The Range of Uncertainty: a Tool for Efficiently Addressing Result Variability Around Clinical Decision Points for Hepatitis C Response-guided Therapy

Paul Jülicher and Jens Dhein

Abbott GmbH and Co KG, Wiesbaden, Germany.

ABSTRACT: Despite efforts to standardize molecular diagnostic tests, performance differences are not rare. Laboratories are challenged in situations where treatment rules have been established using a reference assay that is different from the assays being used in daily practice. Assessing the viral load status of patients with chronic hepatitis C under modern triple therapy is a recognized example. We demonstrate the use of the range of uncertainty as an easy and efficient tool that can provide information on the certainty of a test result's interpretation in the context of making hepatitis C virus treatment decisions.

KEYWORDS: in vitro diagnostic tests, standardization, precision, hepatitis C, viral load, response guided therapy

CORRESPONDENCE: paul.juelicher@abbott.com

Introduction

Diagnostic tests that are used as reference standards do not provide error-free results. The level of error contributed by a reference standard (reference assay) can either positively or negatively bias the sensitivity, specificity, and the overall accuracy of a comparison test.[1] Despite various efforts for standardizing assays, performance differences are not particularly rare with respect to molecular diagnostic assays.[2,3] The pertinent question is how the analytical differences found with different diagnostic assays impact patient management or, ultimately, patient outcomes. The diagnostic risk, which is the uncertainty about the true health status of a patient, might become unpredictable in situations where treatment rules have been established based on test results obtained with a reference assay that is different from those assays being used in daily practice.[4] The reference assay does not necessarily reflect the true disease status, but a reference status that can be used to evaluate and define treatment rules. This scenario does not appear to be rare in cases where advanced treatment algorithms require regular monitoring of a biomarker, so as to balance patient benefit and harm, and to also avoid the unnecessary use of resources.

Modern triple therapy algorithms for patients with chronic hepatitis C virus (HCV) provide perfect examples of this scenario.[5] Combination therapy with pegylated interferon alpha (PEG-IFN) and ribavirin serves as the basis of treatment in patients infected with HCV. Following the recent approval of the HCV NS3/4 protease inhibitors, boceprevir, and telaprevir in the US and Europe, triple therapy together with PEG-IFN and ribavirin has become the new standard of care in many countries for HCV genotype 1-infected patients.[6] A key parameter for the management of individualized antiviral therapy is HCV ribonucleic acid (RNA). The HCV genotype, in combination with HCV RNA concentrations based on single HCV RNA measurements at baseline and on treatment, is used in order to terminate treatment early in virologic nonresponders, and to define the treatment duration for patients responding to antiviral therapy (response-guided therapy).[5] The clinical decision points are located at the low end of the linear dynamic range of quantitative HCV RNA assays (for example, at 1,000 IU/mL, 100 IU/mL, and at the limit of quantification and/or the limit of detection with HCV RNA "detected" versus HCV RNA "undetected"). The

rules for response-guided therapy were established using a single assay for the measurement of HCV RNA.[7] This assay does not appear to be widely used in clinical practice due to the availability of newer commercial assays with a higher level of automation. Differences in the analytical performance between the HCV RNA tests available on the market have been reported previously, with precision, accuracy, and sensitivity being the main performance indicators.[8,9] However, little is known about the impact of the analytical differences in HCV RNA assay performance on response-guided therapy and sustained virological response assessments at the low end of the measurement range.

Three comparative studies evaluated low HCV RNA viremia using various commercially available HCV RNA assays.[10–12] Although the real-time polymerase chain reaction (PCR) assays used in these studies were standardized against the World Health Organization International Standard for Hepatitis C Virus RNA, and given that they all have broad linear measurement ranges and low detection limits, head-to-head comparison studies revealed that there were differences in precision, quantitation, and detectability between the methods used.[10–12] Different clinical decisions would be made from single HCV RNA determinations yielding "HCV RNA detected" by one method and "HCV RNA not detected" by another method.[10,11] Low HCV viremic samples were tested in ten independent runs using one replicate each across five different assays.[11] At HCV RNA concentrations around 100 IU/mL and 1,000 IU/mL, one study showed that the quantification between the assays and the interrun variation expressed as the percentage coefficient of variation was quite different.[11] Accordingly, HCV RNA could be measured above or below the 100 IU/mL or 1,000 IU/mL clinical decision point, depending on the assay used.[11] Taken together, these results suggested that there were method-related implications for clinical decisions based on response-guided therapy and sustained virological response assessments.

