002;30:1292–305.
- Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 2001;29:2002–7.
- Manaresi E, Gallinella G, Zuffi E, Bonvicini F, Zerbini M, Musiani M. Diagnosis and quantitative evaluation of parvovirus B19 infections by real-time PCR in the clinical laboratory. *J Med Virol* 2002;74:275–81.
- Harder TC, Hufnagel M, Zahn K, Beutel K, Schmitt H-J, Ullmann U, et al. New LightCycler PCR for rapid and sensitive quantification of parvovirus B19 DNA guides therapeutic decision-making in relapsing infections. *J Clin Microbiol* 2001;39:4413–9.
- Aberham C, Pendl C, Gross P, Zerlauth G, Gessner M. A quantitative, internally controlled real-time PCR Assay for the detection of parvovirus B19 DNA. *J Virol Methods* 2001;92:183–91.
- Gruber F, Falkner FG, Dorner F, Hammerle T. Quantitation of viral DNA by real-time PCR applying duplex amplification, internal standardization, and two-color fluorescence detection. *Appl Environ Microbiol* 2001;67:2837–9.
- International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). Guidelines for validation of analytical procedures (Q2A) and methodology (Q2B). <http://www.ich.org>. (Q2A, October 1995; Q2B, November 1996).
- Saldanha J, Lelie N, Yu MW, Heath A. Establishment of the first World Health Organization International Standard for human parvovirus B19 DNA nucleic acid amplification techniques. *Vox Sang* 2002;82:24–31.
- Gallinella G, Moretti E, Nardi G, Zuffi E, Bonvicini F, Bucci E, et al. Analysis of B19 virus contamination in plasma pools for manufacturing, by using a competitive polymerase chain reaction assay. *Vox Sang* 2002;83:324–31.
- Siegl S, Cassinotti P. Presence and significance of parvovirus B19 in blood and blood products. *Biologicals* 1998;26:89–94.

DOI: 10.1373/clinchem.2003.027292

Reevaluation of the Power of Error Detection of Westgard Multirules, *Graham R.D. Jones* (Department of Chemical Pathology, St. Vincent's Hospital, Darlinghurst, NSW, 2010, Australia; fax 61-2-8382-2489, e-mail giones@stvincents.com.au)

The multirules system for optimizing assay error detection originally developed by James Westgard and others is now widely used in clinical pathology laboratories (1). The power of various combinations of rules to detect changes in assay performance can be assessed by use of power function charts (2), some of which are freely available on the Westgard QC website (3). In this report I reassess the power of error detection of rules that require data from more than one quality-control (QC) run.

Variables that can be adjusted in setting a QC protocol, and modeled in power function charts, include the choice

of rules, the number of QC samples in a run (n), and the number of runs over which data are to be assessed (R). In the development of power function charts, other variables must also be clearly defined so that the charts represent the actual practice in the laboratory.

Westgard's rules can be divided into within-run rules, where the data for decision-making are available from within one QC run, and cross-run rules, where data from more than one QC run are required ($R > 1$). Using approved terminology (4), examples of within-run rules for $n = 2$ include 1_{3s} and 2_{2s} , and examples of cross-run rules for $n = 2$ include 4_{1s} and 10_x . Within-run QC rules are clearly preferable to cross-run rules because they will allow detection of changes in assay performance as soon as possible after the change. Cross-run rules are commonly added to within-run rules with the aim of gaining additional power of error detection to meet quality specifications.

In this report I use a spreadsheet application to develop power function charts with the aim of assessing the power of error detection of combinations of rules when cross-run rules are included. Specifically, I review the assumptions made in generating currently available power function charts (3) and reassess the power of error detection for these combinations of rules.

