Statistical Modelling of Performance Data for Molecular Amplification Methods in Diagnostic Virology

LLenalia García Fernández, BSc, MSc.

A Thesis submitted in partial fulfilment of the requirements
of the University of Abertay Dundee
for the degree of Doctor of Philosophy.

I certify that this thesis is the true and accurate version of the thesis approved by the examiners.

Signed: (Principal Supervisor) Date:

April 2012

Declaration

I, Llenalia García Fernández, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I can confirm that this has been indicated in the thesis.

Signed: Date

Abstract

Nucleic Acid Technology (NAT), introduced in the late 90s, is a molecular amplification method that can be used for the diagnosis and management of patients with infectious diseases. NAT test results are obtained quicker and are quantified, providing greater information than the positive/negative results available from traditional techniques. However, NATs are technically demanding, susceptible to contamination and hence results from associated diagnostic tests may be inaccurate. External Quality Assessment (EQA) services are programmes developed to assess and advance the quality performance of laboratories that use NAT kits to diagnose, manage and control human diseases. Quality Control for Molecular Diagnostics (QCMD), an organisation that provides EQA, uses proficiency panels designed with samples containing no, weak, medium and strong microbial loads. The panels are distributed to participating laboratories who analyse them knowing the pathogen but blind to the microbial load.

In this thesis, factors which are significantly associated with EQA participants' performance are identified. In particular, rigorous statistical methods are used and developed to interrogate, for the first time, the large reservoir of QCMD data and model participants' performance over time for different pathogens. Furthermore, new scoring schemes are developed to assess individual participants' performance on individual panels.

Existing scoring schemes do not take into account known prior information about the sample viral load. We propose, using Bayesian techniques, to score participants with respect to a 'Bayesian mean' value obtained from prior information available to QCMD and the values from 'reference' laboratories with high reputation. For qualitative (presence/absence) diagnosis, logistic regression models from a Bayesian perspective are developed to fit historical and current data in order to identify factors which are significantly associated with participant performance. For quantitative (estimate of sample microbial load) diagnosis, Generalised Linear Models (GLM) from a Bayesian perspective are developed to fit the data and find significant factors associated with participants' estimates of the sample microbial load. A more natural parameter inference is made from a Bayesian perspective using the distributions of the parameters given the data. Model validation and robustness are also investigated. Some responses in the quantitative diagnosis are given as censored data, so a GLM which allows the inclusion of the censored observations is introduced and developed in order to obtain a more accurate model to fit these data. Also, a variation of an existing model comparison tool, the Deviance Information Criterion (DIC), is developed in order to discriminate between different suggested models. Extensive use is made of Markov Chain Monte Carlo (MCMC) methods using R statistical software to obtain model estimates.

The benefits of adopting this approach are the full use of data from panels for the same pathogen over time, above/below limit of detection data and a more accurate target value. These provide a better measure of participant performance, so the advice given to participants about the best technology to be used improves. The techniques developed in this thesis can be applied to other research areas- especially those where GLM for censored observation are used, such as survival analysis in medical research and industrial experiments on reliability.

Acknowledgments

I would like to express my gratitude to my supervisor Dr. Nat Jack for his professional skills in the supervision of this project and for the helpful feedback and guidance related to this thesis. Also, I would like to thank Dr. Phil Collier for his support. I am very grateful to Dr. Harry Staines, my former supervisor who continued helping me voluntarily after his commitment at the University of Abertay Dundee finished. He shared with me his experience and knowledge about how to progress with my project. His expertise and encouraging feedback on my work gave to the project a dimension for which I cannot be grateful enough.

I am also grateful to the Dundee Business School for giving me the opportunity to do my PhD at the University of Abertay Dundee. I would like to thank Quality Control for Molecular Diagnostics (QCMD) for the provision of the data. Special thanks to Dr. Paul Wallace from QCMD for his patience in explaining me related background knowledge and the many discussions with him that were very important for the outcome of this work. I spent most of my study and research time at the QCMD office in Glasgow, which I will always remember as a warm, friendly and creative place. Many people made this place so special for me, both from a research and from a personal point of view. I am particularly grateful for the help and support received from Sue, Colin, Callum, William and Margaret.

I would also like to express my gratitude to some of my friends, Ricardo, Marco, Carlos, Eva and Elisabeth for their emotional support and advice during my PhD studies. Finally, my most important gratitude is to my partner, Tilman, for his expertise and statistical support; his suggestions, his patience while reading the thesis and his creativity providing feedback to help me to improve the thesis. Furthermore, I would like to thank him for his emotional support, continuous efforts in cheering me up during the stressful moments and encouragement in difficult situations. Special gratitude goes to my family for their support, patience and understanding during all my studies. Producing this thesis would not have been possible without the continuous encouragement of my parents, Maruchi and Juan, my siblings, Marian, Belinda and Hilario and my partner, Tilman.

Contents

List of Tables

List of Figures

Abbreviations

Common Abbreviations used through the Report

AIC: Akaike's Information Criterion
BBV: Blood Borne Viruses
bDNA: bDNA technologies
DNA: Deoxyribonucleic acid
BIC: Bayesian Information Criterion
BMA: Bayesian Model Averaging
CBM: Censored Bayesian Model
CC: Conventional Commercial technologies
CIH: Conventional In-house technologies
DIC: Deviance Information Criterion
EQA: External Quality Assessment
EV: Enteroviruses
GLM: Generalised Linear Model
GP: General Practitioner
HBV: Hepatitis B Virus
HC: Hybrid Capture technologies
IU: International Units
MCMC: Markov Chain Monte Carlo
NASBA: NASBA technologies
NAT: Nucleic Acid Technology
PT: Proficiency testing
QC: Quality Control
QCMD: Quality Control for Molecular Diagnostics
QLBM: Qualitative Bayesian Model
QTBM: Quantitative Bayesian Model
RNA: Ribonucleic acid
RTIH: Real Time In-house technologies
RTC: Real Time Commercial technologies
SE: Standard Error
SD: Standard Deviation
TMA: Transcription mediated amplification technologies

Chapter 1

Introduction

The motivation of this research starts with a very common scenario. A general practitioner (GP), consulted by an unwell patient, may suspect that the patient is suffering from a disease caused by a pathogen, hepatitis B (HBV), for example. To obtain a diagnosis the doctor takes a sample (such as a blood sample) from the patient and sends it to a clinical laboratory to be tested for a specific virus. Traditional diagnostic tests take two weeks to provide a positive or negative diagnosis for having HBV. More recent molecular techniques provide an estimate of the microbial load within a much shorter period, often on the same day.

The shorter time to obtain the results from molecular diagnostic techniques has a number of advantages. Treatment, if required, can begin much earlier when the microbial load may be lower. In addition, there are some circumstances, such as transplant patients, where an infection may be life threatening and therefore waiting for two weeks for a diagnosis would be too long. Furthermore, there are some applications, such as in clinics that treat sexually transmitted diseases (for example HIV), where the patient may not return to collect their diagnosis. It is advantageous in these cases for the clinician to obtain the results of a diagnostic test when the patient is still present at the consultation. Point of care, molecular diagnostic kits are currently being designed for such circumstances.

The ability of molecular diagnostic kits to produce an estimate of the microbial load possesses also a series of advantages, such as the rapid availability of the results allowing the clinician to monitor pathogen load that can be used to determine and adjust the required dose by the patient. This is an important consideration in diseases with drug resistant pathogens, such as HIV.

The advantages above assume that the clinical laboratory using molecular diagnostic kits provide clinicians with accurate results. Therefore, it is of interest to find out how well laboratories analyse the samples, to which degree the laboratory does so correctly and the reason why different laboratories may provide different estimates of the sample microbial load for the same sample.

1.1 Background

1.1.1 Molecular Amplification Methods

Molecular Amplification Methods in Diagnostic Virology are used for the diagnosis and management of patients with infectious diseases such as hepatitis B (HBV), hepatitis C (HCV) and Enterovirus (EV). New technologies, in particular Nucleic Acid Technology (NAT), have been recently introduced and used by laboratories worldwide (Heid, et al., 1996; Fleige and Pfaffl, 2006). These technologies have the advantages that results can be obtained rapidly and can be quantified. Also, it is claimed that NATs have fewer false positives (more sensitive) and false negatives (more specific) than traditional techniques.

However, NATs are technically demanding, susceptible to contamination and hence results of diagnostic tests provided by these technologies may be inaccurate (Germann and Telenti, 1995; BioTecniques, 2007). Therefore, there is a need to monitor and improve laboratory quality, and one method used for this is undertaking External Quality Assessment (EQA), known in the United States as Proficiency Testing (PT) (ISO/IEC Guides, 1997a).

Nucleic acid amplification techniques are tests that directly detect the genetic material of a microbe. These techniques vary depending on the amplification method. One method is conventional Polymerase Chain Reaction (PCR), which is based on the amplification of a single or few copies of a piece of DNA in order to generate thousands of copies of a particular DNA sequence. A similar method is real time PCR based on the PCR method to amplify a DNA sequence. However, real time PCR simultaneously quantifies a target DNA molecule in real time as the reaction progresses, whilst with conventional PCR the product of the reaction is detected at its end. Nucleic Acid Sequence Based Amplification (NASBA) is a NAT method used to amplify RNA sequences. Amongst other methods are branched DNA (bDNA) techniques, which use a molecule that links to the specific genetic material, and Transcription Mediated Amplification (TMA). It should be noted that laboratories can use a commercial system or can develop a system in-house (Apfalter et al., 2005).

1.1.2 External Quality Assessment

EQA has long been considered the most important way of monitoring laboratory quality. It allows a laboratory to monitor independently its performance and provides feedback to identify and investigate potential areas of concern. Specific guidelines and general principles are common to most EQA schemes (ILAC-G13, 2000; ISO/IEC Guides, 1997b). However, there are many different approaches to EQA depending on the clinical, analytical, and regulatory goals, which require variations in the design, implementation and reporting mechanism of the EQA programme. Thus, the identification of appropriate "performance indicators" and

their statistical analysis vary across EQA providers. Clinical chemistry has taken the lead on laboratory quality issues but the traditional approaches to quality and EQA are limited and difficult to apply for molecular diagnostics. Therefore, there is a need to adapt current methods to define, develop and implement suitable performance indicators.

Within clinical virology and microbiology, EQA organisations providing schemes for molecular diagnostics of infectious diseases have focused on the traditional subjective approach to EQA with peer group review and consensus analyses used to measure the results from all participants. They have also defined performance indicators by simple and immediate measures of the participants' performance based on functions of the error or deviation (NEQAS, 2010; QCMD, 2010). These performance indicators can be used as a relative measure to compare laboratories, and are easy to compute and interpret. However, their statistical distributions are unknown, and so it is difficult to establish limits to identify participants who are performing satisfactorily.

1.1.3 Quality Control for Molecular Diagnosis and brief introduction to the data

Quality Control for Molecular Diagnosis (QCMD) is an organization based in Glasgow which provides an EQA service for molecular diagnostic kits users and which aims "to assess and advance the quality of NAT for diagnosis, management and control of human diseases with particular reference to infectious diseases". QCMD offers a variety of independent programmes and nowadays covers 29 programmes of different pathogens. Each programme provides a panel of samples with different pathogen loads to be tested with respect to a target pathogen. A worldwide range of laboratories participate, often voluntarily, on the EQA programmes offered by QCMD so that they can be assessed on their performance.

Using the results obtained by each participating laboratory of the EQA programme of a particular pathogen, QCMD provides an individual score which allows the participants to compare its results with those from other participants.

The QCMD organisation provides EQA programmes for a large range of viruses, such as, Blood Borne Viruses (BBV) (e.g. Hepatitis B, Hepatitis C and HIV) and Enterovirus (EV)(QCMD, 2010). QCMD designs panels of a single known pathogen containing samples of no, weak, medium and strong microbial load. The target samples' microbial or viral loads are estimated by QCMD technicians. The estimated samples' target values, determined by QCMD, are based on the design of the panels. In addition, prior to distribution, the panels are also sent to selected laboratories ('reference' laboratories) with a good reputation for working with the pathogen. These 'reference' laboratories provide estimates of the samples' microbial load that QCMD uses in order to check for possible inconsistencies. However, this information is not currently used when scoring participants.

The panels and a questionnaire related to the entire laboratory practice are distributed to participating laboratories. Participating laboratories analyse a panel knowing the pathogen, but are blind to the microbial loads estimated by QCMD. Then, the results obtained from the participants are compared with the consensus of other participants estimates of the sample loads.

Previously, QCMD data have been used to provide individual performance indicator scores and a general report summarising participants' performance for individual pathogens for an individual panel only. However, due to the lack of adequate statistical methods in the field of quality control for molecular diagnosis, QCMD was not able to provide participants of EQA programmes with suitable feedback and information about how their laboratory practices affected their performance.

To address this, the present study uses a Bayesian approach to develop a statistical framework suitable for the analysis of QCMD data. Panels for three pathogens over four years (2002-2005) will be used for model development and testing.

1.2 Structure of the Thesis

The first and second chapter of the thesis describe the background of the project and the data to be analysed. In Chapter 1 a general overview of the statistical techniques and contexts that will be used in this thesis to develop an appropriate model for the data is given. It is also introduced the statistical methodology used and the two schools into which the discipline of statistics is divided. In Chapter 2 the data of interest are described and an exploratory analysis carried out.

Chapter 3 describes a newly developed and improved scoring system which assesses individual participant's performance. It also shows a comparison of the results obtained from the improved scoring system applied to Hepatitis B virus data from 2005 with the current scoring system used by QCMD.

In Chapter 4 and 5, new approaches for the analysis of qualitative and quantitative responses in EQA application are. A qualitative response is the detection (or not) of the pathogen, whilst a quantitative response is the estimation of the microbial load. The statistical models developed from a Bayesian perspective are described and their applications to Enterovirus, Hepatitis B virus and Hepatitis C virus data are shown.

Due to the characteristics of the quantitative data, a new model that allows the use the responses outside the limits of detection of the assay are refined and developed in order to find information about participants' performance. The improved model for quantitative responses is described and shown in Chapter 6. Chapter 7 describes the results obtained from a simulation study carried out in order to test and extend the improved model to other data sources and areas of research.

Finally, Chapter 8 provides a discussion of this project and proposes further work. The appendices contain information about the probability distributions used throughout the thesis and tables of results from the model application, as well as some classical tests used to study the robustness of the models.

1.3 Statistical Data Modelling

Mathematical models are the realization of a real problem under study via equations that explain or describe the problem itself. However, the mechanisms described by mathematical models are often influenced by external factors that cannot be easily treated mathematically, such as the behaviour of biological organisms or atmospheric conditions. These systems cannot be exactly expressed by a deterministic equation. In the current application, the quantification of viral load from a blood sample may, for example, be influenced by the technician who analyses the sample, the technology used or the laboratory procedure. Therefore, observational data may be collected and valid analysis techniques incorporating probabilities about the observed data into the mathematical equations turn deterministic models into statistical models (Krzanowski, 1998). In this way, variability using probability distributions accommodating both random and systematic variations are represented.

Statistical data modelling can be seen as a tool that enables the extrapolation from the observed information from a sample of the population to the general population under study. Thus, the statistical modelling process should involve a scientific team which provides guidance and advice on how to represent reality by equations and probability functions. Considerations of how the model is built and matched to the data need to be taken into account and this can lead to the production of alternative models. This is the process of model checking and criticism (Morgan, 2000).

Therefore, statistical models are about what we can learn from data. The art of modelling lies in finding and providing a 'good' technique to describe the model, which explains the real problem, and answers questions proposed in the most sensible and possibly less complex way. Model complexity depends on the problem to be solved and the type of the answer required (Davison, 2003).

1.3.1 Bayesian Models

Statistical modelling of data seeks to quantify some uncertainty, and so it is reasonable to do it by using probability. Since physical randomness induces uncertainty, it seems sensible to describe this uncertainty in terms of random events instead of fixing it with frequencies from repeated measurements under the same conditions. Also, decision-making based on statistical inference implies that the uncertainty must be represented in terms of probability (Bernardo and Smith, 1994). The basic principle in Bayesian statistics is that probability is a measure of uncertainty. If y is the known information and θ the unknown information, then Bayes' theorem (Bayes, 1763) can be formulated as follows:

$$f(\theta|y) \propto f(y|\theta)f(\theta), \tag{1.1}$$

where $f(y|\theta)$ is the likelihood function of y given θ, $f(\theta)$ denotes our prior beliefs about the unknown information and $f(\theta|y)$ is the posterior density of the unknown information given the known data.

The normalising factor in equation (1.1) is the inverse of the likelihood accumulated over all possible prior values,

$$f(y) = \int f(\theta) f(y|\theta) d\theta. \tag{1.2}$$

In Bayesian modelling the posterior density of the unknown information is determined by updating the assumed prior information with the data. Thus, information on the probability distribution and characteristics of the unknown information is obtained. As a result of this, Bayesian methods provide a comprehensive and robust approach to model estimation. They are not dependent on the assumption of asymptotic normality as much as in classical statistical modelling. In addition, with a Bayesian approach we are able to incorporate and combine different sources of information, leading to a potential improvement of the precision of the estimates (Berger, 2000). Furthermore, the use of 'non-informative' prior information in Bayesian models (objective Bayesian) often leads to equivalent estimates to those obtained in the classical approach with the advantage of obtaining posterior distributions of random parameters (Gelman et al., 2004).

The Bayesian approach to modelling can be divided into 'full' and 'empirical' Bayes estimation. In this context 'empirical' is referred to when the data are used to estimate the prior parameters whilst 'full' refers to estimating the prior parameters of the model independently of the data (Carlin and Louis, 2001).

Hierarchical Bayesian modelling is a natural way to fit complex data structures when the information is available on several different levels of observational units. If instead a non hierarchical model is used with many parameters then the model produced may fit the data

well, but predictions for new data are inferior (overfitted). In turn, hierarchical models use a predefined distribution to impose some dependence into parameters, this allows more parameters to be included without facing problems of 'overfitting' (Gelman et al., 2004).

Bayesian modelling can be used to deal with the complex situations arising from missing or censored data, which are often impossible or difficult to handle in a frequentist framework. Commonly, in the classical framework, missing data are discarded from the QCMD scoring analysis and due to asymptotic problems censored data may make the statistical analysis difficult to handle. One advantage of the use of Bayesian models is that they are not based on asymptotic theory and with the use of probability distributions for missing and censored information the data can be fitted in a more appropriate manner without discarding important information.

So, the main advantages of Bayesian models can be summarised as follows: the models take into account the uncertainty in multiparameter settings; no particular point estimate is required, since uncertainty is measure in terms of probability; missing data imputation is handled using the missing at random assumption; and hierarchical modelling avoids problems of overfitting (David et al., 1986; Lindley, 1965; Liu, 1995; Little and Rubin, 1992). However, the complex computational issues when fitting the models (Gamerman and Lopes, 2006) including the time needed to obtain the results from the model fitting, the expertise required to choose appropriate models and the difficulties deriving their results make Bayesian statistics less attractive to researchers, particularly to non-statisticians. Although nowadays software has been developed to provide applicants of Bayesian statistics with a simpler tool to model their data (The BUGS Project, 1996-2004), it is still necessary to have a strong knowledge of computing programming languages and statistics to interpret the results. In addition to implementing general models using software, specific issues with the data have to be handled manually and new models have to be developed and programmed (Gilks et al., 1996), which is time consuming and requires a broad knowledge of computing and statistics.

1.3.2 Modelling from Frequentist Perspective versus Bayesian Perspective

The discipline of statistics is divided between the frequentist and the Bayesian schools. The frequentist approach is based on the concept of 'frequency probability', which is derived from the observed frequency distribution or proportions of populations (population parameters are fixed) assuming the experiment could be repeated under the same conditions arbitrarily often.

In contrast, the Bayesian approach is based on the representation of the uncertainty under study by the use of probability distributions where the parameters to be estimated are treated as random variables (Gelman et al., 2004). Therefore, the Bayesian approach does not require the assumption of repeatability of the experiment under the same conditions.

Moreover, in testing an hypothesis, while frequentist inference obtains the probability of the data given that the null hypothesis H_0 is true, Bayesian inference obtains the probability of H_0 given the observed data which is, some argue, more in line with commonsense interpretations.

Furthermore, Bayesian techniques allow the incorporation of prior information in the analysis, which is updated by the information obtained in the experiment. Thus, Bayesian inference has formalised the process of learning from data by updating prior beliefs with our recent knowledge (Congdon, 2001). However, not all statisticians agree about incorporating prior assumption via a prior density, claiming that doing so is subjective. In turn, they believe that the observed data information should be kept separate from prior assumption, which typically reduces Bayesian estimation to likelihood estimation theory (frequentist perspective) (Davison, 2003). Bayesian techniques from an objective perspective, where the prior information is not subjective, have been developed and a summary of this can be found in the literature (Berger, 2000).

Nevertheless, all statistical methods are subjective, in the sense of relying on idealizations of the world, in a form of a specified likelihood. In addition, all inferential processes require a priori scientific judgment that motivates the design and the study itself.

1.3.3 Generalised Linear Models

In this thesis participants' EQA performance is analysed based on qualitative or quantitative responses depending on whether they test for detection of the microbe or quantification of the sample microbial load. Generalised Linear Models (GLMs) (Dobson, 1990) are fitted to the data with possible associated covariates included in the models.

GLMs are statistical models that allow the response variable to follow any distribution that is a member of the "exponential family", which include the Normal, Binomial, Poisson and Gamma distribution. The distribution of the response variable, denoted by y, can be mathematically described by (Davison, 2003):

$$f(y|\theta,\phi) = \exp\left\{\frac{y\theta - b(\theta)}{\phi} + c(y,\phi)\right\},$$

where θ depends on the linear predictor, and ϕ is the dispersion parameter. That is, the density of the response belongs to the exponential family, where the functions $b(.)$ and $c(.)$ are specific functions corresponding to the type of exponential family.

In addition, under the GLMs the mean of the response variable can be modelled as a linear function of the explanatory variables via a link function, i.e.: η is a linear predictor,

$$\eta = g(\mu) = \vec{x}^T\vec{\beta}, \tag{1.3}$$

where $\mu = E(y)$, $\vec{x}$ is the vector of explanatory variables and $\vec{\beta}$ the vector of parameters. The function g is a monotone link function that relates μ to the linear predictor η.

Linear Regression Models

The simplest form of the GLM occurs when the response variable y is continuous and its density is from a normal distribution with mean μ and variance σ^2 (Dobson, 1990). In this case, the mean can be expressed as follows: $E(y) = \mu = \vec{x}^T\vec{\beta}$.

Thus, the link function g in equation (1.3) is the identity function since $\eta = \mu$, and therefore the regression model is called Linear Regression Model.

Logistic Regression Models

The logistic regression model is a GLM when the response variable y is a binary response that follows a Bernoulli distribution. In that case, $E(y) = p$, which is the probability of success, and the natural link function is given by the logit function (Fahrmeir and Tutz, 2001), i.e.:

$$\eta = g(p) = \text{logit}(p) = \log\left(\frac{p}{1-p}\right). \tag{1.4}$$

1.3.4 Generalised Linear Models from a Bayesian Perspective

GLMs can be approached from a classical point of view or from a Bayesian perspective. From a classical point of view, parameters are estimated as unknown constants. The parameter inference takes the form of point estimates, confidence intervals, hypothesis tests, predictions or decisions. Therefore, the problem is to take into account the random nature of the data and give an interpretation of the results.

The probability for the parameters cannot be discussed because the only random elements in the model are the data. However, GLMs from a Bayesian perspective estimate the parameters as random quantities, which fits better the natural approach of using probability distributions as a measure of uncertainty. Parameter inference is carried out by the use of the distributions of the parameters given the data, which implies a probabilistic and natural way of making inference. On the other hand, to achieve this it is necessary to make prior beliefs about the parameters to estimate in order to obtain their posterior distribution. Thus, the posterior distribution represents the prior beliefs of the parameters given the data (Lee, 2004).

In addition, the Bayesian approach to GLMs allows the inclusion of information which GLMs in classical analysis would discard due to the partial lack of information, as for example censored data, and/or avoids asymptotic problems from arrising frequentist theory (McCulloch and Searle, 2001; Spanos, 1999; McCullagh and Nelder, 1989). Despite the difference between the classical and Bayesian point of view of GLMs, similar conclusions are often found from both approaches for simple analyses. However, Bayesian methods allow for more flexible assumptions to be considered when modelling data and can be extended to more complicated models (Gelman et al., 2004).

1.3.4.1 Model Selection Strategy

When developing a multivariable model a list of prespecified variables are included in the regression model. However, in the desire to develop a concise model and to avoid collinearity between variables, when presenting the results 'insignificant' regression coefficients are normally removed. In this case, to assist with the selection of the significant variables for the model, 'stepwise selection methods' are commonly used (Harrell, 2001).

The stepwise method is a very popular technique which has been developed to identify appropriate subsets of models (with a reduced number of predicted variables from the prespecified list proposed at the beginning of a study), with considerable less computing issues and efforts than when all possible variables are included. The forward selection procedure chooses the variables by adding one at a time starting with the one that accounts for the largest amount of variation in the dependent variable. The backward elimination procedure starts with the full model and eliminates at each step the one variable that causes the residual sum of squares to increase the least. Neither of these two methods takes into account the effect that the addition or elimination of a variable can have on the other variables. The stepwise selection procedure is a combination of forward and backward elimination criteria by adding or deleting (depending on the method), sequentially, the variable that has the greatest impact on the residual sum of squares (Rawlings et al., 1998).

The use of stepwise selection methods has advantages and disadvantages. Sometimes a backward elimination procedure is preferred because it performs better than forward stepwise methods when collinearity is present and it examines the full model fitted. This provides accurate standard errors and p-values since full models have the advantage of providing meaningful confidence intervals using standard formulae (Harrell, 2001). However, economists normally use the strategy of deleting only those variables that are not significant, which have some issues in some biological problems by setting certain regression coefficients to zero.

Regression models from a Bayesian perspective are developed in this study. Ideally all relevant information should be included in a statistical model, which in regression means including all possible explanatory variables. Since by applying hierarchical structure in Bayesian analysis the problem of overfitting is of less concern, we consider first the full model with all possible covariates. This approach is based on the Bayesian perspective that to study a full model is the main objective even if some variables are not statistically significant (Gelman et al., 2004, p. 263). Then, the full model is reduced to a simple one to show how

from a Bayesian perspective model reduction and variable selection are performed. In order to do so, a backward elimination procedure was applied. In particular, a combination of a backward selection procedure and decision based on the conditional posterior distribution obtained by modeling the data is used to reduce the full model to the best simple model. In the simple model (reduced model), the 'best' list of independently important predictors are selected only. The decision of which covariate will be removed from the model is based on the conditional posterior distribution and the probability that the parameter associated to that covariate takes the value zero (as an equivalent to what economist frequently do by reducing those covariates with non-significant parameter).

Multiple Testing

One area of concern in classical theory is the interpretation of significance levels of multiple tests. In the case, a multiple testing adjustment is applied and an adjusted p-value needs to be specified. In Bayesian analysis, such adjustments are not necessary because of the use of posterior predictive checks based on the posterior distributions. A particular aspect of the data, that is expected to appear in replications, is studied via the posterior predictive distribution of the parameter: being checked if in future replications it is possible that the parameter takes on the value 0. Thus, in Bayesian application there is no concern about the p-value (defined as the probability of falsely rejecting the null hypothesis when it is true) as with classical theory. In fact, the posterior checks are used 'to understand the limits of its applicability in realistic applications' (Gelman et al., 2004).

1.3.4.2 Model Checking

The goodness of fit and robustness of the model is assessed through the examination of residuals. In the Bayesian approach, examining the residuals is equivalent to estimating the posterior predictive distributions of the residuals, in contrast to the classical estimated residual which is based on a point estimate.

The distributions of the residuals are checked, and the 95% highest density interval lying within -2 and 2 indicates a good fit of the model (Gelman et al., 2000).

To assess a model, rather than fixing the test statistic at some point estimate, the posterior predictive distribution of an appropriate test statistic is obtained. Thus, Bayesian techniques do not use the asymptotic distribution of a test statistic based on large-sample approximation and do not employ the classical theory for a test statistic and hypothesis testing. Furthermore, in Bayesian methodology, the test statistic used to produce a valid p-value has no sample size related restriction (Gelman et al., 2004).

One basic technique used to check the fit of a model is to define a 'test quantity' which measures the discrepancy between the replicated data from the model and the observed data. In Bayesian analysis, the test quantity has the same role as a statistical test in frequentist data analysis. Note that there is no general agreement on how to choose the function for the 'test quantity'. This function should be defined in a sensible way according to the characteristic of the data to be analysed (Gelman et al., 2004).

The main difference of a test statistic in the Bayesian framework compared to the classical perspective is the use of the model parameters under their posterior distributions, as well as the data, to summarise discrepancies between the model and the data. Thus, because of the use of the posterior distributions of the model parameters, the test quantity in the Bayesian framework is more flexible than a test quantity in the classical approach.

The lack of fit of the model to the data can be measured by the tail-area probability of the posterior predicted distribution (the equivalent to p-value in classical statistics). In Bayesian analysis, the probability that replicated data could be more extreme than observed data is defined as the Bayesian 'p-value'. To obtain the Bayesian 'p-value', a predictive replicate of

the data, y^{rep}, is drawn from the predictive distribution conditional on each simulated parameter θ. Thus, the draws are considered from the joint posterior distribution, $p(y^{rep}, \theta|y)$. Then, the density of the test quantity based on the replicated data (predictive test quantity) is obtained. This is used to confirm that the distribution of the test quantity contains the observed test value, which is employed to obtain the Bayesian 'p-value'.

The test quantity of the observed data is now compared with the predictive test quantity. Thus, the tail-area probability of the posterior predictive distribution is the proportion of the simulations for which the test quantity exceeds the observed value of the test. Tail-area probabilities close to 0 or 1 indicate that the observed data are unlikely to be replicated data if the model is true. In that case, the position of the observed test quantity is in the tail of the posterior predictive distribution of the test statistic.

1.3.4.3 Model Comparison

Model comparison is an important issue when choosing between nested models. Usually, a larger model has the advantage of fitting the data better, since the features of the data are explained by a larger range of variables. Larger models are also more appropriate when the aim of the analysis is to find out influential variables. On the other hand, larger models have the disadvantage of being more difficult to interpret, compute and are liable to contain more covariates with missing values. Also, larger models may be overfitted, i.e., the model fits very well the observed data but it cannot detect the underlying process and so predictions are not good. Therefore, if the purpose of the model is not only to find factors that are significantly associated with participants' performance, but to obtain future predictions, then a reduced model is more suitable, easier to interpret, to compute and the measure of the effect of an exploratory variable on the results is more precise.

Bayesian model comparison does not aim to tell the "true" model, but rather to find the model that fits better given the data and the information provided. In a Bayesian framework there are several methods that can be used for model comparison. These include posterior model probabilities, Bayes factor and approximations such as the Bayesian Information Criterion (BIC), the Deviance Information Criterion (DIC), and Bayesian Model Averaging (BMA) (Congdon, 2001).

In the present application, posterior probabilities, the Bayesian Information Criterion (BIC) for censored models (Volinsky and Raftery, 2000) and the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002) are used for model comparison. The posterior model probabilities are used to compare nested models (comparing models with different regression parameters, Chapter 4 and 5). For model comparison, a smaller value of BIC or DIC indicates a better fit. Therefore, the selection procedure is similar to the one based on the Akaike's Information Criterion (AIC). A new approach based on the DIC, which takes into account model fit and number of parameters in the model, is used to select the best model amongst two different proposed models (Chapter 6 and 7).

The BIC is defined as

$$BIC = logL(\bar{\theta}|y) - p/(2log(c)),$$

where θ is the p-dimensional vector of parameters, $\bar{\theta}$ the posterior mean of θ, c the number of non-censored observations and p the number of parameters in the model. The BIC formula represents the log of the likelihood function evaluated at the posterior mean of the vector of parameters θ and a penalisation based on the number of parameters and the total number of non-censored observations.

The DIC is calculated as distances measures

$$DIC = \text{“goodness of fit”} + \text{“complexity”}$$

$$DIC = Mean[-2logL(\theta|y)] + \{Mean[-2logL(\theta|y)] - (-2logL(\bar{\theta}|y))\}. \quad (1.5)$$

To obtain the DIC, the average of the log-likelihoods at each iteration of the Markov Chain Monte Carlo (MCMC, which will be presented later on) and the log-likelihood given the posterior means of the parameters are calculated. It can be shown that the DIC is a measure of the predictive accuracy $Mean[-2\log(L(\theta|y))]$ and the ‘effective number of parameters’ of the model given by $p_D = Mean[-2\log(L(\theta|y))] - (-2\log(L(\bar{\theta}|y)))$ (Spiegelhalter et al., 2002).

1.3.4.4 Model Validation: Sensitivity Analysis

Sensitivity analysis is needed in order to check for uncertainty in posterior inferences. This uncertainty may be due to the existence of alternative models. That is, other reasonable models can fit the observed data equally well, but posterior inferences may be different, so the model proposed in that case would not be robust. In other words, changes on the model assumption may produce different results.

The basic method of sensitivity analysis is to fit different models to the same problem (Gelman et al., 2004) and check the posterior distributions (posterior inferences). Alternative models can differ in the likelihood function, prior distribution or in both, likelihood and prior. In this thesis, our main interest is to investigate if changes of prior knowledge would affect the posterior conclusions, therefore several models with different prior distribution are proposed and applied to the data in order to check for the robustness of results.

Note that changes of the likelihood function are not considered since the proposed ones are the most reliable and appropriate, and model checking provided valid results. Therefore a sensitivity analysis focusing on the prior distribution was the approach taken to study model sensitivity.

1.3.5 Markov Chain Monte Carlo Methods

Bayesian inference and computation have the problem of finding the normalising factor given in equation (1.2) to ensure that the integration of equation (1.1) equals one. That is, to integrate the product of the likelihood and the prior over the space of elements of θ given by (1.2), which in a multidimensional space is a highly non-trivial numerical problem. This is mainly because in a multidimensional parameter space we may want to know, for example, the marginal distribution of each component or the posterior mean, which implies a complex numerical integration problem. When it is not possible to work out the closed form of the marginal posterior distribution because we are not able to analytically integrate the posterior density, the use of methods for the calculation of the conditional and marginal distributions and their moments are needed (Geyer, 1992).

