Combining Qualitative and Quantitative Diagnostic Tests with no Gold Standard and with Missing Data: GBV-C Viremia as an Example

Suhong Zhang[1,*,†], Kathryn Chaloner[1,2], Jack T. Stapleton[3,4]

[1] *Department of Biostatistics, University of Iowa, Iowa City, IA*

[2] *Department of Statistics and Actuarial Science, University of Iowa, Iowa City, IA*

[3] *Department of Internal Medicine, University of Iowa, Iowa City, IA*

[4] *Iowa City VA Medical Center, Iowa City, IA*

SUMMARY

Using multiple methods to detect a virus in clinical samples, when no standard test exists, introduces several potential problems. This paper describes how discrepancies from multiple tests with missing data can be evaluated and reconciled statistically. Two novel aspects are addressed: 1) tests can be quantitative or qualitative and 2) not all tests are done on all samples. Quantitative test results are categorized into ordinal responses, with sensitivities and specificities defined by category. Bayesian latent class analysis is used to model the responses from the different tests. The model is

*Correspondence to: Suhong Zhang, Department of Biostatistics, College of Public Health, E176 General Hospital 230, 200 Hawkins Drive, Iowa City, IA 52242

†E-mail: suhong-zhang@uiowa.edu

Contract/grant sponsor: NIH/NIAID; contract/grant number: R01 058740

parameterized by the prevalence, sensitivity and specificity of each test, and probability of each test being missing.

KEY WORDS: Classification; Bayesian methods; Diagnostic tests; GB virus type C; Latent class analysis; Negative predictive value; Positive predictive value; Reverse transcription polymerase chain reaction (RT-PCR); Real time RT-PCR; Sensitivity; Specificity

1. INTRODUCTION

Diagnostic testing plays a significant role in health care and medical research. It is therefore important to evaluate the accuracies of each diagnostic test by sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). However, a gold standard, which is one hundred percent sensitive and specific, does not necessarily exist for all situations. Under this limitation, it is still important to have the best possible estimate of the sensitivity, specificity, PPV, NPV of a specific diagnostic test, and of the prevalence of the disease or condition in the population. In addition, classifying each individual based on the combination of imperfect tests is necessary for the appropriate action to be taken.

A latent class approach models the unobservable condition as a categorical latent variable. Under the assumption that the diagnostic tests are conditionally independent given the latent variable, the model is parameterized by the conditional probability distribution of each diagnostic test given the latent variable, and the probability of the condition itself (prevalence). This model readily produces estimates for the properties of each diagnostic test.

Latent class analysis was introduced in 1950 by Lazarsfeld [1], who used the technique as a tool for building typologies based on observed dichotomous variables. It was referred to as

"latent class analysis" by Kaldor and Clayton [2], and Walter and Irwig [3]. Espeland and Handelman [4], Uebersax and Grove [5], and Garrett et al. [6], among others, apply latent class model to various studies. Evans et al. [7], Gyorkos and Coupal [8], Dendukuri and Joseph [9] implement Bayesian analyses of several latent class models with prior distributions on unknown parameters.

Pepe [10] describes a discrepant resolution approach, which resolves the discrepant results between the new diagnostic test and the imperfect reference test by a resolver test. Alonzo and Pepe [11] propose a method defining a composite reference standard test on the basis of multiple imperfect reference tests. See also Kawkins et al. [12].

In this paper we extend latent class analysis to incorporate not only qualitative, but also quantitative diagnostic tests and, in addition, the absence of a test result (missingness) is taken into consideration. It is not unusual that not all the tests planned in practice are performed as the volume of available specimen may be limited. These two novel aspects are addressed in a motivating example of RT-PCR test results for GB virus type C (GBV-C).

The remainder of this paper is organized as follows. Section 2 motivates the problem of multiple tests for GBV-C. Section 3 describes the latent class model and how the Bayesian approach is incorporated in the latent class model. Section 4 introduces the extended latent class analysis that combines both qualitative and quantitative tests, with possibly missing data. Section 5 presents the results for the GBV-C study. Section 6 concludes with discussion. The complete model specification is given in Appendix A.

2. MOTIVATING EXAMPLE

Persistent co-infection with GBV-C is associated with prolonged survival among individuals also infected with HIV [13]. In different HIV-infected cohorts, GBV-C viremia has been detected in 14% to 43% of individuals [14]. Discordant results on the same sample were commonly found in the same laboratory when testing for GBV-C viremia using reverse transcription polymerase chain reaction (RT-PCR) methods employing four different primers (E2, NS3, NS5A, 5'NTR) [15], presumably related to the diversity in nucleotide sequence common to RNA viruses. Studies in other laboratories demonstrate similar discrepancies and also variability between laboratories [16, 17, 18]. There is no standard test for GBV-C RNA detection [15], and similar variability was previously seen in RT-PCR tests for hepatitis C virus [19].

RT-PCR works by first copying the RNA genome into its DNA complement (cDNA) by a method called reverse transcription. The cDNA is then copied in a process called the polymerase chain reaction (PCR)[20]. This process amplifies specific parts of a DNA molecule through the temperature mediated enzyme DNA polymerase and DNA primers [20]. Real time RT-PCR is a technique used to simultaneously amplify and quantify a specific part of a RNA molecule. The initial reverse transcription process transcribing RNA to cDNA is identical to that in RT-PCR, but the second stage of real time RT-PCR uses fluorescent probes to measure PCR amplification in real time [21].

In our study, a total of 381 serum samples obtained from HIV positive subjects were studied. Four different RT-PCR methods amplifying four separate regions (E2, NS3, NS5A and 5'NTR) of the GBV-C RNA genome were used, although not all of the four tests were done on all

samples. In addition, real time RT-PCR was performed on all samples, and thresholds are set for the result to be classified into three ordinal categories. The qualitative and categorized ordinal quantitative test results are then combined using Bayesian latent class analysis. A missing test of any kind is considered as an additional response category.

3. CLASSICAL LATENT CLASS ANALYSIS AND THE BAYESIAN APPROACH

Let X represent the latent disease status, and C the number of the latent classes. Let Y_t represent the result of each of the T observed diagnostic tests, $1 \leq t \leq T$. The variables Y_t, called manifest variables, are assumed to have D_t levels. Let $\mathbf{Y}_i$ denote the vector $(Y_{i1}, \cdots, Y_{it})^T$ for the i^{th} sample.

