Quantitative Inference of Commutability for Clinical Viral Load Testing

L. Tang,^a Y. Su,^a S. Pounds,^a R. T. Hayden^b

^aDepartment of Biostatistics, St. Jude Children's Research Hospital, Memphis, Tennessee, USA

^bDepartment of Pathology, St. Jude Children's Research Hospital, Memphis, Tennessee, USA

KEYWORDS commutability, EBV, quantitative PCR, viral load

Viral load testing is integral to the clinical care of immunocompromised patients, guiding initiation of preemptive therapy and therapeutic endpoints, while allowing assessment of therapeutic response. Over time, numerous assays have been developed for clinical testing; however, a high degree of interassay result variability remains an issue. Although development of international quantitative standards has helped address such variability, in some cases suboptimal commutability has reduced the effectiveness of such reference materials. Commutability has been assessed via prediction interval approaches (1), correspondence analysis (2), and other similar approaches based on assessing whether reference materials tested by two or more assays show the same quantitative relationship as human specimens tested by the same methods. Because the traditional approaches to assessment of commutability are subject to several limitations, such as lacking a quantitative measure of commutability and failing to directly connect to clinical utility or formal statistical testing, we recently proposed a new quantitative metric of commutability. This metric, termed the “deviation-from-ideal” (DFI) metric (3), is based on the use of digital PCR as a reference standard (4). DFI offers an absolute (rather than relative) metric of commutability, takes variability of measures into account, provides the possibility of normalizing the commutability assessment to clinically meaningful units of measure, and should allow linkage of commutability with clinical predictive value of an assay system. One major advantage of DFI lies in the associated intuitive graphic illustration (see Fig. 2 in reference 3), which shows the driving force of a large DFI via three-color graphs. (The use of the term “assay system” here refers to all components of sample preparation, processing, and analysis, any aspect of which might affect commutability in the context of a given instance of reference material.)

A major limitation of the proposed DFI approach was that there was no statistical testing procedure proposed, which hinders making formal comparisons between assays and connecting with clinical assessment. Here, we provide methods to complete formal statistical inferences on the DFI metric, allowing quantitative comparisons of the commutability of reference materials among various assay systems and statistical evaluations of whether a given reference material/assay system shows sufficient commutability to reach a prespecified clinical utility threshold. To achieve these aims, bootstrapping (5) is used to compute 95% confidence intervals (CI) (6) of DFI values. Additional statistical details are available as a supplemental document (see Text S1 in the supplemental material) for those who are more interested in mathematical details, as well as in a forthcoming manuscript. Each assay's 95% CI of DFI is plotted with a reference line representing a prespecified clinical threshold. If the entire CI exceeds this threshold, the reference material is defined as noncommutable in the context of the tested assay system and will yield insufficient agreement or insufficiently comparable results. If the entire CI is below the threshold, there is adequate commutability. If the

Accepted manuscript posted online 28
March 2018

Citation Tang L, Su Y, Pounds S, Hayden RT.
2018. Quantitative inference of commutability
for clinical viral load testing. *J Clin Microbiol*
56:e00146-18. <https://doi.org/10.1128/JCM.00146-18>.

Editor Yi-Wei Tang, Memorial Sloan Kettering
Cancer Center

Copyright © 2018 American Society for
Microbiology. All Rights Reserved.

Address correspondence to R. T. Hayden,
randall.hayden@stjude.org.

FIG 1 DFI values with 95% bootstrap confidence intervals determined for four real-time EBV PCRs (ELITech, Focus, Luminex, and LDT). Technical details about computing 95% confidence interval of DFI values can be found in Text S1 and S3.

CI overlaps the threshold, then there are insufficient data to make a firm conclusion regarding the adequacy of the commutability. The CIs from several assays can be plotted alongside one another to determine which combination of assay system and reference material is most commutable.