Recent developments in the computing environment facilitate the application of Bayesian methods in the fields of biostatistics and elsewhere. This is characterised by the use of computer intensive sampling methods for parameter estimation (Gilks et al., 1993; Heath, 1997). Such an example are the Markov Chain Monte Carlo (MCMC) methods (Metropolis et al., 1953; Hastings, 1970; Geman and Geman, 1984) that are widely used to solve complex Bayesian inference problems, by allowing the handling of complicated distributions in multidimensional space. In general, Markov Chain simulations are complicated, but for some models, such as hierarchical models, they are still a feasible way to obtain consistent results (Gelman et al., 2004).

Markov Chain Monte Carlo (MCMC) is a class of simulation algorithms that has been largely developed during the nineties. MCMC methods may be used to simulate from complex distributions and can be explained as a combination of two mathematical and computational tools: Markov Chain and Monte Carlo methods. The Monte Carlo method (Fishman, 1996) is used in the draws of the posterior distributions obtained from the Markov Chain in order to extract properties of the desired posterior distributions of the parameters to be estimated.

The Monte Carlo integration method (Kloek and van Dijk, 1978) solves the integration problem by estimating the complex integral with realisations of repeated random simulations from the posterior distribution. In a general context and notation Monte Carlo integration works as follows: If X is a random variable with density function $f(x)$, it is of interest to evaluate $E(g(X))$ for some function $g(X)$. Then by definition

$$E(g(X))) = \int_X g(x)f(x)dx. \tag{1.6}$$

If $x_1, .., x_n$ are simulated realisations of X, the integral in equation (1.6) can be approximated by

$$E(g(X)) \approx \frac{1}{n}\sum_{i=1}^{n} g(x_i).$$

Furthermore, when it is not possible to simulate realisations of X, but instead to simulate realisations of a random variable Y, $y_1, .., y_n$, with the same support space as X and density function $h(y)$, then

$$E(g(X)) = \int_X g(x)f(x)dx = \int_X \frac{g(x)f(x)}{h(x)}h(x)dx,$$

which can be approximated by

$$E(g(X)) = \frac{1}{n}\sum_{i=1}^{n} \frac{g(y_i)f(y_i)}{h(y_i)}.$$

In order to use the Monte Carlo method to obtain approximations of marginal posterior distributions and their moments, simulation of random quantities from some standard distributions are required. By using a Markov chain (Gamerman and Lopes, 2006), which is a stochastic process with the property that the future states are independent of the past states given the present states (that is the past states do not provide information about the future state if the present state is known), simulated realisations from posterior distributions can be obtained. The following formula represents the general equation to express a first order Markov chain

$$\pi(\theta^{n+1}|\theta^n, \theta^{n-1}, .., \theta^0) = \pi(\theta^{n+1}|\theta^n).$$

In this thesis a Markov chain is defined as the process of obtaining successive simulation quantities from the conditional posterior distributions of the parameters to estimate, depending only on their immediate predecessors. The realization of the chain is iterated until convergence to a stationary distribution is achieved. This stationary distribution is the posterior distribution of the parameters to be estimated. Therefore, after convergence, the draws obtained in the iterations are in equilibrium, and they can be used as a sample of the target distribution (Gamerman and Lopes, 2006). Then, the Monte Carlo method is used to sample the draws of the posterior distribution obtained from the Markov chain in order to calculate properties of the desired posterior distribution of the parameters to be estimated.

There is a wide range of MCMC algorithms available, the choice of which depends on the model to be fitted (Gilks et al., 1996). The priority when choosing an algorithm is to be efficient in computational aspects, in terms of computational costs and the number of iterations it takes for the chain to converge (Gamerman and Lopes, 2006). In this thesis, the well-known Metropolis-Hasting (M-H) algorithm is used for simulating properties from a density of interest (Gamerman and Lopes, 2006). This algorithm generates a chain where at each stage a new value is simulated from a proposal distribution (a distribution which it is easy to simulate a value from). This generated value is either accepted, in which case the chain moves, or rejected, in which case the chain stays where it is. Whether or not

the value is accepted or rejected depends on an acceptance probability which depends on the relationship between the density of interest and the proposal distribution. A general description of the M-H algorithm is provided below: If $\pi(\theta)$ is the density of interest and $q(\theta, \phi)$ is a proposal distribution from which it is easy to simulate

- 1. Initialise the interaction counter of the chain to j=1, and intialise the chain to θ^0.
- 2. Generate a proposed value ϕ using $q(\theta^{j-1}, \phi)$, called the 'candidate value'.
- 3. Evaluate the acceptance probability $\alpha(\theta^{j-1}, \phi)$ of the proposed move, where

$$\alpha(\theta, \phi) = min\left\{1, \frac{\pi(\phi)q(\phi, \theta)}{\pi(\theta)q(\theta, \phi)}\right\}.$$

- 4. Put $\theta^j = \phi$ with probability $\alpha(\theta^{j-1}, \phi)$, and put $\theta^j = \theta^{j-1}$ otherwise.
- Change the counter of the chain from j to $j+1$ and return to step 2 until convergence is reached.

The samples obtained from the MCMC, under mild regularity conditions, converge to a sample from the target posterior distribution. Therefore, the draws from the algorithm can be considered realisations from the desired distribution and can be used to obtain information about the target distribution (Gamerman and Lopes, 2006).

As explained above, the idea of MCMC is to create a Markov chain for the simulation of values form the conditional posterior distribution and treat them as draws from the target distribution (posterior distribution). To make this possible, the state space for the Markov chain has to be the parameter space and the posterior distribution has to be the limit distribution. Samples from the chain may only be considered realisations from the target distribution if the following two conditions hold:

- Invariance: the Markov chain moves should leave the target distribution invariant.
- Irreducibility: all the states in which the posterior probability is positive have to be obtainable from the values simulated in the Markov chain.

When these two conditions hold, the convergence to posterior expectations are assured. This theoretical result and other distributional results are available in Barndorff-Nielsen, Cox and Kluepperlberg (2001).

The introduction presented here provides the background to Bayesian computation and to the general concept of MCMC methodology used throughout this thesis. More information may be found in the literature cited here. Further information on the concrete algorithm used here is given in Chapter 7.

1.4 Project Aim

The project aim is to develop a suitable statistical framework to assess users of molecular amplification techniques. In particular, appropriate scoring schemes are devised to assess an individual EQA participant's performance on individual panels. Furthermore, statistical models are adapted and developed in order to find factors that are significantly associated with participants' performance over time for different pathogens. The large reservoir of QCMD data, which has not been analysed previously, is used for the development and testing of the statistical methods.

Classical statistical methods, based on asymptotic theory, failed to fit the data adequately due to the amount of missing information and the random character of the observations because of the measuring technique. For this reason, the statistical approach is carried out from a Bayesian perspective. The Bayesian statistical methods have been adapted, refined and coded to fit the specific requirements of the data and the goals of the investigation.

Using the methods developed in this research project, QCMD can now interrogate the data gathered over time to improve the design of future EQA programmes to advance the quality of diagnosis of NATs users and hence improve patients' health.The techniques developed may be also applied to a broad range of problems within other areas of research, such as clinical medicine, veterinary medicine and clinical chemistry.

Chapter 2

Exploratory Data Analysis

This chapter provides a description of the QCMD data, which are used in this project for model development and testing. In particular, the datasets used are from participants of the Enterovirus (Gastrointestinal virus), Hepatitis B virus and Hepatitis C virus (Blood Borne Viruses) QCMD quality control panels from 2002 to 2005 inclusive. These panels are chosen for the following reasons:

- Different pathogens are needed to test if the proposed models work independently of the type of pathogen chosen. Pathogens which infect different areas of the human body and in a different way (via blood or food, etc.) are needed, so tests can be performed to determine whether the models are robust enough to distinguish among different types of pathogens.
- The programmes have to be run for different observation years in order to conduct comparisons over time.
- The number of participants across programmes has to be large enough to be representative and to provide enough information about laboratory practice.

QC programmes are carried out over time and information about participants' performance is gathered by QCMD. Participating laboratories vary over time, since many laboratories enroll voluntarily on QCMD quality control programmes.

The proficiency panels are provided to participants together with a questionnaire. Information about the entire laboratory is asked for, such as the technology used to analyse the sample, the method of analysis, the use of anti-contamination system, the inhibition test performed, the laboratory type, the accreditation status and the experience of analysing other specimens (biopsies, swabs...). Note that responses are frequently missing.

The application of more than one validated method of routine analysis may be used when a single determination analysis has failed, often because the variability of the assay method has not been acceptable. Thus, when precision and accuracy do not fall within acceptable tolerance limits, duplicate or even triplicate analysis may be performed to obtain a better estimate (Niazi, 2007). When a duplicate or triplicate analysis was used, this was recorded as the method of analysis in the questionnaire.

One common concern of molecular amplification methods is contamination. In this case, a highly sensitive molecular test may result in a false positive and so an increased microbial load may be reported if an anti-contamination system is not used. Furthermore, inhibition may occur since the sample can contain substances that interfere with the molecular reaction, and the lack of formation of amplified gene products that inhibit the DNA polymerase enzyme may result in a false negative or decreased microbial load reported without the use of an inhibition test. Consequently, specific anti-contamination strategies are essential to minimise the chance of anti-contamination and an inhibition test may be added to the sample to determine if an interfering substance has caused inhibition of the enzymatic reaction (Dennis Lo et al., 2006; Burtis and Ashwood, 2007).

Laboratory accreditation determines the technical competence to perform a specific type of testing. It provides a recognition of competent laboratories and it is renewed annually (ILAC, 2010). Participants may be or may not be accredited, but not all participants provide that information. Participants are classified by type of laboratory as: hospital labs, public labs, private labs, reference labs, manufacture labs and research/clinical labs.

2.1 Exploratory Qualitative Data Analysis

A description of the qualitative data analysed for this project is provided in this subsection. The description mainly involves the qualitative performance provided by participants of Enterovirus and Hepatitis B virus programmes over time, as they are a representative sample of qualitative QCMD programmes over time. Note that quantitative estimates of pathogen load are not provided by participants for Enterovirus panels.

2.1.1 Description of Enterovirus Programmes Qualitative Data

Enterovirus (EV) is a group of viruses that may infect the gastrointestinal tract and can spread to other areas such as the nervous system of humans and animals.

Here, a general description of the EV data from 2002 to 2005 is presented. EV proficiency panels consisted of 12 samples with a varying number of negative, non-EV and positive EV samples. Negative samples do not contain EV viral load, non-EV samples contain viral load of a different pathogen to EV and positive EV samples contain different EV viral loads across years. Samples are grouped by sample dilution series, non-EV and negative samples. Table 2.1 summarises the panel composition over time showing the number of samples included in the panel per sample group and year.

Table 2.1: *EV panel composition over time.*

Number of samples per group and year		**2002**	**2003**	**2004**	**2005**
Sample group by dilution series	**1x10^{-3}**	1	0	0	1
	1x10^{-4}	0	0	1	1
	1x10^{-5}	2	3	3	2
	1x10^{-6}	4	3	3	2
	1x10^{-7}	2	1	1	2
	1x10^{-8}	1	1	1	1
Non-EV		0	2	1	2
Negative		2	2	2	1

The total number of datasets returned by participants are 100, 89, 116 and 107 from 2002 to 2005, respectively. Results provided by participants are positive, negative and not determined. Not determined is reported if the assay shows an equivocal result. For the purpose of the analysis carried out for this project, participants' responses are classified as correct or incorrect depending on whether the laboratory detects the sample rightly or wrongly. Throughout this thesis, to be consistent with the approach taken by QCMD, not determined responses are interpreted as incorrect.

Table 2.2 shows the percentage of correct results per year and sample group. For the negative, non-EV and strongest (10^{-3}) sample groups, the percentages of correct results are higher than for the rest of the sample groups. In general, lower percentages of correct results per positive sample are obtained as the sample viral load decreases. It is observed that the overall highest percentage of correct results are obtained for 2003, which may be due to the fact that performance of sample groups with the lower dilution series, $1x10^{-7}$ and $1x10^{-8}$, are better than for other years.

According to the NAT method and laboratory's system, the technology used to analyse the EV samples are classified as follows: CC- Conventional PCR Commercial technologies, RTC- Real Time PCR Commercial technologies, CIH- PCR Conventional In-house technologies, RTIH- Real Time PCR In-house technologies and NASBA-NASBA technologies.

Table 2.2: *EV percentage of correct results over time per sample group.*

% Correct results		2002	2003	2004	2005	Total
Sample group by dilution series	**1x10^{-3}**	96.00	-	-	97.19	96.62
	1x10^{-4}	-	-	73.27	80.37	76.68
	1x10^{-5}	94.00	94.75	87.64	85.04	90.18
	1x10^{-6}	76.25	80.52	75.57	80.84	77.79
	1x10^{-7}	45.00	82.02	73.27	56.07	59.45
	1x10^{-8}	23.00	39.32	29.31	22.42	28.16
Non-EV		-	82.58	87.06	88.32	86.02
Negative		94.50	93.25	94.39	96.26	94.42
Total		74.25	83.25	78.44	76.40	

Table 2.3 shows the percentages of datasets analysed per technology group and the percentages of correct results per sample dilution series for each technology group during the period 2002-2005. It is observed that the most common technologies used by participants are conventional in-house technologies. In contrast, the least popular technology used by participants is NASBA. The proportions of correct results by users of real time technologies are higher than the proportions of conventional technologies for negative and non-EV samples. However, these proportions are lower for the sample with the strongest and medium dilution series when comparing RTC and CC technologies. All participants using NASBA technologies detect correctly negative, non-EV samples and samples with stronger dilution series. The smallest percentage of correct results for the sample with the weakest dilution series is obtained for the group of participants using NASBA technologies.

The data returned by participants in the questionnaire issued with the panel are summarised in Tables 2.4 to 2.6 inclusive. Table 2.4 shows the number of datasets per year and the percentage of results from participants who used a single, duplicate or other method of analysis. The percentage of results from participants who used an anti-contamination system, and the percentages of results from participants who performed an inhibition test on all samples, only negatives or did not perform an inhibition test are shown in Table 2.4. Note that, as previously stated, not all participants answered those questions in the questionnaire.

Table 2.3: *EV percentages of datasets analysed and correct results per sample group classified by technology group from 2002 to 2005.*

% Correct results		**CIH**	**CC**	**RTC**	**RTIH**	**NASBA**
% Datasets analysed per sample		59.71	4.37	3.64	30.10	2.18
% Correct results	**1x10^{-3}**	95.97	100.00	88.89	98.33	100.00
	1x10^{-4}	79.51	100.00	53.84	72.73	100.00
	1x10^{-5}	92.34	97.87	63.89	87.50	95.00
	1x10^{-6}	78.93	86.79	58.34	76.59	70.00
	1x10^{-7}	62.70	48.00	25.00	59.78	50.00
	1x10^{-8}	28.45	16.67	26.67	30.64	11.11
Non-EV		81.41	76.00	96.15	92.22	100.00
Negative		93.45	93.75	100.00	95.61	100.00

Table 2.4 also provides information about the percentage of returned results from participants who did not answer those questions. The majority of results are given by participants who performed single and duplicate analysis methods. Approximately 5% of the results from participants are missing since they did not reply to the question about the method of analysis. It is observed that for 2004 and 2005 no participant used other methods of analysis, and for these two years higher percentages of missing information for the analysis method were found. With respect to the use of an anti-contamination system, most of the results are given by participants who did not use an anti-contamination system and roughly 3.2% of the results are from participants who did not provide information about this issue. Almost 99% of the results are from participants who provided information about performing an inhibition test, and the majority of those results are from participants who did not perform an inhibition test on the samples.

Table 2.5 shows the number of datasets per year and the percentage of results from participants categorised by laboratory type and accreditation status. As in previous questions, a participant may not provide information about the type of laboratory, so Table 2.5 also reflects the percentages of results from those participants who did not answer these questions.

Table 2.4: *EV percentage of participants' results per year classified by method of analysis, use of an anti-contamination system and performance of an inhibition test. Number of datasets per year is given by n.*

Total number and percentages of participants' results		**2002**	**2003**	**2004**	**2005**	**Total**
		n=100	n=89	n=116	n=107	n=412
Analysis	**Single**	35.00	46.10	48.30	45.80	43.94
	Duplicate	58.00	47.2	42.2	47.7	48.54
	Other	6.00	5.60	0.00	0.00	2.67
	Not answered	1.00	1.10	9.50	6.50	4.85
Anti-contamination	**Yes**	16.00	19.10	24.10	20.60	20.15
	No	84.00	78.70	71.60	73.80	76.70
	Not answered	0.00	2.20	4.30	5.60	3.15
Inhibition test	**Yes**	27.00	36.10	39.70	42.10	36.41
	No	66.00	52.80	56.00	53.30	57.04
	Only Negatives	7.00	9.00	3.40	2.80	5.34
	Not answered	0.00	2.20	0.90	1.90	1.21

Table 2.5: *EV percentage of participants' results per year classified by laboratory type and accreditation status. Number of datasets per year is given by n.*

Total number and percentages of participants' results		**2002**	**2003**	**2004**	**2005**	**Total**
		n=100	n=89	n=116	n=107	n=412
Laboratory	**Hospital**	58.00	49.40	40.50	50.50	49.27
	Public	15.00	15.70	12.90	8.40	12.86
	Private	4.00	4.50	4.30	2.80	3.88
	Reference	6.00	0.00	5.20	0.90	3.15
	Manufacture	6.00	1.10	0.90	1.90	2.43
	Research/Clinic	4.00	2.20	3.40	1.90	2.91
	Not answered	7.70	24.00	38.00	36.00	25.48
Accreditation	**Yes**	45.00	32.60	31.90	31.80	35.19
	No	42.00	39.30	38.80	35.50	38.83
	Not answered	13.00	28.10	29.30	32.70	25.97

Roughly 50% of results are given by participants from a hospital laboratory followed by public laboratories with almost 13% of results. The percentages of results returned by participants from private, reference, manufactures and research laboratories are around 3% for each laboratory type. There is a high percentage, 25.48%, of results from participants who did not provide information about the laboratory type. The lowest percentage of missing information about participants' laboratory type is found for 2002. Information about the accreditation status of the participants is also missing for almost 30% of the results, with the proportions of results from accredited and non accredited participants being similar over time.

Table 2.6 shows the number of datasets per year and the percentages of participants grouped according to performing annual tests of different types of specimens, such as a plasma test, cerebrospinal fluid (CSF), serum, biopsies and swabs. Between 25% and 30% of participants did not provide information about performing annual tests of the different types of specimens. Participants performing a plasma test are classified by the number of tests performed annually. Approximately 29% of the results are given by participants who performed between 101 and 1,000 tests annually, and only 1.4% of the results are from participants who performed more than 10,000 tests annually. More than 67% of the results are from participants who had experience with testing CSF samples. The percentages of results from participants with experience with swabs, serum and biopsies tests are around 35%. In general, the percentage of missing information about performing annual tests of different types of specimens increases over time.

Table 2.6: *EV percentage of participants' results per year classified by specimen test experience. Number of datasets per year is given by n.*

Total number and percentages of participants' results		**2002**	**2003**	**2004**	**2005**	**Total**
		n=100	n=89	n=116	n=107	n=412
Number of plasma test	**0-10**	0.00	0.00	31.00	28.00	16.02
	11-100	13.00	19.10	20.70	19.60	18.20
	101-1,000	65.00	38.20	8.60	7.50	28.40
	1,001-2,000	5.00	11.20	6.00	5.60	6.80
	2,001-10,000	6.00	4.50	1.70	3.70	3.40
	>10,000	0.00	0.00	2.60	2.80	1.45
	Not answered	11.00	27.00	29.30	32.70	30.58
CSF	**Yes**	87.00	65.2	58.60	59.80	67.23
	No	5.00	7.90	12.10	7.50	8.25
	Not answered	8.00	27.00	29.30	32.70	24.51
Serum	**Yes**	39.00	28.10	34.50	42.10	36.16
	No	53.00	44.90	33.60	25.20	38.59
	Not answered	8.00	27.00	31.90	32.70	25.24
Biopsies	**Yes**	30.00	27.10	36.40	34.60	32.28
	No	62.00	46.00	34.50	32.70	43.69
	Not answered	8.00	27.00	29.30	32.70	24.51
Swabs	**Yes**	37.00	27.00	34.50	40.20	34.95
	No	55.00	46.00	36.20	27.10	40.53
	Not answered	8.00	27.00	29.30	32.70	24.51

2.1.2 Description of Hepatitis B Virus Programmes Qualitative Data

Hepatitis B virus (HBV) is the causative agent of viral hepatitis type B, a form of liver inflammation. Here, a general description of data of Hepatitis B virus (HBV) QC programmes from 2002 to 2005 is provided. Note that participants may return qualitative and quantitative responses for this pathogen.

The HBV proficiency panels consist of eight samples per year. There is one negative sample per year and the positive sample are classified depending on the viral load as: sample groups 6, 5, 4, 3.5, 3, and 2.3 $\log_{10}$ copies/ml. The positive samples are either of subtype A or D.

Table 2.7 shows the panel composition over time, the number of samples per group, N, and the sample subtype. The total number of datasets per year and sample are 96, 87, 107 and 123 from 2002 to 2005, respectively.

Table 2.7: *HBV panel composition over time. Number and subtype of samples included in the panel each year by group of viral load.*

Number of samples per group and year		**2002**		**2003**		**2004**		**2005**	
		N	Subtype	N	Subtype	N	Subtype	N	Subtype
Sample group $\log_{10}$ **Copies/ml**	6	1	D	1	A	1	A	1	D
	5	2	A/D	2	A/D	2	A/D	2	A/D
	4	1	A	1	A	1	A	2	D
	3.5	2	A	1	A	0	-	0	-
	3	0	-	1	A	2	A/D	2	A/D
	2.3	1	A	1	A	1	A	0	-
Negative		1	-	1	-	1	-	1	-

Table 2.8 shows that the percentage of correct results per sample group and year generally increases over time and as the sample viral load increases. Less than 70% of the overall results are correct for the sample group with the weakest viral load. Around 88% of the results are correct for the sample groups of 3 and 3.5 log_{10} copies/ml viral load. Less than 5% of the results are incorrect for the stronger samples.

Table 2.8: *HBV percentage of correct results over time per sample group.*

% Correct results		**2002**	**2003**	**2004**	**2005**	**Total**
Sample group $\log_{10}$ **Copies/ml**	6	98.96	100.00	97.20	99.19	98.79
	5	91.67	97.13	96.26	96.34	95.76
	4	87.50	93.10	96.26	96.34	94.22
	3.5	86.46	91.95	-	-	89.07
	3	-	87.36	92.52	83.33	87.57
	2.3	70.83	62.07	72.90	-	68.97
Negative		97.92	97.70	96.26	95.93	96.85
Total		88.93	90.80	92.64	93.60	

The technologies used to analyse the samples are classified in different groups. Participants' results of HBV programmes are grouped according to the following technologies: CC- Conventional PCR Commercial technologies, RTC- Real time PCR Commercial technologies, CIH- Conventional PCR In-house technologies, RTIH- Real time PCR In-house technologies, bDNA- bDNA technologies, HC- Hybrid Capture technologies and TMA- Transcription mediated amplification technologies.

Table 2.9 shows the percentage of datasets per sample analysed by each technology type and the percentage of correct results per technology and sample group. Less than 5% of the results are provided by HC and TMA technologies users. Between 15% and 20% of the results are returned by in-house technologies users, and a similar percentage of results are given by real time commercial technologies users. The most widely used technology is CC, almost 38% of the results are provided by CC users. The proportion of correct results obtained with real time technologies is higher than the proportion with conventional technologies for all sample groups except for the negative samples. bDNA and HC technologies users obtained lower percentages of correct results than PCR users, although they are amongst the less popular methods used by participants.

Table 2.9: *HBV percentages of datasets analysed and correct results per sample group classified by technology group from 2002 to 2005.*

% Correct results		**CIH**	**CC**	**RTC**	**RTIH**	**bDNA**	**HC**	**TMA**
% Datasets analysed		15.49	37.77	18.64	19.61	5.56	2.41	0.48
% Correct results	6	97.40	98.72	100.00	100.00	95.65	100.00	100.00
	5	96.75	98.08	99.22	98.77	69.57	65.00	100.00
	4	94.38	97.95	99.03	97.14	76.67	8.33	100.00
	3.5	89.71	100.00	90.91	98.21	20.00	7.69	100.00
	3	83.72	93.50	93.16	86.79	58.06	14.29	0.00
	2.3	63.08	76.92	84.00	78.95	0.00	12.50	100.00
Negative		97.40	99.36	95.31	95.06	91.30	90.00	100.00

Participants in HBV programmes are asked to fill in a questionnaire about their laboratory practice, but, in general, not all participants provide all the information. Table 2.10 shows for each of the four years, the number of datasets per sample and the percentages of participants' results classified by method of analysis, use of an anti-contamination system and performance of an inhibition test, as defined in the previous section. Less than 3% of the results are from participants who failed to answer the questions. The majority of the results are from participants who performed a single method of analysis. More than 50% of the results are given by participants who used an anti-contamination system and performed an inhibition test.

Table 2.10: *HBV percentage of participants' results per year classified by method of analysis, use of an anti-contamination and performance of an inhibition test. Number of datasets per year is given by n.*

Total number and percentages of participants' results		**2002** n=96	**2003** n=87	**2004** n=107	**2005** n=123	**Total** n=413
Analysis	**Single**	60.40	63.20	69.20	74.00	67.31
	Duplicate	33.30	32.20	25.20	20.30	27.12
	Other	6.30	3.40	4.70	4.10	4.60
	Not answered	0.00	1.10	1.90	1.60	1.21
Anti-contamination	**Yes**	50.00	49.40	56.10	58.50	53.99
	No	43.80	47.10	42.10	40.70	43.10
	Not answered	6.30	3.40	1.90	0.80	2.90
Inhibition test	**Yes**	50.00	62.10	64.50	56.90	58.35
	No	40.60	27.60	30.80	34.10	33.41
	Only Negatives	5.30	8.00	1.90	7.30	5.81
	Not answered	3.10	2.30	2.80	1.60	2.42

Table 2.11 shows the number of datasets per sample and the percentages of results per year from participants by laboratory type and accreditation status. Approximately 42% of the results are from accredited participants and 36% from participants who were not accredited. However, more than 20% of the results are given by participants who failed to answer the question. Participants are classified by type of laboratory as in previous section. Almost 22% of the results are given by participants who did not provide information about the laboratory type.

Almost half of the results are from hospital laboratories. Only 2% of the results are from public labs and approximately 4% of the results are given by participants from reference, manufacture or research labs.

Table 2.11: *HBV percentage of participants' results per year classified by laboratory type and accreditation status. Number of datasets per year is given by n.*

Total number and percentages and of participants' results		**2002** n=96	**2003** n=87	**2004** n=107	**2005** n=123	**Total** n=413
Laboratory	**Hospital**	53.10	52.90	44.90	44.70	48.43
	Public	12.50	9.20	8.40	6.50	2.18
	Private	12.50	9.20	11.20	13.00	11.62
	Reference	5.20	2.30	0.90	2.40	3.87
	Manufacture	4.20	3.40	0.00	4.90	3.38
	Research/Clinic	0.00	2.30	7.50	4.10	3.63
	Not answered	12.50	20.70	27.10	24.40	21.55
Accreditation	**Yes**	43.80	39.10	43.90	35.80	41.65
	No	42.70	37.90	29.00	39.80	36.08
	Not answered	13.50	23.00	27.10	24.40	22.28

Participants are classified according to the information that they provide about their experience testing samples of different types of specimens. In Table 2.12 the number of datasets, n, and the percentages of participants' results classified by the experience of the participant performing annual tests of different types of specimens, such as, plasma test, serum and others (biopsies, swabs,..) are provided per year. The experience performing tests on plasma and serum samples are classified by the number of samples analysed annually. More than 20% of the results are from participants who failed to answer the questions about their experience performing tests on samples of different types of specimens. More than 50% of the results are given by participants who tested up to 1,000 samples of serum and plasma annually. Around 10% of the results are from participants who tested between 2,001 and 10,000 samples of serum and plasma annually. Approximately 55% of the results are provided by participants who did not have experience testing samples of other specimens such as biopsies and swab.

Table 2.12: *HBV percentage of participants' results per year classified by specimen test experience. Number of datasets per year is given by n.*

Total number and percentages and **of participants' results**		**2002**	**2003**	**2004**	**2005**	**Total**
		n=96	n=87	n=107	n=123	n=413
Number of plasma test	**0-10**	32.30	29.90	25.20	22.80	27.12
	11-100	9.40	17.20	5.60	4.90	8.72
	101-1,000	28.10	26.40	17.80	18.70	22.28
	1,001-2,000	10.40	4.60	3.70	9.80	7.26
	2,001-10,000	9.40	0.00	15.90	14.60	10.65
	>10,000	0.00	1.10	6.50	4.90	3.39
	Not answered	10.40	20.70	25.20	24.40	20.58
Number of serum test	**0-10**	27.10	27.60	11.20	13.00	18.89
	11-100	11.50	12.60	7.50	13.80	11.38
	101-1,000	33.30	29.90	25.20	17.90	25.91
	1,001-2,000	14.60	5.70	11.20	8.10	9.93
	2,001-10,000	5.20	3.40	15.90	17.90	11.38
	>10,000	0.00	0.00	1.90	4.10	1.69
	Not answered	8.30	20.70	27.10	25.20	32.20
Other Specimen	**Yes**	18.80	12.60	30.80	30.90	24.21
	No	72.90	66.70	42.10	44.70	55.20
	Not answered	8.30	20.70	27.10	24.40	20.58

2.2 Exploratory Quantitative Data Analysis

2.2.1 Description of Hepatitis B Virus Programmes Quantitative Data

In the previous section, the HBV qualitative data have been described. The panel composition and the percentages of correct results have been provided, and some characteristics of the participating laboratories' practices such as technology used, method of analysis and accreditation status have been summarised. In this section, the quantitative HBV data (estimation of viral load) for the positive samples are summarised.

Table 2.7 provides information about the panel composition over time. There are six groups of samples defined by sample viral load. Participants are asked to return the estimated viral loads of the samples. However, not all participants are able to provide that information; some participants only provide the qualitative information about the samples (as described in previous sections); others provide both, qualitative and quantitative results, but, not all of the participants who provided quantitative results are able to provide an exact estimate of the viral load. Instead they produce a response, for example, such as load greater than 4.5 $\log_{10}$ copies/ml (a threshold of the sample viral load).

The threshold of the sample viral load is provided when the participant has detected a viral load below or above the limit of detection of the assay used to analyse the sample. Therefore, the information that this participant has about the viral load of the sample is incomplete, since the participant cannot provide an exact estimated value for the viral load of the sample. This partial information is statistically defined as a 'censored observation'. Thus, the censored observation provided by a participant represents the threshold of the viral load for the sample (limit of detection of the assay). The exact estimated viral load of the sample is a value below ($c \geq$ value) or above ($c \leq$ value) the threshold of the viral load reported by the participant (denoted by c).

For the purposes of the quantitative analysis, the results from participants who provided estimated viral loads of 0 are discarded since that would mean that these participants do not detect the sample correctly. The participants' results are classified as exact values, when they provide exact estimates of viral loads; left censored values ($c \geq$), when they provide information about below the limit of detection of the assay used; and right censored values ($c \leq$), when they provide information about above the limit of detection of the assay used.

Table 2.13 shows the total number n of quantitative results returned by participants, the total number of censored data n_c (without taking into account the direction of censoring) and the percentage of censored data rounded to the nearest integer, $\%_c$, per year and sample group. The sample groups are classified as weak, medium and strong depending on the sample viral load. The weakest sample group corresponds to the 2.3 $\log_{10}$ copies/ml sample group and the strongest sample group corresponds to the 6 $\log_{10}$ copies/ml sample group. The sample groups of medium viral load such as 4 and 5 $\log_{10}$ copies/ml have the lowest percentages of censored data.

Table 2.13: *HBV total and percentage of censored datasets returned by participants over time per sample group. The total number of quantitative results per year is given by n. The number of results out of the total per year that are censored is given by n_c.*

Datasets per year		**2002**			**2003**			**2004**			**2005**			**Total**		
		n	n_c	$\%_c$	n	n_c	$\%_c$	n	n_c	$\%_c$	n	n_c	$\%_c$	n	n_c	$\%_c$
Sample	6	65	8	12	62	9	15	82	19	23	100	14	14	309	50	16
group	5	130	15	12	124	5	4	164	0	0	200	3	2	618	23	4
$\log_{10}$	4	64	9	14	61	4	7	82	0	0	199	2	1	406	15	4
Copies/ml	3.5	130	19	15	61	4	7	-	-	-	-	-	-	191	23	12
	3	-	-	-	61	7	11	162	7	4	190	22	12	413	36	9
	2.3	63	17	27	58	26	45	79	21	27	-	-	-	200	64	32
Total		452	68	15	427	55	13	569	47	8	689	41	6	2137	211	10

Overall, 10% of the results are censored. As an illustrative example of the participants' quantitative results, Figure 2.1 shows the participants' reported values for all years combined by technology group for sample groups 2.3 and 6 $\log_{10}$ copies/ml. Blue triangles correspond to left censored responses and red squares indicate the right censored responses.