The contribution of the ith individual to the likelihood is:

$$P(\mathbf{Y}_i = \mathbf{y}_i) \quad = \quad \sum_{c=1}^{C} P(X_i = c) P(\mathbf{Y}_i = \mathbf{y}_i | X_i = c), \tag{1}$$

where the dependence of the probabilities above on unknown parameters has been omitted.

3.1. Classical Latent Class Analysis

In the classical latent class model, the assumption of conditional independence is made. Specifically, within each latent class, the T manifest variables are assumed to be mutually independent conditional on the latent variable:

$$P(\mathbf{Y}_i = \mathbf{y}_i | X_i = c) = \prod_{t=1}^{T} P(Y_{it} = y_{it} | X_i = c) \tag{2}$$

where $y_{it} = 1, 2, \cdots, D_t$. Combining equations (1) and (2) yields the following:

$$P(\mathbf{Y}_i = \mathbf{y}_i) \quad = \quad \sum_{c=1}^{C} P(X_i = c) \prod_{t=1}^{T} P(Y_{it} = y_{it} | X_i = c) \tag{3}$$

This latent class model is well suited for estimating the disease prevalence, sensitivity and specificity for each of the diagnostic tests, since the model is parameterized in terms of the probabilities that define the sensitivities, specificities and the prevalence.

The prevalence, sensitivities and specificities, can be estimated by maximizing the likelihood function $\mathcal{L} = \prod_{i=1}^{N} P(\mathbf{Y}_i = \mathbf{y}_i)$ for N samples with respect to model parameters to give the maximum likelihood estimates (MLE). The variance-covariance matrix can be approximated using the Hessian matrix evaluated at the MLE. A popular method for solving the MLE in latent class model is the Expectation-Maximization (EM) algorithm [22]. It is well suited for fitting latent class models by the method of maximum likelihood because the models are naturally formulated in terms of latent (i.e. incomplete) data.

One of the problems in the estimation of latent class models using maximum likelihood is that the parameters may be non-identifiable. Non-identifiability means that different sets of parameter values yield the same maximum of the log-likelihood function, and so there is no unique set of MLE. For example, with only two diagnostic tests, there is non-identifiability, see Joseph et al. [8].

3.2. Bayesian Approach

The Bayesian approach constructs a joint prior distribution over the unknown quantities. The data, through the likelihood function, are then combined with the prior distribution to produce the posterior distribution. The posterior distribution updates the distribution of the model

parameters, taking into account the information provided by the data. Prior distributions are useful to incorporate knowledge about unknown quantities. One advantage of the Bayesian approach is that if there is non-identifiability in the likelihood, the posterior distribution is proper and well-defined. Anderson [23], and Johnson et al. [24] discuss how the Bayesian estimates are impacted in these situations.

Given the complexity of the model, it is not possible to obtain the marginal distributions for the parameters analytically. The Gibbs sampler can be used to obtain samples from the marginal posterior distribution of each parameter. The Gibbs sampler is also used by Joseph et al. [8] for one or two diagnostic tests, and also by Branscum et al. [25] who use WinGUGS [26] for up to three diagnostic tests.

4. ANALYSIS OF GBV-C TESTS

4.1. Model Setting

The approach is illustrated through the GBV-C data set. Let X represent the latent GBV-C status: $X = 1$ if GBV-C present, $X = 0$ otherwise. Let $Y_1, \cdots, Y_4$ denote the four qualitative tests, and Y_5 the quantitative test.

There are substantial missing data for each of the four qualitative tests, although each subject has at least one qualitative test available. To take advantage of all available information, all samples should be included in the model. A missing test result is considered to be an additional response category for each qualitative test.

In contrast, the quantitative valued test Y_5, real time RT-PCR, is available on all samples. Y_5 could be dichotomized and combined with the other tests, with consequent loss of information.

The common assessment of continuous diagnostic tests is through the Receiver Operating Characteristic (ROC) curve, where the true positive rate against the false positive rate for the different possible thresholds of a diagnostic test are investigated. In this example, Y_5 is categorized into three levels: "high", "medium" and "low or none" (see Figure 1b). Specifically, let

$$Y_t = \begin{cases} 1 & t^{th} \text{ test result positive} \\ 0 & t^{th} \text{ test result negative} \\ NA & t^{th} \text{ test result missing} \end{cases}$$

where $t = 1, 2, 3, 4$, and

$$Y_5 = \begin{cases} 2 & 5^{th} \text{ test result} \geq 10^6 \text{ copies/ml (high)} \\ 1 & 5^{th} \text{ test result} \in [10^3, 10^6) \text{ copies/ml (medium)} \\ 0 & 5^{th} \text{ test result} < 10^3 \text{ copies/ml (low or none).} \end{cases}$$

We assume the following:

(1) The probability that each of $Y_1, \cdots, Y_4$ is missing is potentially different for each test, and does not depend on latent variable X, the true GBV-C status.

(2) Conditional on the latent variable X, the variables $Y_1, \cdots, Y_5$ are independent.

Suppose N samples are collected and y_{it} is the tth test result for the ith subject. From equation (3), the likelihood can be written as:

$$\prod_{i=1}^{N} [\sum_{c_i=0}^{1} P(X_i = c_i) \prod_{t=1}^{5} P(Y_{it} = y_{it} | X_i = c_i)] \tag{4}$$

where $y_{it} = 0, 1, NA$ for $t = 1, \cdots, 4$ and $y_{i5} = 0, 1, 2$.

Components in equation (4) are parameterized through: the prevalence of latent GBV-C status X, denoted by θ; the probabilities of each qualitative test being missing, denoted by M_t for $t = 1, \cdots, 4$; and the sensitivities and specificities of each test. For $t = 1, \cdots, 4$, denote the sensitivities and specificities by S_t and C_t respectively. For $t = 5$, the sensitivity of a high result ($Y_5 = 2$) and a medium result ($Y_5 = 1$) are denoted by SH_5 and SI_5. Correspondingly, the specificity of a low result ($Y_5 = 0$) and a medium result ($Y_5 = 1$) are denoted by CI_5 and CL_5. All sensitivities and specificities are conditional on the test being performed (not missing).

$$\begin{aligned}
\theta &= P(X = 1) \\
M_t &= P(Y_t = NA) \qquad t = 1, \cdots, 4 \\
S_t &= P(Y_t = 1 | X = 1, Y_t \neq NA) \qquad t = 1, \cdots, 4 \\
C_t &= P(Y_t = 0 | X = 0, Y_t \neq NA) \qquad t = 1, \cdots, 4 \\
SH_5 &= P(Y_5 = 2 | X = 1) \\
SI_5 &= P(Y_5 = 1 | X = 1) \\
CI_5 &= P(Y_5 = 1 | X = 0) \\
CL_5 &= P(Y_5 = 0 | X = 0)
\end{aligned} \tag{5}$$

To incorporate the constraint that the sum of SH_5 and SI_5 is less than 1, the conditional sensitivity SI_5^* is defined as below, conditional on the results not being "high". $CL_5^\dagger$ is defined

for a similar reason.

$$\begin{aligned} SI_5^* &= P(Y_5 = 1|X = 1, Y_5 \neq 2) \\ CL_5^\dagger &= P(Y_5 = 0|X = 0, Y_5 \neq 1) \end{aligned}$$

Under the parameterization in terms of SI_5^* and $CL_5^\dagger$ instead of SI_5 and CL_5, no constraints are required: they can each take any value in $[0, 1]$.