Power function charts were produced with an application developed in Microsoft Excel (Microsoft Corporation). QC results were simulated with the Excel random number generator function with a gaussian distribution. Changes in assay bias were modeled by adding various constants to the output of the random number generator. QC rules and combinations of rules were evaluated by the frequency with which the rules were triggered for each change in bias. For $n = 2$, 1_{3s} , 2_{2s} , 4_{1s} and 10_x and for $n = 4$, 1_{3s} , 2_{2s} , 4_{1s} and 8_x were evaluated for their ability to detect changes in assay bias. For $n = 2$, the 2_{2s} , 4_{1s} and 10_x rules were all applied across both sets of QC data. For $n = 4$, 2_{2s} was applied only to the same QC material and 4_{1s} and 8_x across the two materials, the same criteria as used on the Westgard website. A 1_{2s} "warning" rule was not used in the models. Data representing 1000 QC runs were used to evaluate each set of rules. I did not investigate the effect of increases in imprecision because the cross-run rules under consideration are designed to detect changes in bias rather than changes in precision. Similarly, I also did not include the R_{4s} rule in the evaluation because only changes in bias were considered. Further information on the spreadsheet model can be found in the Data Supplement that accompanies the online version of this Technical Brief at <http://www.clinchem.org/content/vol50/issue4/>. No approval was sought from the institutional ethics committee because no patient samples or data derived from patient materials were involved in the study.

The spreadsheet application was used initially to reproduce the power function chart for multirules with $n = 2$ (Fig. 1A) as shown on the Westgard website. The legend shows the number of QC runs (R) for each of the sets of rules as appears on the website diagram. When I was

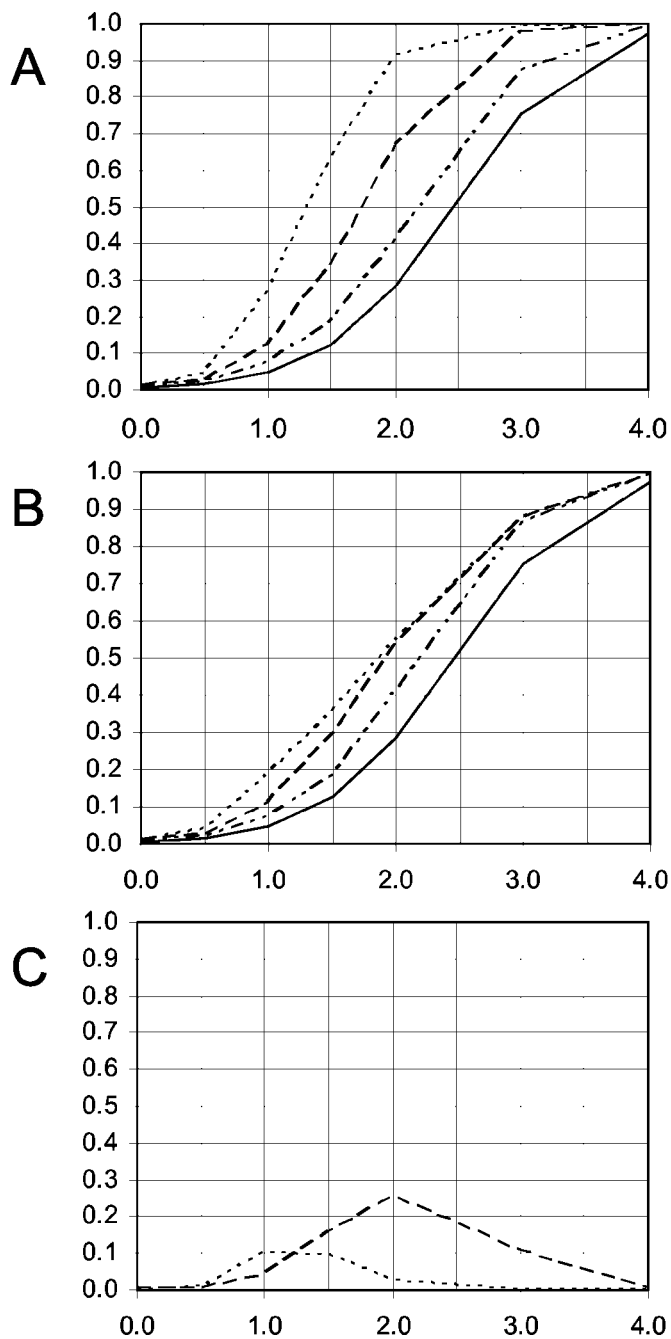


Fig. 1. Power function charts derived as described in the text, using $n = 2$.