Figure 2.2 shows the box plots for participants' estimates of viral loads from all years combined by technology group for sample groups 2.3 and 6 $\log_{10}$ copies/ml. The box plots are obtained once the censored observations have been removed. Then, the censored observations are superimposed on the boxplots. The variability of the responses per technology changes depending on the sample group analysed. Thus, for the sample group 2.3 $\log_{10}$ copies/ml the

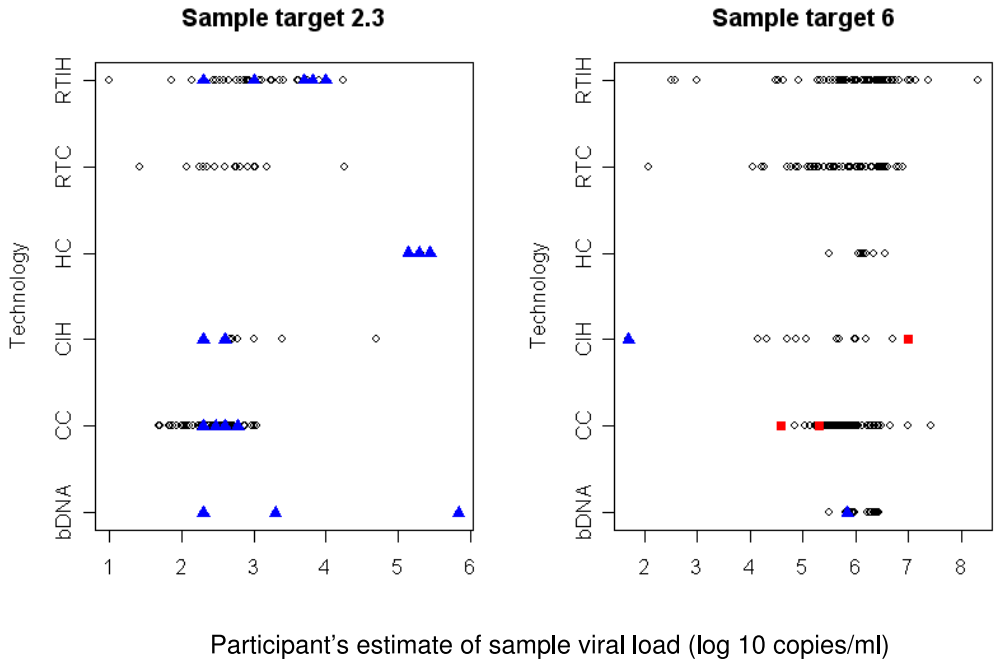


Figure 2.1: *HBV participants' reported values for all years combined, ordered by estimated viral load, for sample groups 2.3 and 6* $\log_{10}$ *copies/ml. Blue triangles and red squares are left and right censored responses, respectively.*

responses from RTC technology users have lower variability than for sample group 6 $\log_{10}$ copies/ml. These variabilities also change amongst technology groups. The majority of the censored observations are outside the box for both sample groups.

Table 2.14 shows the mean of the observed sample viral loads provided by participants (consensus mean) per sample group over time. To obtain the consensus mean, participants reporting censored observations have been excluded.

In addition to some of the technologies used for analysing HBV qualitative data, which are CC- Conventional Commercial, RTC- Real time Commercial, CIH- Conventional Commercial and RTIH- Real time In-house technologies, some participants used bDNA and HC- Hybrid Capture technologies when analysing the samples to report quantitative estimates of viral loads.

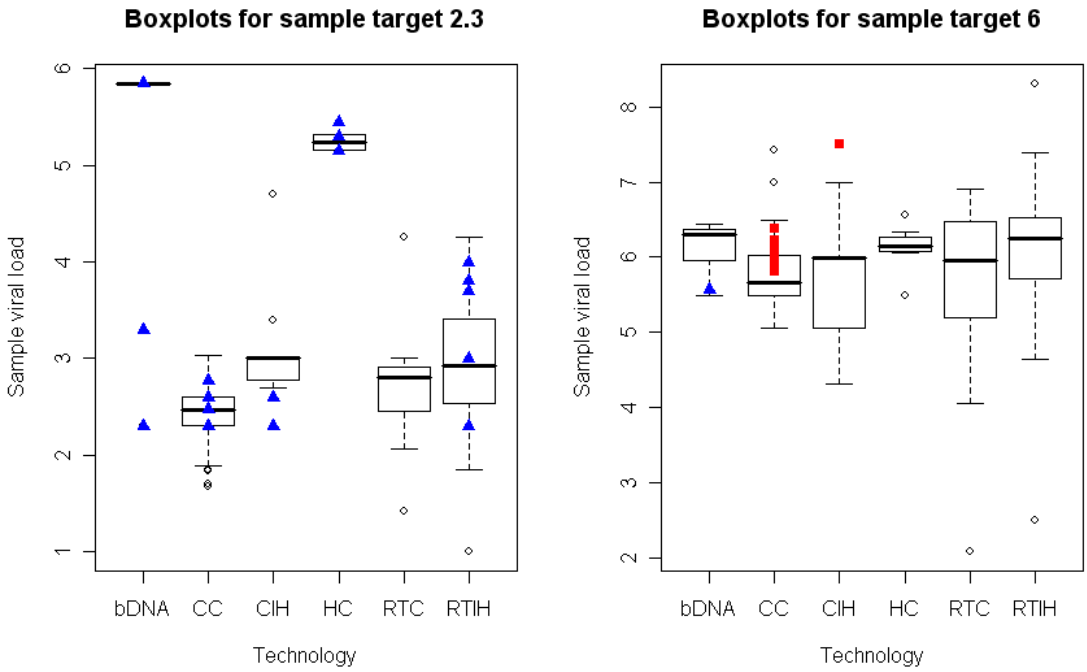


Figure 2.2: *HBV box plots of participants' reported values for all years combined by technology group for sample groups 2.3 and 6* $\log_{10}$ *copies/ml. Blue triangles and red squares are left and right censored responses respectively.*

Table 2.15 shows the number of datasets analysed and the percentage of censored information per technology and sample group. The highest percentages of censored data are provided by participants using bDNA and HC technologies with an overall of 42% and 54% of the censored observations, respectively. However, participants using those technologies return fewer censored observations than conventional technologies users for the strongest sample viral load. No more than 5% of the overall censored observations are returned by participants using CIH, RTC and RTIH technologies.

Table 2.16 shows the consensus mean per sample group by technology used. Table 2.17 shows the numbers and percentages of datasets per year by method of analysis, use of an anti-contamination system and performance of an inhibition test. Less than 3% of the re-

Table 2.14: *HBV consensus mean of estimated sample viral loads over time per sample group.*

Consensus mean by year		**2002**	**2003**	**2004**	**2005**	**Total**
Sample group	6	6.03	5.92	5.96	5.53	5.83
$\log_{10}$	5	5.14	4.94	5.05	4.71	4.93
Copies/ml	4	4.21	4.04	4.03	3.83	3.96
	3.5	4.06	3.46	-	-	3.86
	3	-	3.01	3.38	2.89	3.13
	2.3	2.80	2.53	2.68	-	2.69

Table 2.15: *HBV number of datasets and percentage of censored information per sample group classified by technology group.*

Datasets		**CC**		**CIH**		**RTC**		**RTIH**		**bDNA**		**HC**	
per Tech. group		n	$\%_c$	n	$\%_c$	n	$\%_c$	n	$\%_c$	n	$\%_c$	n	$\%_c$
Sample	6	137	34	21	10	54	0	66	0	23	4	8	0
group	5	274	1	42	0	108	0	132	0	46	30	16	25
$\log_{10}$	4	169	28	25	0	89	0	85	1	30	23	8	88
Copies/ml	3.5	103	0	15	0	2	0	46	4	15	80	10	90
	3	171	2	25	0	100	1	82	13	31	55	4	75
	2.3	104	30	13	23	18	0	43	21	16	100	6	83
Total		958	14	141	4	371	0	454	5	161	42	52	54

sults are returned by participants who failed to answer these questions. In total, almost 70% of the results are from participants who performed single analysis methods when analysing the samples. Roughly 64% of the overall results are from participants who used an anti-contamination system and a similar percentage did not perform an inhibition test. The use of an anti-contamination system and single analysis methods increases over time.

Table 2.18 summarises the number and percentages of datasets returned by participants per year classified by laboratory type and accreditation status. More than 20% of the results are returned by participants who failed to answer the questions. Almost 50% of the results are from hospital laboratories, and participation of private laboratories increases over time, so that in 2005 around 4% more of the results are returned by them compared to 2002. Almost 44% of the results are from accredited participants, with the highest percentage of results returned by accredited participants occurring in 2003.

Table 2.16: *HBV consensus mean per sample group classified by technology group.*

Consensus mean by technology group		**CC**	**CIH**	**RTC**	**RTIH**	**bDNA**	**HC**
Sample group $\log_{10}$ **Copies/ml**	6	5.74	5.70	5.68	5.96	6.15	6.13
	5	4.97	4.63	4.81	4.93	5.29	5.32
	4	4.01	3.62	3.80	4.05	4.23	3.92
	3.5	3.72	3.41	3.66	4.23	5.54	3.74
	3	3.09	2.98	3.01	3.25	3.45	5.34
	2.3	2.43	3.08	2.70	3.05	-	5.32

Table 2.17: *HBV number and percentage of datasets per year classified by method of analysis, use of an anti-contamination and performance of an inhibition test.*

Number and percentage of participants' results		**2002**		**2003**		**2004**		**2005**		**Total**	
		n	%	n	%	n	%	n	%	n	%
Analysis	**Single**	286	63.27	258	60.42	378	64.43	527	76.49	1449	67.81
	Duplicate	145	32.08	148	34.66	165	29.00	134	19.45	592	27.70
	Other	21	4.65	21	4.92	26	4.57	21	3.05	89	4.16
	Not answered	0	0.00	0	0.00	0	0.00	7	1.01	7	0.33
Anticont.	**Yes**	257	56.86	277	64.87	363	63.80	473	68.65	1370	64.11
	No	160	35.40	143	33.49	199	34.97	209	30.33	711	33.27
	Not answered	35	7.43	7	1.64	7	1.23	7	1.02	56	2.62
Inhibition Test	**Yes**	139	30.75	83	19.44	168	29.52	198	28.73	588	21.51
	No	279	61.72	110	72.60	373	65.55	414	60.09	1376	64.39
	Only Neg.	20	4.42	27	6.32	14	2.46	63	9.14	124	5.80
	Not answered	14	3.10	7	1.64	14	2.46	14	2.03	49	2.29

Table 2.19 shows the number and percentages of datasets returned depending on the experience of participants testing different specimens per year. More than 19% of the results are returned by participants who failed to answer these questions. Approximately 27% of the results are from participants who had little experience performing a plasma test (between 0 and 10 tests annually), and less than 4% of the results are from participants who performed more than 10,000 plasma tests annually. In total, 24.3% of the results are from participants who performed annually between 101 and 1,000 plasma tests. More than 50% of the results are returned by participants who had experience performing less than 1,000 serum tests annually. More than 50% of the results are from participants who did not have experience performing tests for other types of specimens.

Table 2.18: *HBV percentage of datasets per year classified by laboratory type and accreditation status.*

Number and percentage of participants' results		**2002**		**2003**		**2004**		**2005**		**Total**	
		n	%	n	%	n	%	n	%	n	%
Lab. type	**Hospital**	250	55.31	223	52.22	269	47.27	323	46.88	1065	49.84
	Public	62	13.72	42	9.84	63	11.07	27	3.92	194	9.08
	Private	35	7.74	31	7.26	63	11.07	84	12.19	213	9.97
	Reference	28	6.19	7	1.64	0	0.00	21	3.05	56	2.62
	Manufacture	21	4.65	20	4.68	0	0.00	42	6.09	83	3.88
	Research/Clin.	0	0.00	14	3.28	42	7.38	35	5.08	91	4.26
	Not answered	56	12.39	90	21.08	132	23.20	157	22.79	435	20.35
Accred.	**Yes**	189	41.81	210	49.18	266	46.75	269	39.04	934	43.71
	No	207	45.80	127	29.74	171	30.05	263	38.17	768	35.94
	Not answered	56	12.39	90	21.80	132	23.20	157	22.79	435	20.35

Table 2.19: *HBV Percentage of datasets per year classified by experience of the participants testing other specimens.*

Number and percentage of participants' results		**2002**		**2003**		**2004**		**2005**		**Total**	
		n	%	n	%	n	%	n	%	n	%
Number Plasma test	**0-10**	145	32.08	129	30.21	146	25.66	155	22.50	575	26.91
	11-100	49	10.84	42	9.84	35	6.15	20	4.06	154	7.21
	101-1,000	125	27.65	132	30.91	109	19.16	153	22.21	519	24.29
	1,001-2,000	49	10.84	27	6.32	14	2.46	77	11.17	167	7.81
	2,001-10,000	42	9.29	0	0.00	112	19.68	91	13.21	245	11.46
	>10,000	0	0.00	7	1.64	35	6.15	28	4.06	70	3.27
	Not answered	42	9.29	90	21.08	118	20.74	157	22.79	407	19.04
Number Serum test	**0-10**	132	29.20	125	29.27	56	9.84	88	12.77	401	18.76
	11-100	28	6.19	42	9.84	42	7.38	104	15.09	216	10.11
	101-1,000	160	35.40	128	29.98	151	26.54	133	19.30	572	26.77
	1,001-2,000	83	18.36	28	6.56	55	9.67	47	6.82	213	9.97
	2,001-10,000	14	3.10	14	3.28	119	20.91	132	19.16	279	13.05
	>10,000	0	0.00	0	0.00	14	2.46	28	4.06	42	1.96
	Not answered	35	7.74	90	21.08	132	23.20	157	22.79	414	19.37
Other Specimen	**Yes**	82	18.14	55	12.88	178	31.28	214	31.06	529	24.75
	No	335	74.11	282	66.04	259	45.52	318	46.15	1194	55.87
	Not answered	35	7.74	90	21.08	132	23.20	157	22.79	414	19.37

2.2.2 Description of Hepatitis C Virus Programmes Quantitative Data

Hepatitis C virus (HCV) is the causative agent of viral hepatitis type C, a form of liver inflammation. A general description of quantitative data of the Hepatitis C virus (HCV) QC programmes from 2002 to 2005 is now provided. Note that participants can return both qualitative and quantitative responses for HCV panels. As in the previous subsection dealing with quantitative HBV data, participants provide a quantitative response measure of the viral load for each positive sample. Table 2.20 shows the panel composition of HCV programmes over time. The HCV proficiency panels consist of eight samples per year. There is one negative sample per year, and the positive sample can be classified depending on the viral load. The positive samples across time are classified per sample group of viral load as follows: sample group 5.9, 4.9, 3.9, 3.5, 3.2 and 2.2 $_{10}$IU/ml. The positive samples are either of genotype 1, 3, 4, or 5. Reported values of 0 and censored observations are treated in the same way as for HBV quantitative data, outlined in Section 2.1.2.

Table 2.20: *HCV panel composition for positive samples over time. N is the number of samples per panel and year.*

Number of samples per group and year		**2002**		**2003**		**2004**		**2005**	
		N	Genotype	N	Genotype	N	Genotype	N	Genotype
Sample group	5.9	1	1	1	1	1	1	0	-
$\log_{10}$	4.9	1	3	2	1/4	2	1/3	2	3/5
IU/ml	3.9	2	1	1	1	1	1	2	1/3
	3.5	1	1	1	1	1	1	0	-
	3.2	1	1	1	1	1	1	2	3/5
	2.2	1	4	1	1	1	1	1	1

The total number of quantitative results returned by participants, n, the number of censored data n_c and the percentage of censored data rounded to the nearest integer $\%_c$, per year and sample group, are shown in Table 2.21. The sample groups can be classified as weak, medium and strong depending on their viral loads. The strongest is the 5.9 $\log_{10}$ IU/ml sample group and the weakest is the 2.2 $\log_{10}$ IU/ml sample group. The sample groups of medium and

strong viral load, such as, 3.5, 3.9, 4.9 and 5.9 $\log_{10}$IU/ml have the lowest percentages of censored data, while the extreme sample group with the weakest viral load has the highest percentage of censored data (62% of censored observations).

Table 2.21: *HCV percentage of censored datasets returned by participants over time per sample group. The total number of quantitative results per year is given by n. The number of results out of the total per year that are censored is given by n_c.*

Datasets per per year		**2002**			**2003**			**2004**			**2005**			**Total**		
		n	n_c	$\%_c$	n	n_c	$\%_c$	n	n_c	$\%_c$	n	n_c	$\%_c$	n	n_c	$\%_c$
Sample group $\log_{10}$ **IU/ml**	5.9	50	0	0	112	10	18	70	15	21	0	0	0	176	25	14
	4.9	50	0	0	56	1	1	138	0	0	177	2	1	477	3	1
	3.9	100	6	6	56	1	2	70	0	0	176	48	27	402	55	14
	3.5	50	7	14	56	2	3	69	3	4	0	0	0	175	12	7
	3.2	45	28	62	54	6	11	69	10	14	166	39	23	334	83	25
	2.2	47	31	66	40	31	77	54	36	67	75	36	48	216	134	62
Total		342	72	21	374	51	14	470	64	14	594	125	21	1780	312	17

Table 2.22 shows the mean of the estimated sample viral loads provided by participants (consensus mean) per sample group over time. To obtain the consensus mean, participants reporting censored observations are excluded.

Table 2.22: *HCV consensus mean of estimated sample viral loads over time per sample group.*

Consensus mean by year		**2002**	**2003**	**2004**	**2005**	**Total**
Sample group $\log_{10}$ **IU/ml**	5.9	5.25	5.92	6.04	-	5.75
	4.9	4.78	4.72	4.64	4.58	4.66
	3.9	3.91	3.78	3.71	3.69	3.76
	3.5	3.18	3.33	3.77	-	3.26
	3.2	2.63	3.01	2.98	2.92	2.93
	2.2	2.69	1.92	2.02	2.54	2.39

The technologies used to analyse the samples are classified as for the HBV quantitative data with the exception that the HC technology is not used by participants when analysing HCV samples. The number of datasets analysed and the percentage of censored information per technology and sample group are shown in Table 2.23. The highest percentages of censored data are returned by participants using CC and bDNA technologies with an overall of 21%

and 22% of censored data, respectively. However, for the sample group with the strongest viral load, 20% and 9%, respectively, of the reported results from users of CC and RTC technologies are censored observations, whilst non-censored observation is returned from bDNA, CIH or RTIH technologies users.

Table 2.23: *HCV number of datasets and percentage of censored information per sample group classified by technology group.*

Datasets per Tech. group		**CC**		**CIH**		**RTC**		**RTIH**		**bDNA**	
		n	$\%_c$	n	$\%_c$	n	$\%_c$	n	$\%_c$	n	$\%_c$
Sample group $\log_{10}$ **IU/ml**	5.9	121	20	6	0	11	9	17	0	21	0
	4.9	305	1	9	0	52	0	47	0	64	0
	3.9	259	20	13	0	42	0	34	3	54	4
	3.5	121	5	6	0	11	0	16	6	21	24
	3.2	210	27	7	29	38	3	30	7	49	45
	2.2	136	76	5	20	22	5	18	11	35	74
Total		1152	21	46	7	176	2	162	4	244	22

As an illustrative example of the participants' quantitative results, Figure 2.3 shows the participants' reported values for all years combined by technology group for sample groups 2.2 and 5.9 $\log_{10}$IU/ml. Blue triangles highlight the left censored responses and red squares indicate the right censored responses.

Figure 2.4 shows the box plots for participants' estimates of viral loads from all years combined by technology group for some selected samples. The variability on the responses changes depending on the sample group and technology used to analyse the samples. This variability is lower for the sample group 5.9 $\log_{10}$IU/ml than for the sample group 2.2 $\log_{10}$IU/ml. It is observed that the majority of censored observations for sample group 2.2 $\log_{10}$IU/ml is outside the box.

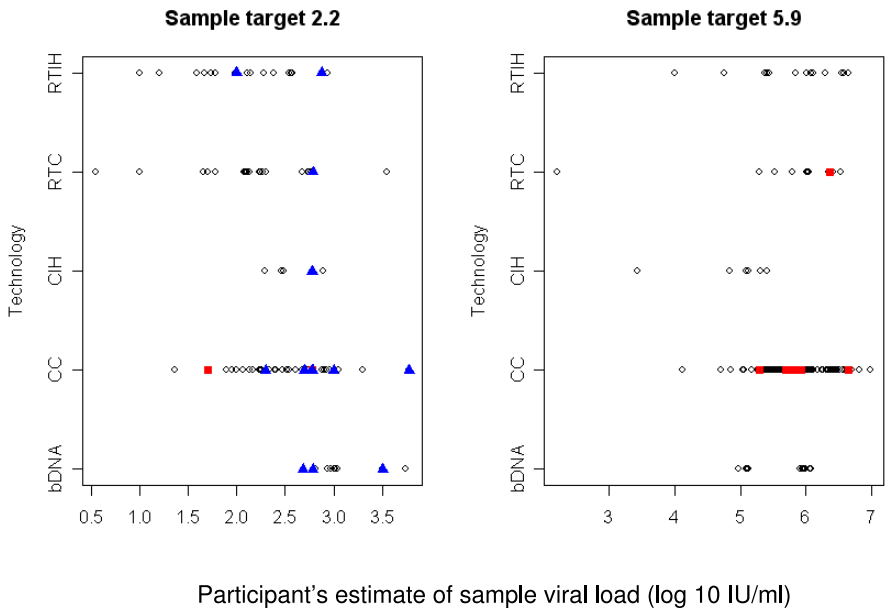


Figure 2.3: *HCV participants' reported values for all years combined, ordered by estimated viral load, for sample groups 2.2 and 5.9* $\log_{10}$*IU/ml. Blue triangles and red squares are left and right censored responses, respectively.*

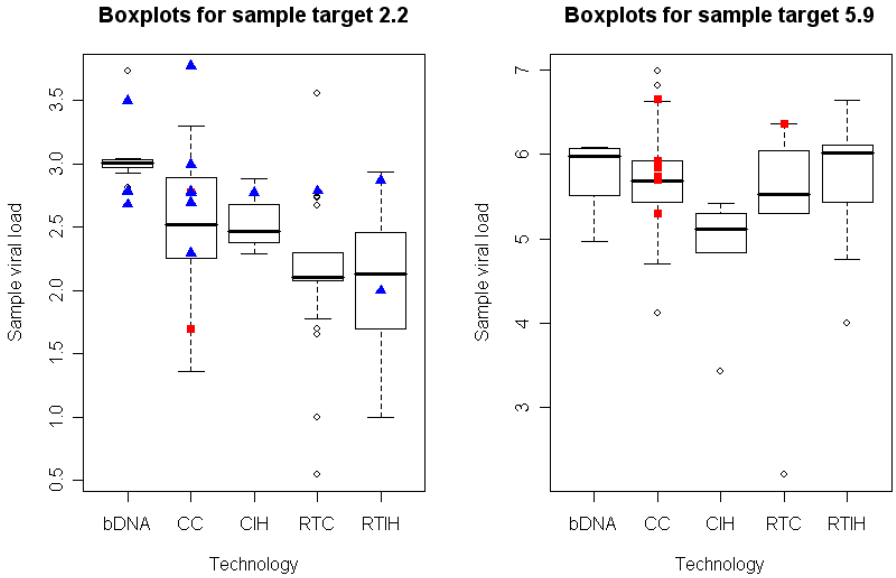


Figure 2.4: *HCV box plots of participants' reported values for all years combined by technology group for sample groups 2.2 and 5.9* $\log_{10}$*IU/ml. Blue triangles and red squares are left and right censored responses respectively.*

Table 2.24 shows the consensus mean per sample group by technology used. Except for the weakest sample group, the consensus means are lower than the target viral load (viral load of the sample group) for all technology groups.

Table 2.24: *HCV consensus mean per sample group classified by technology group.*

Consensus mean by technology group		**CC**	**CIH**	**RTC**	**RTIH**	**bDNA**
Sample group $\log_{10}$ **IU/ml**	5.9	5.78	4.86	5.62	5.83	5.78
	4.9	4.79	3.98	4.44	4.34	4.49
	3.9	3.89	3.40	3.60	3.60	3.48
	3.5	3.32	-	3.45	3.45	3.13
	3.2	3.01	2.42	2.63	2.84	3.02
	2.2	2.54	2.53	2.10	1.97	3.10

Table 2.25 shows the numbers and percentages of datasets per year by method of analysis, use of an anti-contamination system and performance of an inhibition test. The question regarding the method of analysis was answered by all participants. Less than 2.5% of the results are returned by participants who failed to answer questions referring to an anti-contamination system and an inhibition test. In total, more than 80% of results are from participants who performed a single analysis method and 65% from participants who used an anti-contamination system. Most of the results are from participants who did not perform an inhibition test, and few results are from participants who performed an inhibition test on negative samples.

The number and percentages of datasets per year returned from participants classified by laboratory type and accreditation status are shown in Table 2.26. More than 25% of the results are from participants who did not answer the questions. In total, the highest percentage of results is returned by hospital laboratories, and less than 3% of the results are from reference, manufacture or research laboratories. Almost 42% of the results are returned by accredited participants, and the highest percentage of results from accredited participants is found for 2002.

Table 2.25: *HCV number and percentage of datasets per year classified by method of analysis, use of an anti-contamination and performance of an inhibition test.*

Number and percentage of participants' results		2002		2003		2004		2005		Total	
		n	%	n	%	n	%	n	%	n	%
Analysis	**Single**	281	82.16	303	81.02	364	77.45	496	83.50	1444	81.12
	Duplicate	61	17.84	71	18.98	99	21.06	85	14.31	316	17.75
	Other	0	0.00	0	0.00	7	1.45	13	2.19	20	1.23
	Not answered	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
Anticont.	**Yes**	237	69.30	241	64.44	307	65.32	373	62.79	1158	65.05
	No	91	26.61	114	30.48	156	33.19	221	37.20	582	32.70
	Not answered	4	4.09	19	5.08	7	1.49	0	0	40	2.25
Inhibition test	**Yes**	66	19.30	80	21.39	99	21.06	155	26.09	400	22.47
	No	234	68.42	282	75.40	332	70.64	411	69.19	1259	70.73
	Only Neg.	28	8.19	12	3.21	19	4.04	28	4.71	87	4.88
	Not answered	14	4.03	0	0.00	20	4.25	0	0.00	34	1.91

Table 2.26: *HCV percentage of datasets per year classified by laboratory type and accreditation status.*

Number and percentage of participants' results		2002		2003		2004		2005		Total	
		n	%	n	%	n	%	n	%	n	%
Lab. type	**Hospital**	218	63.74	185	49.46	229	48.72	264	44.44	896	50.34
	Public	42	12.88	47	12.57	27	5.74	21	3.53	137	7.70
	Private	40	11.69	14	3.74	40	8.51	74	12.46	168	9.44
	Reference	14	4.09	0	0.00	7	1.45	14	2.36	35	1.97
	Manufacture	0	0.00	6	1.60	0	0.00	21	3.53	27	1.52
	Research/Clin.	7	2.05	7	1.87	7	1.49	22	3.70	43	2.41
	Not answered	21	6.14	115	30.75	160	34.04	178	29.97	474	26.63
Accred.	**Yes**	175	51.17	148	39.57	167	35.53	244	41.08	734	41.23
	No	146	42.69	104	27.81	143	30.42	172	28.96	565	31.74
	Not answered	21	6.14	122	32.62	160	34.04	178	29.97	481	27.02

Finally, Table 2.27 describes the number and percentages of datasets per year returned depending on the experience of participants testing different specimens. More than 26% of the results are from participants who failed to answer questions about testing different specimens. More than 15% of the results are returned by participants who had experience testing between 2,001 and 10,000 plasma and serum samples annually. In general, the percentages of results from participants who tested annually more than 100 plasma samples and more than 100 serum samples are around 48% and 43%, respectively.

Table 2.27: *HCV percentage of datasets per year classified by the experience of participants testing others specimens.*

Number and percentage of participants' results		**2002**		**2003**		**2004**		**2005**		**Total**	
		n	%	n	%	n	%	n	%	n	%
Number Plasma test	**0-10**	113	33.04	52	13.90	68	14.47	100	16.83	333	18.71
	11-100	7	2.04	18	4.81	18	3.83	32	5.39	75	4.21
	101-1,000	68	19.88	106	28.34	116	24.68	109	18.35	399	22.41
	1,001-2,000	77	22.51	28	7.49	21	4.47	34	5.72	160	8.99
	2,001-10,000	56	16.37	49	13.01	61	12.98	108	18.18	274	15.39
	>10,000	0	0.00	6	1.60	26	5.53	33	5.55	65	3.65
	Not answered	21	6.14	115	30.75	160	34.04	178	29.97	474	26.63
Number Serum test	**0-10**	82	23.98	107	28.61	60	12.76	118	19.86	367	20.62
	11-100	14	4.09	14	3.74	32	6.80	53	8.92	113	6.35
	101-1,000	47	13.74	79	21.12	79	16.80	79	13.30	284	15.95
	1,001-2,000	102	29.82	31	8.29	48	10.21	47	7.91	228	12.81
	2,001-10,000	69	20.17	28	7.49	91	19.36	92	15.49	280	15.73
	>10,000	6	1.75	0	0.00	0	0.00	27	4.54	33	1.85
	Not answered	22	6.43	115	30.75	160	34.04	178	29.97	475	26.68
Other Specimen	**Yes**	55	16.08	36	9.62	88	18.72	134	22.56	313	17.58
	No	259	75.73	223	59.62	222	47.23	282	47.47	986	55.39
	Not answered	28	8.19	115	30.75	160	34.04	178	29.97	481	27.02

Chapter 3

Monitoring Quantitative Performance of Molecular Diagnostic Assays Users

External quality assurance (EQA) programmes allow a laboratory to monitor independently its performance and provide feedback to identify and investigate potential areas of concern. Recently EQA providers for molecular diagnostic kit users have started incorporating information about the technology used to their score scheme (Staines et al., 2009). However, none of the EQA providers make use of their prior information about the proficiency panel, provided by the knowledge of 'reference' laboratories and the internal EQA working team, in order to monitor participants' performance independently and give feedback.

In this chapter a new scoring system based on well-known statistical principles is developed. The system is simple, flexible and easy to interpret. In addition, it can be incorporated as additional information to the existing scoring system and can be used to measure performance for single samples or across a panel to provide useful and meaningful information to participants in EQA programmes. The newly proposed scoring system is compared to the current scores using the 2005 QCMD Hepatitis B Virus Proficiency programme.

3.1 Introduction

External quality assessment or proficiency testing is considered the most important way of monitoring participants' quality and a complement to their internal quality control. By participating in EQA programmes it allows laboratories to monitor their performance and provides them with part of the laboratory's quality system requirements to gain the ISO (International Standard Organisation) certification or national accreditation schemes. However, although there are specific guidelines (ISO/IEC Guides, 1997a; ISO/IEC Guides, 1997b; ILAC-G13, 2000) and general principles, which are common to most EQA schemes, there are many different approaches to EQA and reporting mechanisms of the EQA programmes.

Therefore, there is a need to provide a common and appropriate way of monitoring participants' performance in order to assess participanting laboratories. Traditionally, EQA organisations providing schemes for molecular diagnosis of infectious diseases within the area of clinical virology and microbiology have used a subjective approach to define indicators based on consensus analysis or peer group review (CEN TC 140 prEN 14136, 2004), rather than on defined performance indicators related to specific analytical and clinical parameters.

Furthermore, the QCMD organisation has historically used a very simple scoring system for their EQA programmes and the individual participant scores were not reported, but used for internal report analysis only. However, as regulatory requirements changed and the number of laboratories asking for a performance score increased, a scoring system based on well-known statistical principles that is simple, flexible and easy to interpret was developed by Staines et al. (2009) and applied by Pogathota (2007). The scoring system proposed by Staines et al. (2009) is based on a standardised quantity depending on the participant's quantitative response, an estimated mean and standard deviation. The estimated mean can be obtained by the sample mean, known as the consensus value. This estimate may be bi-

ased towards the mean of the modal assay used and may be influenced by poorly performing laboratories (Westgard, 2004; Wong, 2005).

In this chapter approaches to performance scoring within the EQA programmes are evaluated. Also, an alternative estimate for the target value, from a Bayesian perspective, based on 'reference' laboratories performance prior to the distribution of the EQA panel is proposed. The information from 'reference' laboratories about the sample load concentration is gathered by QCMD before delivering the panel to participants for each programme. This information is used internally to measure panel quality before the QC programme starts, but not used to score participants. The primary aim is to establish a suitable mechanism for monitoring participant performance that gives an appropriate representation of a participant's result, which is independent from other participants' results and can be used in a clinical virological and microbiological settings.

The proposed way of scoring provides laboratories with an indicator as a measure of quality that is simple, easy to interpret and has the ability to include cumulative information from previous EQA programmes or 'reference' laboratories. Furthermore, it will provide useful, meaningful and independent information to laboratories which take part in EQA programmes. Here, a summary of the proposed scoring system by Staines et al. (2009) is presented, the new proposed estimation for obtaining the scores based on Bayesian methods is explained, and the application of the scores to HBV data from QCMD-EQA programmes in 2005 is shown.

3.2 Proposed Scoring System for Quantitative Participants' Performance

Performance indicators for individual samples within a panel were investigated by Staines et al. (2009) and extended to an overall performance score for a panel.

In their study, Staines et al. (2009) provided performance indicators for the quantitative estimate of viral or microbial load and sample reproducibility. If the EQA panel consists of J samples, the j^{th} $(j = 1, ..., J)$ sample is assumed to have an estimated 'target' load, μ_j. They also assumed that participant i, $(i = 1, ..., I)$ has reported a viral or microbial load of w_{ij} for sample j on a log_{10} scale quantitative measurement.

In this section the performance indicator for quantitative responses using QCMD's current scoring scheme and a scheme based on a Bayesian approach, developed in this thesis, are described. This is followed by the description of a performance indicator of within-participant consistency and the performance indicators for a panel.

3.2.1 Existing Quantitative Performance Indicators for Individual Samples

Simple and immediate measures of the performance of participant i for sample j are available based on functions of deviation, $d_{ij} = (w_{ij} - \mu_j)$. Commonly used functions of d_{ij} include the absolute deviation, $|d_{ij}|$, the squared deviation, d_{ij}^2 and the percentage absolute deviation $100\,|d_{ij}|\,/\mu_j$ (Westgard, 2004). These metrics can be used as a relative measure to compare laboratories and are easy to compute and interpret. However, their statistical distributions are not known, and so it is difficult to determine limits to identify participants that are performing satisfactorily. An alternative approach is to set acceptance limits, typically $\mu_j \pm 0.5$ (Burtis and Ashwood, 2007). However, for this measure, all values within a given range

are regarded as equally good and those outside the limits as uniformly bad. For example, if the target value, on a log_{10} scale, were 3, then all the values between 2.5 and 3.5 are acceptable and all others are not. So a value of 2.51 is acceptable and scored the same as the target value of 3 and a value 3.49. The values of 0, 3.51 and 6 are regarded as equally unacceptable, even though their clinical significance may be very different.