We denote the set of parameters

$$\{(M_t, S_t, C_t, SH_5, SI_5^*, CL_5^\dagger, CL_5), t = 1, \cdots, 4\}$$

by $\mathbf{\Theta}$. The likelihood expressed in equation (4) can be parametrized by $\mathbf{\Theta}$. Appendix A gives details. One of the benefits of this parameterization strategy is that the model is directly expressed by the sensitivity and specificity of each test, the quantities of primary interest.

In addition, we define the test based on the high cutoff of 10^6 copies/ml as RT(H), where the test is considered positive if $Y_5 = 2$ and negative otherwise. Similarly define RT(M) as positive if $Y_5 \geq 1$, and negative if $Y_5 = 0$, then the sensitivity and specificity of using the two cutoffs are easily expressed as functions of the parameters above. The sensitivities, S_{5H}, S_{5M} and specificities, C_{5H}, C_{5M}, of these two thresholds are:

$$\begin{aligned} S_{5H} &= P(Y_5 = 2|X = 1) = SH_5 \\ C_{5H} &= P(Y_5 = 0 \text{ or } 1|X = 0) = CL_5^\dagger(1 - CI_5) + CI_5 \\ S_{5M} &= P(Y_5 = 1 \text{ or } 2|X = 1) = SI_5^*(1 - SH_5) + SH_5 \\ C_{5M} &= P(Y_5 = 0|X = 0) = CL_5^\dagger(1 - CI_5) \end{aligned}$$

The expression of PPV and NPV of each test, function of the prevalence, sensitivity and specificity of the same kind, can be found in Appendix A.

4.2. Bayesian approach

A prior distribution for the unknown parameters defined in (5) is proposed. All are assumed independent of each other and each has a Beta distribution, with possibly different parameters:

$$\begin{aligned}
\theta &\sim Beta(\alpha_\theta, \beta_\theta) \\
M_t &\sim Beta(\alpha_{M_t}, \beta_{M_t}) \qquad t = 1, \cdots, 4 \\
S_t &\sim Beta(\alpha_{S_t}, \beta_{S_t}) \qquad t = 1, \cdots, 4 \\
C_t &\sim Beta(\alpha_{C_t}, \beta_{C_t}) \qquad t = 1, \cdots, 4 \\
SH_5 &\sim Beta(\alpha_{SH_5}, \beta_{SH_5}) \\
SI_5^* &\sim Beta(\alpha_{SI_5^*}, \beta_{SI_5^*}) \\
CI_5 &\sim Beta(\alpha_{CI_5}, \beta_{CI_5}) \\
CL_5^\dagger &\sim Beta(\alpha_{CL_5^\dagger}, \beta_{CL_5^\dagger})
\end{aligned}$$

Two different prior distributions are used. One specifies independent Beta distributions centered at the estimates from a previous study in the same laboratory [15], with the variance adjusted such that the prior belief is equivalent to 10 samples. For example, the estimated prevalence of GBV-C in [15] is 27.9%. In our model, the prior distribution for θ is therefore $Beta(2.79, 7.21)$, which has a mean of 0.279 and $2.79 + 7.21 = 10$ [27]. The detailed specifications of the prior distributions are given in Appendices B and C. Although these prior distributions are informative, considerable uncertainty is present. The alternative prior distribution specifies independent uniform prior distributions in the range $[0, 1]$, which are $Beta(1, 1)$ distributions and have more uncertainty.

The WinBUGS program [26] is used for performing the Gibbs Sampler. The parameters

of primary interest include the prevalence θ, the sensitivities of each test conditional on the test being performed: $S_1, \cdots, S_4$ and also SH_5, SI_5, as well as the corresponding specificities $C_1, \cdots, C_4$ and also CI_5, CL_5. The WinBUGS code is in Appendix B of an online technical report.

4.3. Classification

The Bayesian decision rule with underling symmetric loss function is used for the classification. Let $d(Y)$ denote the decision made on the true GBV-C status after observing Y. The decision set D is therefore $\{0, 1\}$. Let $L(X, d(Y))$ define the loss function. The symmetric loss function is:

$$L(X, d(Y)) = \begin{cases} 0 & d(Y) = X \\ k & d(Y) \neq X \end{cases}$$

where k is any positive real number. The expected loss function, i.e, the risk function for classifying the ith individual is:

$$\mathrm{E}L(X_i, d(Y)) = \mathrm{E}\sum_{c=0}^{1} L(X_i = c, d(Y))P(X_i = c|Y),$$

with the expectation taken over the posterior distribution of the parameters.

The best decision $d^*(Y)$ minimizes the risk function. For the symmetric loss function L,

$$d^*(y) = \begin{cases} 1 & P(X_i = 1|\mathbf{Y}) > P(X_i = 0|\mathbf{Y}) \\ 0 & otherwise \end{cases}$$

i.e., if $P(X_i = 1|\mathbf{Y}) > P(X_i = 0|\mathbf{Y})$ [27], the individual sample is classified as positive;

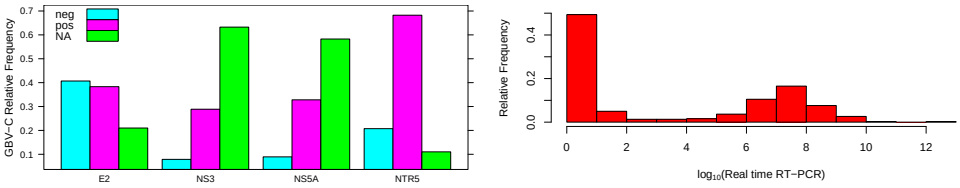

Figure 1. 1a (Left): The relative frequency of GBV-C being negative, positive, and missing by each of the qualitative test 5'NTR, E2, NS3, NS5A. 1b (Right): The relative frequency of log transformation of real time RT-PCR. The zeros represent undetectable GBV-C.

otherwise negative.