(A), recreation of the power function charts as described on the Westgard website. The values of R are taken from the website. (·····), $1_{3s}/2_{2s}/1_{4s}/10_x$ ($R = 5$); (---), $1_{3s}/2_{2s}/4_{1s}$ ($R = 2$); (- · - · -), $1_{3s}/2_{2s}$ ($R = 1$); (—), 1_{3s} ($R = 1$). (B), the same rules as in A except that cross-run rules add to error detection only on those occasions when a change in bias has not been detected in the QC runs required for application of the cross-run rules. For cross-run rules, error detection is shown for the first run in which each rule has sufficient data to operate. (·····), $1_{3s}/2_{2s}/4_{1s}/10_x$; (---), $1_{3s}/2_{2s}/4_{1s}$; (- · - · -), $1_{3s}/2_{2s}$ ($R = 1$); (—), 1_{3s} ($R = 1$). (C), the power of error detection of the individual cross-run rules for errors undetected by shorter-run rules. (·····), 10_x ; (---), 4_{1s} .

choosing the parameters required to replicate the original data, it became clear that the meaning of R is somewhat complex when applied to multirules that include one or more cross-run rules. For example, a $1_{3s}/2_{2s}/4_{1s}$ combination of rules is listed as $R = 2$, when this value of R applies only to the 4_{1s} rule because the within-run rules are applied only to the latest set of QC data ($R = 1$). Similarly, when the 10_x rule is added, the data line indicated with $R = 5$ represents different data collections over 1, 2, and 5 QC runs amalgamated in the same line. An important finding was that the available charts for cross-run rules do not consider the probability that a change in assay performance may have been detected by other rules during the period of data accumulation.

I believe that the version of run definitions implied in current power function charts does not reflect actual practice. A QC run can either detect a change in assay performance that has occurred since the previous QC run or pick up a change that has persisted for longer but has not previously been identified. Cross-run rules have the ability to act only in the latter manner; however, their power of error detection should be limited to changes that have not been detected by other rules that have been run during the period of data accumulation. Fig. 1B shows the effect of adjusting the model to take this limitation into account. Specifically, the 4_{1s} rule was triggered only if a shift had been present for two QC runs and the within-run rules did not detect the shift during the previous QC run. Similarly, the 10_x rule was triggered only if a shift had been present for five QC runs and the shift had not been identified by the within-run rules or the 4_{1s} rule in the previous four runs. It can be seen that the power of error detection for the cross-run rules changes markedly with this alteration to the model, showing reduced probability of error detection for all changes in bias between 0.5 and 4.0 times the SD of the assay. A full description of this power function chart would be the probability of detecting a new bias with the within-run rules together with the probability of identifying a previously undetected preexisting bias with the cross-run rules.

One method of summarizing the data in power function charts is the shift that can be detected with 90% certainty. This parameter is a marker of the change in assay bias that the QC protocol can reliably detect. As seen in Fig. 1A, the change in bias in an assay that Westgard multirules ($1_{3s}/2_{2s}/4_{1s}/10_x$; $n = 2$; $R = 5$) can apparently detect with 90% certainty is ~ 2.0 times the SD of the assay. When the model is adjusted so that previously detected shifts are removed from the data, the assay bias that can be detected with the same degree of certainty is increased from 2.0 to 3.3 times the assay SD, no different from the change that can be detected by the within-run rules in use. Fig. 1C shows the additional error detection provided individually by the 4_{1s} rule beyond that provided by the within-run rules, and 10_x beyond that of the shorter-run rules. It can be seen that the additional error detection of these cross-run rules peaks and then decreases. This decrease is caused by the increased chances that a large change in

assay bias will be detected by within-run rules before the application of the cross-run rules.

Data were also modeled for $n = 4$, and the error detection for full multirules was lower than described on the Westgard site for all changes in bias from 1.0 to 3.0 times the SD when previously detected shifts were removed from the 8_x rule (data not shown). The change in bias that can be detected with 90% certainty decreases from ~ 1.8 times the SD, as shown on the Westgard website, to ~ 2.5 times the SD, the same change in bias that can be detected with the use of within-run rules only.