Since current treatment regimens frequently cause severe adverse effects, stopping treatments early is supposed to avoid toxicity in patients who are unlikely to achieve treatment success. On the other hand, an effective selection of patients who may potentially respond to treatment could help to efficiently deploy health care resources to patients with a high chance for viral clearance and cure.[13]

It has been recommended that stopping rules should be applied with particular caution when the HCV RNA values fall within the assay's variability range of the decision thresholds.[7] Given the fact that these HCV drugs have been approved using specific rules that were based on one specific HCV RNA test system, it therefore remains a challenge how laboratories could assist physicians with health status information that are highly confident while using different HCV RNA test systems.

In this short communication, we applied a conceptual approach in addressing the range of uncertainty (RoU) in a hepatitis C monitoring scenario, and we presented the formula used to calculate assay-specific RoU intervals with the objective of providing additional information on the certainty of a test result's interpretation.

Methods

We assumed six hypothetical assays with different analytical performances (Table 1). The clinical cutoff (CO) was set to 100 IU/mL, using a drug-specific therapy-stopping rule as an example.[5] The assay used for establishing the stopping rules was assigned the name "Reference." The hypothetical results of the other assays were interpreted in relation to the "Reference." The statistical analysis for evaluating a RoU was based on precision and the mean differences of different test systems compared to the "Reference" assay results. The calculations of the limits of the range of uncertainty (RL) were derived from a general confidence limit formula that assumed a normal distribution and that approached the clinical decision point from lower or higher viral loads.[14] A significance level of 0.05 (reflecting a 95% confidence limit) was chosen for all calculations.

$$RL_{Low} = \frac{CO_A}{1 + (z \times \frac{CV\%}{\sqrt{n}})}, \quad RL_{Up} = \frac{CO_A}{1 - (z \times \frac{CV\%}{\sqrt{n}})} \quad (1)$$

with z being the value of the standard normal distribution at the respective probability level (here, 1.645), and n being the number of replicates. CV% reflects the assay-specific coefficient of variation at the respective cutoff level. CO_A is the assay-specific equivalent of the CO, which considers the accuracy of the assay relative to the reference (mean log difference, MD):

Table 1. Test characteristics of hypothetical HCV RNA assays.

	TEST CHARACTERIZATION	MEAN LOG DIFFERENCE TO REFERENCE (MD)	CV%
Reference	Test used to evaluate treatment rules	0	20%
Test A	Test quantifies similar to reference; high imprecision	0	30%
Test B	Test quantifies higher than reference; moderate imprecision	0.2	20%
Test C	Test quantifies higher than reference; high imprecision	0.2	30%
Test D	Test quantifies lower than reference; low imprecision	−0.2	10%
Test E	Test quantifies lower than reference; high imprecision	−0.2	30%

$$CO_A = 10^{[Log(CO)+MD]} \qquad (2)$$

Results and Discussion

Wachtel et al[15] introduced a "range of uncertainty," which reflects a diagnostic overlap zone between populations with and without a suspected disease. In order to reduce the risk of false positive and false negative results, the authors suggested two cutoff points beyond which the risk for false results was very low, and between which the result was not diagnostic or uncertain. The interval between these two cutoffs was called the "range of uncertainty." The concept proposed by Wachtel et al[15] implied that the use of sequential testing only occur for patients whose results are in the RoU. The authors demonstrated the usefulness of the concept in the workup of patients who may not require a bone marrow examination for the diagnosis of an iron deficiency. They also raised the problem of defining an appropriate RoU. Cutoffs for defining the RoU were based on distributions for iron-deficient and noniron-deficient cases, and an arbitrarily chosen sensitivity and specificity of 90%, respectively. The RoU concept identified three groups of patients: highest probability of having the disease; highest probability of not having the disease; and uncertain cases that require further workup.