The predictive distribution of the latent variable $X_i = 1$ given the observed variables Y, $P(X_i = 1|\mathbf{Y})$, is the predictive distribution $P(X_i = 1|\mathbf{Y}, \mathbf{\Theta})$ averaged over the posterior distribution of $\mathbf{\Theta}|\mathbf{Y}$. Note that $P(X_i = 0|\mathbf{Y}) = 1 - P(X_i = 1|\mathbf{Y})$.

The estimate of the predictive posterior distribution can be easily achieved during the Markov Chain Monte Carlo (MCMC) sampling procedure. Suppose M Markov Chain Monte Carlo iterations are saved and $\mathbf{\Theta}^{(m)}$ is the sample from the m^{th} iteration. The predictive posterior distribution can be approximated by the mean of $Pr(X_i = 1|\mathbf{Y}, \mathbf{\Theta}^{(m)})$, over M iterations:

$$P(X_i = 1|\mathbf{Y}) \approx \frac{1}{M} \sum_{m=1}^{M} P(X_i = 1|\mathbf{Y}, \mathbf{\Theta}^{(m)}).$$

5. RESULTS FROM GBV-C EXAMPLE

5.1. Summary of Original Data

The proportion of positive results by individual E2, NS3, NS5A and 5'NTR tests, given that the test is done, is 48.5%, 78.6%, 78.6% and 76.7%, respectively. These prevalence estimates from

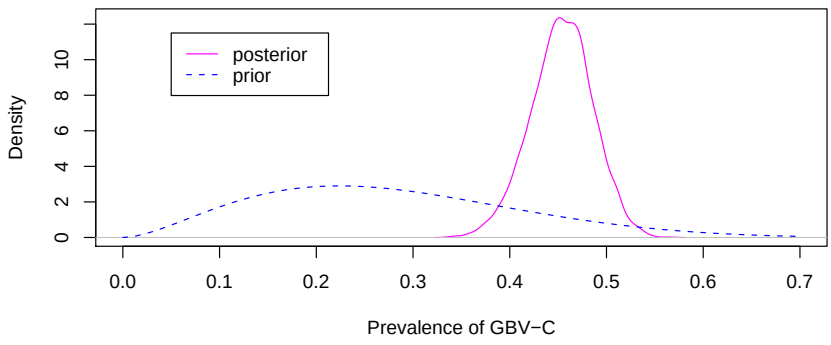

Figure 2. The distribution of GBV-C prevalence: the dashed line is the prior distribution and the solid line is the posterior distribution.

the last three primer tests are higher than the highest prevalence reported in the literature. Figure 1a shows that the corresponding proportion of missing results are approximately 21%, 63%, 58% and 11%, respectively. The primer test 5'NTR shows 77% positive results and is missing for only 11% of the samples. For real time RT-PCR, the proportion of positive results using a threshold of 10^3 copies/ml or 10^6 copies/ml is 44.4% and 37.8%, respectively. Figure 1b shows the real time RT-PCR result is approximately normally shaped in the log scale, but with an inflated frequency for low values.

5.2. Model Based Estimates

To fit the Bayesian extension of the latent class model to the GBV-C data set, the first 900 iterations of the MCMC sample are discarded and the approximation of posterior distribution is based on the subsequent 10,000 iterations. The prior distributions introduced in section 4 are used and the results from the first are given below. Similar results are found when uniform prior distributions are employed.

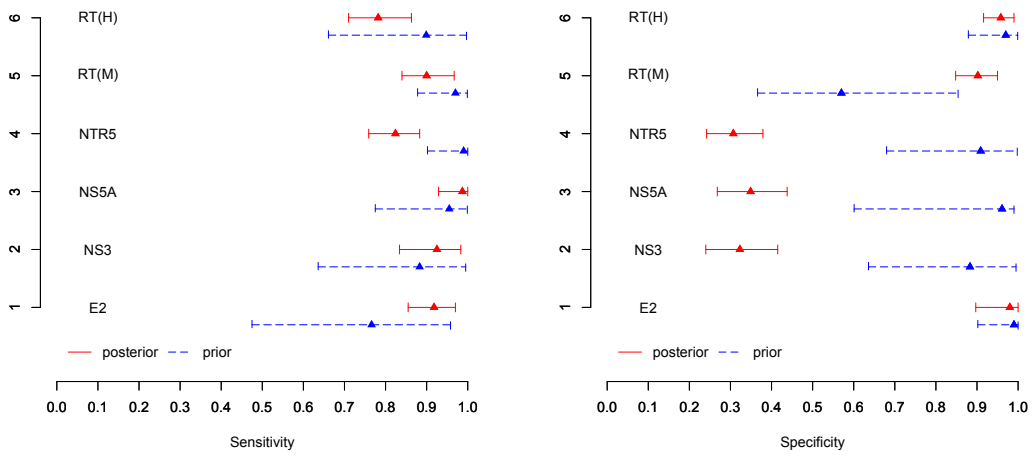

Figure 3. Posterior mean and 95% credible region for sensitivity and specificity of each diagnostic test.

5.2.1. Prevalence Figure 2 shows the prior and posterior distributions of GBV-C prevalence. The posterior mean of GBV-C prevalence is 45.4% and the 95% credible region is $[38.7\%, 51.4\%]$.

5.2.2. Sensitivity, Specificity, PPV and NPV In Figure 3 and Tables II and III of Appendix C in the online technical report, prior and posterior means and 95% credible regions of the sensitivity and specificity of each of the five tests are shown. Specifically, the sensitivity of RT(M) is the sensitivity of real time RT-PCR if the lower cutpoint (10^3 copies/ml) is set, and the sensitivity of RT(H) is the analog when the higher cutpoint (10^6 copies/ml) is set. The specificity, PPV and NPV of RT(M) and RT(H) are defined similarly. See Appendix A.