Therefore, all the scoring schemes proposed so far by other authors are not totally appropriate and a new approach needs to be taken in order to obtain a scoring scheme, which can score participants' performance satisfactorily, independently of all other participants' results and comparable across years or EQA programmes.

3.2.2 Proposed Quantitative Performance Indicator for Individual Samples

The proposed quantitative performance indicators for individual samples is based on the standardised score of a participant estimated viral or microbial load for sample j, w_{ij}, from a set of data with known assigned value (mean μ_j) and known assigned standard deviation σ_j, defined as

$$z_{ij} = \frac{(w_{ij} - \mu_j)}{\sigma_j}. \tag{3.1}$$

The proposed score for sample j for the i^{th} participant is defined as

$$z^*_{ij} = min(3, integer\,[|z_{ij}|]). \tag{3.2}$$

The absolute value of z_{ij} is used since it is assumed that an underestimation and overestimation by the same amount indicates equally poor performance. The integer function with a maximum of three helps interpretation by participants. The value three was chosen since z_{ij} is a standardised value, which in theory has low probability of being greater than 3.

The possible values of z^* are 0, 1, 2 and 3. The scores 0 to 3 are presented to participants as 'highly satisfactory', 'satisfactory', 'unsatisfactory' and 'highly unsatisfactory' and may

be visualised with an associated colour code, for example: green, yellow, orange and red, respectively. Note that, in general, the mean, μ_j, and the standard deviation, σ_j, of the viral or microbial load of sample j are not known. They may be estimated in different ways which are explained in the next subsection.

3.2.2.1 Procedure to calculate Quantitative Performance Indicators for Individual Samples

The procedure to calculate quantitative performance indicators for individual samples are summarised in four steps:

- Step 1: Check the data for outliers and remove them before proceeding with the next steps.
- Step 2: Estimate the standard deviation σ_j in formula (3.1).
- Step 3: Estimate the mean μ_j in formula (3.1).
- Step 4: Obtain the score for each individual sample using the formula (3.2) based on the estimated mean and standard deviation from step 2 and 3.

Three possible estimates for the mean μ_j are presented in this section and applied in the following section. For the standard deviation, two possibles estimates are presented below in order to obtain the score depending on the estimated mean used.

Outlier Detection

Once participants have submitted their results and data have been cleaned (based on a pre-defined Standard Operation Procedure), participants may be classified into K mutually exclusive and exhaustive strata (e.g. based on technology used). For those strata with at least 5 observations the standardised residuals are calculated based on the strata mean of observations and the standard deviation of the observed values. Outliers are defined as those

values with a standardised residual with absolute value greater than 3 (Zar, 1999). When the stratum has less than five observations, outliers are defined as those values greater than 1.5 times the interquartile range from the relevant quartile (Upton and Cook, 2002). Outliers and strata with less than four observations are removed from the data when calculating $\hat{\sigma}_j$, the estimate of the standard deviation, and the mean used to obtain the performance indicator z^*_{ij}.

Standard Deviation Estimation

The standard deviation, σ_j, may be estimated by using s_j, the sample standard deviation. Assuming that participants are a random sample of all diagnostic technologies users, then s_j^2 is an unbiased estimate of σ_j^2. However, since participants may be classified into K mutually exclusive and exhaustive strata (e.g. based on the technology used), a more accurate estimate of the pooled standard deviation, σ_j, may be found by considering the strata sizes and within strata standard deviations as shown below.

If there are n_k participants within stratum k $(k = 1, 2, ..., K)$ and their standard deviation for sample j is s_{jk}, then an unbiased estimate for σ_j^2 is found from

$$\hat{\sigma}_j^2 = \frac{\sum_k (n_k - 1) s_{jk}^2}{\sum_k (n_k - 1)}. \tag{3.3}$$

This is the value of the mean square for the error in the analysis of variance (ANOVA) Table when the response is the participant's result and the factor levels are the strata (Zar, 1999).

Note that for the case when participants on the specific stratum k are scored with respect to participants in the same stratum, then the estimated standard deviation for the sample j within the stratum k, s_{jk}, can be used in formula (3.1) instead of the estimated pooled standard deviation $\hat{\sigma}_j^2$ in formula (3.3).

Mean Estimation from a Classical Perspective

The assigned value, the mean μ_j, may be estimated by the sample mean, $\mu_j^c = \bar{w}_j = \frac{\sum_{i \in I} w_{ij}}{I}$, known as the consensus value. For participants in the same stratum k, the estimated stratum mean of observations, $\mu_j^k = \frac{\sum_{i \in k} w_{ij}}{n_k}$, can be obtained and used as the assigned mean value μ_j.

The mean μ_j may also be estimated from the use of a limited number of 'reference' laboratories prior to the distribution of the panel, but this estimate may be inaccurate and biased towards the technology used by them. Hence, the use of a more robust estimate such as a trimmed mean may be more appropriate.

Mean Estimation from a Bayesian Perspective

A Bayesian approach is proposed to provide a more accurate and appropriate mean estimate that makes use of a prior estimate 'target' viral or microbial load or sample 'target concentration', ϑ_j, for the j^{th} sample. The prior sample 'target' concentration may be available prior to the panel distribution by the EQA organisation. The proposed estimate is based on the prior 'target' information updated with estimates provided by 'reference' laboratories to obtain the 'posterior information'. This is the distribution around the most likely true concentration target based on the information available.

Here, a classical approach of Bayesian inference for the normal distribution with conjugated prior is adopted to estimate the mean, which will be used to calculate the scoring system.

The prior information represents the distribution of the unknown sample 'target' concentration, μ_j. The observed information represents the estimated sample 'target' concentration by 'reference' laboratories, y_{rj} for the laboratory r with $r = 1, ..., R$ and sample j.

In the proposed performance indicator of individual samples, it is assumed that the prior and observed distributions are normal.

The prior distribution of μ_j is assumed to be $N(\vartheta_j, \tau_j^2)$ (see Figure 3.1). The mean ϑ_j is a defined prior 'target' concentration for sample j and the variance τ_j^2 is chosen to be 0.0625 for all samples since this ensures that 95% of the prior distribution lies within the interval $\vartheta_j \pm 0.5$ (Valentine-Thon et al., 2001). The distribution of y_{rj} is defined as $N(\mu_j, \zeta_j^2)$ where ζ_j^2 is an unknown parameter having an Inverse Gamma distribution with parameters a and b, $InvGamma(a, b)$. Since there is no proper prior information about ζ_j^2, the parameters a and b are taken to be 0.0001 corresponding to a non-informative prior distribution. Note that other distributions can be used for the unknown ζ_j^2 (Gelman et al., 2004).

The conditional posterior distribution for the 'target' concentration μ_j is the normal distribution (Gelman et al., 2004),

$$\mu_j \sim N\left(\frac{\zeta_j^2\vartheta_j + \tau_j^2\sum_r^R y_{rj}}{\zeta_j^2 + \tau_j^2 R}, \frac{\tau_j^2\zeta_j^2}{\zeta_j^2 + R\tau_j^2}\right).$$

Therefore, the Bayesian mean, μ_j^b is estimated by

$$\hat{\mu_j} = \frac{\zeta_j^2\vartheta_j + \tau_j^2\sum_r^R y_{rj}}{\zeta_j^2 + \tau_j^2 R}. \tag{3.4}$$

And the Bayesian estimate for the variance ζ_j^2 is

$$\hat{\zeta_j^2} = \frac{2b + \sum_r^R y_{rj}}{R + 2(a-1)}, \tag{3.5}$$

where a=0.0001, b=0.0001 and $\tau_j^2 = 0.0625$. The Bayesian estimate for the variance in formula (3.5) is only used and needed to calculate the mean estimate given by the previous formula (3.4) but it is not used to calculate the score in formula (3.1). The standard deviation used in the score formula (3.1) is specified in the next subsection.

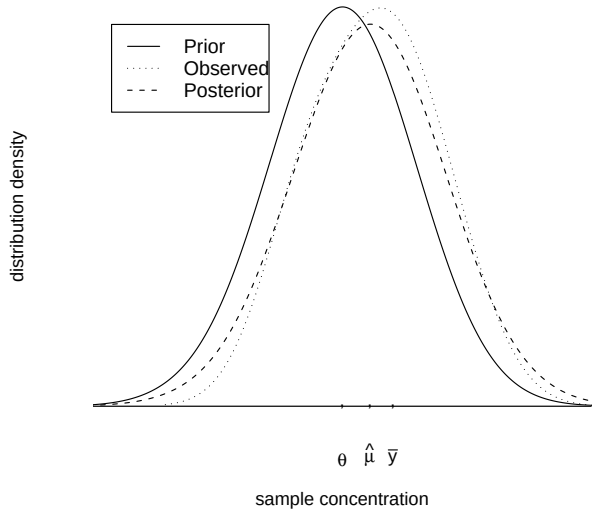


Figure 3.1: *Prior, observed and posterior distributions for a sample with prior sample 'target' concentration.*

Three different ways of scoring

Depending on the objectives of the EQA, participants may be assessed with respect to the Bayesian mean concentration, consensus value or stratum (technology) consensus value. Thus, if the objective of the EQA is to compare the results of participants with the general performance of laboratories then the consensus mean value can be used to provide scores. However if the aim of the EQA is to compare participants using a specific technology then scores based on the technology consensus mean value are more appropriate. If the objective of the EQA is to provide an independent score for each laboratory which can be used to compare results within-participants over time then the Bayesian mean value is the most suitable one.

When scoring participants, the assigned mean μ_j in formula (3.1) may be replaced by the consensus mean, μ_j^c, and the standard deviation used is as defined in formula (3.3). However, if the interest is in scoring participants within a certain stratum k the stratum consensus mean, μ_j^k, should replace μ_j, and the standard deviation used is s_{jk}. In case of scoring in-

dependently of other participants' results the Bayesian mean, μ_j^b can be used to replace the assigned mean, and the pooled standard deviation in formula (3.3) is used. These changes of the estimated mean provide three different ways of scoring which may depend on the aim of the EQA and interest of participants' laboratories.

3.2.3 Proposed Performance Indicator of within Participant Consistency

Here a new procedure is suggested to assess the performance of within participant consistency:

- Calculate the difference of a participant's results for two samples, d_i.
- Estimate the standard deviation of the differences, $\hat{\sigma_d}$, as it was calculated for the quantitative score
$$\hat{\sigma_d^2} = \frac{\sum_k (n_k - 1) s_{dk}^2}{\sum_k (n_k - 1)},$$
where s_{dk} is the standard deviation, of the difference of estimated viral or microbial load for the two samples, for the stratum k.
- Calculate the score, Z_d^*, based on the previous score formula , with
$$z_{d_i}^* = min(3, integer \left[\left| \frac{(d_i - \hat{d})}{\hat{\sigma_d}} \right| \right]),$$
where $\hat{d}$ is the known or estimated mean for the difference of viral or microbial load for the two samples.

The interpretation of the score is equivalent to the quantitative score.

Reproducibility is a special case to monitor participant consistency with d=0. The reproducibility is defined as the extend to which a participant can produce the same estimate viral or microbial load for two identical samples within a panel. Note that reproducibility scores must not be added to obtain the panel score since it is not independent of the scores of individual samples of the panel. The reproducibility score provides extra information about the ability to reproduce sample concentration loads.

3.2.4 Proposed Performance Indicator for a Panel

The proposed performance indicator for an individual sample ranges from 0 (highly satisfactory) to 3 (highly unsatisfactory). One measure of overall performance for a panel is to sum a participant's score for those samples where a value is reported. The distribution of this score is not known and will vary according to the number of samples reported. Participants are classified using the method outline below.

Assuming normality and independence, the proposed score for an individual sample takes the values 0 to 3 with probabilities 0.683, 0.272, 0.043 and 0.002 respectively. J columns, one for each sample, of 10,000 Monte Carlo simulations from the previous probability mass function are found to generate 10,000 virtual participants. The sums of the J columns for each Monte Carlo simulation are obtain. The frequencies of the sums are found and used as follows. For consistency with the scoring for individual samples, participants that reported J samples are given score 0 (classified as 'highly satisfactory') if their sum is in the smallest 68.3% of the simulated values, score 1 ('satisfactory') in the next 27.2%, score 2 ('unsatisfactory') in the following 4.3% and score 3 ('highly unsatisfactory') in the highest 0.2%. Therefore, depending on the sum of individual scores a panel score can be otained as described above.

Table 3.1: *Panel score table for panels with up to 15 samples.*

Number of samples per panel	Panel Score			
	0	1	2	3
1 sample	0	1	2	3
2 samples	0	1	2-3	4+
3 samples	0-1	2	3-4	5+
4 samples	0-1	2-3	4-5	6+
5 samples	0-1	2-4	5-6	7+
6 samples	0-2	3-4	5-7	8+
7 samples	0-2	3-5	6-7	8+
8 samples	0-3	4-5	6-8	9+
9 samples	0-3	4-6	7-8	9+
10 samples	0-4	5-6	7-9	10+
11 samples	0-4	5-7	8-10	11+
12 samples	0-4	5-7	8-10	11+
13 samples	0-5	6-8	9-11	12+
14 samples	0-5	6-8	9-12	13+
15 samples	0-6	7-9	10-13	14+

Table 3.1 gives the range of total scores corresponding to each panel score for a given number of samples (Staines et al.,2009). For example, a participant who has a total score of six or seven from seven samples is scored 2 (i.e.'unsatisfactory').

3.3 Application

An application of the proposed performance indicators to the 2005 QCMD Hepatitis B Virus Proficiency Programme (QCMD, 2010) is presented in this section. The Panel composition is shown in Table 3.2. Seven independent 'reference' laboratories analysed the panel before it was sent to participants. There were 116 participants from 27 countries of which 102 returned panel results. The total number of datasets submitted was 122 of which 21 datasets provided only qualitative information on the detection of the virus and 101 datasets provided quantitative and qualitative information on sample viral load.

The datasets are classified per technology group used to analyse the panel: Conventional Commercial PCR (CC, n=38), Conventional In-house PCR (CIH, n=12), Real time Commercial PCR (RTC, n=39), Real time In-house PCR (RTIH, n=24), bDNA (bDNA, n=7) and Hybrid Capture (HC, n=2). However, the total number of datasets reporting quantitative results per technology group varies depending on the sample analysed.

Table 3.2: *2005 QCMD HBV panel composition.*

Sample	Sample Type	QCMD Sample Conc. Log Copies/ml	Defined Sample Status
HBV01	D	5.00	P
HBV02	A	4.00	P
HBV03	D	3.00	WP
HBV04	D	4.00	P
HBV05	A	3.00	WP
HBV06	D	6.00	SP
HBV07	N/A	0.00	N
HBV08	A	5.00	P

D-virus subtype D; A-virus subtype A; N/A-no applicable

P-positive sample; N-negative sample

WP- weak positive sample; SP-strong positive sample

3.3.1 Quantitative Analysis

A summary of the quantitative score obtained with respect to consensus mean and technology group consensus mean sample concentration is shown in this section. Note that negative samples, a zero microbial load reported, and values reported as outside the detection limits of the assays are excluded from the quantitative analysis.

To illustrate how the score is calculated for laboratories the results for laboratories of technology group RTC are considered. In this example, two laboratories are used, participant 1 (Lab1) and participant 2 (Lab2), and one sample (HBV02) with target concentration 4 log_{10} copies/ml. The results provided by these two laboratories are 3.509 and 1.826, respectively,

which give standardised residuals -0.658 and -3.520, respectively. Note that for the calculation of standardised residuals the procedure followed was as described in Section 3.2.2.1.

Lab2 is detected as an outlier as its standardised residual is outside the interval $(-3, 3)$. Then, the new consensus mean and standard deviation are calculated and the results are 3.988 and 0.473, respectively, once outliers have been removed. Therefore, the score is calculated in the following way:

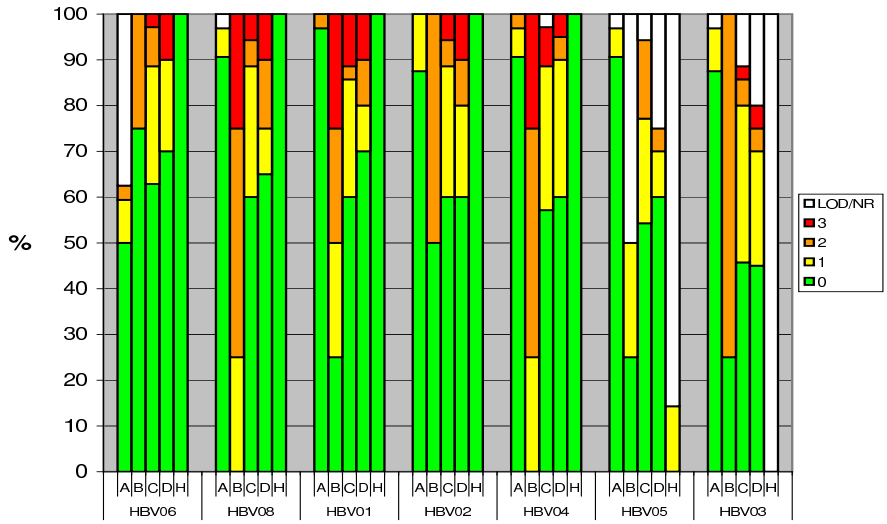
Lab1 $z_{lab1} = \frac{3.509-3.988}{0.473} = -1.013$, $z^*_{lab1} = 1$.

Lab2 $z_{lab2} = \frac{1.826-3.988}{0.473} = -4.571$, $z^*_{lab2} = 3$.

3.3.1.1 Score with Respect to Consensus Mean

The consensus mean $\bar{w}_j$ and standard deviation, $\hat{\sigma_j}$, for each sample $j = 1, ..., 6, 8$ are calculated (datasets provided by laboratories using HC technologies and outliers are not included for the calculations). Four technology groups, CC, RTC, RTIH and bDNA, are used to estimate the standard deviation for samples HBV01, HBV02, HBV04, HBV06 and HBV08. However, there are not enough datasets (≥ 5) to take into account the bDNA group for samples HBV03 and HBV05 since some estimates are outside the limit of detection for the assay.

Table 3.3 (illustrated in Figure 3.2) shows the estimated consensus mean and standard deviation of the log_{10} viral load, arranged in decreasing consensus mean, and the frequency of z^* scores for each sample by technology. For example, 16 participants out of 32 that used technology CC-Commercial PCR have score 0 ('highly satisfactory') for the sample HBV06 with target viral load of 6. For this case, the consensus mean was 5.702 and the pooled standard deviation obtained as described in Section 3.2.2 was 0,607. When the result reported by a participant was outside the limit of detection or missing the participant was



A=CC, B=CIH, C=RTC, D=RTIH, H=bDNA

Figure 3.2: *Percentage of datasets scoring 0, 1, 2 and 3, with respect to consensus mean, per sample and technology group.*

classified as LOD/NR. The score amongst participants withing RTC or RTIH technologies appear to be more variable than the score amongst participants of other technology groups. Most of participants using CC technology have score 0 or 1, while participants using RTC have scores dispersed around 0 or 1.

Note that all laboratories except those with results outside the limit of detection of the assay are scored even if they are not included in the calculation of the mean or standard deviation.

Table 3.3: *Summary participants' score with respect to consensus mean.*

	Consensus		CC n=32					CIH n=4				
	Mean	σ_j	0	1	2	3	LOD/NR*	0	1	2	3	LOD/NR*
HBV06	5.702	0.607	16	3	1	0	12	3	0	1	0	0
HBV08	4.887	0.511	29	2	0	0	1	0	1	2	1	0
HBV01	4.789	0.576	31	0	1	0	0	1	1	1	1	0
HBV02	3.988	0.473	28	4	0	0	0	2	0	2	0	0
HBV04	3.834	0.544	29	2	1	0	0	0	1	2	1	0
HBV05	2.952	0.576	29	2	0	0	1	1	1	0	0	2
HBV03	2.879	0.453	28	3	0	0	1	1	0	3	0	0

	Consensus		RTC n=35					RTIH n=20				
	Mean	σ_j	0	1	2	3	LOD/NR*	0	1	2	3	LOD/NR*
HBV06	5.702	0.607	22	9	3	1	0	14	4	0	2	0
HBV08	4.887	0.511	21	10	2	2	0	13	2	3	2	0
HBV01	4.789	0.576	21	9	1	4	0	14	2	2	2	0
HBV02	3.988	0.473	21	10	2	2	0	12	4	2	2	0
HBV04	3.834	0.544	20	11	0	3	1	12	6	1	1	0
HBV05	2.952	0.576	19	8	6	0	2	12	2	1	0	5
HBV03	2.879	0.453	16	12	2	1	4	9	5	1	1	4

	Consensus		bDNA n=7					HC n=2				
	Mean	σ_j	0	1	2	3	LOD/NR*	0	1	2	3	LOD/NR*
HBV06	5.702	0.607	7	0	0	0	0	2	0	0	0	0
HBV08	4.887	0.511	7	0	0	0	0	0	0	0	0	2
HBV01	4.789	0.576	7	0	0	0	0	0	0	0	0	2
HBV02	3.988	0.473	7	0	0	0	0	0	0	0	0	2
HBV04	3.834	0.544	7	0	0	0	0	0	0	0	0	2
HBV05	2.952	0.576	0	1	0	0	6	0	0	0	0	2
HBV03	2.879	0.453	0	0	0	0	7	0	0	0	1	1

*LOD/NR: Result reported as lower limit detection or upper limit detection/no value or no result reported

CC=Commercial PCR, CIH=Conventional In-house PCR, RTC=Real time Commercial PCR

RTIH=Real time In-house PCR, bDNA and HC=Hybrid Capture

3.3.1.2 Score with Respect to Technology Consensus Mean

The consensus mean for each technology group, with at least four observations once outliers have been removed, is calculated. Each participants score with respect to the stratum consensus mean and standard deviation, where the stratum refers to the group of participants using the same type of technology. Thus, participants can be classified by technology group or stratum depending on the technology that was used to analyse the samples. CIH and HC technology groups do not satisfy the requirement of having at least 5 datasets before and 4 datasets after the removal of outliers. Therefore, laboratories using these technologies are not scored with respect to their technology group.

Table 3.4 shows the estimated technology consensus mean and standard deviation of the log_{10} viral load and the frequency of z^* scores for each sample by technology. As in previous section it is observed that 16 participants out of 32 that used technology CC have score 0 for the sample HBV06, however the estimated mean and the standard deviation are different since now the way of scoring is with respect to the technology group mean and standard deviation. Those values were calculated using only the information provided by the 32 participants that used CC technology to analyse the samples. RTC participants are scoring better with respect to their own technology than with respect to the consensus mean (see Table 3.3).

Figure 3.3 shows the percentages of overall datasets scoring 0, 1, 2 and 3 per sample with respect to technology consensus mean (LOD/NR: result reported as outside the limit of detection/ no value or no result reported; Conventional Commercial PCR (CC); Real Time Commercial (RTC); Real Time In-house (RTIH); bDNA).

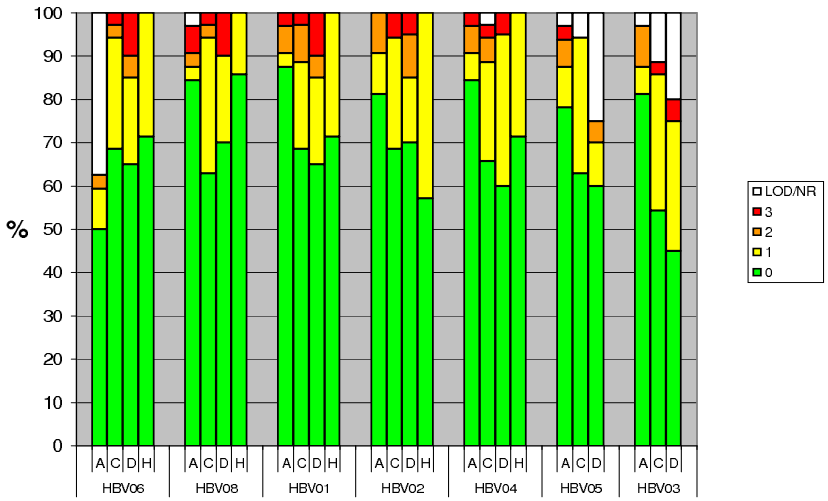
Table 3.4: *Summary participants' score with respect to technology consensus mean.*

	Technology		CC n=32					Technology		bDNA n=7				
	Mean	σ_j	0	1	2	3	LOD/NR*	Mean	σ_j	0	1	2	3	LOD/NR*
HBV06	5.721	0.496	16	3	1	0	12	5.940	0.035	5	2	0	0	0
HBV08	5.068	0.241	27	1	1	2	1	5.161	0.092	6	1	0	0	0
HBV01	4.967	0.281	28	1	2	1	0	5.004	0.069	5	2	0	0	0
HBV02	4.121	0.253	26	3	3	0	0	4.163	0.140	4	3	0	0	0
HBV04	4.017	0.291	27	2	2	1	0	3.988	0.125	5	2	0	0	0
HBV05	3.102	0.257	25	3	2	1	1	-	-					
HBV03	2.976	0.255	26	2	3	0	1	-	-					

	Technology		RTC n=35					Technology		RTIH n=20				
	Mean	σ_j	0	1	2	3	LOD/NR*	Mean	σ_j	0	1	2	3	LOD/NR*
HBV06	5.637	0.720	24	9	1	1	0	5.676	0.574	13	4	1	2	0
HBV08	4.874	0.679	22	11	1	1	0	4.700	0.568	14	4	0	2	0
HBV01	4.697	0.799	24	7	3	1	0	4.641	0.563	13	4	1	2	0
HBV02	3.957	0.593	24	9	0	2	0	3.788	0.577	14	3	2	1	0
HBV04	3.763	0.716	23	8	2	1	1	3.806	0.609	12	7	0	1	0
HBV05	2.794	0.754	22	11	0	0	2	2.937	0.601	12	2	1	0	5
HBV03	2.836	0.564	19	11	0	1	4	2.855	0.525	9	6	0	1	4

*LOD/NR: Result reported as lower limit detection or upper limit detection/no value or not result reported

CC=Commercial PCR, CIH=Conventional In-house PCR, RTC=Real time Commercial PCR RTIH=Real time In-house PCR bDNA and HC=Hybrid Capture



A=CC, C=RTC, D=RTIH, H=bDNA

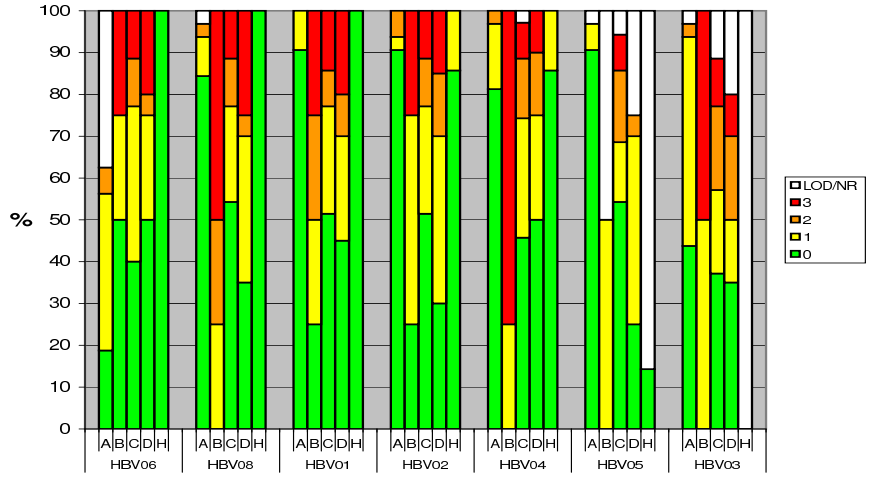
Figure 3.3: *Percentage of datasets scoring 0, 1, 2 and 3, with respect to technology consensus mean, per sample and technology group.*

3.3.1.3 Score with Respect to Bayesian Mean

In order to obtain laboratories' score with respect to the Bayesian mean, the target value is updated from estimates provided by 'reference' laboratories.

Using the Bayesian model defined in Section 3.2.2.1 coded in WinBUGS (Project, 1996-2004), the 'target' sample concentrations are estimated (see Table 3.5).

Table 3.5 and Figure 3.4 show the total number and percentage of datasets scoring 0, 1, 2 and 3 with respect to the Bayesian mean, per sample and technology used. It is observed that all samples have Bayesian means higher than the corresponding consensus mean, perhaps because most 'reference' laboratories analysed the samples using CC or RTC technologies. For the strongest viral load sample group, 6 and 14 datasets are scored 0 with respect to the Bayesian mean for CC and RTC technologies, whilst 16 and 22 datasets are scored 0 with respect to consensus mean for those technologies, respectively. Participants' scores are more scattered when scoring with respect to Bayesian mean than those with respect to consensus mean.



A=CC, B=CIH, C=RTC, D=RTIH, H=bDNA

Figure 3.4: *Percentage of datasets scoring 0, 1, 2 and 3, with respect to Bayesian mean, per sample and technology group.*

Table 3.5: *Summary participants' score with respect to Bayesian mean.*

	Bayesian		CC n=32					CIH n=4				
	Mean	σ_j	0	1	2	3	LOD/NR	0	1	2	3	LOD/NR
HBV06	6.355	0.607	6	12	2	0	12	2	1	0	1	0
HBV08	5.440	0.511	27	3	1	0	1	0	1	1	2	0
HBV01	5.368	0.576	29	3	0	0	0	1	1	1	1	0
HBV02	4.469	0.473	29	1	2	0	0	1	2	0	1	0
HBV04	4.431	0.544	26	5	1	0	0	0	1	0	3	0
HBV05	3.399	0.576	29	2	0	0	1	0	2	0	0	2
HBV03	3.444	0.453	14	16	1	0	1	0	2	0	2	0
	Target		RTC n=35					RTIH n=20				
	Mean	σ_j	0	1	2	3	LOD/NR	0	1	2	3	LOD/NR
HBV06	6.355	0.607	14	13	4	4	0	10	5	1	4	0
HBV08	5.440	0.511	19	8	4	4	0	7	7	1	5	0
HBV01	5.368	0.576	18	9	3	5	0	9	5	2	4	0
HBV02	4.469	0.473	18	9	4	4	0	6	8	3	3	0
HBV04	4.431	0.544	16	10	5	3	1	10	5	3	2	0
HBV05	3.399	0.576	19	5	6	3	2	5	9	1	0	5
HBV03	3.444	0.453	13	7	7	4	4	7	3	4	2	4
	Bayesian		bDNA n=7					HC n=2				
	Mean	σ_j	0	1	2	3	LOD/NR	0	1	2	3	LOD/NR
HBV06	6.355	0.607	7	0	0	0	0	0	2	0	0	0
HBV08	5.440	0.511	7	0	0	0	0	0	0	0	0	2
HBV01	5.368	0.576	7	0	0	0	0	0	0	0	0	2
HBV02	4.469	0.473	6	1	0	0	0	0	0	0	0	2
HBV04	4.431	0.544	6	1	0	0	0	0	0	0	0	2
HBV05	3.399	0.576	1	0	0	0	6	0	0	0	0	2
HBV03	3.444	0.453	0	0	0	0	7	0	0	0	1	1

*LOD/NR: Result reported as lower limit detection or upper limit detection/no value or not result reported

CC=Commercial PCR, CIH=Conventional In-house PCR, RTC=Real time Commercial PCR

RTIH=Real time In-house PCR, bDNA and HC=Hybrid Capture

3.3.2 Scoring Performance within Participant Consistency

As an illustrative example, sample HBV01 and HBV04, both of which are subtype D with QCMD sample concentration loads of 5.00 and 4.00 log_{10} copies/ml respectively, are considered. The consensus mean and standard deviation of the difference after removal of outliers are 0.942 and 0.230 respectively. Therefore, the difference, z value and the score for consistency of a participant with observations 4.954 and 3.923 for samples HBV01 and HBV04, respectively, are given by:

$d = 1.031$, $z_d = \frac{1.031-0.942}{0.230} = 0.387$ and $z_d^* = 0$

3.3.3 Panel Score

Panel scores for quantitative analysis are calculated for laboratories which return estimates for all seven positive samples. It is found that only 56 participants returned complete quantitative datasets of which 47 datasets receive a panel score 0, 2 datasets obtain score 1, 3 datasets obtain score 2 and 4 datasets obtain score 3 (scores obtained with respect to consensus mean).

With respect to Bayesian mean, it is found that 32 datasets obtain a panel score of 0, 2 datasets obtain a score of 1 and 11 datasets receive a score of 2 or 3. The results on the panel score show that with respect to the consensus mean more participants are classified as high satisfactory (a score of 0), than with respect to the Bayesian mean. Scoring with respect to the Bayesian mean classified more participants with scores 2 and 3 than scoring with respect to the consensus mean.

3.4 Summary and Conclusions

The Bayesian mean estimation used to obtain the score allows to incorporate the QCMD prior information about the sample load concentration, which is updated with information provided by selected 'reference' laboratories.

Since the Bayesian mean estimation is obtained based on probability theory (the most likely estimate for the mean) and on independent laboratories from the QC programme, it may be considered to be more objective and realistic than using the consensus mean, which is based on performances from participating laboratories. In fact, scoring with respect to Bayesian mean provided external quality assessment based on an independent mechanism to the participants' performances from the EQA programme. The proposed scoring system based on a Bayesian mean is a step forward in providing appropriate performance indicators to monitor quantitative performance of molecular diagnostic users.

In contrast to the use of consensus or technology mean, the Bayesian mean is independent of participants' results. Thus, the use of a Bayesian mean provides an appropriate and independent indicator as a measure of external quality.