The idea of the RoU could be transferred from a lack of discrimination between populations to several other diagnostic scenarios showing uncertainty or variation around a cutoff point which is, to our best knowledge, first shown in this paper. Here, we applied the RoU concept to the stopping rules of hypothetical HCV RNA assay performances, and we provided a formula to calculate a RoU based on assay-specific analytical performance information (imprecision and accuracy, relative to a reference).

The RoU comprises an interval determined by an upper limit (UL) and a lower limit (LL) in relation to the CO. Both limits refer to the value, at which point a test result does not cross the clinical decision point with a confidence level of 95%. Therefore, any test result that falls within the RoU does not provide sufficient confidence to guide a course of action (if 95% confidence is assumed to be appropriate). Following this concept, it has been recommended that those results that fall into the RoU require further work-up to evaluate the true response according to treatment rules.[15] The principles of the RoU approach are illustrated in Figure 1. The width and position of the RoU are driven by test system imprecision and quantification differences relative to the reference. While assay-specific imprecision is a performance parameter that is typically assessed during the product development of an in-vitro diagnostic assay, quantification differences could only be taken from direct comparison studies of the reference. Discrepancies in viral load results between different molecular diagnostic assays might be driven by the use of different international standards, different nucleic acid extraction methods, different target regions for PCR primers and
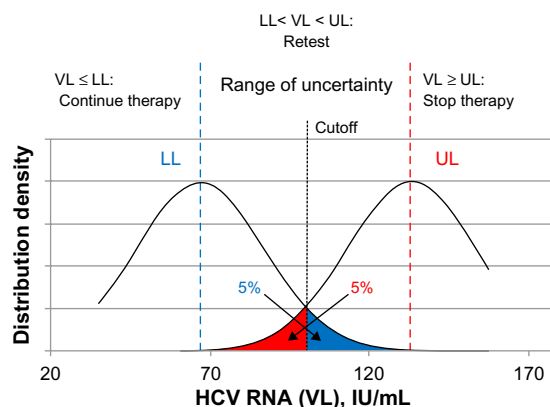

Figure 1. Schematic illustration of the range of uncertainty concept.
Notes: LL is defined by the highest value below the cutoff that does not statistically cross the cutoff with an error rate of less than 5%. UL is defined by the lowest value above the cutoff that does not statistically cross the cutoff with an error rate of less than 5%.
Abbreviations: LL, lower limit of the RoU; UL, upper limit of the RoU.

probes, different cycling conditions for target amplification, and different calibration methods.

The RoU for different assays in relation to the stopping rules evaluated with the reference system is shown in Table 2 and Figure 2. The reference points (indicated by black diamonds) reflect assay-specific quantification differences when compared to the reference assay value. The bars in the chart show three categories for viral load test interpretation: results that fall into the blue areas provide a high probability that the viral load status is truly below the clinical threshold. Results observed in red areas are highly indicative of a viral load status that is above the threshold. If results fall into the grey zone, which represents the assay-specific RuO, the true viral load status remains unclear given the fact that random error might have caused a result to cross the threshold in both directions. Although a single measurement might be affected by random error (imprecision) of the test system, as long as the observed result falls into either the blue or red area, the position of the result relative to the cutoff (hence, the viral load status) is not misleading. Consequently, and according to response-guided therapy schemes, a clinical decision to continue or to stop

Table 2. Range of uncertainty intervals for different assay characteristics, as described in Table 1.

	REFERENCE VALUE (CO$_A$)	LL	UL
Reference	100	75	149
Test A	100	67	197
Test B	158	119	236
Test C	158	106	313
Test D	63	54	76
Test E	63	42	125

Note: Reference values (CO$_A$) refer to the assay-specific equivalent to the cutoff established with the "Reference."
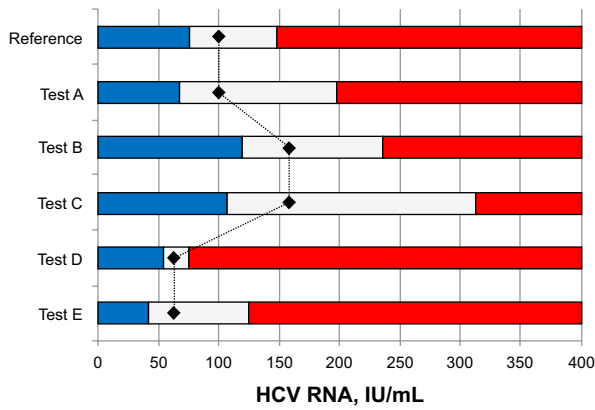**Abbreviations:** LL, lower limit; UL, upper limit; all numbers in IU/mL.