The analysis indicates that NS3, NS5A and 5'NTR produce too many false positives, and have low specificities. E2 has high specificity and reasonably high sensitivity. RT(M) has slightly higher sensitivity compared to RT(H), and slightly lower specificity. Similar patterns

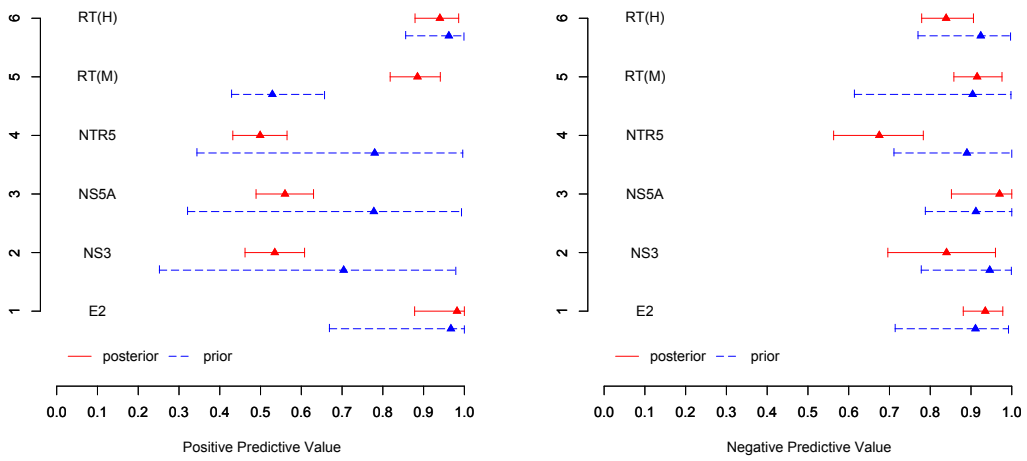

Figure 4. Posterior mean and 95% credible region for positive predictive value and negative predictive value of each diagnostic test.

are observed for positive predictive values and negative predictive values in Figure 4.

5.3. Classification

Using the Bayesian decision rule and symmetric loss function, 175 out of 381 samples are classified as positive. The value of Cohen's Kappa between this new classification and each primer test is given in Table I. E2 has the greatest agreement with the new classification. Table I also gives the relative sensitivity and specificity of each primer test, compared to the new classification. For the real time RT-PCR, the lower cutpoint (10^3 copies/ml) has higher sensitivity (0.909) than the higher cutpoint (10^6 copies/ml), and has reasonably good specificity (0.952).

Table I. Cohen's Kappa between the new classification and each primer test

	E2	NS3	NS5A	5'NTR	RT(M)	RT(H)
Cohen's Kappa	0.927	0.145	0.205	0.089	0.862	0.780
$\hat{S}^*$	0.930	0.949	1.000	0.815	0.909	0.794
$\hat{C}^*$	1.000	0.277	0.304	0.277	0.952	0.976

$\hat{S}^*$: sensitivities compared to the new classification.
$\hat{C}^*$: specificities compared to the new classification.

6. DISCUSSION

In the analysis of the GBV-C data set, the estimated posterior prevalence is about 45%, which is not very different from other studies in the literature. E2 is shown to be best single primer test. The specificities of 5'NTR, NS5A and NS3 are low, leading to PPVs close to the value 0.5 which corresponds to random guessing. The NPVs are more informative. For the real time RT-PCR, the trade off between sensitivity and specificity in using a cutoff of 10^3 or 10^6 copies/ml can be seen by comparing the estimates for RT(M) and RT(H).

The reason for the low specificity of three of the RT-PCR tests is unclear. The final classification is close to that ignoring these three tests (Table I). In other studies the prevalence based on these three tests is lower [15]. A conjecture is that these primers may amplify non-viral DNA from these samples. GBV-C virus has only been of interest relatively recently, and so tests for the presence of the virus are not standardized. Our method provides a mechanism for reconciling different test results in a systematic way.

Although the model has been developed here with four quantitative tests and one qualitative test, the methods easily generalize to arbitrary numbers of tests.

A limitation of the methods here are two critical assumptions. First the conditional independence assumption and second the assumption that missingness is independent of the

latent variable. Relaxing these assumptions should be further investigated. The conditional independence assumption has been criticized, see for example [10, 12]. Recent work has extended models for multiple diagnostic tests to correlated binary tests [28, 29, 30, 31, 32].

Advantages of the Bayesian approach include: appropriate incorporation of non-identifiability in the likelihood; readily accessible posterior estimates of uncertainty rather than asymptotic standard errors; the ability to make decisions on classification using Bayesian decision theory with different loss functions; the ability to incorporate the results of other studies through the prior distribution; easy implementation through WinBUGS or other programs.

This case study needs further development to investigate other methods to incorporate real time RT-PCR and combine with qualitative RT-PCR. It would be preferable to develop a method to incorporate the quantitative result directly rather than reduce to ordered categories. However, categorizing the quantitative test into ordinal categories makes combining all tests straightforward. In addition, missing quantitative test results are straightforward to incorporate. The relationship between the quantitative result and the results of the qualitative RT-PCR tests should also be examined.

In summary the method described here is a very feasible and practical way of combining the results of imperfect quantitative and qualitative diagnostic tests, especially when not all tests are performed on all samples.

ACKNOWLEDGEMENTS

This research was supported by NIH/NIAID (R01 058740).

REFERENCES

1. Lazarsfeld PF. The logical and mathematical foundations of latent structure analysis. In *Studies in Social Psychology in World War II. Vol. IV, Measurement and Prediction*, by Stouffer SA, Guttman L, *et al.* Princeton University Press: Princeton, 1950.
2. Kaldor J, Calyton D. Latent class analysis in chronic disease epidemiology. *Statistics in Medicine* 1985; **4**:327-335.
3. Walter SD, Irwig LM. Estimation of test error rates, disease prevalence and relative risk from misclassified data: a review. *Journal of clinical epidemiology* 1988; **41**:923-937.
4. Espeland MA, Handelman SL. Using latent class models to characterize and assess relative error in discrete measurements. *Biometrics* 1989; **45**:587-599.
5. Uebersax JS, Grove WM. Latent class analysis of diagnostic agreement. *Statistics in Medicine* 1990; **9**:559-572.
6. Garrett ES, Eaton WW, Zeger S. Methods for evaluating the performance of diagnostic tests in the absence of a gold standard: a latent class model approach. *Statistics in Medicine* 2002; **21**:1289-1307.
7. Evans MJ, Gilula Z, Guttman, I. Latent Class Analysis of Two-Way Contingency Tables by Bayesian Methods. *Biometrika* 1989; **76**:557-563.
8. Joseph L, Gyorkos TW, Coupal L. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *American Journal of Epidemiology* 1995; **141**:263-272.
9. Dendukuri N, Joseph L. Bayesian Approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics* 2001; **57**:158-167.
10. Pepe MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction* Oxford Statistical Science Series, Oxford University Press: Oxford, 2003.
11. Alonzo TA, Pepe MS. Using a combination of reference tests to assess the accuracy of a new diagnostic test. *Statistics in Medicine* 1999; **18**:2987-3003.
12. Hawkins DM, Garrett JA, Stephenson B. Some issues in resolution of diagnostic tests using an imperfect gold standard. *Statistics in Medicine* 2001; **20**:1987-2001.
13. Zhang W, Chaloner K, Tillmann HS, Williams CF, Stapleton JT. Effect of early and late GBV-C viraemia on survival of HIV-infected individuals: a meta-analysis. *HIV Medicine* 2006; **7**:173-80.
14. Stapleton JT. GB virus type C/hepatitis G virus. *Semin Liver Disease* 2003; **23**:137-148.