The scoring system has been applied with a variety of assigned values. The results obtained, when the scoring system has been applied to consensus, technology and Bayesian mean estimations, show that the percentages of laboratories with better score varies with the assigned value used. The scores obtained with the consensus mean are better than those obtained from the estimated Bayesian mean. Although these differences in the percentages are expected, the results state that when laboratories are assessed based on consensus mean obtained from their own results then they obtain a better score than when independent means (Bayesian means) are used.

Although a direct statistical comparison of the three scoring methods is possible it does not appear to provide further insight as the 'real target' value is unknown. That is, in the absence of the 'real target' value it is impossible to make a judgment in terms of which of the three scores is the best indicator from a statistical point of view. However, conclusions as to how participants' performances are assessed from a descriptive and methodological point of view can be made as shown in this chapter.

The application of these performance indicators to the 2005 QCMD Hepatitis B Virus Proficiency programme highlights the flexible use and desirable properties of the proposed scoring system for assessing various aspects of participant performance. However, since it is based on an updating process of information provided from 'reference' laboratories, their choice is crucial. Therefore, the selecting procedure of 'reference' laboratories should be considered and carefully done. Furthermore, the use of the Bayesian mean and its interpretation should be studied carefully in some cases such as when all references used the same technology for analysing the samples and thus no variability on technology used are represented on the reference laboratories.

The proposed score provides a flexible and mathematically rigorous metric to assess participant performance for molecular diagnostic kit users. However, there are some drawbacks when considering quantitative results.

Firstly, the score requires participants to provide their estimates of the viral or microbial load. Sometimes, participants report values outside the detectable limits of the assay they use. These have been ignored for the purposes of this chapter. Possible approaches to this problem using the frequentist approach are to include censored value techniques or to replace the value by either the limit of detection or half of this value. In Chapter 4, a Bayesian alternative to this approach will be introduced.

Secondly, a Bayesian approach to estimating the target value has been suggested. The Bayesian estimate is the value suggested by the EQA provider (from internal investigations and previous panels) updated by estimates from 'reference' laboratories. Although these may contain assay bias due to the way of choice of 'reference' laboratories, care can be taken to ensure a range of assays are covered by the 'reference' laboratories.

Finally, the score assumes normality, which is almost certainly not valid for scoring participants who report false positives from negative samples. Hence the Bayesian mean estimate is only recommended for positive samples since negative samples do not have a microbial or viral load to be quantified. Further discussion can be found in Chapter 6.

Chapter 4

Modelling Qualitative Performance of Participants of QCMD Programmes over Time

In the previous chapter, performance indicators for microbial load estimation by molecular diagnostic assay users have been proposed. In this chapter, risk factors associated with participants qualitative performance (specificity and sensitivity) are studied in order to gain further knowledge for the design of future EQA programmes and to provide better feedback to participants.

This chapter proposes a new model to investigate which of the exploratory variables, defined in chapter 2, are related to qualitative performance of participants of QCMD programmes for one pathogen over time. The model will be applied to qualitative responses of participants' performance which provide information about the correct detection of sample microbe. The relation between qualitative responses and the participating laboratory practice is studied with the proposed model.

In this chapter the pathogens Enterovirus (EV) and hepatitis B virus (HBV) are considered for the model application to real data.

4.1 Introduction

The statistical tools used to analyse qualitative data provided by EQA participants in molecular diagnostics, have previously been based on frequentist methods such as Generalised Linear Models (GLM). While Bayesian approaches to analyse data have been considered recently in some areas, such as clinical chemistry and veterinary medicine (Conraths and Schare, 2006; Geurden et al., 2004), these methods have not been used yet to identify significant factors associated with performance of molecular diagnostic assays, such as sensitivity and specificity.

The analysis of factors of interest that are associated with individual participant's performance requires the inclusion of multiple parameters in the model. Due to the large number of parameters, the missing information from covariates and the nature of the data, it is beneficial to approach the analysis of qualitative performance from a Bayesian perspective. The Bayesian approach can easily handle multiple parameters and missing values by employing a hierarchical structure of the model. Based on prior probability distributions, such a Bayesian model can be used to estimate the posterior distribution of the parameters and missing values.

The approach proposed here is to model the qualitative responses of participants' performance using a GLM, in particular, a logistic regression model from a Bayesian perspective. The model, which will be called the 'Qualitative Bayesian Model' (QLBM), enables the identification of significant factors associated with qualitative participants' performance. While the theory of logistic regression models and Bayesian data analysis is well known, combining the two statistical tools for the modelling of biomedical data is still under development.

Logistic regression models and Bayesian data analysis are combined in this thesis in order to derive a model for the qualitative data. The model will derive the posterior information needed to answer several research questions, relevant for measuring the qualitative performance of EQA participants.

The proposed QLBM model takes into account the peculiarities of the present data conditions. In particular, it allows the inclusion of those datasets which would be discarded using a classical GLM due to missing covariate information. The model uses the observed data to derive the probability distribution of the missing observations. Thus, the QLBM not only enables the estimation of parameters related to the factors under study, but it also allows the estimation of missing observations using the learning process inherent within the Bayesian framework. Therefore, knowledge is gained about the data by not discarding incomplete datasets.

The developed QLBM is programmed using the statistical software WinBUGS (Project, 1996-2004), and is applied to the 2002 to 2005 QCMD EV and HBV programmes. The code to carry out the model estimations have been developed by the author and it can be found in the CD attached to the thesis.

Before proposing the Bayesian model (QLBM), a classical GLM approach to analyzing qualitative EQA responses based on the logistic regression is presented.

4.2 Logistic regression GLM: basic notation and model formulation for the EQA qualitative data

If Y_{si} denotes the i^{th} participant's response for the sample group s, then under the logistic regression model, this is assumed to follow a Bernoulli distribution with parameter p_{si}.

Thus, p_{si} is the probability that the i^{th} participant's response is correct for the sample group s. Note that the participant's response takes on the value 0 - incorrect if the participant fails to detect the virus in a positive sample or reports the detection of the virus in a negative sample, and it takes on the value 1 - correct if the participant detects the virus in a positive sample or reports no detection of the virus in a negative sample.

This probability is assumed to depend lineraly on the covariates under investigation through the 'logit' link function (1.4) in the following way (Dugard et al., 2010):

$$logit(p_{si}) = \log(\frac{p_{si}}{1-p_{si}}) = \vec{x}_{si}\vec{\beta}_s. \tag{4.1}$$

As a result:

$$p_{si} = \frac{\exp(\vec{x}_{si}\vec{\beta}_s)}{1+\exp(\vec{x}_{si}\vec{\beta}_s)} \tag{4.2}$$

where:

- i is the i^{th} observation in the sample group s with *i=1,..,*n_s.
- n_s is the total number of observations within sample group s with *s=1,..,*l.
- l is the total number of sample groups.
- $\vec{x}_{si} = (x_{si1}, .., x_{sir})$ is the r-dimensional vector of covariates for the i^{th} observation of sample group s with $x_{si1} = 1$ as the intercept.
- The covariate matrix $\mathbf{X}_s$ for each sample group is the matrix with r columns and n_s rows. Each column corresponds to the covariate for participants in sample group s. Thus, each row is the r-dimensional vector of covariates $\vec{x}_{si} = (x_{si1}, .., x_{sir})$ for the i^{th} observation of sample group s.
- $\vec{\beta}_s$ is the r-dimensional column vector of regression coefficients, $\vec{\beta}_s = (\beta_{s1}, .., \beta_{sr})^T$.

Note that the subscript si is used for indicating the participant i in sample group s instead of s_i which is reserved for allowing different variances on subgroups of participants' responses within sample group s (see Chapter 5 and 6).

To illustrate how the model described by equation (4.1) applies to the EV dataset, a simple example using a subset of covariates is presented in what follows. For this purpose, we chose the fifth response for the first sample group (1), $y_{15} = 1$, which is a correct result meaning that the participant detected the virus correctly. This response was returned by an accredited participant (Accred.) using CC technology (CC) for a sample from the year 2004 (2004). This is represented here by a vector of dummy variables $\vec{x}_{15}$ corresponding to the covariates associated with the response, i.e. $\vec{x}_{15} = (x_{15(Id)}, x_{15(Accred.)}, x_{15(CC)}, x_{15(2004)})$, where Id is the variable set to value 1 to allow for an intercept to be included into the model ($x_{15(Id)} = 1$). Using equation (4.1), the probability of a correct response for a sample of this group is given by:

$$logit(p_{15}) = \beta_{11} * Id + \beta_{12} * (Accred. = yes) + \beta_{13} * (CC = yes) + \beta_{14} * (2004 = yes)$$

which is equivalent to the mathematical expression

$$logit(p_{15}) = \beta_{11} * x_{15(Id)} + \beta_{12} * x_{15(Accred.=yes)} + \beta_{13} * x_{15(CC=yes)} + \beta_{14} * x_{15(2004=yes)}.$$

4.3 Problems Arising with Classical GLM when Analysing Qualitative Responses

A logistic regression GLM as described in the previous section was applied to the data, and it was observed that the classical techniques failed to model the EQA qualitative data for several reasons, which are briefly discussed in what follows:

- The classical GLM cannot deal with observations containing missing covariates.
- For some sample groups all participants using a specific technology provide correct results. Then, since GLM is based on asymptotic theory it fails to fit the data with no variability within the responses. Consequently, the model fails to provide an estimate for the corresponding parameter. Therefore, in classical analysis all observations from those technologies with all correct results have to be discarded for investigating the effect of technology used when analysing the sample.
- Some covariates have such a high proportion of correct or wrong results that the estimated parameters have large estimated standard errors. It is therefore not possible to draw appropriate conclusions about the effect of those covariates on the response when GLM is applied.

The following examples do not provide a complete analysis, but rather aim to highlight some of the major problems arising when using a classical GLM approach in this application. Table 4.1 and Table 4.3 show the results when the classical GLM is applied for the detection of the virus in the strongest and weakest (viral loads) samples, of the EV programmes over time (see Chapter 2 for information about the EV data). The covariates are chosen to illustrate in a simple way the problems that appear when applying GLM to the type of data dealt in this project.

In Table 4.1 the strongest sample group has samples with dilution series of 1×10^{-3}. When applying the classical GLM to the responses the software returns an error message because all responses from CC and NASBA are correct results. Hence, the model cannot obtain estimates for these parameters. Thus, for studying the association of technologies with the correct detection of the virus, the subgroup of responses from participants using these two technologies has to be discarded (which roughly corresponds to 7.5% of the dataset, see Chapter 2).

The classical GLM is also applied to study the association of year, use of an anti-contamination system and accreditation status with the correct detection of the virus. The results in Table 4.1 show that the estimated parameter for the use of an anti-contamination system has a very large standard error suggesting that this estimate is not reliable (in Table 4.2 we can see that 100% of the participants who used anti-contamination system provided correct results. In this case, it is difficult to provide a reliable model estimate using classical GLM). Furthermore, as the GLM cannot handle missing covariates the dataset has to be reduced from 207 to 155 observations. Therefore, by using a standard classical GLM it is not possible to analyse the complete dataset.

Table 4.1: *Results from the classical GLM applied to the strongest EV sample group.*

Parameters	Estimate	SE	p-value
Intercept	4.746	1.327	0.0003
Year 2002 (baseline 2005)	-0.966	1.139	0.396
Anti-contamination Yes (baseline No)	16.017	1934.027	0.993
Accreditation Yes (baseline No)	-1.324	1.136	0.244

Note that the strongest EV sample group was only found in the panels of 2002 and 2005. Therefore, parameters for other years are not estimated.

Results for the weakest positive EV sample group with dilution series of 1×10^{-8} are presented in Table 4.3. The classical GLM is applied to study the association of year, technology, use of an anti-contamination system and accreditation status with the correct detection of the virus.

The results in Table 4.3 show that the estimated parameters are reliable, in the sense that unlike in the previous example we do not get large standard errors. However, as in the previous example, since the GLM cannot handle missing covariates the dataset has to be reduced from 412 to 300 observations.

Table 4.2: *Percentages of correct results per sample group classified by covariate level for the EV programmes.*

Covariates	Percentages of correct results							
	$1x10^{-3}$	$1x10^{-4}$	$1x10^{-5}$	$1x10^{-6}$	$1x10^{-7}$	$1x10^{-8}$	Non-EV	Negative
Year								
Year 2002	96.00	-	94.00	76.25	45.00	23.00	-	94.50
Year 2003	-	-	94.75	80.52	82.02	39.32	82.58	93.25
Year 2004	-	73.27	87.64	75.57	73.27	29.31	87.06	94.39
Year 2005	97.19	80.37	85.04	80.84	56.07	22.42	88.32	96.26
Technology group								
Tech CIH	95.97	79.51	92.34	78.93	62.70	28.45	81.41	93.45
Tech CC	100.00	100.00	97.87	86.79	48.00	16.67	76.00	93.75
Tech RTIH	98.33	72.73	87.50	76.59	59.78	30.64	92.22	95.61
Tech NASBA	100.00	100.00	95.00	70.00	50.00	11.11	100.00	100.00
Tech RTC	88.89	53.84	63.89	58.34	25.00	26.67	96.15	100.00
Anticontamination								
Yes	100.00	86.00	91.94	76.95	57.85	28.91	88.67	95.13
No	95.70	73.45	89.68	78.38	59.91	29.11	85.82	94.03
Not answered	100.00	87.81	90.90	66.66	57.89	0.00	76.19	100.00
Accreditation								
Yes	94.93	73.23	91.01	76.23	59.82	25.51	87.11	93.75
No	98.75	79.51	87.75	77.47	55.83	26.87	87.95	93.61
Not answered	95.83	76.81	92.67	80.60	64.51	33.64	82.46	96.64
CSF								
Yes	97.35	75.75	89.85	77.04	56.54	25.63	87.50	94.28
No	92.30	81.81	85.39	76.76	65.95	32.35	88.23	90.00
Not answered	95.34	76.81	92.69	80.43	65.97	33.66	82.23	96.40
Serum								
Yes	96.42	70.58	89.80	77.55	58.79	30.20	90.00	93.28
No	97.50	84.84	89.16	76.73	56.06	22.64	84.97	94.15
Not answered	95.37	76.38	92.19	80.00	65.98	33.65	82.58	96.53
Swab								
Yes	96.25	73.49	91.47	80.75	62.50	27.08	88.50	94.28
No	97.61	80.28	87.52	74.00	52.98	25.74	86.81	93.44
Not answered	95.34	76.81	92.69	80.43	65.97	33.66	82.23	96.40
Biopsies								
Yes	97.01	70.88	92.16	79.59	62.50	27.06	90.85	93.88
No	96.90	82.66	87.18	75.22	53.81	25.84	84.89	93.76
Not answered	95.34	76.81	92.69	80.43	65.97	33.66	82.23	96.40
Analysis method								
Analysis Singly	95.23	70.47	88.67	75.80	58.86	30.93	86.01	93.29
Analysis Duplicated	97.24	83.00	90.23	77.59	57.28	24.00	85.95	94.84
Analysis Other	100.00	-	100.00	92.30	76.47	36.36	100.00	100.00
Not answered	100.00	77.77	94.23	88.88	78.57	40.00	81.48	96.96
Plasma								
Group 0: 0-10	100.00	84.84	83.92	76.78	66.66	19.69	85.41	94.11
Group 1: 11-100	94.11	68.88	85.66	70.96	51.37	21.33	91.00	93.02
Group 2: 101-1,000	95.89	55.55	92.08	79.16	56.84	29.05	86.17	95.13
Group 3: 1,001-2,000	100.00	92.30	97.26	75.90	61.53	28.57	87.17	90.00
Group 4: 2,001-10,000	100.00	66.66	94.73	82.00	50.00	37.50	94.44	89.28
Group 5: > 10,000	100.00	83.33	86.66	80.00	44.44	66.66	77.77	100.00
Not answered	95.65	76.81	92.85	81.25	66.00	33.65	82.23	96.53
Labtype								
Hospital	96.42	79.20	87.12	74.71	54.92	24.63	86.83	94.88
Public Health	100.00	75.00	95.55	75.75	58.44	24.52	88.52	89.69
Private	85.71	25.00	85.36	77.55	47.82	50.00	84.21	96.55
Reference	100.00	85.71	93.75	88.63	65.00	23.07	100.00	92.00
Manufacture	100.00	100.00	100.00	97.05	72.22	20.00	100.00	100.00
Research	100.00	83.33	93.33	84.21	77.77	41.66	83.33	86.36
Not answered	95.34	77.02	92.27	80.76	66.89	33.33	82.91	96.55
Inhibition Test								
Inhibition test No	95.12	80.32	91.23	79.97	61.73	30.63	83.15	94.18
Inhbition test Yes	98.61	74.72	88.09	73.84	53.60	26.00	91.50	96.86
Inhibiton test only Negative samples	100.00	57.14	92.85	78.57	75.00	22.72	73.07	80.48
Not answered	100.00	33.33	92.30	84.61	57.14	0.00	88.89	100.00

It is concluded that the classical approach to analyse the qualitative data is not appropriate. Therefore, a model needs to be developed that fits the data appropriately based on techniques other than the classical approach.

Table 4.3: *Results from the classical GLM applied to the weakest EV sample group.*

Parameters	Estimate	SE	p-value
Intercept	-0.918	0.361	0.011
Year 2002 (baseline 2005)	-0.288	0.405	0.477
Year 2003 (baseline 2005)	0.624	0.389	0.109
Year 2004 (baseline 2005)	0.020	0.389	0.957
Tech. CC (baseline CIH)	-1.656	1.066	0.121
Tech. RTC (baseline CIH)	-0.152	0.721	0.833
Tech. NASBA (baseline CIH)	-1.165	1.086	0.283
Tech. RTIH (baseline CIH)	-0.212	0.319	0.507
Anti-contamination Yes (baseline No)	-0.063	0.337	0.850
Accreditation Yes (baseline No)	-0.044	0.272	0.871

4.4 Proposed Model for the Qualitative Responses based on Bayesian Methods

The logistic regression model formulated in Section 4.2 under the classical approach will be further use to develop the Bayesian framework. Under this new framework a prior distribution for the probability of detecting correctly the virus in a sample needs to be specified in advance. When information is given on several different levels of observational units a hierarchical model approach should be used (Gelman et al., 2004). The data have a hierarchical structure mainly due to the different number of sample groups and differences in participants' laboratory practices.

When fitting models to the data, the use of a non-hierarchical model is inappropriate for hierarchical data since many parameters have to be estimated, and non-hierarchical models tend to overfit these data (Gelman et al., 2004). Overfitting occurs when the model fits the data well, but it leads to inferior prediction for new data. By using a hierarchical model instead, it is possible to use the probability distributions to structure some dependence into the parameters and avoid problems of overfitting.

For this reason, the QLBM was developed as a hierarchical model with prior and hyper-prior distributions for the parameters and missing covariates to be estimated. In this way, inferences on the probabilities can be derived from the conditional posterior distributions of the parameters and missing covariates.

A general derivation of the conditional posterior distribution similar to the one presented in this chapter can be found in advanced statistics text books (Gilks et al., 1996, Gelman et al., 2004, Banerjee et al., 2004). Since a different combination of likelihood and priors is used than those which can be found in articles and books, the particular equations necessary for the present application are derived and shown in what follows.

4.4.1 Likelihood Function

Let y_{si} be a realization of the i^{th} participant's response observed for sample group s. The random variable Y_{si} follows a Bernoulli distribution with parameter p_{si}. As a result, the probability mass function is given by

$$f(y_{si}|p_{si}) = (p_{si})^{y_{si}}(1-p_{si})^{(1-y_{si})}.$$

Assuming independence of y_{si}, the likelihood of p_{si} after observing the data can be written as:

$$L(p|y) = \prod_{si} f(y_{si}|p_{si}) = \prod_{si}(p_{si})^{y_{si}}(1-p_{si})^{(1-y_{si})},$$

where the multiplication $\prod_{si} = \prod_{s=1}^{l}\prod_{i=1}^{n_s}$ is the product over all observations over the l groups $(s = 1, .., l; i = 1, .., n_s)$.

4.4.2 Regression Model

In order to fit the logistic regression model described by equation (4.2), the likelihood is expressed as a function of the vector of parameters $\vec{\beta}_s$ and the vector of covariates $\vec{x}_{si}$ as:

$$L(p|y) = \prod_{si} f(y_{si}|p_{si}) = \prod_{si} f(y_{si}|\vec{x}_{si}\vec{\beta}_s) = L(\beta, x|y). \tag{4.3}$$

The last term on the right hand side of equation (4.3) can be expressed as

$$L(\beta, x|y) = \prod_{si} f(y_{si}|\vec{x}_{si}\vec{\beta}_s) = \prod_{si} \left(\frac{\exp(\vec{x}_{si}\vec{\beta}_s)}{1+\exp(\vec{x}_{si}\vec{\beta}_s)}\right)^{y_{si}} \left(1 - \frac{\exp(\vec{x}_{si}\vec{\beta}_s)}{1+\exp(\vec{x}_{si}\vec{\beta}_s)}\right)^{(1-y_{si})}$$

$$= \prod_{si} \frac{\exp(\vec{x}_{si}\vec{\beta}_s y_{si})}{(1+\exp(\vec{x}_{si}\vec{\beta}_s))^{y_{si}}} \frac{1}{(1+\exp(\vec{x}_{si}\vec{\beta}_s))^{(1-y_{si})}}.$$

Thus,

$$L(p|y) = L(\beta, x|y) = \prod_{si} \frac{\exp(\vec{x}_{si}\vec{\beta}_s y_{si})}{1+\exp(\vec{x}_{si}\vec{\beta}_s)}.$$

4.4.3 Bayesian Framework

A particular feature of the Bayesian approach is the use of prior distributions for the parameters to be estimated and for missing covariates. In the QLBM the following prior distributions are used:

- For the vector parameter $\vec{\beta}_s^T = (\beta_{s1}, .., \beta_{sr})$, with $s = 1, .., l$, $r - 1$ being the total number of explanatory variables included in the model for sample s and β_{s1} being the parameter associated with the identity vector to allow for the inclusion of an intercept into the model, the prior is given by the normal distribution:

$$\vec{\beta}_s|\vec{\beta}_{s0}, \mathbf{V}_{s0} \sim N_p(\vec{\beta}_{s0}, \mathbf{V}_{s0}),$$

  with mean vector β_{s0} and covariance matrix $\mathbf{V}_{s0}$.

- For the missing covariates x_{sij}, with j taking values from *2,..,r* and r being the total number of covariates for sample s, the priors are chosen based on the type of the covariate as follows:

  - If the j^{th}-covariate is a binary variable, then

    $$x_{sij}|b_j \sim Bernoulli(b_j),$$

    where b_j is the probability of success.

  - If the j^{th}-covariate is a categorical variable (with more than two categories), then

    $$x_{sij}|\vec{g}_j[] \sim Categorical(\vec{g}_j[]),$$

    where $\vec{g}_j[]$ is the vector of assigned probabilities to each category of the variable. That is the $Prob(x_{sij}) = g_j[x_{sij}]$ with $x_{sij} = 1, .., dim(\vec{g}_j)$ such that $0 \leq g_j[x_{sij}] \leq 1$ and $\sum_{i=1}^{dim(\vec{g}_j)} g_j[i] = 1$. The Categorical distribution is a discrete univariate distribution for a random variable that measure one possible outcome out of several categories. It is like an extension of a Bernoulli distribution but instead of having failure and success as possible outcomes of the variable, in the Categorical the outcomes are more than two categories, i.e. type of laboratory where more than two options of laboratory type are available. The random variable that follows a Categorical distribution in this case has different probabilities of being each category of possible outcome as described above. Application of this discrete distribution to data can be found in Roche et al (1975) and Thissen (1986) (p.71) (Lunn et al., 2000). This should not be confused with a multinomial distribution where the sum of several outcomes with different categories are measured (Gelman et al., 2004).

In the model proposed here, $\vec{\beta}_{s0}\,, \mathbf{V}_{s0}\,, b_j$ and $\vec{g}_j[]$ are the hyperparameters. The hyperparameters may be estimated using only the data, known as the 'empirical Bayes' approach or they may also be given a prior distribution, which is known as the 'full Bayes' approach. The 'full Bayes' approach with prior distributions for the hyperparameters, described in the following, is used.

Since there is no specific information available a priori about the regression hyperparameters and variances, their corresponding prior distributions can be set up as follows:

- Prior distributions for the elements of the vector $\vec{\beta}_{s0}$ are set up as uniform flat distributions

$$\beta_{s01}, .., \beta_{s0r} \sim U[-100, 100].$$

- Prior distributions for the elements of the diagonal matrix V_{s0} are set up as inverse-gamma flat distributions

$$V_{s01}, .., V_{s0r} \sim InvGamma(0.001, 0.001).$$

- Prior distributions for b_j are set up as Beta distributions

$$b_j \sim Beta(a, b),$$

  where a and b take positive values; in particular $a = 2$ and $b = 2$ are used to obtain an informative symmetric hyper-prior distribution with mean 0.5 or $a = 1$ and $b = 1$ to obtain a flat hyper-prior distribution.

- Prior distributions for $\vec{g}_j[]$ are set up as Dirichlet distributions

$$\vec{g}_j[] \sim Dirichlet(\vec{\alpha}[]),$$

which is the multivariate distribution corresponding to the beta distribution. The vector of parameters $\vec{\alpha}[]$ is defined positive (each component is greater than 0), so that each of the components of the vector $\vec{g}_j[]$ has equal expectation.

The prior distribution are well known priors appropriate for the hyperparameters, which were also chosen to provide conjugate posterior distribution where it is possible (Gelman et al., 2004).

4.4.3.1 Posterior Distributions

The posterior distribution for the parameters to be estimated is derived from the likelihood and the prior distributions as:

$$\pi(\beta, x|y) \propto likelihood \times prior \propto L(\beta, x)\pi(\beta|\beta_0, V_0)\pi(x|b, g[]).$$

Since $\vec{\beta}_s$ and $\vec{x}_{si}$ are independent, the prior density functions for β and x can be written as:

$$\pi(\beta|\beta_0, V_0) = \prod_s \pi(\vec{\beta}_s|\vec{\beta}_{s0}, \mathbf{V}_{s0}) \sim \prod_s N_p(\vec{\beta}_{s0}, \mathbf{V}_{s0})$$

and

$$\pi(x|b, g) = \prod_{sij} \pi(x_{sij}|b_j, \vec{g}_j[]) \sim \prod_{sij} Bernoulli(b_j)^{(I_j)} Categorical(\vec{g}_j[])^{(1-I_j)},$$

where I_j is an indicator variable for the missing covariates and the product $\prod_{sij} = \prod_{s=1}^{l} \prod_{i=1}^{n_s} \prod_{j=1}^{r}$ is over all covariates j for all participants' responses i and all sample group s.

The posterior distribution of each parameter can be expressed in terms of its posterior conditional distribution, which will be used in the estimation procedure. Although the vector of parameters $\vec{\beta}_s$ is of primary interest, conditional posterior distributions for $\vec{\beta}_s$ and the missing covariates x_{sij} are also presented here.

- The conditional posterior distribution for $\vec{\beta}_s$ with $s = 1, .., l$ is given by:

$$\pi(\vec{\beta}_s|y, x, \vec{\beta}_{-s}) \propto L(\vec{\beta}_s, x)\pi(\vec{\beta}_s|\vec{\beta}_{s0}, \mathbf{V}_{s0})$$

$$\pi(\vec{\beta}_s|y, x_{s.}, \vec{\beta}_{-s}) \propto \prod_{i_{(s)}} \frac{\exp(\vec{x}_{si}\vec{\beta}_s y_{si})}{1 + \exp(\vec{x}_{si}\vec{\beta}_s)} \times \exp\left\{-\frac{1}{2}(\vec{\beta}_s - \vec{\beta}_{s0})^T \mathbf{V}_{s0}^{-1}(\vec{\beta}_s - \vec{\beta}_{s0})\right\},$$

  where $i_{(s)}$ is the subgroup of responses from participants for sample group s, and $\vec{\beta}_{-s}$ are the vectors of regression parameters for each sample group except for sample group s (that is, except $\vec{\beta}_s$).

- The conditional posterior distribution for missing variables x_{sij} where j can take on values from *2,..,r* are given as follows:

  – For missing binary variables

$$\pi(x_{sij}|y_{si}, x_{-sij}, \beta_{sj}) \propto L(\beta_{sj}, x_{sij})\pi(x_{sij}|b_j)$$

$$\pi(x_{sij}|y_{si}, x_{-sij}, \beta_{sj}) \propto \frac{\exp(x_{sij}\beta_{sj}y_{si})}{1 + \exp(x_{sij}\beta_{sj})} \times (b_j)^{x_{sij}}(1 - b_j)^{(1-x_{sij})}.$$

  – For missing categorical variables

$$\pi(x_{sij}|y_{si}, x_{-sij}, \beta_{sj}) \propto L(\beta_{sj}, x_{sij})\pi(x_{sij}|g_j[x_{sij}])$$

$$\pi(x_{sij}|y_{si}, x_{-sij}, \beta_{sj}) \propto \frac{\exp(x_{sij}\beta_{sj}y_{si})}{1 + \exp(x_{sij}\beta_{sj})} \times g_j[x_{sij}],$$

  where x_{-sij} are all the covariate values except the observed j^{th} covariate value for the i^{th} observation in sample group s.

4.4.4 Model Selection Procedure

As described in Chapter 1, for model selection, a backwards elimination procedure based on the conditional posterior distributions for the estimated parameters was applied. The 95% highest density intervals for the means of the conditional posterior distributions of the parameters were obtained and used to perform model selection. In addition to the model selection procedure, possible confounders, interactions and correlated parameters were studied.

In a first step, the full model with all covariates is fitted. Then, using a backward selection procedure, the number of parameters in the model is reduced to obtain conclusions about the effect of the significant covariates on the estimated probabilities of correct results. Covariates which are furthest from being significant at the two-sided 5% level are successively removed. In each step, it is checked that the change of the model does not affect the values of all other parameters by more than 30% of their previous values (as a rule of thumb defined by other authors) (Miettinen and Cook, 1981), otherwise the removed covariate in the previous step is returned back to the model (confounder variable) (Hak et al., 2002).

The parameters associated with each of the variables are tested at the two-sided 5% significance level. The test is based on the posterior probabilities of the parameter estimates being greater than zero. Posterior probabilities from the conditional posterior distributions of the parameters are obtained. If the posterior probability is close to zero (smaller than 0.025), zero is on the upper tail of the distribution, and thus the parameter is concluded to be significant at the one-sided 2.5% level. If the posterior probability is close to 1 (higher than 0.975), zero is on the lower tail of the distribution and it is concluded that the parameter is significant at the one-sided 2.5% level. Finally, if the parameter is not significant at the one-sided 2.5% level, the covariate corresponding to that parameter is removed from the model.

A further check is made to ensure that changes to the model do not interfere on the significance of other parameters. If a high correlation between parameters of covariates is found then the covariate is not removed from the model (highly correlated is defined as a correlation coefficient with modulus higher than 0.7) (Cohen, 1988).

Pair-wise interactions between covariates are studied and taken into account when reducing the parameters in the model. The pair-wise interactions studied are chosen by theoretical knowledge and practical interest. Given that the covariate Technology is the exploratory variable of main interest, in general all pair-wise interactions studied are between Technology and other covariates.

The results from the full and reduced models are presented in Section 4.5, together with the 95% highest density intervals (or confidence intervals), defined as the most likely (with probability of 0.95) estimates of the parameter under investigation. If the value 0 is within the interval then the parameter is assumed to be 0, and the associated covariate of the parameter is removed from the model for not being significantly different from 0. When reporting the results, the estimated probability refers to the expected probability for the distribution of participants' results (incorrect/correct detection of the virus). This expectation is obtained from the posterior distribution of the probabilities for detecting the virus correctly otained using MCMC method.

4.4.5 Model Checking

Statistical models should be checked for adequacy of their fit to the data, thus model checking should be included in the statistical modelling analysis. To check if a model is consistent with the data, the posterior predictive results are assessed. If the model fits the data well, then replicated data under model conditions should look similar to the observed data.

For qualitative response variables, which are under investigation in this chapter, the statistical test and the procedure to assess predictive posterior distribution are described as follows:

- Let $T(y_{si_k}, \theta_s)$ be a test statistic, a summary measure of the data y_{si_k} and the parameters θ_s for the sample group s and the subgroup of observations of the year k, i_k.
- $T(y_{si_k}, \theta) = \frac{\sum_{i_k} y_{si}}{n_{sk}}$ is defined as the sum of the responses within sample group s and year k divided by the total number of responses, n_{sk}, for that sample group and year.
- For each simulation from the posterior distribution of θ_s one replicated dataset of responses from the predictive distribution for the sample group s and year k, $y^r_{si_k}$, is generated. From those replicated data, the joint posterior predictive distribution $p(y^r_{si_k}, \theta_s | y_{si_k})$ is obtained.
- Then, the observed test quantity, $T(y_{si_k}, \theta_s)$, is compared with the predicted test quantities, $T(y^r_{si_k}, \theta_s)$.
- The estimated Bayesian 'p-value' is the proportion of those simulations for which the predictive test quantity is equal or exceeds the observed value, i.e. the proportion of simulation such that $T(y^r_{si_k}, \theta_s) \geq T(y_{si_k}, \theta_s)$.

4.4.6 Model Comparison

In order to determine if the reduced model fits the data well, methods based on comparing the posterior distributions of the full and reduced models are used. Therefore, the estimated posterior probabilities of detecting the virus correctly and their 95% confidence intervals are obtained for both full and reduced models. Then, the estimated posterior probabilities and their 95% confidence intervals of the full model are plotted, and the estimated posterior probabilities of the reduced model are added to the plot.

If the reduced model fits the data well, it is expected that the posterior probabilities estimated by the reduced model lie within the 95% confidence intervals of the posterior probabilities obtained with the full model.

An advantage of the Bayesian approach is that the posterior predictive distribution can be computed for any data summary, as it has been shown with the estimated probabilities. Using simulation, the posterior predictive probabilities are computed and the means of the distributions of the estimated probabilities for the reduced model are checked to lie within the distribution of estimated probabilities for the full model.