In order to illustrate the framework, we revisited the recently published Epstein-Barr virus (EBV) commutability study of the WHO quantitative standard (3). Well-documented SAS and R code libraries for this approach are provided in the supplemental material, and they can be readily used by investigators without their needing to be fully versed in the method's technical details. A description of R functions and SAS macros is provided in Text S2, with programs provided in Text S3 and Text S4. Sample data sets are also included in Text S5 (R) and Text 6 (SAS), which can be directly input into the provided programs included to generate Fig. 1. Using these methods, the 95% CIs of DFI were computed for the previously tested assays (ELITech, Focus, Luminex, and LDT). These CIs were then plotted and compared against the two hypothetical clinical interassay agreement thresholds of 0.3 and 0.6 $\log_{10}$ IU/ml (Fig. 1). Although a threshold of 0.5 $\log_{10}$ IU/ml is more common in practice, here we selected 0.3 and 0.6 as thresholds to illustrate how the proposal works in a general setting. In other words, if the discrepancy between measures of human specimens and those of reference material is over the threshold in a given assay, we conclude that the reference material, when used in the evaluated assay system, might not bring about the desired degree of interassay agreement and therefore might not be clinically useful. Figure 1 indicates that, using 0.3 as the threshold, no assay was commutable enough, possibly because the prespecified threshold was not practical. The results also suggested that the choice of a threshold reflects clinical tolerance. Prespecifying a threshold of 0.6, the LDT method was the most commutable in this trial, with the lowest DFI value (0.29 $\log_{10}$ IU/ml) and a 95% CI of 0.15 to 0.43 at below 0.6 $\log_{10}$ IU/ml. The Focus assay had an intermediate DFI value (0.51 $\log_{10}$ IU/ml) and a 95% CI of 0.43 to 0.58, which was also acceptable. The ELITech and Luminex methods showed significantly poorer commutability, with DFI values of 0.91 (95% CI, 0.85 to 0.98) and 0.92 (95% CI, 0.83 to 1.01), respectively. It should be noted although the LDT method had the lowest DFI value, its variation was large. An advantage of the confidence interval-based approach is that it takes variability into account to make valid inferences. In other words, the error bar of a DFI value determines its confidence interval width, which is crucial in making statistical comparisons or inferences.

The proposed framework represents recognition of the fact that a given instance of reference material may behave differently across different assay systems, making it

possible to compare the behaviors of the reference material across such systems. It allows quantitative evaluation of commutability, given a desired level of interassay agreement, adding a critical link between the commutability of a standard material-assay system and that system's practical and clinically meaningful utility. In practice, we suggest the use of both Fig. 1 and the graphic illustration of DFI value (see Fig. 2 in reference 3) to gain a comprehensive evaluation of commutability of the given material in an assay system. One limitation should be noted: LDT and digital PCR (dPCR) in the example cited utilized the same primers and probes, potentially explaining the fact that LDT had the lowest DFI value. For illustrative purposes, it was not felt that this hindered readers' understanding of the principles of the proposed approach.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/JCM.00146-18>.

SUPPLEMENTAL FILE 1, PDF file, 0.3 MB.

SUPPLEMENTAL FILE 2, PDF file, 0.8 MB.

SUPPLEMENTAL FILE 3, PDF file, 0.2 MB.

SUPPLEMENTAL FILE 4, PDF file, 0.2 MB.

SUPPLEMENTAL FILE 5, PDF file, 0.1 MB.

SUPPLEMENTAL FILE 6, PDF file, 0.1 MB.

ACKNOWLEDGMENTS

R.T.H. has served as an advisory board member for Roche Molecular and for Abbott Molecular.

REFERENCES

1. Vesper HW, Miller WG, Myers GL. 2007. Reference materials and commutability. *Clin Biochem Rev* 28:139–147.
2. Bretaudiere JP, Dumont G, Rej R, Bailly M. 1981. Suitability of control materials: general principles and methods of investigation. *Clin Chem* 27:798–805.
3. Tang L, Sun Y, Buelow D, Gu Z, Caliendo AM, Pounds S, Hayden RT. 2016. Quantitative assessment of commutability for clinical viral load testing using a digital PCR-based reference standard. *J Clin Microbiol* 54:1616–1623. <https://doi.org/10.1128/JCM.03346-15>.
4. Sedlak RS, Jerome KR. 2013. Viral diagnostics in the era of digital PCR. *Diagn Microbiol Infect Dis* 75:1–4. <https://doi.org/10.1016/j.diagmicrobio.2012.10.009>.
5. Efron B. 1979. Bootstrap methods: another look at the jackknife. *Ann Statist* 7:1–26. <https://doi.org/10.1214/aos/1176344552>.
6. Efron B. 1987. Better bootstrap confidence intervals. *J Am Stat Assoc* 82:171–200. <https://doi.org/10.2307/2289144>.