Figure 2. Range of uncertainty for different hypothetical assays in relation to a reference.

Notes: Blue and red areas represent areas of ≥95% confidence that do not cross the decision threshold of the 100 IU/mL reference assay analogue. Diamonds refer to the assay-specific values that are equivalent to the reference threshold value (CO_A). Grey areas refer to the RoU.

treatment would be highly appropriate for results that fall in the blue or red area, respectively.

According to Wachtel et al[15] the RoU is used to select patients that need to undergo subsequent investigation (Fig. 1). Retesting or additional sampling at another time point have been suggested to improve the assessment of treatment response.[12] These could also be regarded as methods of choice for assessing unclear results in the grey zone or RoU. The RoU formula shown in this paper indicates that the UL and LL of the RoU are actually affected by the number of replicates. If precision and quantification differences to the reference are known, a laboratory could easily adjust the RoU to the requested level of confidence.

Accuracy and precision systematically affect diagnostic test results. If not factored in, both could lead to significantly misleading interpretations in response-guided therapy schemes (Fig. 2, "Test A–E" versus "Reference"): a result of 110 IU/mL might be interpreted as a viral load status above the clinical threshold. Observing this result with "Test B," the RoU, however, points to a high probability that the viral load truly falls below the threshold. A "Test D" result of 110 IU/mL would accurately reflect a viral load status that falls above the established cutoff, whereas a "Test A" or "Test E" result of 110 IU/mL would fall into the RoU, and the patient or sample should undergo further investigation. These scenarios exemplify potentially controversial clinical decisions that might be made based on the results obtained from different test systems.

Evaluating the RoU with its limits below and above the clinical threshold would introduce two additional assay-specific decision points. Results below the assay-specific LL would represent a health status that falls below the reference threshold with high confidence, indicating a therapy responder in the current example. Results above the UL would, on the

other hand, reflect a viral load that falls above the reference threshold, indicating a nonresponder status. The RoU concept would mitigate uncertainty issues with different test systems in comparison to a reference. With known precision data and known quantification differences that are compared to a reference assay used for defining clinical decision rules, the RoU would estimate a laboratory's specific probability of supporting appropriate decisions and mitigating the risk of inappropriate decisions. Thus, the RoU concept might be used as a practical guide for laboratories in the context of defined decision rules.

The clinical consequences ultimately depend on the number of affected patients. Apparently, narrowing the RoU (for example, by improving precision, "Test B" versus "Test C" or "Test D" versus "Test E" in Figure 2) would reduce the number of potentially affected patients, thus reducing the need for subsequent action or sequential testing, and improving diagnostic and therapeutic yield.[16] Using viral load information at respective time points, a future model could simulate clinical and economic consequences, and this may help create a hypothesis for a future study validating the RoU concept. The RoU may also be used to define clinical specifications for diagnostic tests and facilitate the comparability of results across laboratories.

Conclusion

In areas where specific test systems have been used to develop treatment rules, differences in assay performance in terms of precision or quantification could result in an additional risk of indicating an inaccurate patient condition (for example, a viral load status). Since inappropriate clinical decisions based on single measurements could result in dramatic consequences or harm for patients under HCV treatment, test results should be reported only with a high level of confidence. Assay-specific RoU could be calculated in each molecular laboratory by using simple formula and performance information for those tests actually used in the laboratory. The RoU described in this paper graphically exhibits assay-specific precision, deviation to the reference assay used for evaluating the treatment rules, and confidence limits for guiding the appropriate or preferred course of clinical action. Given the established treatment guidelines for specific drugs, and until assay-specific decision points have been fully evaluated, the RoU might be an easy and efficient tool to support treatment decisions, while providing additional information regarding the certainty of a test result's interpretation.