4.5 Model Application

The results presented in this section are obtained for the following specifications of the QLBM:

- The hyperparameters b_j are estimated from a Beta distribution with shape and scale parameter 2.
- $\vec{g}_j$ are estimated from a Dirichlet distribution as described in the previous section.
- The $\vec{\beta}_{s0}$ are set to be 0 and $\mathbf{V}_{s0}$ defined as the identity diagonal matrix in order to provide an informative prior distribution to the parameters with equal information for all sample groups.

Variations of the estimates from the QLBM with different choices of prior distributions for the hyperparmeters are studied, and the results are summarised in Section 4.6.

4.5.1 Modelling EV Qualitative Data

Application of the QLBM to the Enterovirus (EV) proficiency panels introduced in Chapter 2, will be presented in this section. Table 4.4 describes the variables included in the model. Note that the covariate plasma is categorised given that participants provided an interval around the number of plasma tests performed annually.

Table 4.4: *Covariates included in the EV analysis.*

Covariate	Description	Values
Year	Year when the sample was analysed	Indicator variables for years 2002 to 2004 compared with year 2005
Technology	Technology used to analyse the sample	Indicator variables for technology groups: CC, RTC, RTIH and NASBA compared with CIH
Anti	Use of anti-contamination system	Indicator variable with No use of anti-contamination system as baseline
Accred	Laboratory accreditation status	Indicator variable with No accredited laboratory as baseline
CSF	Experience on CSF samples performance	Indicator variable with No experience performing CSF test as baseline
Swab	Experience on swab samples performance	Indicator variable with No experience performing swab test as baseline
Biopsies	Experience on biopsies samples performance	Indicator variable with No experience performing biopsies test as baseline
Serum	Experience on serum samples performance	Indicator variable with No experience performing serum test as baseline
Plasma	Annual number of plasma tests performed by the participant	Indicators variables per group of number of plasma test: 0-10 baseline, 11-100- group 1, 101-1,000 - group 2, 1,001-2,000- group 3, 2,001-10,000- group 4, $> 10,000$- group 5
Analysis	Method of analysis used by the participant	Indicator variables for analysis method. The baseline of singular analysis compared with Duplicate and other analysis methods
Inhibition	Performance of inhibition test by the participant	Indicator variables with non-performance of inhibition tests as baseline compared with performance of inhibition test and performance of inhibition test only in negative samples
Labtype	Laboratory type where the sample was analysed	Indicator variables with hospital laboratories as baseline compared with public health laboratory, private laboratories, reference laboratories, manufactures laboratories and research laboratories

Technologies: CC=Conventional Commercial, RTC=Real Time Commercial, CIH=Conventional Commercial, RTIH=Real Time In-house and NASBA=NASBA.

The baseline of a non-ordinal covariate is chosen based mainly on the most frequent level of the covariate. For an ordinal covariate the baseline is chosen to be the lowest category. For the covariate year the baseline is chosen to be the last year, in order to compare and draw conclusions on the responses for samples from the last year with respect to the responses for samples from previous years (a priori a better performance is expected in more recent years). In a previous study some covariates were assumed to behave linearly and were considered to be continuous variables (García-Fernández et al., 2007). However, this consideration of linearity assumption led to inappropriate results for those covariates.

4.5.1.1 Full EV Model

The QLBM is fitted to the data in order to check the influence of significant covariates associated with the correct detection of the virus. A summary of the results obtained from applying the QLBM to the full data set for each sample group dilution series is presented in this section (Table 4.5). Significant refers to significance at the two-sided 5% level.

Sample group dilution series 1×10^{-3}

The strongest samples, with dilution of 1×10^{-3}, are only included in the panels of 2002 and 2005. A summary of the results obtained from the full model for this sample group can be found in Table 4.6.

Participants using CC, RTIH and NASBA technologies are more likely to detect the virus correctly than those using CIH technology. In contrast, participants using RTC technology are less likely to detect the virus correctly than those using CIH technology, although, these differences are not significant at the two-sided 5% level.

Table 4.5: *Mean and SD of the parameter estimates from the full QLBM for EV sample groups. The results in bold are significant at the two-sided 5% level.*

Covariates	Mean (SD)							
	$1x10^{-3}$	$1x10^{-4}$	$1x10^{-5}$	$1x10^{-6}$	$1x10^{-7}$	$1x10^{-8}$	Non-EV	Negative
Intercept	1.499 (0.744)	1.669 (0.524)	1.053 (0.408)	1.056 (0.342)	0.008 (0.412)	-0.948 (0.468)	0.888 (0.446)	1.895 (0.556)
Year- baseline "2005"								
Year 2002	-0.090 (0.687)		0.579 (0.395)	**-0.602 (0.269)**	**-0.846 (0.282)**	-0.069 (0.377)		-0.137 (0.487)
Year 2003			**0.941 (0.359)**	-0.078 (0.265)	**1.323 (0.336)**	**0.810 (0.348)**	-0.325 (0.336)	-0.167 (0.456)
Year 2004		-0.467 (0.347)	-0.023 (0.266)	-0.374 (0.219)	**0.739 (0.269)**	0.273 (0.307)	0.095 (0.353)	0.157 (0.431)
Technology group- baseline "CIH"								
Tech CC-CIH	0.261 (0.924)	0.951 (0.789)	0.788 (0.647)	0.580 (0.403)	-0.839 (0.440)	-0.540 (0.561)	-0.434 (0.500)	-0.153 (0.660)
Tech RTIH-CIH	0.585 (0.717)	-0.642 (0.390)	-0.519 (0.260)	-0.064 (0.177)	-0.095 (0.223)	0.192 (0.275)	**0.890 (0.329)**	0.307 (0.398)
Tech NASBA-CIH	0.191 (0.937)	0.242 (0.929)	-0.064 (0.783)	-0.728 (0.469)	-0.411 (0.550)	-0.568 (0.737)	0.559 (0.863)	0.657 (0.839)
Tech RTC-CIH	-0.349 (0.844)	-0.860 (0.604)	-1.418 (0.448)	-0.728 (0.375)	**-1.368 (0.474)**	-0.062 (0.571)	0.772 (0.662)	0.644 (0.807)
Anti- baseline "No"	0.756 (0.813)	**0.919 (0.472)**	0.584 (0.312)	0.012 (0.195)	-0.102 (0.244)	0.026 (0.308)	0.035 (0.364)	0.036 (0.445)
Accred- baseline "No"	-0.170 (0.668)	-0.438 (0.406)	0.173 (0.276)	-0.156 (0.194)	0.314 (0.230)	-0.280 (0.293)	0.094 (0.337)	0.164 (0.388)
CSF- baseline "No"	**1.763 (0.735)**	0.871 (0.558)	0.743 (0.367)	0.591 (0.298)	0.354 (0.371)	-0.263 (0.389)	0.116 (0.417)	0.773 (0.470)
Serum- baseline "No"	-0.060 (0.732)	-0.472 (0.504)	-0.205 (0.323)	-0.216 (0.227)	-0.067 (0.278)	0.526 (0.344)	0.305 (0.379)	-0.389 (0.403)
Swab- baseline "No"	-0.134 (0.740)	-0.127 (0.530)	0.061 (0.381)	0.201 (0.289)	0.360 (0.341)	0.023 (0.406)	-0.243 (0.430)	0.307 (0.460)
Biopsies- baseline "No"	0.541 (0.729)	-0.131 (0.511)	0.637 (0.349)	0.277 (0.270)	0.308 (0.333)	-0.174 (0.410)	0.612 (0.447)	0.179 (0.467)
Analysis method- baseline Singly								
Analysis Duplicated	0.489 (0.630)	**0.818 (0.377)**	0.091 (0.236)	0.060 (0.156)	0.033 (0.195)	-0.334 (0.248)	-0.130 (0.289)	0.239 (0.346)
Analysis Other	0.267 (0.921)	0.041 (0.993)	0.915 (0.749)	**1.291 (0.512)**	1.043 (0.555)	0.068 (0.620)	1.114 (0.753)	1.091 (0.763)
Plasma- baseline 0-10								
Group 1: 11-100	-0.635 (0.748)	-0.530 (0.458)	-0.259 (0.319)	-0.399 (0.273)	-0.640 (0.327)	-0.640 (0.327)	0.422 (0.433)	-0.027 (0.464)
Group 2: 101-1,000	0.163 (0.740)	-0.734 (0.541)	0.067 (0.369)	0.145 (0.292)	-0.063 (0.356)	-0.063 (0.356)	0.244 (0.412)	0.544 (0.479)
Group 3: 1,001-2,000	0.402 (0.901)	1.133 (0.698)	0.878 (0.582)	-0.321 (0.358)	-0.562 (0.449)	-0.562 (0.449)	0.178 (0.529)	-0.173 (0.561)
Group 4: 2,001-10,000	0.328 (0.885)	-0.068 (0.745)	0.551 (0.625)	0.094 (0.447)	-0.635 (0.506)	-0.635 (0.506)	0.385 (0.725)	-0.441 (0.664)
Group 5: > 10,000	0.164 (0.954)	0.632 (0.772)	0.210 (0.679)	0.071 (0.614)	-0.679 (0.667)	-0.679 (0.667)	-0.349 (0.713)	0.415 (0.901)
Labtype- baseline Hospital								
Public Health	0.890 (0.803)	-0.016 (0.512)	**1.010 (0.413)**	0.306 (0.228)	0.069 (0.304)	0.067 (0.372)	0.429 (0.428)	-0.392 (0.425)
Private	-0.569 (0.890)	**-1.734 (0.661)**	-0.368 (0.472)	0.085 (0.387)	-0.434 (0.458)	0.753 (0.503)	-0.057 (0.579)	0.228 (0.728)
Reference	0.302 (0.912)	0.130 (0.783)	0.457 (0.621)	0.777 (0.462)	0.178 (0.498)	0.004 (0.596)	0.733 (0.788)	-0.191 (0.694)
Manufacture	0.547 (0.854)	0.794 (0.832)	**1.443 (0.710)**	**2.098 (0.568)**	**1.468 (0.535)**	0.369 (0.686)	0.595 (0.808)	0.824 (0.787)
Research	0.440 (0.882)	0.522 (0.765)	1.015 (0.595)	0.781 (0.443)	1.002 (0.555)	0.465 (0.582)	0.059 (0.657)	-0.085 (0.631)
Inhibition Test-baseline "No"								
Inhbition test Yes	0.872 (0.709)	-0.543 (0.401)	-0.137 (0.256)	-0.301 (0.173)	-0.280 (0.217)	-0.251 (0.273)	**0.752 (0.324)**	0.675 (0.406)
Inhibiton test only Negative samples	0.452 (0.877)	-0.853 (0.709)	0.023 (0.525)	-0.108 (0.326)	0.712 (0.445)	-0.512 (0.500)	-0.494 (0.494)	-1.095 (0.484)

Results in bold are significant at the two-sided 5% level.

The results show that participants with experience performing CSF tests are significantly more likely to detect the virus correctly than those without experience performing CSF tests.

Experience performing duplicate or other methods of analysis, such as triplicate methods, tends to improve participants' performance, but this finding is not significant at the two-sided 5% level.

Experience performing plasma or biopsies tests has a positive influence when detecting the virus correctly, however this tendency is not significant at the two-sided 5% level.

Only private laboratories are less likely to detect the virus correctly than hospital laboratories, and performing inhibition tests indicate a positive tendency for detecting the virus correctly, but these findings are not significant at the two-sided 5% level.

As an illustrative example, the estimated probability of providing a correct result, using the QLBM model, is obtained for a participant with a particular combination of laboratory practices. Then, the estimated probability is compared with the observed probability of correct results from participants fulfilling the same conditions.

For this example, consider a participant with the following laboratory practice for analysing a sample of dilution series 1×10^{-3} from 2002: The participant

- used RTIH technology to analyse the sample,
- was an accredited laboratory,
- did not use an anti-contamination system (baseline),
- had experience performing a CSF test,

- had experience performing a serum test,
- did not have any experience with biopsies and swab (baseline),
- was using a duplicated method of analysis,
- has experience in testing between 101 and 1,000 plasma tests annually,
- was a hospital laboratory (baseline),
- did not perform any inhibition test (baseline).

For such a participant, the probability of having a correct result is derived from the estimated means given in Table 4.6 as follows:

$$logit(\hat{p}) = 1.499 - 0.09 \times (\textit{Year=2002}) + 0.585 \times (\textit{RTIH=yes}) - 0.17 \times (\textit{Accred.=yes}) +$$

$$1.763 \times (\textit{CSF=yes}) - 0.06 \times (\textit{Serum=yes}) + 0.489 \times (\textit{Duplic.=yes}) + 0.163 \times (\textit{Plasma=yes}).$$

Since the covariates are indicator variables for non-baseline information, the above expression can be rewritten as

$$logit(\hat{p}) = 1.499 - 0.09 + 0.585 - 0.17 + 1.763 - 0.06 + 0.489 + 0.163 = 4.179$$

$$\hat{p} = \frac{\exp(4.179)}{(1 + \exp(4.179))} = 0.99.$$

Thus, the participant in this example has a probability of 0.99 to return the result correctly. A comparison with the observed data shows that all participants with the same laboratory practice provided their results correctly for the sample from 2002.

A participant with the same characteristics, but who had no experience with CSF tests would have:

$$logit(\hat{p}) = 1.499 - 0.09 + 0.585 - 0.17 + 1.763 \times 0 - 0.06 + 0.489 + 0.163 = 2.416$$

and hence the estimated probability

$$\hat{p} = \frac{\exp(2.416)}{(1 + \exp(2.416))} = 0.918.$$

So participants having experience performing CSF test are $\frac{0.99}{0.918} = 1.078$ times more likely to detect the virus correctly than participants with no experience performing CSF tests.

Table 4.6: *Summary statistics for the parameter estimates from the full QLBM for EV sample group dilution series* 1×10^{-3}*: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	1.499	0.744	0.091	1.494	3.004		
Year- baseline "2005"							
Year 2002	-0.090	0.687	-1.440	-0.094	1.293	-	No
Technology group- baseline "CIH"							
Tech CC-CIH	0.261	0.924	-1.541	0.263	2.114	+	No
Tech RTIH-CIH	0.585	0.717	-0.770	0.571	2.051	+	No
Tech NASBA-CIH	0.191	0.937	-1.603	0.175	2.046	+	No
Tech RTC-CIH	-0.349	0.844	-1.984	-0.367	1.384	-	No
Anti- baseline "No"	0.756	0.813	-0.746	0.726	2.406	+	No
Accred- baseline "No"	-0.170	0.668	-1.505	-0.165	1.123	-	No
CSF- baseline "No"	1.763	0.735	0.329	1.755	3.230	+	Yes
Serum- baseline "No"	-0.060	0.732	-1.431	-0.074	1.385	-	No
Swab- baseline "No"	-0.134	0.740	-1.595	-0.130	1.331	-	No
Biopsies- baseline "No"	0.541	0.729	-0.919	0.542	1.971	+	No
Analysis method- baseline Singly							
Analysis Duplicated	0.489	0.630	-0.725	0.489	1.740	+	No
Analysis Other	0.267	0.921	-1.550	0.262	2.093	+	No
Plasma- baseline 0-10							
Group 1: 11-100	-0.635	0.748	-2.087	-0.636	0.852	-	No
Group 2: 101-1,000	0.163	0.740	-1.284	0.172	1.619	+	No
Group 3: 1,001-2,000	0.402	0.901	-1.360	0.390	2.178	+	No
Group 4: 2,001-10,000	0.328	0.885	-1.359	0.317	2.102	+	No
Group 5: > 10,000	0.164	0.954	-1.681	0.167	2.058	+	No
Labtype- baseline Hospital							
Public Health	0.890	0.803	-0.655	0.879	2.520	+	No
Private	-0.569	0.890	-2.285	-0.583	1.214	-	No
Reference	0.302	0.912	-1.455	0.286	2.108	+	No
Manufacture	0.547	0.854	-1.116	0.526	2.263	+	No
Research	0.440	0.882	-1.232	0.423	2.211	+	No
Inhibition Test-baseline "No"							
Inhbition test Yes	0.872	0.709	-0.516	0.854	2.281	+	No
Inhibiton test only Negative samples	0.452	0.877	-1.214	0.435	2.202	+	No

Consider now the example of a participant with the following laboratory practice for analysing a sample of dilution series 1×10^{-4} from 2004:

- used RTIH technology to analyse the sample,
- was an accredited laboratory,
- did not use an anti-contamination system (baseline),
- had experience performing CSF tests,
- had no experience performing serum tests (baseline),
- did not have any experience with biopsies and swab (baseline),
- was using a duplicated method of analysis,
- had no experience in testing plasma samples annually (baseline),
- was a hospital laboratory (baseline),
- provided a missing value for performing any inhibition test.

For the missing value of the covariate 'performing any inhibition test', the model estimates the most likely value given the knowledge of the other covariates. In this example, the estimation obtained by the model is that the participant performed an inhibition test. Based on this result, in a second step, the model estimates the probability of a correct result, which in the given example is $p = 0.779$.

The probability of having a correct result for a participant with the same characteristics, but who provided the information of performing an inhibition test (see Table 4.5, more detailed results are presented in Appendix B), is given by:

$$logit(\hat{p}) = 1.669 - 0.467 \times (\textit{Year=2004}) - 0.642 \times (\textit{RTIH=yes}) - 0.438 \times (\textit{Accred.=yes}) +$$

$$0.871 \times (\textit{CSF=yes}) + 0.818 \times (\textit{Duplic.=yes}) - 0.543 \times (\textit{Inh. Test=yes}).$$

$$logit(\hat{p}) = 1.268$$

$$\hat{p} = \frac{\exp(1.268)}{(1 + \exp(1.268))} = 0.780.$$

Note how the model adjusts the probability of a correct result for the participant who provided missing information. The adjustment for this probability with respect to the probability of a correct result for a participant who provided the complete information is not high, although the model assigned slightly less probability of correctly detecting the virus to the participant with missing data.

By checking the observed data, the proportion of correct results from participants with those characteristics is 0.8, which is in agreement with the results provided by the model. The probability of providing a correct result for a participant with the same characteristics but not performing an inhibition test is 0.86, and for a participant only performing an inhibition test in negative samples is 0.72. This indicates that the estimate from the model for the participant with missing covariate is appropriate. The estimated probability from the model is closer to the probability for a participant who performed an inhibition test than to the probability for a participant not performing inhibition tests or performing inhibition tests only in negative samples.

Sample group dilution series 1×10^{-4}

A summary of results for this sample group is presented in Table 4.5. This dilution series is only contained in the panels of 2004 and 2005. No significant differences were found between samples from 2004 and 2005, although for samples from 2004 participants are less likely to detect the virus correctly than for samples from 2005.

There are no significant differences between technologies, but the following tendencies are observed: participants using CC and NASBA technologies are more likely to detect the virus correctly than those using CIH technology, whilst participants using RTC and RTIH technologies are less likely to detect the virus correctly than participants using CIH technology.

Consistent with the results shown in Table 4.2, the use of an anti-contamination system is a significant factor positively associated with the correct detection of the virus.

Experience performing CSF tests tends to improve participants' performance. Those participants are more likely to detect the virus correctly, but it is not significant at the two-sided 5% level.

Participants performing duplicate analysis methods are significantly more likely to detect the virus correctly than those performing single analysis methods.

Participants from private laboratories are significantly less likely to detect the virus correctly than those from hospital laboratories.

Participants who did not perform inhibition tests are more likely to detect the virus correctly than those performing inhibition tests, but the differences on their performance are not statistically significant.

Sample Group Dilution Series 1×10^{-5}

A summary of results for this sample group is presented in Table 4.5. Participants are significantly more likely to detect the virus correctly for samples from 2003 than for samples from 2005.

Participants using CC technology are more likely to detect the virus correctly than those using CIH, although these differences are not significant. Participants using RTIH, RTC and NASBA technologies are less likely to detect the virus correctly than CIH technology users, but these differences between their performance are not significant.

In agreement with Table 4.2, the use of an anti-contamination system tends to improve participants' performance.

Although not statistically significant, experience performing CSF, swabs, biopsies or plasma tests tends to improve the performance of participants.

Public health and manufactures laboratories are significantly more likely to detect the virus correctly than hospital laboratories, while private laboratories are less likely to detect the virus correctly than hospital laboratories.

Sample Group Dilution Series 1×10^{-6}

A summary of results for this sample group is presented in Table 4.5. For samples from 2005 participants are significantly more likely to detect the virus correctly than for samples from 2002.

Although the use of different technologies is not significantly associated with participants' performance, the users of CC technology are more likely to detect the virus correctly than CIH technology users.

Participants using other methods of analysis are significantly more likely to detect the virus correctly than participants using single analysis methods.

Manufacture laboratories are significantly more likely to detect the virus correctly than hospital laboratories, which is consistent with the observed percentages of correct results in Table 4.2. In contrast with previous samples, private laboratories tend to detect the virus correctly as well as hospital laboratories.

Sample Group Dilution Series 1×10^{-7}

A summary of results for this sample group is presented in Table 4.5. For samples from 2002 participants are significantly less likely to detect the virus correctly than for samples from 2005. In contrast, for samples from 2003 and 2004 participants are significantly more likely to provide a correct result than for samples from 2005, which is consistent with the percentages provided in Table 4.2.

Participants using RTC technology are significantly less likely to detect the virus correctly than CIH technology users.

The use of duplicate or other methods of analysis such as triplicate methods tends to improve participants' performance, but this improvement is not significant.

Manufacture laboratories are significantly more likely to provide a correct result than hospital laboratories. Private laboratories are less likely to detect the virus correctly than hospital laboratories, but the differences are not statistically significant.

Sample Group Dilution Series 1×10^{-8}

A summary of results for this sample group is presented in Table 4.5. The samples of dilution series 1×10^{-8} are the weakest positive samples included in the panels across years. In agreement with the information provided in Table 4.2, the results show that for samples from 2003 participants are significantly more likely to detect the virus correctly than for samples from 2005. Although not statistically significant, the same tendency is observed for participants' performance for samples from 2004 with respect 2005.

Users of commercial technologies (RTC, NASBA and CC) are less likely to detect the virus correctly than CIH technology users. However, the differences are not significant.

No significant differences were found between the results from different laboratory types, although all of them are more likely to provide a correct result than hospital laboratories.

Negative Sample Group

A summary of results for this sample group is presented in Table 4.5. None of the covariates included in the model are significantly associated with participants' performance for negative samples.

In particular, no significant differences were found between technology groups or laboratory types. However, the tendencies of the parameters indicate that users of RTIH, NASBA and RTC technologies are more likely to provide the correct response than CIH technology users, as observed in Table 4.2.

The use of duplicate or other method of analysis, such as triplicate, tends to improve participants' performance, but the differences between the methods of analysis are not significant at the two-sided 5% level.

Private and manufacture laboratories are more likely to provide a correct results than hospital laboratories.

4.5.1.2 Non-EV Sample Group

A summary of results for this sample group is presented in Table 4.5. The results show that for samples from 2003 participants are less likely to provide a correct result than for samples from 2005, but this finding is not statistically significant.

Users of RTIH technology are significantly more likely to provide a correct result than CIH users, which is consistent with the observed results in Table 4.2. The same tendency is observed for the results of NASBA and RTC users, however these differences are not significant.

The general tendency of experience testing other specimens, such as plasma, CSF and biopsies, is to improve participants' performance, but it is not statistically significant.

Participants from private laboratories are the worst in providing correct results, but not significantly different from hospital laboratories.

Participants performing an inhibition test are significantly more likely to provide a correct result than those not performing any inhibition test.

4.5.1.3 Reduced EV Model

In order to simplify the full model those covariates, per sample group, that are not significant at the two-sided 5% level are removed. However, covariates behaving as confounders are not excluded from the model even if they are not significant. A covariate is considered to be a cofounder if the estimated parameters, for all other covariates of the reduced model, change more than 30% of their values from the previous model.

Interactions between covariates and correlations of the estimated parameters were also studied. Interaction between parameters were not found. Correlations between the parameters were checked when reducing the model. The estimated parameters were not correlated, in most cases, the absolute values of the correlation coefficients being lower than 0.25.

The results of the reduced model are summarised in the next subsections (more detailed results can be found in Tables B.8 to B.15 in Appendix B). Table 4.7 shows the mean and standard deviation for each of the parameter estimates from the reduced model. The results in bold are the significant estimates at the two-sided 5% level.

Sample Group Dilution Series 1×10^{-3}

The reduced model for the sample group dilution series 1×10^{-3} shows that only the experience performing an CSF test is significantly associated with the correct detection of the virus. Those participants with experience performing CSF tests have $\exp(1.591) = 4.91$ times higher odds of detecting the virus correctly than those with no experience performing CSF tests.

Table 4.7: *Mean and SD of the parameter estimates from the reduced QLBM for EV sample groups with different dilutions series*

Covariates	Mean (SD)							
	$1x10^{-3}$	$1x10^{-4}$	$1x10^{-5}$	$1x10^{-6}$	$1x10^{-7}$	$1x10^{-8}$	Non-EV	Negative
Intercept	1.861 (0.518)	1.156 (0.494)	1.068 (0.372)	0.852 (0.324)	0.070 (0.378)	-1.061 (0.389)	1.156 (0.375)	2.004 (0.349)
Year- baseline "2005"								
Year 2002			0.616 (0.395)	**-0.558 (0.251)**	**-0.762 (0.276)**	0.027 (0.322)		
Year 2003			**0.959 (0.359)**	-0.051 (0.251)	**1.322 (0.335)**	**0.795 (0.307)**		
Year 2004			-0.023 (0.266)	-0.361 (0.214)	**0.705 (0.263)**	0.265 (0.294)		
Technology group- baseline "CIH"								
Tech CC-CIH		0.843 (0.799)	0.760 (0.644)	0.439 (0.398)	**-0.880 (0.429)**	-0.657 (0.535)	-0.481 (0.482)	
Tech RTIH-CIH		-0.604 (0.384)	**-0.562 (0.250)**	-0.116 (0.170)	-0.231 (0.212)	0.114 (0.249)	**0.827 (0.315)**	
Tech NASBA-CIH		0.244 (0.926)	-0.094 (0.765)	-0.891 (0.460)	-0.507 (0.546)	-0.573 (0.703)	0.667 (0.827)	
Tech RTC-CIH		-0.897 (0.596)	**-1.437 (0.419)**	**-0.850 (0.369)**	**-1.448 (0.477)**	-0.110 (0.546)	0.824 (0.653)	
Anti- baseline "No"		0.739 (0.455)	0.561 (0.302)					
Accred- baseline "No"								
CSF- baseline "No"	**1.591 (0.608)**	0.649 (0.544)	**0.740 (0.356)**	0.604 (0.293)	0.381 (0.379)	-0.437 (0.385)	-0.042 (0.402)	**0.870 (0.388)**
Serum- baseline "No"		-0.484 (0.498)	-0.179 (0.302)	-0.210 (0.217)	-0.034 (0.273)	0.499 (0.318)	0.351 (0.363)	
Swab- baseline "No"		0.054 (0.545)	0.128 (0.377)	0.268 (0.285)	0.369 (0.356)	0.035 (0.388)	-0.200 (0.415)	
Biopsies- baseline "No"		-0.208 (0.520)	0.592 (0.369)	0.229 (0.284)	0.266 (0.353)	-0.106 (0.388)	0.592 (0.416)	
Analysis method- baseline Singly								
Analysis Duplicated		**0.737 (0.366)**	0.077 (0.229)	0.052 (0.156)	-0.023 (0.195)			
Analysis Other		0.020 (0.994)	0.909 (0.769)	**1.279 (0.511)**	0.984 (0.551)			
Plasma- baseline 0-10								
Group 1: 11-100		-0.474 (0.452)	-0.269 (0.311)	-0.384 (0.264)	**-0.630 (0.315)**			
Group 2: 101-1,000		-0.738 (0.532)	0.028 (0.370)	0.115 (0.300)	-0.081 (0.348)			
Group 3: 1,001-2,000		1.112 (0.711)	0.891 (0.593)	-0.294 (0.367)	-0.500 (0.445)			
Group 4: 2,001-10,000		-0.060 (0.740)	0.540 (0.624)	0.060 (0.458)	-0.639 (0.511)			
Group 5: > 10,000		0.610 (0.769)	0.237 (0.672)	0.046 (0.627)	-0.654 (0.650)			
Labtype- baseline Hospital								
Public Health		-0.053 (0.500)	**1.033 (0.403)**	0.305 (0.227)	0.185 (0.294)			
Private		**-1.764 (0.640)**	-0.312 (0.461)	0.082 (0.383)	-0.388 (0.444)			
Reference		0.168 (0.762)	0.497 (0.611)	0.823 (0.457)	0.278 (0.485)			
Manufacture		0.728 (0.849)	**1.397 (0.704)**	**2.086 (0.561)**	**1.416 (0.516)**			
Research		0.361 (0.760)	1.027 (0.574)	0.823 (0.441)	**1.040 (0.534)**			
Inhibition Test-baseline "No"								
Inhbition test Yes							**0.676 (0.306)**	0.662 (0.379)
Inhibiton test only Negative samples							-0.461 (0.450)	**-1.075 (0.433)**

Results in bold are significant at two-sided 5% level.

Sample Group Dilution Series 1×10^{-4}

Performing serum, plasma, cerebrospinal fluid, biopsies and swabs tests are not significant. However, they are confounders for other covariates in the model. Therefore, it is preferable to include them into the final model to gain precision in prediction. As the technology variable is a confounder it is also included in the final model.

Participants performing duplicate analysis methods have $\exp(0.737) = 2.08$ times higher odds of detecting the virus correctly than those performing only single analysis.

Private laboratories have $1/\exp(-1.764) = 1/0.17 = 5.83$ times lower odds of detecting the virus correctly than hospital laboratories.

Sample Group Dilution Series 1×10^{-5}

The odds of detecting this virus correctly for samples from 2003 is $\exp(0.959) = 2.6$ times higher than the odds for samples from 2005.

Users of RTC and RTIH technologies have, respectively, 4.2 and 1.75 times lower odds of detecting the virus correctly than participants using CIH technology.

Using an anti-contamination system does not significantly influence participants' performance, although those participants using it have almost twice the odds of detecting the virus correctly than those not using an anti-contamination system ($\exp(0.561) = 1.75$).

Experience in plasma, biopsies and swab tests are not significantly associated with the correct detection of the virus; however, this experience has an influence on others variables such as those related to performing a serum test or related to the method of analysis. Thus, it is more appropriate to include them into the final model.

Participants from public health and manufacture laboratories have 2.8 (exp(1.033)) and 4 (exp(1.397)) times higher, respectively, odds of detecting the virus correctly than hospital laboratories.

Sample Group Dilution Series 1×10^{-6}

For samples from 2002 participants have $1/\exp(-0.558) = 1.747$ times lower odds of detecting the virus correctly than for samples from 2005.

RTC, NASBA and RTIH technologies users are less likely to detect the virus correctly than CIH users. For example, the odds of detecting the virus correctly for RTC users is $1/\exp(-0.85) = 2.33$ times lower than the odds for CIH users.

Participants performing other methods of analysis, such as triplicate, are significantly more likely to detect the virus correctly than those performing single analysis methods, with an odds ratio of $\exp(1.279) = 3.59$.

Participants from manufacture laboratories have $\exp(2.086) = 8.05$ times higher odds of detecting the virus correctly than hospital laboratories.

Sample Group Dilution Series 1×10^{-7}

It was found that for samples from 2002, participants have 2.14 times lower odds of detecting the virus correctly than for samples from 2005. For samples from 2003 and 2004 participants are significantly more likely to detect the virus correctly than for samples from 2005 (with odds ratios of 3.7 and 2.02, respectively).

All other technology users are less likely to detect the virus correctly than CIH users, for example the odds for RTC users are 4.22 times lower than the odds for CIH users.

As in previous samples groups, the exploratory variables experience performing tests from serum, biopsies, CSF, swabs and method of analysis are taken into account in the model, since their exclusion implies distortion between them and experience performing plasma tests. However, they are not significant at the two-sided 5% level.

The odds for participants from manufacture and research laboratories are 4 times higher than the odds for hospital laboratories.

Sample Group Dilution Series 1×10^{-8}

For the weakest concentration sample group no significant differences were found between users of technology groups. However, the odds of detecting the virus correctly for samples from 2003 is twice the odds for samples from 2005.

Negative Sample Group

Participants with experience performing CSF tests have 2.38 times higher odds of detecting the virus correctly than those with no experience performing CSF tests, at the two-side 5% significance level.

Although not statistically significant, participants performing inhibition tests are more likely to detect the virus correctly than those not performing inhibition test (with an approximately odds ratio of 2). However, the odds for those participants who performed only inhibition tests in negative samples is twice lower than the odds for participants who did not perform any inhibition test, at the two-sided 5% significance level.

Non-EV Sample Group

In contrast to positive EV sample groups, for the non-EV sample group RTC, NASBA and RTIH technologies users are more likely to provide a correct result than CIH users, with the odds for RTIH users 2.27 times higher than the odds for CIH users, at the two-sided 5% significance level.

Those laboratories performing inhibition tests have almost twice the odds of detecting the virus correctly than laboratories who did not perform any inhibition test.

4.5.1.4 Model Checking

Figure 4.1 shows the density function of the test statistic T obtained from 10,000 simulated datasets from the model for the negative sample in 2004. The black line represents the observed data value of T, $T(y_{si_k}, \theta_s)$. The area under the density curve on the right tail represents the Bayesian 'p-value' or the probability that the predictive posterior statistics T from replicated data is more extreme than the T from observed data.

Table 4.8 shows the probability of $T(y^r_{si_k}, \theta_s) \geq T(y_{si_k}, \theta_s)$ for each sample group and year. That is the Bayesian 'p-value' used for assessing the statistical significance of discrepancies between the observed data and the predicted data. An extreme Bayesian 'p-value' indicates a conflict between the observed data and an aspect of the model.