15. Souza IE, Allen JB, Xiang J, Klinzman D, Diaz R, Zhang S, Chaloner K, Zdunek D, Hess G, Williams CF, Benning L, Stapleton JT. Effect of primer selection on estimations of GB Virus C (GBV-C) prevalence and response to antiretroviral therapy for optimal testing for GBV-C viremia. *Journal of Clinical Microbiology* 2006; **44**:3105-3113. (Erratum, 2006, 44:4630)
16. Bogard M, Buffet-Janvresse C, Cantaloube JF, *et al.* GEMHEP multicenter quality control study of PCR detection of GB virus C/Hepatitis G virus RNA in serum. *Journal of Clinical Microbiology* 1997; **35**:3298-3300.
17. Kunkel U, Hohne M, Berg T, Hopf U, Kekule AS, Frosner G, Pauli G, Schreiner E. Quality control study on the performance of GB virus C/hepatitis G virus PCR. *Journal of Hepatology* 1998; **28**:978-984.
18. Lefrere JJ, Lerable J, Mariotti M, Bogard M, *et al.* Lessons from a multicenter study of the detectability of viral genomes based on a two round quality control iof GB virus C (GBV-C)/hepatitis G virus (HGV) polymerase chain reaction assay. *Journal of Virological Methods* 2000; **85**:117-124.
19. French Study Group for the Standardization of hepatitis C virus PCR. Improvement of hepatitis C virus RNA polymerase chain reaction through a multicenter quality control study. *Journal of Virological Methods* 1994; **49**:79-88.
20. Saiki RK, Bugawan TL, Horn GT, Mullis KB, Erlich HA. Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes. *Nature* 1986; **324**:163-166.
21. Mackay IM, Arden KE, Nitsche A. Real-time PCR in virology. *Nucleic Acids Research* 2002; **30**:1292-1305.
22. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B* 1977; **39**:1-38.
23. Andersen S. Re: "Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard." (Letter). *Am J Epidemiol.* 1997; **145**:290-301.
24. Johnson WO, Gastwirth JL, Pearson LM. Screening without a "gold standard": the Hui-Walter paradigm revisited. *Am J Epidemiol.* 2001; **153**:921-924.
25. Branscum AJ, Gardner IA, Johnson WO. Estimation of diagnostic-test sensitivity and specificity through Bayesian modeling. *Preventive Veterinary Medicine* 2005; **68**:145-163.
26. Spiegelhalter D, Thomas A, Best N, Lunn D. WinBUGS User Manual, 2003. URL http://www.mrc-bsu.cam.ac.uk/bugs.
27. DeGroot MH. *Optimal Statistical Decisions* New York: McGraw-Hill, 1970.
28. Qu Y, Hadgu A. A model for evaluating sensitivity and specificity for correlated diagnostic tests in efficacy studies with an imperfect reference test. *Journal of the American Statistical Association* 1998; **93**:920-928.

29. Shih JH, Albert PS. Latent model for correlated binary data with diagnostic error. *Biometrics* 1999; **55**:1232-1235.

30. Dendukuri N, Joseph L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics* 2001; **57**:158-167.

31. Black MA, Craig BA. Estimating disease prevalence in the absence of a gold standard. *Statistics in Medicine* 2002; **21**:2653-2669.

32. Toft N, Jrgensen E, Hjsgaard S. Diagnosing diagnostic tests: evaluating the assumptions underlying the estimation of sensitivity and specificity in the absence of a gold standard. *Preventive Veterinary Medicine* 2005; **68**:19-33.

Appendix

A: Components of the Likelihood

Specification of (4) requires conditional probabilities of each test taking any possible value, including missing results, given X. The connection between these conditional probabilities and the parameters defined in (4) are as below:
For $t = 1, \cdots, 4$,

$$\begin{aligned}
P(Y_t = NA|X = 1) &= P(Y_t = NA|X = 0) \;=\; M_t \\
P(Y_t = 1|X = 1) &= P(Y_t = 1|X = 1, Y_t \neq NA)P(Y_t \neq NA|X = 1) = S_t(1 - M_t) \\
P(Y_t = 0|X = 1) &= P(Y_t = 0|X = 1, Y_t \neq NA)P(Y_t \neq NA|X = 1) = (1 - S_t)(1 - M_t) \\
P(Y_t = 1|X = 0) &= P(Y_t = 1|X = 0, Y_t \neq NA)P(Y_t \neq NA|X = 0) = (1 - C_t)(1 - M_t) \\
P(Y_t = 0|X = 0) &= P(Y_t = 0|X = 0, Y_t \neq NA)P(Y_t \neq NA|X = 0) = C_t(1 - M_t).
\end{aligned}$$

For $t = 5$, there is no missing test result and

$$\begin{aligned}
P(Y_5 = 2|X = 1) &= SH_5 \\
P(Y_5 = 1|X = 1) &= P(Y_5 = 1|X = 1, Y_5 \neq 2)P(Y_5 \neq 2|X = 1) \\
&= SI_5^*(1 - SH_5) \\
P(Y_5 = 0|X = 1) &= 1 - P(Y_5 = 2|X = 1) - P(Y_5 = 1|X = 1) \\
&= (1 - SI_5^*)(1 - SH_5) \\
P(Y_5 = 2|X = 0) &= 1 - P(Y_5 = 1|X = 0) - P(Y_5 = 0|X = 0) \\
&= (1 - CI_5)(1 - CL_5^\dagger) \\
P(Y_5 = 1|X = 0) &= CI_5 \\
P(Y_5 = 0|X = 0) &= P(Y_5 = 0|X = 0, Y_5 \neq 1)P(Y_5 \neq 1|X = 0) \\
&= CL_5^\dagger(1 - CI_5).
\end{aligned}$$