Author Contributions

Developed the statistical tools: PJ. Wrote the first draft of the manuscript: PJ, JD. Agree with manuscript results and

conclusions: PJ, JD. Made critical revisions: PJ, JD. All authors reviewed and approved of the final manuscript.

DISCLOSURES AND ETHICS

REFERENCES

1. Reitsma JB, Rutjes AWS, Khan KS, Coomarasamy A, Bossuyt PM. A review of solutions for diagnostic accuracy studies with an imperfect or missing reference standard. *J Clin Epidemiol*. 2009;62(8):797–806.
2. Holden MJ, Madej RM, Minor P, Kalman LV. Molecular diagnostics: harmonization through reference materials, documentary standards and proficiency testing. *Expert Rev Mol Diagn*. 2011;11(7):741–55.
3. Panteghini M. Traceability as a unique tool to improve standardization in laboratory medicine. *Clin Biochem*. 2009;42(4–5):236–40.
4. Pauker SG, Kassirer JP. Therapeutic decision making: a cost-benefit analysis. *N Engl J Med*. 1975;293(5):229–34.
5. Yee HS, Chang MF, Pocha C, et al. Department of Veterans Affairs Hepatitis C Resource Center Program; National Hepatitis C Program Office. Update on the management and treatment of hepatitis C virus infection: recommendations from the Department of Veterans Affairs Hepatitis C Resource Center Program and the National Hepatitis C Program Office. *Am J Gastroenterol*. 2012;107(5):669–89.
6. Wedemeyer, H. On the fast track towards IFN-free therapy for hepatitis C? *Nat Rev Gastroenterol Hepatol*. 2013;10(2):76–8.
7. Jacobson IM, Marcellin P, Zeuzem S, et al. Refinement of stopping rules during treatment of hepatitis C genotype 1 infection with boceprevir and peginterferon/ribavirin. *Hepatology*. 2012;56(2):567–75.
8. Pisani G, Cristiano K, Marino F, et al. Quantification of hepatitis C virus (HCV) RNA in a multicenter study: implications for management of HCV genotype 1-infected patients. *J Clin Microbiol*. 2009;47(9):2931–6.
9. de Leuw P, Sarrazin C, Zeuzem S. How to use virological tools for the optimal management of chronic hepatitis C. *Liver Int*. 2011;31 Suppl 1:3–12.
10. Wedemeyer H, Kessler HH, Michel-Treil V, et al. The ExPECT Study: an international multicenter clinical trial evaluating low HCV RNA viremia with the COBAS® AmpliPrep/COBAS® TaqMan® HCV Test, v2.0 compared to assays used in routine clinical practice and drug trials. Poster presented at: AASLD Meeting, November 9–13, 2012; Boston, MA.
11. Braun P, Naeth G, Berger A, et al. Comparison of viral load assays for the quantification of HCV genotype 1 in the setting of low viremic samples. Poster presented at: CROI 2013; Atlanta, GA.
12. Maasoumy B, Cobb B, Bremer B, et al. A clinical trial evaluating low HCV RNA viremia in patients treated with triple therapy including Telaprevir: implications for management using different assays in clinical practice. Poster presented at: EASL Meeting 2013; April 24–28, 2013; Amsterdam, the Netherlands.
13. Liu S, Cipriano LE, Holodniy M, Owens DK, Goldhaber-Fiebert JD. New protease inhibitors for treatment of chronic hepatitis C. a cost-effectiveness analysis. *Ann Intern Med*. 2012;156(4):279–90.
14. Dekking FM, Kraaikamp C, Lopuhaä HP, Meester LE. *A modern introduction to probability and statistics*. London: Springer; 2010.
15. Wachtel TJ, Reuben DB, Fulton JP, Wartman SA. The use of the "range of uncertainty". *Med Decis Making*. 1985;5(3):325–34.
16. Ferrante di Ruffano L, Hyde CJ, McCaffery KJ, Bossuyt PM, Deeks JJ. Assessing the value of diagnostic tests: a framework for designing and evaluating trials. *BMJ*. 2012;344:e686.