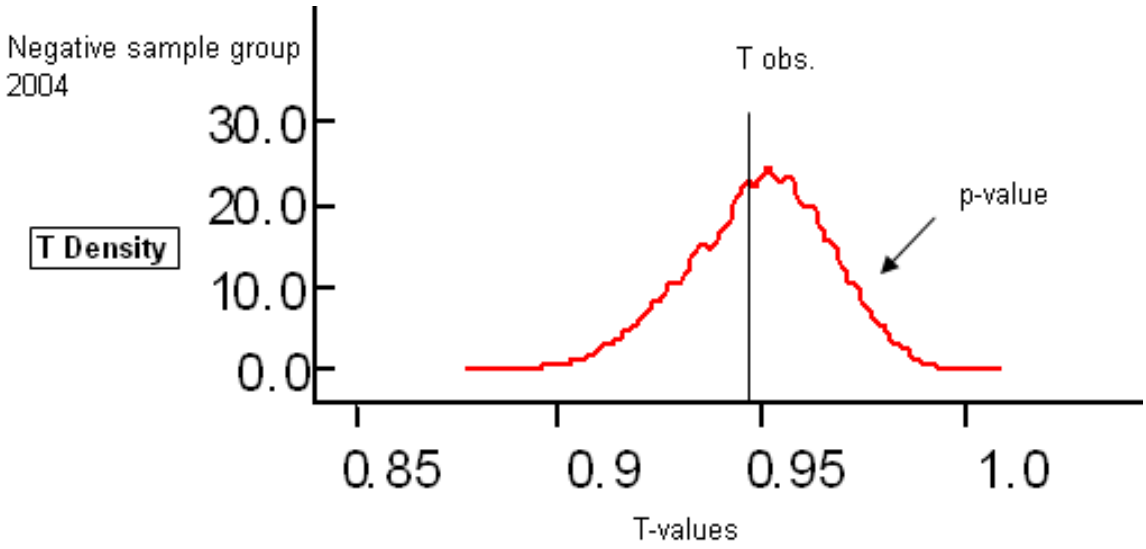


Figure 4.1: *Density function of the T test for the Negative sample in 2004 from the reduced QLBM.*

Table 4.8: *EV Bayesian 'p-values' from the reduced QLBM per year and sample group.*

Test	Bayesian P-value			
Sample group	Year 2002	Year 2003	Year 2004	Year 2005
$1x10^{-3}$	0.643	0.880	0.339	0.322
$1x10^{-4}$	-	0.505	0.818	0.171
$1x10^{-5}$	0.534	0.486	0.540	0.565
$1x10^{-6}$	0.551	0.545	0.550	0.415
$1x10^{-7}$	0.545	0.462	0.500	0.577
$1x10^{-8}$	0.509	0.482	0.512	0.656
Negative	0.479	0.684	0.528	0.323
Non-EV	-	0.740	0.388	0.388

No discrepancies were found for any of the sample groups and years. Therefore, using this approach it is concluded that the model fits the data adequately.

4.5.1.5 Model Comparison

Figure 4.2 shows the estimated posterior probabilities of detecting the virus correctly for two randomly selected samples, together with their 95% confidence intervals from the full model and the estimated probabilities from the reduced model. For illustration, a sample of dilution series 1×10^{-4} from 2004 and a non-EV sample from 2003 are selected. The plots are ordered by decreasing posterior probabilities. It is observed that the posterior predicted probabilities from the reduced model are within the 95% confidence limits obtained from the full model, indicating a good fit of the reduced model to the data for those sample groups.

An observed common feature for all sample groups is that the estimated probability of detecting the virus correctly decreases when the sample dilution series decreases (see Figure B.1). For the sample group dilution series 1×10^{-7}, it is observed that the probabilities of getting a correct result for samples from 2002 are lower than for samples from subsequent years. The negative and the strongest sample group dilution series 1×10^{-3} are the most likely samples to be detected correctly by participants across years.

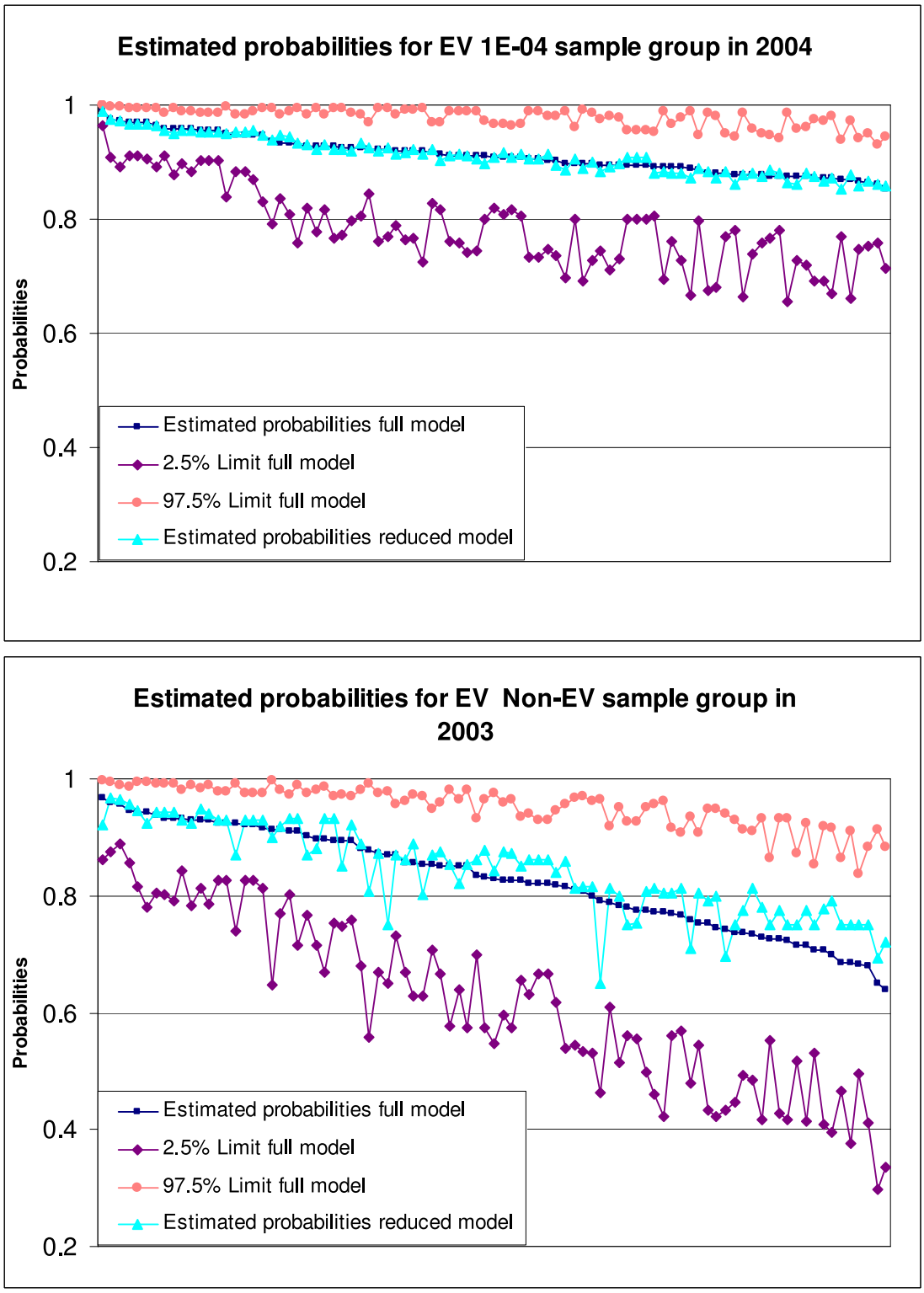


Figure 4.2: *EV estimated probabilities from the full and reduced QLBM for the sample group dilution series* 1×10^{-4} *in 2004 and for the Non-EV sample group in 2003. Results are ordered by decreasing estimated probabilities. The x-axis represents an identification number of participants' results.*

4.5.2 Modelling Hepatitis B Virus Qualitative Data

In order to check if the QLBM fits well the data from other pathogens, the model is applied to the Hepatitis B Virus data. Since the steps involved to reduce the full model are the same as in the previous subsection, the details of the results obtained by applying the full model are not shown. Instead, the results from the final reduced model along with the conclusions from model comparison and checking are presented.

Hepatitis B virus (HBV) proficiency panels consist of 8 samples per year with different subtype and viral load. Section 1.3.1 in Chapter 1 contains a summary of the HBV data. Table 4.10 shows the percentages of correct results per sample group and covariate level.

Table 4.9: *Changes on covariates for the analysis of HBV qualitative data.*

Covariate	Description	Values
Technology	Technology used to analyse the sample	Indicator variables for technology groups: CIH, RTC, RTIH, bDNA, HC and TMA compare with CC
OtherSp.	Experience on other specimen sample performance such as biopsies, swabs, etc..	Indicator variable with No experience performing test of other specimens as baseline
Serum	Annual number of serum tests performed by the participant	Indicator variables per group of number of plasma test: 0-10 baseline, 11-100- group 1, 101-1,000 - group 2, 1,001-2,000- group 3, 2,001-10,000- group 4, $> 10,000$- group 5

Technologies: CC=Conventional Commercial, RTC=Real Time Commercial, CIH=Conventional Commercial, RTIH=Real Time In-house, bDNA=bDNA, HC=Hybrid Capture and TMA=TMA.

The variables included in the model are described in Table 4.4 except for minor changes for some of the covariates shown in Table 4.9. Note that the baseline for the technology variable is CC technology in this analysis. The baseline technology method is changed to the most frequent technology used.

Table 4.10: *Percentages of correct results per sample group classified by covariate level for the HBV programmes. Sample viral load is given in* $\log_{10}$ *copies/ml.*

Covariates	Percentages of correct results						
	6	5	4	3.5	3	2.3	Negative
Year							
Year 2002	98.96	91.67	87.50	86.46	-	70.83	97.92
Year 2003	100.00	97.13	93.10	91.95	87.36	62.07	97.70
Year 2004	97.20	96.26	96.26	-	92.52	72.90	96.26
Year 2005	99.19	96.34	96.34	-	83.33	-	95.93
Subtype							
A	98.45	95.88	93.17	89.07	88.11	68.97	-
D	99.08	95.64	95.93	-	86.52	-	-
Technology group							
Tech CC	98.72	98.08	97.95	100.00	93.50	76.92	99.36
Tech CIH	97.40	96.75	94.38	89.71	83.72	63.08	97.40
Tech RTIH	100.00	98.77	97.14	98.21	86.79	78.95	95.06
Tech RTC	100.00	99.22	99.03	90.91	93.16	84.00	95.31
Tech bDNA	95.65	69.57	76.67	20.00	58.06	0.00	91.30
Tech HC	100.00	65.00	8.33	7.69	14.29	12.50	90.00
Tech TMA	100.00	100.00	100.00	100.00	0.00	100.00	100.00
Anticontamination							
Yes	99.10	97.08	95.96	93.40	94.00	74.83	98.20
No	98.87	95.22	91.85	85.54	80.14	62.50	95.50
Not answered	91.66	79.16	79.16	77.77	66.66	64.51	91.66
Accreditation							
Yes	99.41	98.25	96.50	92.76	88.46	73.17	97.67
No	100.00	95.63	91.27	87.83	87.03	66.66	98.65
Not answered	95.65	91.30	91.30	83.33	86.70	64.51	92.39
Other Specimens							
Yes	100.00	98.50	98.00	94.82	84.14	74.19	95.00
No	99.56	96.05	92.32	89.06	89.87	68.78	99.56
Not answered	95.29	91.76	91.76	82.69	86.36	63.63	91.76
Analysis method							
Analysis Singly	97.82	94.56	92.39	89.68	80.80	68.44	95.65
Analysis Duplicated	99.58	96.47	94.60	89.70	91.45	66.66	98.34
Analysis Other	100.00	97.91	95.83	84.61	88.88	100.00	91.66
Not answered	90.00	90.00	80.00	80.00	71.42	0.00	90.00
Plasma							
Group 0: 0-10	100.00	95.08	88.83	85.96	80.86	67.85	99.10
Group 1: 11-100	100.00	97.22	97.22	95.83	98.14	66.66	97.22
Group 2: 101-1,000	98.91	98.36	96.73	94.00	89.23	69.56	98.91
Group 3: 1,001-2,000	100.00	95.00	96.66	89.28	100.00	61.11	100.00
Group 4: 2,001-10,000	100.00	97.72	95.45	77.77	88.57	80.76	93.18
Group 5: > 10,000	100.00	100.00	100.00	100.00	85.71	75.00	100.00
Not answered	95.29	91.76	91.76	83.92	86.00	67.27	91.76
Serum							
Group 0: 0-10	100.00	97.43	92.94	92.00	89.43	69.35	97.43
Group 1: 11-100	97.87	97.87	97.87	95.45	97.22	73.33	100.00
Group 2: 101-1,000	100.00	96.26	92.52	89.65	88.00	75.29	98.13
Group 3: 1,001-2,000	100.00	96.34	97.56	92.10	90.74	67.74	100.00
Group 4: 2,001-10,000	100.00	95.74	91.48	62.50	80.95	52.00	95.74
Group 5: > 10,000	100.00	100.00	100.00	-	57.14	50.00	100.00
Not answered	95.35	91.86	91.86	82.69	86.14	65.45	91.86
Labtype							
Hospital	99.50	96.75	94.75	90.72	90.26	68.96	97.50
Public Health	100.00	98.64	97.29	97.50	88.00	82.75	100.00
Private	100.00	96.87	89.58	80.00	79.16	68.75	100.00
Reference	100.00	90.90	90.90	85.71	83.33	75.00	100.00
Manufacture	100.00	100.00	100.00	100.00	100.00	85.71	92.30
Research	100.00	100.00	93.33	100.00	80.00	50.00	100.00
Not answered	95.50	91.01	91.01	81.66	86.36	62.71	92.13
Inhibition Test							
Inhibition test No	97.82	94.56	92.39	89.68	80.80	62.50	95.65
Inhbition test Yes	99.58	96.47	94.60	89.70	91.45	73.68	98.34
Inhibiton test only Negative samples	100.00	97.91	95.83	84.61	88.88	80.00	91.66
Not answered	90.00	90.00	80.00	80.00	71.42	25.00	90.00

The association of the covariates in Table 4.9 to the correct detection of the virus is checked by applying the QLBM to the HBV data. The full QLBM is fitted to the HBV data and the parameter associated with each variable is tested at the two-sided 5% significance level (test based on the posterior distribution of the estimated parameters as in the previous EV data analysis). Covariates that are not significant, or do not interact with others and/or do not behave as confounders in the model are discarded. Thus, the full model fitted to HBV data is simplified to obtain a reduced model, as in the previous EV data analysis. Only the results and tables that summarise the reduced model fitted to HBV data are presented in this section (the summary of results obtained from the full model can be found in Appendix B, Table B.16).

4.5.2.1 Reduced HBV Model

As in the previous EV data analysis, the same selection procedure as described in Section 4.4 is used to select the reduced model. The reduced model includes those covariates, per sample group, that are significant at the two-sided 5% level. In addition, confounding covariates, interactions between them and correlations of the parameter estimates were checked and confounding covariates were retained in the model as well as covariates showing correlation between parameter estimates. In general, no significant interactions were found.

Tables B.17 to B.23 in Appendix B show the results obtained from the reduced model for each sample group. Table 4.11 summarises the results with the mean and standard deviation of the estimated parameters from the reduced HBV model. The results in bold shown in Table 4.11 are the significant parameters. Experience performing serum and plasma tests, laboratory type and performing inhibition test are not included in the reduced model for any sample group, therefore they are not shown in the table.

Table 4.11: *Mean and SD of the parameter estimates from the reduced QLBM for HBV sample groups.*

Covariates	Mean (SD)						
	6	5	4	3.5	3	2.3	Negative
Intercept	3.870 (0.325)	2.528 (0.289)	1.635 (0.389)	2.189 (0.348)	0.784 (0.353)	0.978 (0.201)	2.825 (0.293)
Year- baseline "2005"							
Year 2002							
Year 2003					0.597 (0.375)		
Year 2004					**1.011 (0.313)**		
Technology group- baseline "CC"							
Tech CIH-CC		**0.180 (0.479)**	0.246 (0.505)	-0.093 (0.468)	-0.022 (0.394)	-0.440 (0.312)	
Tech RTIH-CC		0.992 (0.558)	0.618 (0.537)	0.909 (0.642)	-0.182 (0.362)	0.352 (0.361)	
Tech RTC-CC		1.049 (0.616)	1.148 (0.619)	-0.050 (0.767)	**0.797 (0.413)**	0.614 (0.497)	
Tech bDNA-CC		**-2.512 (0.438)**	**-1.584 (0.534)**	**-3.158 (0.546)**	**-1.335 (0.447)**	**-2.669 (0.584)**	
Tech HC-CC		**-2.064 (0.523)**	**-3.383 (0.599)**	**-3.285 (0.527)**	**-1.679 (0.666)**	**-1.674 (0.632)**	
Tech TMA-CC		0.296 (0.892)	0.238 (0.911)	0.292 (0.903)	-	0.484 (0.864)	
Anti- baseline "No"			**0.922 (0.417)**	**0.863 (0.424)**	**1.235 (0.316)**		**1.036 (0.492)**
Accred- baseline "No"		**1.505 (0.454)**	**1.482 (0.450)**		0.573 (0.316)		
OthrSpc.- baseline "No"		**1.072 (0.518)**	**1.399 (0.534)**				
Analysis method- baseline Singly							
Analysis Duplicated						-0.102 (0.294)	
Analysis Other						**1.434 (0.693)**	

Results in bold are significant at the two-sided 5% level.

Sample Group 6 $\log_{10}$ Copies/ml Viral Load

No significant difference was found between participants' performance for samples of subtype A and subtype D. The results show that participants perform similarly, independent of the technology used, experience testing other specimens, type of laboratory, accreditation status and use of an anti-contamination system.

The estimated probability of a correct result is $(\exp(3.870))/(1+\exp(3.870)) = 0.98$. Thus, as a general conclusion, the 2% error for the correct detection of the virus is not significantly associated with any of the covariates included in the model.

Sample Group 5 $\log_{10}$ Copies/ml Viral Load

Participants using bDNA or HC technology are significantly less likely to detect the virus correctly than CC technology users (with odds of $1/\exp(-2.512) = 12.35$ and $1/\exp(-2.069) = 7.9$, respectively). Although not statistically significant, RTIH and RTC users have 2.5 and 2.8 times higher odds of detecting the virus correctly than CC users, respectively.

The odds for accredited participants is 4.5 times higher than the odds for non-accredited participants.

Participants with experience performing other sample tests, such as biopsies, have almost 3 times the odds of participants without experience.

Sample Group 4 $\log_{10}$ Copies/ml Viral Load

No significant differences were found between participants' performance for samples of subtype A and subtype D.

The results show that bDNA and HC technology users are significantly less likely to detect the virus correctly than CC technology users (with inverse of odds ratios of 5 and 30, respectively).

Participants using an anti-contamination system have 2.5 times higher odds of detecting the virus correctly than those not using an anti-contamination system.

Accredited participants or participants with experience performing other specimen tests are significantly more likely to detect the virus correctly than non-accredited participants or participants with lack of experience (with odds ratios of 4.4 and 4, respectively).

Sample Group 3.5 $\log_{10}$ Copies/ml Viral Load

The results of participants from 2003 are not significantly different from the results of participants from 2002.

The users of bDNA and HC technologies are significantly less likely to detect the virus correctly than CC users (the inverse of the odds ratios are more than 20).

The use of an anti-contamination system tends to improve participants' performance, the odds for participants using an anti-contamination system is 2.3 times higher than the odds for participants not using an anti-contamination system.

Sample Group 3 $\log_{10}$ Copies/ml Viral Load

No significant difference was found between participants' performance for samples of subtype A and subtype D.

The users of bDNA and HC technologies are significantly less likely to detect the virus correctly than CC technology users; their inverse odds ratios of detecting the virus correctly are 3.7 and 5.3, respectively. The odds for users of RTC technology is twice the odds for CC users.

The use of an anti-contamination system tends to improve participants' performance, participants using an anti-contamination system having almost 3.5 times higher odds than participants not using any anti-contamination system.

Accredited participants are more likely to detect the virus correctly than non-accredited participants, being the odds for accredited participants almost twice the odds for non-accredited participants.

Sample Group 2.3 $\log_{10}$ Copies/ml Viral Load

It was found that participants using bDNA and HC technologies have 14.4 and 5.3 times higher odds than CC technology users.

The odds for participants performing analysis methods in triplicate or more is 4.2 times the odds for participants performing single analysis methods.

Negative Sample Group

No significant differences were found between the performance of participants using different technologies. However, participants using an anti-contamination system have almost 3 times higher odds of a correct result than participants not using any anti-contamination system.

4.5.2.2 Model Comparison

As in the previous analysis of EV data, model comparison tools are applied in order to determinate if the reduced model fits the data as well as the full model. The approach taken to model comparison is as described in Section 4.4.

Figure 4.3 shows the graphs for two randomly selected samples, the results are ordered by decreasing estimated probability. It is observed that the majority of the posterior probabilities of detecting the virus correctly from the reduced model lie within the confidence intervals (which is expected since we are providing the 95% confidence limits).

However, for the weakest sample group of 2.3 $\log_{10}$ copies/ml, the estimated probabilities from the reduced model are underestimated with respect to the probabilities from the full model when those are above 0.75 (see Figure B.2). Nevertheless, these probabilities are still within the range of the confidence limits, so it is concluded that the reduced model fits the data appropriately.

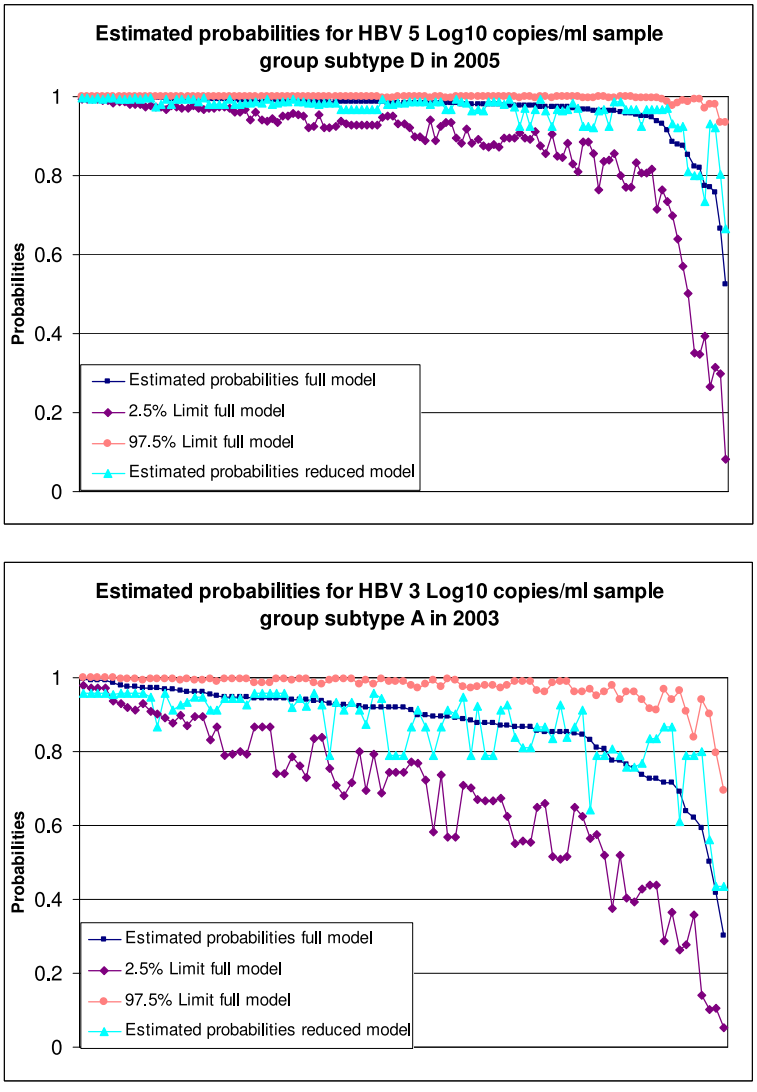


Figure 4.3: *HBV estimated probabilities from the full and reduced QLBM for the sample group 3* $\log_{10}$ *copies/ml in 2002 and for the sample group 5* $\log_{10}$ *copies/ml in 2005. Results are ordered by decreasing estimated probabilities. The x-axis represents an identification number of participants' results.*

Generally speaking, the estimated probability of detecting the virus correctly decreases when the viral load of the sample to be tested decreases. No differences are observed between the estimated probabilities for samples with different subtypes across years (see Figure B.2).

4.5.2.3 Model Checking

To test if the model is consistent with the data the posterior predictive results are assessed. The test quantity used to study model consistency is the same as the one defined in Section 4.4. Lack of fit of the data is assessed by the tail-area probability of the posterior predicted distribution.

Figure 4.4 shows the density of the test statistic T for the sample in group 3 $\log_{10}$ copies/ml and year 2005. It is observed that the area under the distribution of T for values higher than the observed test quantity is bigger than 0.025 and lower than 0.975, so it is concluded that the model fits the data well at the 5% significant level.

Table 4.12 shows the Bayesian 'p-value' that assesses the significance of discrepancies between the observed and the predicted data. An extreme Bayesian 'p-value' indicates disagreement between the observed data and the proposed model.

Observing the probabilities shown in Table 4.12, no discrepancies were found for any of the sample groups and years. Therefore, it is concluded that the model fits the data adequately.

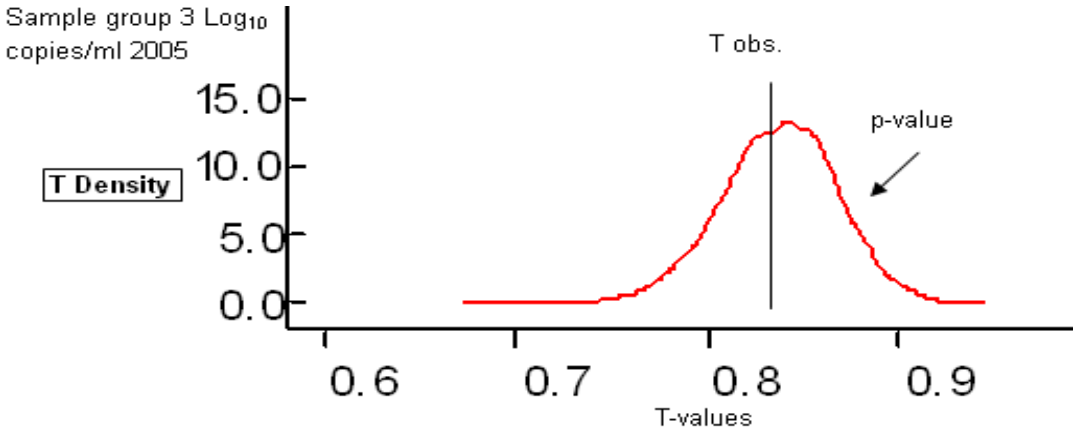


Figure 4.4: *Density function of the T test for the sample group 3* $\log_{10}$ *copies/ml in 2005 from the reduced QLBM.*

Table 4.12: *HBV Bayesian 'p-values' from the reduced QLBM per year and sample group.*

Test	Bayesian P-value			
Sample group	Year 2002	Year 2003	Year 2004	Year 2005
6	0.419	0.172	0.784	0.303
5	0.865	0.128	0.495	0.377
4	0.906	0.446	0.468	0.257
3.5	0.617	0.200	-	-
3	-	0.505	0.485	0.595
2.3	0.268	0.880	0.339	-
Negative	0.303	0.355	0.626	0.681

4.6 Model Prior Sensitivity Analysis

The proposed model, QLBM, estimates the parameters using Bayesian techniques which employ prior distributions to obtain posterior estimates. As described previously, the prior distribution is part of the model within the Bayesian framework, but the precise choice of the prior distribution is to some extend subjective.

Therefore, in order to study the robustness of the model subject to different priors, a sensitivity analysis is performed for a range of alternative prior distributions. This sensitivity analysis assesses the effect of alternative prior models on the posterior inferences.

To measure the effect and sensitivity of the chosen prior distributions on the estimated parameters and covariates, three QLB models are fitted with different prior distributions. In the previous application, the QLBM with normal prior distributions for the regression parameters and low informative priors for the covariates and hyperparameters was considered.

In Bayesian analysis, if the model is unaffected by external information (or it is believed so), then there is not a strong prior knowledge to be taken into account, so it is used what is called a non-informative prior distribution. This distribution has only a weak impact on the posterior distribution since the density is described as vague or flat.

From another perspective, prior distributions can be divided by proper (informative and non-informative) and improper distributions. Improper distributions are those that violate the assumption that probabilities sum up to 1. While, proper distributions do not violate any axiom of probability theory.

In this section, the results obtained from alternative models with different prior distributions are summarised. Since there is no informative knowledge about the parameters models with more informative priors than the proposed for the QLBM are not in this study. All chosen alternative priors are less informative than the QLBM, and the following choices are decreasing in the prior information considered (from more informative to less informative):

- Model 1 with Prior 1: 'Non-informative' or flat prior distributions, Beta(1,1), for the hyperparameters of the covariates to be estimated.

- Model 2 with Prior 2: A hierarchical model with hyperparameters from a flat uniform and inverse-gamma distribution for some of the primary parameters of interest θ.
- Model 3 with Prior 3: Full hierarchical model with non-informative distributions for the hyperparameters of the regression parameters to be estimated and flat prior distributions for the hyperparameters of the covariates to be estimated.

For the sensitivity analysis, the prior distributions are non-informative proper priors in order to make sure that the posterior distributions are proper posterior distributions.

A graphical summary of the posterior checks is presented as an illustration of the differences of the posterior probabilities obtained when applying several models. A Bayesian statistics test can be developed to formally test these differences. In this case, however, a graphical representation gives more insight into possible structural differences between the models than a formal test. In order not to interrupt the flow of the Bayesian spirit of this thesis at this point, some formal classical test procedures are shown in Appendix E for the reader who is more familiar with classical analysis.

4.6.1 Sensitivity Analysis of QLBM Applied to EV Data

The three alternative models using the priors described in the previous subsection, were applied to EV data. No differences were found between the results obtained from the alternative models and the results from the applied QLBM to the EV data, confirming the robustness of the QLBM approach.

As an illustrating example, Figure 4.5 shows the differences between the estimated posterior probabilities from the three models with respect to the estimated probabilities from the QLBM for the sample group 10^{-6} in 2004. The structure of the differences does not have any particular pattern. It is observed that the differences are close to zero for each participant.

There is not a systematic pattern in the differences, suggesting that all three new models agree with the initial QLBM.

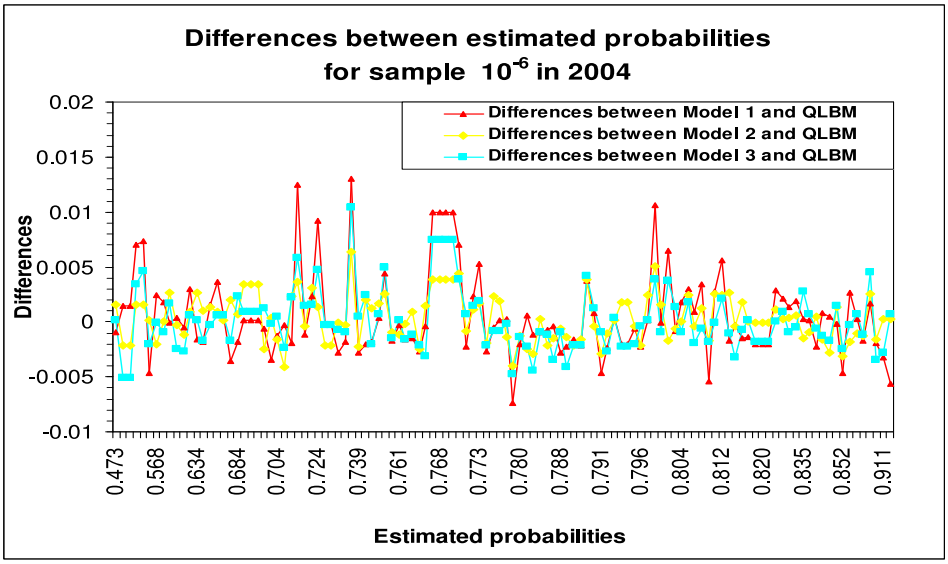


Figure 4.5: *Differences between EV estimated probabilities from the QLBM. The x-axis represents the estimated probabilities of correct detection of the sample.*

4.6.2 Sensitivity Analysis of QLBM Applied to HBV Data

As in the previous subsection, the robustness of the QLBM applied to HBV data is checked by a sensitivity analysis. No differences were found between the estimated posterior probabilities from the three alternative models and the estimated probabilities from the QLBM for any of the sample groups and years.

Figure 4.6 shows the differences between the estimated probabilities from the alternative models and the applied QLBM for the sample group 5 $\log_{10}$ copies/ml and year 2005.

As a general conclusion, based on the sensitivity analysis conducted, no differences were found between the means of the estimated posterior probabilities from the alternative models and the estimated probabilities from the QLBM.

This result is in agreement with the idea of Bayesian analysis, in the sense, that if a great amount of information from observations is available, the prior knowledge has little influence on posterior estimates.