The PPV and NPV for RT(M) are denoted PPV_{5M} and NPV_{5M}, where

$$\begin{aligned} PPV_{5M} &= P(X=1|Y_5=1 \text{ or } 2) \\ &= \frac{[SH_5 + SI_5^*(1-SH_5)]\theta}{[SH_5 + SI_5^*(1-SH_5)]\theta + [1-CL_5^\dagger(1-CI_5)](1-\theta)} \\ NPV_{5M} &= P(X=0|Y_5=0) \\ &= \frac{CL_5^\dagger(1-CI_5)(1-\theta)}{CL_5^\dagger(1-CI_5)(1-\theta) + (1-SI_5^*)(1-SH_5)\theta}. \end{aligned}$$

The PPV and NPV for RT(H) are denoted PPV_{5H} and NPV_{5H}, where

$$\begin{aligned} PPV_{5H} &= P(X=1|Y_5=2) \\ &= \frac{SH_5\theta}{SH_5\theta + (1-CI_5)(1-CL_5^\dagger)(1-\theta)} \\ NPV_{5H} &= P(X=0|Y_5=0 \text{ or } 1) \\ &= \frac{[CL_5^\dagger(1-CI_5) + CI_5](1-\theta)}{[CL_5^\dagger(1-CI_5) + CI_5](1-\theta) + (1-SH_5)\theta}. \end{aligned}$$

B: WinBUGS Code.

C: Tables of posterior and prior estimates for the different tests.

B: WinBUGS Code

The model cannot be specified directly in WinBUGS, but the following code specifies the likelihood (4) and prior distributions:

```
#############################################################################
# Bayesian Latent Class Analysis
#
# This program specifies the prior distribution and likelihood. The WinBUGS
# program is used to implement the Bayesian approach in the latent class model.
#
############################################################
# The observed or latent variables are defined as follows:#
############################################################
# X: latent class variable. X=1,0
# Y[1:5]: 5 tests taken for each person.
# Y[t]=0,1 or NA for t=1:4; Y[5]=0,1,2
#
############################################################
# The parameters modeled are defined as follows:          #
############################################################
# prev : prevalence of the medical condition, i.e. P(X=1)
# pNA[1:4]=P(Y[t]=NA): Probabilities that tests are missing
# S[t]=P(Y[t]=1|X=1,Y[t]!=NA) :  Sensitivities of tests 1,2,3,4
# C[t]=P(y[t]=0|X=0,Y[t]!=NA):  Specificities of tests 1,2,3,4
# S5y2=P(Y[5]=2|X=1): Sensitivity of Y5=2
# S5y1=P(Y[5]=1|X=1): Sensitivity of Y5=1
# S5Y1not2=P(Y[5]=1|X=1,Y[5]!=2): Sensitivity of y5=1 given than Y5!=2
# C5y1=P(Y[5]=1|X=0): Specificity of Y5=1
# C5y0=P(Y[5]=0|X=0): Specificity of Y5=0
# C5Y0not1=P(Y[5]=0|X=0,Y[5]!=1):Specificity of y5=0 given than Y5!=1
# S5H=P(Y[5]=2|X=1): Sensitivity of Y5 if a cutoff of 10^6 is used.
# S5M=P(Y[5]=1 or 2|X=1): Sensitivity of Y5 if a cutoff of 10^3 is used.
# C5H=P(Y[5]=0 or 1|X=0): Specificity of Y5 if a cutoff of 10^6 is used.
# C5M=P(Y[5]=0|X=0): Specificity of Y5 if a cutoff of 10^3 is used.
#
#############################################################################

model
{

  ###### priors ######

  prev ~ dbeta(alpha.prev, beta.prev)
  for (t in 1:4){
     pNA[t] ~ dbeta(alpha.NA[t],beta.NA[t])
  }
  for (t in 1:4){
       S[t] ~ dbeta(alpha.S[t], beta.S[t])
       C[t] ~ dbeta(alpha.C[t], beta.C[t])
  }

  S5y2 ~ dbeta(alpha.S5y2, beta.S5y2)
  S5y1not2 ~ dbeta(alpha.S5y1not2, beta.S5y1not2)
  C5y1 ~ dbeta(alpha.C5y1, beta.C5y1)
  C5y0not1 ~ dbeta(alpha.C5y0not1, beta.C5y0not1)
```

```
S5y1 <-  S5y1not2*(1-S5y2)
C5y0 <- C5y0not1*(1-C5y1)

S5H <- S5y2
S5M <- S5y1 + S5y2
C5H <- C5y0 + C5y1
C5M <- C5y0

###### likelihood ######

## Conditional probabilities of Y1 through Y4, given X.
for (t in 1:4){
    CPy1.X1[t] <- (1-pNA[t])*S[t]
    CPy0.X1[t] <- (1-pNA[t])*(1-S[t])
    CPyNA.X1[t] <- pNA[t]

    CPy1.X0[t] <- (1-pNA[t])*(1-C[t])
    CPy0.X0[t] <- (1-pNA[t])*C[t]
    CPyNA.X0[t] <- pNA[t]
}
## Conditional probabilities of Y5, given X.
CPy52.X1 <- S5y2
CPy51.X1 <- (1-S5y2)*S5y1not2
CPy50.X1 <- (1-S5y2)*(1-S5y1not2)

CPy52.X0 <- (1-C5y1)*(1-C5y0not1)
CPy51.X0 <- C5y1
CPy50.X0 <- (1-C5y1)*C5y0not1

## Specify the specific likelihood through a trick of using Bernoulli probability.
## The idea is that we observed a sample of 1's with the target individual likelihood
## from model. L(i) is the target individual likelihood.
for (i in 1:N) {
    for (t in 1:4){
 CPyX1[i,t] <-  CPy1.X1[t] *equals(Y[i,t],1)+CPy0.X1[t]*equals(Y[i,t],0)+
                       CPyNA.X1[t]*equals(Y[i,t],99)
    }
    for (t in 1:4){
 CPyX0[i,t] <-  CPy1.X0[t]*equals(Y[i,t],1)+CPy0.X0[t]*equals(Y[i,t],0)+
                      CPyNA.X0[t]*equals(Y[i,t],99)
    }
    CPyX1[i,5] <-  CPy52.X1*equals(Y[i,5],2)+ CPy51.X1*equals(Y[i,5],1)+
                   CPy50.X1*equals(Y[i,5],0)

   CPyX0[i,5] <-  CPy52.X0*equals(Y[i,5],2)+ CPy51.X0*equals(Y[i,5],1)+
                  CPy50.X0*equals(Y[i,5],0)

    L[i] <-    prev*CPyX1[i,1]*CPyX1[i,2]*CPyX1[i,3]*CPyX1[i,4]*CPyX1[i,5]+
           (1-prev)*CPyX0[i,1]*CPyX0[i,2]*CPyX0[i,3]*CPyX0[i,4]*CPyX0[i,5]

# Trick to specify a new sampling distribution with individual likelihood L(i).
   ones[i] <- 1
      p[i] <- L[i]
   ones[i] ~ dbern(p[i])
}
```

```
  ###### PPVs and NPVs each of 5 tests ######
  for (t in 1:4)
  {
    CPX1.y1[t] <- prev*CPy1.X1[t]    / (prev*CPy1.X1[t] + (1-prev)*CPy1.X0[t])
    CPX0.y0[t] <- (1-prev)*CPy0.X0[t]/ (prev*CPy0.X1[t] +(1-prev)*CPy0.X0[t])
  }
  CPX0.yNA <- 1-prev
  CPX1.yNA <- prev

  CPX0.y52 <- CPy52.X0*(1-prev)/(CPy52.X0*(1-prev) + CPy52.X1*prev)
  CPX0.y51 <- CPy51.X0*(1-prev)/(CPy51.X0*(1-prev) + CPy51.X1*prev)
  CPX0.y50 <- CPy50.X0*(1-prev)/(CPy50.X0*(1-prev) + CPy50.X1*prev)

  CPX1.y52 <- CPy52.X1*prev/(CPy52.X1*prev + CPy52.X0*(1-prev))
  CPX1.y51 <- CPy51.X1*prev/(CPy51.X1*prev + CPy51.X0*(1-prev))
  CPX1.y50 <- CPy50.X1*prev/(CPy50.X1*prev + CPy50.X0*(1-prev))

  PPV5H <-  CPy52.X1*prev/(CPy52.X1*prev + CPy52.X0*(1-prev))
  PPV5M <- (1-CPy50.X1)*prev/((1-CPy50.X1)*prev+ (1-CPy50.X0)*(1-prev))
  NPV5H <- (1-CPy52.X0)*(1-prev)/((1-CPy52.X0)*(1-prev)+ (1-CPy52.X1)*prev )
  NPV5M <- CPy50.X0*(1-prev)/(CPy50.X0*(1-prev) + CPy50.X1*prev)
  }

#########################################################
# Hyper-parameters for the Beta prior distributions
#########################################################

     alpha.prev <- 2.79
      beta.prev <- 7.21

      alpha.S <- c(7.66,8.83,9.55,9.90)
       beta.S <- c(2.34,1.17,0.45,0.10)

      alpha.C <- c(9.90,8.61,8.99,9.09)
       beta.C <- c(0.10,1.39,1.01,0.91)

      alpha.S5y2 <- 9.0
       beta.S5y2 <- 1.0

      alpha.S5y1not2 <- 7.0
       beta.S5y1not2 <- 3.0

      alpha.C5y1 <- 7.0
       beta.C5y1 <- 3.0

      alpha.C5y0not1 <- 9.0
       beta.C5y0not1 <- 1.0

      alpha.NA <- c(2,5,5,2)
       beta.NA <- c(8,5,5,8)
```