The above data suggest that the true power of error detection of cross-run rules is markedly less than has previously been claimed. It can be seen that the action of the cross-run rules conforms with previous descriptions as warning rules for detecting smaller shifts in assay bias than are easily detectable by common within-run rules. However, these rules cannot detect these minor shifts with any certainty. For example, the 4_{1s} rule will detect assay changes a maximum of 25% of the occasions an otherwise unidentified change occurs. However, the present study was limited to evaluation of changes in bias at a single point in time. It is likely that cross-run rules may be more useful in detecting a gradual change in bias over time because such changes may be less likely to be detected previously by within-run rules.

It is important that laboratory scientists be aware of the performance characteristics of the QC protocols in use in their laboratories. Reliance on within-run rules is preferable to rules requiring data from multiple runs because errors can be detected sooner after they have occurred. However, if cross-run rules are to be used they are of value only if they add to the error detection generated by the rules that have already evaluated the data that are required for them to fire. By excluding previously detected changes in assay bias, I have demonstrated that the cross-run rules 10_x and 4_{1s} with $n = 2$ and 8_x with $n = 4$ have a lower probability of error detection than is shown in the data on the Westgard website and do not add to error detection at the 90% confidence level derived by use of within-run rules.

References

1. Westgard JO, Barry PL, Hunt MR, Groth T, Burnett RW, Hainline A, et al. A multi-rule Shewart chart for quality control in clinical chemistry. *Clin Chem* 1981;27:493–501.
2. Westgard JO. Westgard QC web site. <http://www.westgard.com> (Accessed July 2003)
3. Westgard JO, Groth T. Power function for statistical control rules. *Clin Chem* 1979;25:863–9.
4. National Committee for Clinical Laboratory Standards. Statistical quality control for quantitative measurements: principles and definitions; approved guideline—second edition. NCCLS Guideline C24–A2. Wayne, PA: NCCLS, 1999.

Improved Detection of Serum Estradiol after Sample Extraction Procedure, Anand S. Dighe and Patrick M. Sluss* (Reproductive Endocrine Unit, Department of Medicine, Massachusetts General Hospital, Boston MA 02114; * author for correspondence: fax 617-726-9330, e-mail psluss@partners.org)

The direct measurement of serum estradiol offers advantages over the more traditional sample extraction and immunoassay methods with respect to technical ease of the assay and turnaround time. However, direct assays have not been demonstrated sufficiently sensitive for some important clinical applications, including the measurement of circulating estradiol in women during the early follicular phase of their menstrual cycle, in postmenopausal women, and in men. We report here a rapid sample extraction method developed to improve the detection limit of automated estradiol assays and thus enhance the potential usefulness of these rapid and nonisotopic systems for measuring low concentrations of serum estradiol.

Serum estradiol concentrations have long been used in conjunction with ultrasonography and other clinical indices to monitor ovarian stimulation by exogenous gonadotropin therapy during in vitro fertilization (IVF) procedures (1–3). With the recent use of gonadotropin-releasing hormone agonists, serum estradiol measurement has also been used to assess the adequacy of ovarian suppression before exogenous gonadotropin stimulation. In addition, it has become increasingly important to measure the low initial concentrations of serum estradiol (<150 ng/L) associated with the earlier stages of ovarian follicular stimulation (4, 5). Clinically important measurements of low estradiol concentrations are also required in assessing ovarian activity in perimenopausal women or women experiencing premature ovarian failure and in clinical investigations of the hypothalamic-pituitary-gonadal axis in men (6–8).

Methods for the direct measurement of serum estradiol are replacing the more traditional RIA methods that use preassay sample extraction followed by RIA with antecedent column chromatography. Direct methods are advantageous because they are rapid and less technically demanding. However, direct assays are often inaccurate, particularly for the measurement of relatively low concentrations of serum estradiol (9–11). The presence of sex-hormone-binding globulin in serum and the difficulty of obtaining adequate buffers have been associated with both interference and matrix effects, which further contribute to the inaccuracy of direct RIAs of estradiol (12, 13). Such effects probably underlie much of the high interinstitutional variability that has been reported with serum estradiol measurements (14). Recently, nonisotopic methods for the accurate direct measurement of serum estradiol have been developed. The Abbott IMx system used here is one such method that is completely automated and allows measurement of serum estradiol in <1 h.

The estradiol for these studies was purchased from