C: Tables of posterior and prior estimates for the different tests.

Table II. Posterior estimates of sensitivities, specificities, PPVs, NPVs of primer tests of GBV-C.

	Sensitivity		Specificity		PPV		NPV	
Parameter	Mean(SE)	95% HDR	Mean(SE)	95% HDR	Mean(SE)	95% HDR	Mean(SE)	95% HDR
E2	0.918(0.030)	[0.918, 0.970]	0.980(0.025)	[0.897, 1.000]	0.982(0.033)	[0.878, 1.000]	0.935(0.025)	[0.881, 0.978]
NS3	0.925(0.039)	[0.925, 0.983]	0.323(0.045)	[0.240, 0.415]	0.535(0.037)	[0.462, 0.608]	0.840(0.072)	[0.696, 0.960]
NS5A	0.987(0.020)	[0.987, 1.000]	0.349(0.046)	[0.268, 0.438]	0.560(0.036)	[0.489, 0.630]	0.970(0.042)	[0.852, 1.000]
5'NTR	0.824(0.031)	[0.824, 0.883]	0.307(0.036)	[0.242, 0.379]	0.499(0.034)	[0.432, 0.565]	0.675(0.055)	[0.563, 0.783]
RT(M)*	0.900(0.032)	[0.900, 0.967]	0.902(0.025)	[0.848, 0.950]	0.885(0.031)	[0.818, 0.941]	0.915(0.029)	[0.858, 0.976]
RT(H)**	0.782(0.039)	[0.782, 0.863]	0.958(0.019)	[0.916, 0.990]	0.940(0.028)	[0.879, 0.986]	0.839(0.033)	[0.779, 0.906]

*: A cutoff of 10^3 copies/ml is used for Real time RT-PCR measurement.

**: A cutoff of 10^6 copies/ml is used for Real time RT-PCR measurement.

Table III. Prior distribution of sensitivities, specificities, PPVs, NPVs of primer tests of GBV-C.

	Sensitivity		Specificity		PPV		NPV	
Parameter	Mean(SE)	95% HDR	Mean(SE)	95% HDR	Mean(SE)	95% HDR	Mean(SE)	95% HDR
E2	0.766(0.128)	[0.475, 0.958]	0.990(0.030)	[0.902 ,1.000]	0.967(0.091)	[0.669, 1.000]	0.911(0.074)	[0.714, 0.992]
NS3	0.883(0.097)	[0.636, 0.995]	0.883(0.097)	[0.636 ,0.995]	0.704(0.200)	[0.252, 0.979]	0.946(0.060)	[0.778, 0.999]
NS5A	0.955(0.063)	[0.775, 0.999]	0.861(0.104)	[0.601 ,0.990]	0.778(0.185)	[0.321, 0.993]	0.912(0.035)	[0.788, 1.000]
5'NTR	0.990(0.030)	[0.902, 1.000]	0.909(0.087)	[0.680 ,0.998]	0.780(0.178)	[0.344, 0.996]	0.890(0.017)	[0.711, 1.000]
RT(M)*	0.970(0.033)	[0.878, 0.999]	0.570(0.128)	[0.366, 0.854]	0.529(0.057)	[0.429, 0.657]	0.904(0.104)	[0.614, 0.998]
RT(H)**	0.899(0.091)	[0.661, 0.997]	0.970(0.033)	[0.879, 0.999]	0.962(0.039)	[0.856, 0.999]	0.924(0.062)	[0.770, 0.997]

*: A cutoff of 10^3 copies/ml is used for Real time RT-PCR measurement.

**: A cutoff of 10^6 copies/ml is used for Real time RT-PCR measurement.