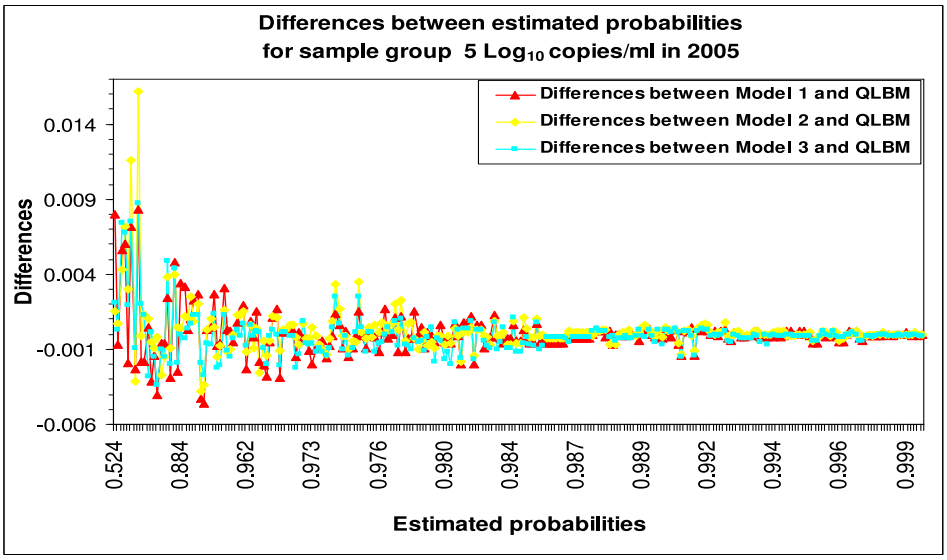


Figure 4.6: *Differences between HBV estimated probabilities from the QLBM. The x-axis represents the estimated probabilities of correct detection of the sample.*

4.7 Summary and Conclusions

The QLBM developed in this study represents a new statistical method that can be used for the identification of factors that are associated with participants' performance of individual samples across time. Unlike the classical approach, the Bayesian framework proposed here allows the inclusion of missing information from participating laboratories. Furthermore, problems related to asymptotic theory that occur in classical models are avoided. The QLBM developed and reported in García-Fernández et al. (2007) (see Appendix F) has been reviewed and adjusted to provide more adequate results and explain the difference on performance in a more appropriate way.

The full QLBM applied to EV and HBV provides a better overview of how the covariates influence the performance. Predictions of the model can be used to inform laboratories about the best procedures to adopt, such as technology to be used, use of an anti-contamination system, experience need to improve their performance, etc. In turn, the reduced model provides estimates of the real impact of the significant covariate on the results, and this can be used for making future predictions. In addition, it provides a measurement for the influence of the laboratory practices over its performance.

As a general overview of the results, it is concluded from the application of the QLBM to EV and HBV data that performance varies depending on the virus to be analysed and the sample load. The best technology used to test the samples varies among the viruses and sample viral loads. bDNA and HC users are less likely to detect the virus correctly than CC for HBV samples independently of the viral load. However, for EV performance, as the dilution series decreases, CC users are less likely to provide a correct result than CIH users. The use of an anti-contamination system tends to improve the correct detection of the virus. Differences on performance between types of labs were found when analysing EV samples, but not when the sample to analyse is of HBV. Private laboratories are less likely to detect the virus correctly than hospital laboratories as the EV sample dilution increases. On the other hand, manufacture and research laboratories are more likely to detect the virus correctly than hospital laboratories as the EV sample dilution decreases. Experience of testing different specimens, such as biopsies, tends to improve (overall) participants' performance.

Both model applications have been checked for goodness of fit and the results have shown that QLBM fits the data appropriately. This was achieved by generating replicate data simulated from the model and performing a test statistics to compare simulated and observed data. The probability distribution used for the data is a well known distribution chosen according to the characteristics and the definition of the data to be analysed. The sensitivity analysis has shown that the QLBM is robust under changes in the prior knowledge.

The proposed model fulfills the requirement of using complete datasets, having no restriction about the number of covariates to be studied, providing appropriate parameter estimates and identifying factors that are associated with participants' performance. Therefore, the QLBM is an appropriate model to fit the data and provide feedback to participants such as information of the best technology to be used when testing the samples depending on the pathogen to be analysed.

It also provides information for the design of future panels including which viral load is more likely to be detected correctly. Although in this thesis no continuous covariate has been used, the QLBM can be used with continuous covariates after some adjustment, as carried out in a previous analysis by García-Fernández et al. (2007). Furthermore, no further care needs to be taken about multiple parameter testing since each parameter has its own probability distribution, and hence no assumption needs to be imposed.

However, the QLBM has some considerations which need to be taken into account. The model assumes that the responses from participants are independent, however, some participants may return several results in the same year. Thus, they may be correlated, and this correlation has not been considered. Another consideration is that some participants are common across years, so results provided by them should be treated as repeated measures. Since the QLBM treats the observations as independent, new considerations about repeated measures and their possible correlations need to be taken into account. However, the independent working assumptions relax the model complexity and computational burden (in this case the model run approximately between 3 to 5 hours, depending on the number of covariates included). Furthermore, the model validation indicates the goodness of fit of the QLBM (for a model detailed discussion see Chapter 8).

Chapter 5

Modelling Quantitative Performance of Participants in QCMD Quality Control Programmes over Time

In the previous chapter the Qualitative Bayesian Model (QLBM) has been proposed to determine risk factors associated with qualitative responses from participants' performance of EQA programmes over time. The QLBM has been applied to two different datasets and the results obtained have been summarised. However, the quantitative responses returned by participants of EQA programmes have not been analysed, i.e. where participants estimate the microbial load within a positive sample. Therefore, a Quantitative Bayesian Model (QTBM) is developed to investigate which of the exploratory variables are related to the quantitative performance of participants of QCMD programmes over time. In this chapter the model is described and applied to two different data sets (HBV and HCV quantitative data described in Chapter 2), then results and conclusions are shown.

5.1 Introduction

The quantitative responses, described in Chapter 2, are analysed from a Bayesian perspective to identify factors significantly associated with participants' quantitative performance.

Similar to the analysis of qualitative data, Generalised Linear Models (GLM) can be applied to analyse quantitative data from EQA participants (as introduced in Chapter 1). A special feature of this type of quantitative data is that the response from a participant using molecular diagnostic assays may be a censored observation of a quantitative measure. Censored observations may arise because some participants are unable to estimate the microbial load when the value is outwith the limits of detection of the assay used. Therefore, there is a need to develop a model that can appropriately incorporate both censored and non-censored observations from participants.

A model that takes into account the censored and non-censored responses from participants' performance is proposed here by using a linear regression model in a Bayesian framework for the $\log_{10}$ transformed estimated microbial load. This model will be referred as 'Quantitative Bayesian Model' (QTBM). As the model is constructed within a Bayesian framework, it has the same benefits as the model for qualitative data presented in the previous chapter. The main advantage is that the large number of parameters and the missing information from covariates can be handled easily and appropriately. In addition, for quantitative data, both, censored and non-censored, observations can be included in the model. Therefore, the QTBM identifies significant factors associated with quantitative participants' performance without discarding valuable information.

The theory of linear regression models (Dobson, 1990) and Bayesian data analysis are well known (Gelman et al., 2004). This knowledge is used to derive a model to fit the peculiarities of these special data conditions. This model includes data that classical linear models would

discard due to the missing covariate information or the censored responses from the participants. The way of approaching these issues here is by assigning probability distributions to them.

Similar to the QLBM approach described in the previous Chapter, for the QTBM model estimates of parameters and missing observations will be provided. In addition, estimates of the censored responses are calculated without having to discard incomplete datasets.

The QTBM developed here is coded in WinBUGS (Project, 1996-2004). The code can be found on the CD attached to this thesis. In this chapter the QTBM will be used and applied to a subset of the large reservoir of QCMD data: the Hepatitis B virus (HBV) and Hepatitis C virus (HCV) QCMD programmes from 2002 to 2005.

5.2 Analysis of Variance

As part of model parameter specification it is important to determine in a first instance whether the response variances per technology group are significantly different. For this purpose, two classical statistical tests are carried out. First, a homogeneity test of variances of the two main technology groups (commercial and in-house) for the overall sample groups is performed and then for each sample group. Since this analysis is approached from a classical point of view, censored observations are discarded.

Levene's test is performed to test for homogeneity of variances (Dugard et al., 2010) among two technology clusters, commercial technologies (CC, bDNA, RTC, HC) and in-house technologies (CIH, RTIH), and the obtained p-value is 0.0115. Therefore, the hypothesis of equal variances between these two clusters is rejected at the two-sided 5% level.

A more detailed analyses is carried out to study the differences of variances among technology types for each sample group. Levene's test for homogeneity of variances is computed and the p-values per sample group are given in Table 5.1 below:

Table 5.1: *P-values from Levene's test. Observed variances are calculated after removing censored observations (Variances C=Commercial group, Variances I=In-house group).*

Sample group $\log_{10}$ copies/ml	6.0	5.0	4.0	3.5	3.0	2.3
P-value	0.032	0.001	0.813	0.348	0.518	0.024
Variances C	0.358	0.228	0.233	0.288	0.263	0.222
Variances I	0.882	0.972	0.817	0.604	0.550	0.415

Based on the p-values of the Levene's test variances among technology groups for samples 6.0, 5.0 and 2.3 are significantly different at the two-sided 5% level. These findings will be incorporated in the formulation of the QTBM model.

5.3 General linear model: basic notation and model formulation for the EQA quantitative data

Let Y_{si} denote the i^{th} quantitative response for the sample group s. Note that the quantitative response measures the viral load of a sample. It is assumed that $w_{si} = \log_{10}(y_{si})$ follows a normal distribution with mean μ_{si} and variance $\sigma^2_{s_i}$ where $s_i = 1, .., q$ with q being the total number of variances within the sample group s. Thus, s_i indicates that the participants' response i of sample group s belongs to the group 1, 2,...,q within the sample group s. For example, the technology used by a participant may be an in-house or commercial technology. The results from in-house assays are expected to be more variable, as this collection of observations are from individually developed assays performed by an individual laboratory. In turn, users of the same commercial molecular diagnostic kit use the same assay and are provided with detailed protocols to prevent different results by different users.

In the previous Section 5.2, it was found that the variances of the responses corresponding to commercial and in house technologies differ. Based on these results we represent the variances from these technology categories by estimating two different variances $s_i = 1, 2$.

In a GLM, the mean μ_{si} is assumed to vary depending on some covariates according to a linear function:

$$\mu_{si} = x_{si}\beta_s$$

where:

- i is the i^{th} observation with i=1,..,n_s.
- n_s is the total number of observations within the sample group s with s=1,..,l.
- l is the total number of sample groups.
- $\vec{x}_{si} = (x_{si1}, .., x_{sir})$ is the r-dimensional vector of covariates for the i^{th} observation of the sample group s with $x_{si1} = 1$.
- The covariate matrix $\mathbf{X}_s$ for each sample group is the matrix with r columns and n_s rows. Each column corresponds to the r^{th} covariate for participants in the sample group s. Thus each row is the r-dimensional vector of covariates $\vec{x}_{si} = (x_{si1}, .., x_{sir})$ for the i^{th} observation of the sample group s.
- β_s is the r-dimensional vector of regression coefficients, $\vec{\beta}_s = (\beta_{s1}, .., \beta_{sr})^T$.

For participants' responses that are censored an estimate from a normal distribution, truncated to the censored observed response, is calculated. For example, if y_{11} is the observed censored response 1 from the sample group 1 and it is right censored, then w_{11} is estimated to be a value greater than $\log_{10}(y_{11})$ from a normal distribution with mean μ_{si} and variance $\sigma^2_{s_i}$.

5.4 Problems Arising with Classical GLMs when Analysing Quantitative Responses

A classical linear model is applied to the $\log_{10}$ transformed microbial load estimates. The following example does not aim to provide a complete analysis, but instead its aim is to highlight some of the problems arising from using classical GLMs.

5.4.1 Results of the GLM Applied to the Strongest HBV Sample Group

Table 5.2 shows the results that are obtained when a classical linear model is applied to the strongest viral load sample group. This sample group is selected because of the high number of censored observations, therefore it is a good example for illustrating the type of problems arising in this context. The strongest sample viral load group has a target viral load of 6 $\log_{10}$ copies/ml. The classical linear model is applied to study the association of year, use of an anti-contamination system and accreditation status with the estimated sample viral load. The total number of returned datasets is 309, but 50 datasets from censored observations and a further 20 from missing covariate values, for accreditation and anti-contamination are discarded. Although the results appear appropriate for the subgroup of data, a lot of information is omitted due to the 70 datasets discarded. The amount of lost information is likely to increase as the number of covariates with missing values increases.

The example above shows that classical techniques fail to model the data appropriately because of missing covariates and censored observations. In addition, the use of asymptotic theory restricts the number of parameters included in the model. Although there are some classical techniques and regression approaches that deal with censored observations such as the Tobit regression model (Tobin, 1958), the theory behind it is still based on asymptotic

Table 5.2: *Results of the classical linear regression model applied to the strongest HBV sample group.*

Parameters	Estimate	SE	p-value
Intercept	5.385	0.173	0.000
Year 2002 (baseline 2005)	0.479	0.134	0.000
Year 2003 (baseline 2005)	0.325	0.136	0.018
Year 2004 (baseline 2005)	0.255	0.121	0.037
Tech. CIH (baseline CC)	-0.095	0.195	0.626
Tech. RTC (baseline CC)	0.419	0.148	0.005
Tech. bDNA (baseline CC)	0.619	0.202	0.002
Tech. RTIH (baseline CC)	0.309	0.127	0.015
Tech. HC (baseline CC)	0.434	0.319	0.174
Anti-contamination Yes (baseline No)	-0.002	0.117	0.984
Accreditation Yes (baseline No)	-0.077	0.100	0.442

assumptions. Therefore, when the covariates are missing, particular care needs to be taken when including parameters in the model.

A similar analysis was carried out for all the sample groups and similar conclusions related to the amount of missing information and the restriction that occurred because of asymptotic theory were found. As a result of this analysis, it can be concluded that classical methods to analyse the quantitative data are inappropriate and there is a need to develop models based on techniques that can handle this complex data situation more efficiently.

5.5 Proposed Model for the Quantitative Responses based on Bayesian Methods

The description of the QTBM and the analytical derivation of the conditional posterior distributions from a Bayesian perspective are presented in this section. A general derivation of the conditional posterior distributions similar to those presented in this chapter can be found in advanced statistics text books (Gilks et al., 1996; Gelman, Carlin, Stern and Ru-

bin, 2004; Banerjee et al., 2004). The QTBM is fitted to the data, and estimates of the means and variances of viral loads are provided under the assumption that the log_{10} of the viral load returned by a participant is from a normal distribution with those parameters. The estimated mean represents the expected mean of sample viral loads that can be obtained by participants under specific laboratory practices. In a Bayesian framework a prior distribution for the mean and one for the variance of the normal distribution need to be specified. For the reasons explained in Chapter 1 and 4, the QTBM is developed as a hierarchical model with prior and hyper-prior distributions for the parameters and missing covariates to be estimated. Therefore, inferences for the means and variances can be derived from the conditional posterior distributions of the parameters and missing covariates.

Since a different combination of likelihood and priors is used, the particular equations necessary for this application are derived and shown here.

5.5.1 Likelihood Function

As stated in Section 5.4, $w_{si} = \log_{10}(y_{si})$ is assumed to follow a normal distribution with mean μ_{si} and variance $\sigma^2_{s_i}$. Then, its probability density function is given by

$$f(w_{si}|\mu_{si}, \sigma^2_{s_i}) = (2\pi(\sigma^2_{s_i}))^{-1/2} \exp\left\{-\frac{(w_{si} - \mu_{si})^2}{2\sigma^2_{s_i}}\right\}.$$

Assuming independence of w_{si}, the likelihood of μ_{si} and $\sigma^2_{s_i}$ can be written as

$$L(\mu, \vec{\sigma}^2|w) = \prod_{si} f(w_{si}|\mu_{si}, \sigma^2_{s_i}) = \prod_{si} (2\pi(\sigma^2_{s_i}))^{-1/2} \exp\left\{-\frac{(w_{si} - \mu_{si})^2}{2\sigma^2_{s_i}}\right\},$$

where multiplication $\prod_{si} = \prod_{s=1}^{l} \prod_{i=1}^{n_s}$ is the product over all observations of l groups ($s = 1, .., l; i = 1, .., n_s$).

5.5.2 Regression Model

In order to fit the linear regression model, $\mu_{si} = \vec{x}_{si}\vec{\beta}_s$, the likelihood is rewritten as a function of the parameters $\vec{\beta}_s$ and the covariate matrix x as follows

$$L(\mu, \vec{\sigma}^2|w) = \prod_{si} f(w_{si}|\vec{x}_{si}\vec{\beta}_s, \sigma^2_{s_i}) = \prod_{si}(2\pi(\sigma^2_{s_i}))^{-1/2} \exp\left\{-\frac{(w_{si} - \vec{x}_{si}\vec{\beta}_s)^2}{2\sigma^2_{s_i}}\right\}.$$

$$L(\mu, \vec{\sigma}^2|w) = L(\beta, x, \vec{\sigma}^2|w).$$

5.5.3 Bayesian Framework

The prior distributions for each $\vec{\beta}_s$ and each missing covariate x_{sij} are considered the same as for the qualitative analysis described in Chapter 4. Additionally to these prior distributions there is a need to introduce the prior distribution for each of the variances $\sigma^2_{s_i}$. It is assumed that $\sigma^2_{s_i}$ follows an inverse-gamma distribution with parameters c_{s_i} and d_{s_i}, which is a common positive defined distribution used to estimate variances (Gelman, Carlin, Stern and Rubin, 2004).

Given that the likelihood for the quantitative responses differs from the likelihood used for the qualitative responses the descriptions of the posterior distributions also differ and are presented in what follows.

5.5.3.1 Posterior Distributions

The posterior distribution for the parameters to estimate is obtained from the likelihood and the prior distributions as follows:

$$\pi(\beta, x, \vec{\sigma}^2|z) \propto likelihood \times prior \propto L(\beta, x, \vec{\sigma}^2)\pi(\beta|\beta_0, V_0)\pi(x|b, g[])\pi(\vec{\sigma}^2|\vec{c}, \vec{d}).$$

Since β, x and $\vec{\sigma}^2$ are assumed to be independent, the prior density functions for $\vec{\beta}_s$, $\vec{x}_{si}$ and $\sigma^2_{s_i}$ can be written as:

$$\pi(\beta|\vec{\beta}_0, \mathbf{V}_0) = \prod_s \pi(\beta_s|\vec{\beta}_{s0}, \mathbf{V}_{s0}) \sim \prod_s N_p(\vec{\beta}_{s0}, \mathbf{V}_{s0})$$

and

$$\pi(x|b, g) = \prod_{sij} \pi(x_{sij}|b_j, \vec{g}_j[]) \sim \prod_{sij} Bernoulli(b_j)^{(I_j)} Categorical(\vec{g}_j[])^{(1-I_j)},$$

where I_j is an indicator variable for the missing covariates and the product $\prod_{sij} = \prod_{s=1}^{l} \prod_{i=1}^{n_s} \prod_{j=1}^{r}$ is over all covariates j, for all participants' responses i and for all sample group s. $\vec{g}_j[]$ is the vector of assigned probabilities to each category of the covariate as defined in Chapter 4.

Using the transformation $\vec{\tau}^2 = 1/\vec{\sigma}^2$, the prior can be re-written as:

$$\pi(\vec{\tau}^2|\vec{c}, \vec{d}) = \prod_{s_i} \pi(\tau^2_{s_i}|c_{s_i}, d_{s_i}) \sim \prod_{s_i} Gamma(c_{s_i}, d_{s_i}).$$

Then, the posterior distribution of each parameter can be expressed in terms of its posterior conditional distribution, which will be used in the estimation procedure. The conditional posterior distributions for $\vec{\beta}_s$, the missing covariates x_{sij} and $\sigma^2_{s_i}$ are shown below:

- Conditional posterior distribution for $\vec{\beta}_s$ with $s = 1, .., l$

$$\pi(\vec{\beta}_s|w, x, \vec{\beta}_{-s}, \vec{\sigma}^2) \propto L(\beta_s, x, \vec{\sigma}^2)\pi(\vec{\beta}_s|\vec{\beta}_{s0}, \mathbf{V}_{s0})$$

$$\pi(\vec{\beta}_s|w, \vec{x}_{s.}, \vec{\beta}_{-s}, \vec{\sigma}^2_{s.}) \propto \prod_{i_{(s)}} \exp\left\{-\frac{(w_{si} - \vec{x}_{si}\vec{\beta}_s)^2}{2\sigma^2_{s_i}}\right\}$$

$$\times \exp\left\{-\frac{1}{2}(\vec{\beta}_s - \vec{\beta}_{s0})^T \mathbf{V}^{-1}_{s0}(\vec{\beta}_s - \vec{\beta}_{s0})\right\},$$

where $i_{(s)}$ is the subgroup of responses from participants of sample group s, and $\vec{\beta}_{-s}$ are the vectors of regression parameters for each sample group except for sample group s (that is, except $\vec{\beta}_s$). The group of covariates from all participants in sample group s is denoted by $x_{s.}$. The variances $\sigma^2_{s.}$ are the group of variances within sample group s.

- Conditional posterior distribution for x_{sij} where j can take on values from *2,..,r*

  - For missing binary variables

$$\pi(x_{sij}|w_{si}, x_{-sij}, \beta_{sj}, \sigma^2_{s_i}) \propto L(\beta_{sj}, x_{sij}, \sigma^2_{s_i})\pi(x_{sij}|b_j)$$

$$\pi(x_{sij}|w_{si}, x_{-sij}, \beta_{sj}, \sigma^2_{s_i}) \propto \exp\left\{-\frac{(w_{si} - x_{sij}\beta_{sj})^2}{2\sigma^2_{s_i}}\right\}$$
$$\times (b_j)^{x_{sij}}(1 - b_j)^{(1-x_{sij})}.$$

  - For missing categorical variables

$$\pi(x_{sij}|w_{si}, x_{-sij}, \beta_{sj}, \sigma^2_{s_i}) \propto L(\beta_{sj}, x_{sij}, \sigma^2_{s_i})\pi(x_{sij}|g_j[x_{sij}])$$

$$\pi(x_{sij}|w_{si}, x_{-sij}, \beta_{sj}, \sigma^2_{s_i}) \propto \exp\left\{-\frac{(w_{si} - x_{sij}\beta_{sj})^2}{2\sigma^2_{s_i}}\right\} \times g_j[x_{sij}],$$

    where x_{-sij} are all the covariate values except the observed j^{th} covariate value for the i^{th} observation in sample group s.

- Conditional posterior distribution for $\sigma^2_{s_i}$ with $s = 1, .., l$ and $s_i = 1, .., q$

$$\pi(\sigma^2_{s_i}|w, \vec{x}_{si}, \vec{\beta}_s, \sigma^2_{-s_i}) \propto L(\vec{\beta}_s, \vec{x}_{si}, \vec{\sigma}^2)\pi(\sigma^2_{s_i}|c_{s_i}, d_{s_i}).$$

  Using the transformation $\tau^2_{s_i} = 1/\sigma^2_{s_i}$

$$\pi(\tau^2_{s_i}|w,\vec{x}_{si},\vec{\beta}_s,\tau^2_{-s_i}) \propto \prod_{i_{(\sigma^2_{s_i})}} (\tau^2_{s_i})^{1/2} \exp\left\{-\frac{\tau^2_{s_i}(w_{si}-\vec{x}_{si}\vec{\beta}_s)^2}{2}\right\}$$

$$\times(\tau^2_{s_i})^{c_{s_i}-1} \exp\left\{-d_{s_i}\tau^2_{s_i}\right\}$$

$$\pi(\tau^2_{s_i}|w,\vec{x}_{si},\vec{\beta}_s,\tau^2_{-s_i}) \propto \prod_{i_{(\sigma^2_{s_i})}} \exp\left\{-\tau^2_{s_i}\left(\frac{(w_{si}-\vec{x}_{si}\vec{\beta}_s)^2}{2}+d_{s_i}\right)\right\} \times (\tau^2_{s_i})^{c_{s_i}-1/2},$$

where $i_{(\sigma^2_{s_i})}$ is the group of observations with the same variance $\sigma^2_{s_i}$.

A special feature of the quantitative data considered in this thesis is censored responses. Since participants' responses are assumed to follow a normal distribution on a $\log_{10}$ scale, the censored response i in sample s is estimated as follows:

- If the participant's response is left censored, the censored response is estimated by a random number w^c_{si} from a normal distribution, such that $w^c_{si} \leq w_{si}$, i.e.:

$$w^c_{si} \sim N(\vec{x}_{si}\vec{\beta}_s, \sigma^2_{s_i})I(0, z_{si}),$$

  where $I(0, w_{si})$ is an indicator function that accepts the generated value from the normal distribution w^c_{si} if $0 < w^c_{si} \leq w_{si}$.

- If the participant's response is right censored, the censored response is estimated by a random number w^c_{si} from a normal distribution, such that $w^c_{si} \geq w_{si}$, i.e.:

$$w^c_{si} \sim N(\vec{x}_{si}\vec{\beta}_s, \sigma^2_{s_i})I(w_{si}, \infty),$$

  where $I(w_{si}, \infty)$ is an indicator function that accepts the generated value from the normal distribution w^c_{si} if $w^c_{si} \geq w_{si}$.

5.5.4 Model Selection Procedure

As described in the previous chapter, for model selection, a backwards selection procedure based on the conditional posterior distribution for the estimated parameters is applied. The 95% highest density intervals for the means of the conditional posterior distribution are obtained and used to determine the model selection procedure. If the confidence interval for the parameter contains the value 0, the corresponding covariate is removed from the model. Additional to the model selection procedure, possible confounders, interactions and correlated parameters are studied using the same approach as in Chapter 4.

5.5.5 Model Checking

As in Chapter 4, it is checked that the model is consistent with the data. If the model fits the data well, then replicated data under model conditions should look similar to the observed data. Using the technique of 'test quantity' described in Chapter 1, the measure of discrepancy between the replicated data, y_{si}^{rep} from the model and the observed data, y_{si} is given by the standardised residual for each observation, which is defined as:

$$T(y_{si}, \mu_{si}, \sigma_{s_i}) = \frac{y_{si} - \mu_{si}}{\sigma_{s_i}}. \tag{5.1}$$

Under the model conditions, replicated data, y^{rep}, and standardised residuals for each of the replicated observations, $T(y_{si}^{rep}, \mu_{si}, \sigma_{s_i})$, are obtained. Thereafter, for each sample group and participant's result, the standardised residual based on the replicated observation is compared with the standardised residual based on the observed response in each simulation step (10,000 data points are simulated). Then, for each sample group s and participant's result i, an indicator is obtained when the standardised residual on the replicated value exceeds the observed standardised residual. That indicator takes on the value 0 if the replicated standardised residual does not exceed the observed standardised residual for each sample s and participant's result i, otherwise the indicator takes on the value 1.

Note that censored observations returned by participants are replaced by their estimates at each simulation step, which is a sensible procedure because there is an interest in checking if replicated data under the model conditions μ and $\vec{\sigma}^2$ are appropriate. Then, in each simulation, the mean of the indicators across participants' results per year and for each sample group s is obtained.

In each step of the simulation the mean of the indicator values across participants' results per year and for each sample group s represents the proportion of participants' results with observed standardised residuals exceeding the replicated standardised residuals. The distribution of these proportions is equivalent to the distribution of the probabilities that the observed standardised residuals for sample group s within year k are equal or exceeded by the predicted standardised residual for the same year and sample group (Bayesian 'p-values' that contrast the observed residuals versus the replicated residuals). Lack of fit to the data occurs when the distribution of these probabilities are lower than 0.025 or higher than 0.975.

The statistical test and the procedure to obtain its predictive posterior distribution are described in what follows:

- Let $T(y_{si}, \mu_s, \sigma_{s_i})$ be a summary measure of the observed data y_{si} and the parameters μ_{si} and σ_{s_i}, for the sample group s and the i^{th}-observation.
- $T(y_{si}, \mu_s, \sigma_{s_i}) = \frac{y_{si}-\mu_{si}}{\sigma_{s_i}}$ is defined as the standardised residual within sample group s and participant's result i.
- For each simulation from the posterior distribution of μ_{si} and σ_{s_i}, one replicated response is obtained from the predicted distribution of the i^{th}-result for the sample group s, y_{si}^{rep}.
- The observed standardised residual, $T(y_{si}, \mu_s, \sigma_{s_i})$, is compared with the predicted standardised residual, $T(y_{si}^{rep}, \mu_s, \sigma_{s_i})$.

- For each simulation, the proportion of the predictive standardised residuals that are equal or exceed the observed standardised residuals is obtained for the group of observations of year k and the sample group s. That is the proportion of observations within year k such as $T(y_{si}^{rep}, \mu_s, \sigma_{s_i}) \geq T(y_{si}, \mu_s, \sigma_{s_i})$.

- Since this proportion is calculated in each simulation step, a sample from the distribution of proportions is obtained.

5.5.6 Model Comparison

As in chapter 4 a model comparison tool based on the posterior distribution is implemented. In particular, a graphical tool to compare the full and reduced model, based on the estimated posterior means of the sample viral loads from both models, is proposed. The estimated means, $\hat{\mu}$, from the reduced model are compared to the estimated means from the full model. The estimated posterior means from the full model are plotted together with their 95% confidence intervals. Then, the estimated posterior means from the reduced model are added to the plot. If the reduced model fits the data well, it is expected that the posterior means estimated by the reduced model lie within the 95% confidence intervals of the posterior means obtained from the full model.

To allow an easier comparison of the means from different sample groups, all means are transformed to a standardised scale, in order to obtain values that lie within the interval (0,1). The function used for standardisation is

$$\frac{\mu_{si} - \min \mu_s}{\max \mu_s - \min \mu_s},$$

where μ_{si} is the estimated mean for the distribution of the participant's result i and sample group s, and μ_s is the group of estimated means for the results of all participants for sample group s.

5.6 Model Application

The results presented in this section are obtained with the same specifications for the hyper-parametes as for the QLBM in Section 4.5. Variations of the QTBM with different choices of prior distributions are studied, and results are summarised in Section 5.7 of this chapter.

When reporting the results from the model in further subsections, the estimated mean refers to the expected mean parameter for the distribution of participants' responses (observed viral loads). This expectation is calculated from the posterior distribution of the means obtained using the MCMC method (see Chapter 1 and 6 for more detailed about MCMC methods).

5.6.1 Modelling HBV Quantitative Data

A description of the HBV panels from 2002 to 2005 can be found in Chapter 2. Table 5.3 shows the mean of participants' responses for each covariate level (excluding censored observations, see Table 2.13 for the percentage of censored data per sample group). The variables included in the model are described in Table 4.9 with the exception of technology group. For the quantitative data the technology groups are: CIH, RTIH, RTC, bDNA and HC compared with the CC technology group.

The full QTBM is fitted to the data, and then a reduced model is found using a backward selection procedure, as described in Chapter 4. The results from the full and reduced models are presented in the next subsection.

Table 5.3: *Mean of participants' results per covariate level for HBV programmes.*

Covariates	Mean of paticipants' results					
	6	5	4	3.5	3	2
Year						
Year 2002	6.032	5.141	4.213	4.064	-	2.799
Year 2003	5.929	4.945	4.044	3.463	3.015	2.531
Year 2004	5.964	5.051	4.035	-	3.384	2.681
Year 2005	5.532	4.712	3.832	-	2.891	-
Subtype						
A	5.943	5.004	4.021	3.860	3.084	2.694
D	5.734	4.870	3.763	-	3.141	-
Technology group						
Tech CC	5.742	4.971	4.012	3.722	3.094	2.439
Tech CIH	5.697	4.635	3.622	3.410	2.982	3.081
Tech RTIH	5.964	4.932	4.057	4.235	3.256	3.054
Tech RTC	5.682	4.815	3.768	3.669	3.019	2.657
Tech bDNA	6.153	5.296	4.230	5.542	3.454	-
Tech HC	6.137	5.328	3.928	3.744	5.342	5.327
Anticontamination						
Yes	5.821	4.945	4.002	3.873	3.120	2.643
No	5.814	4.913	3.850	3.854	3.062	2.780
Not answered	6.003	4.974	4.084	3.769	3.574	2.623
Accreditation						
Yes	5.830	4.963	3.973	3.701	3.066	2.601
No	5.901	4.921	4.012	3.980	3.220	2.764
Not answered	5.680	4.890	3.834	4.042	3.021	2.482
Other Specimens						
Yes	5.834	4.938	3.982	3.996	3.134	2.664
No	5.881	4.950	3.990	3.804	3.132	2.761
Not answered	5.657	4.881	3.824	4.010	3.026	2.483
Analysis method						
Analysis Singly	5.847	4.957	3.981	3.812	3.081	2.635
Analysis Duplicated	5.819	4.879	3.890	3.890	3.184	2.775
Analysis Other	5.613	5.013	3.964	4.212	3.168	2.723
Not answered	5.514	4.758	3.773	-	2.859	-
Plasma						
Group 0: 0-10	5.868	4.943	3.982	3.852	3.290	2.972
Group 1: 11-100	5.634	4.753	3.764	3.803	2.903	2.614
Group 2: 101-1,000	5.787	4.871	3.947	3.721	2.991	2.581
Group 3: 1,001-2,000	6.042	5.094	4.169	3.956	3.214	2.687
Group 4: 2,001-10,000	5.966	5.076	4.051	4.274	3.915	2.658
Group 5: > 10,000	6.019	5.054	4.070	4.552	3.714	2.543
Not answered	5.663	4.880	3.823	4.041	3.014	2.502
Serum						
Group 0: 0-10	5.819	4.975	4.013	3.853	3.104	2.637
Group 1: 11-100	5.923	4.918	3.976	3.534	3.076	2.916
Group 2: 101-1,000	5.865	4.924	4.001	3.957	3.123	2.720
Group 3: 1,001-2,000	5.713	4.723	3.763	3.639	3.010	2.763
Group 4: 2,001-10,000	5.952	5.090	4.074	4.140	3.252	2.813
Group 5: > 10,000	6.179	5.243	4.231	-	3.444	-
Not answered	5.644	4.879	3.816	4.003	3.018	2.493
Labtype						
Hospital	5.831	4.896	3.950	3.820	3.144	2.695
Public Health	6.046	5.175	4.162	3.987	3.360	2.781
Private	5.863	4.923	3.873	4.093	2.971	2.792
Reference	5.490	4.819	3.914	3.416	3.002	2.653
Manufacture	6.161	5.213	4.413	3.813	3.431	2.291
Research	5.782	4.934	3.976	3.469	2.901	3.790
Not answered	5.668	4.881	3.818	3.956	3.021	2.481
Inhibition Test						
Inhibition test No	5.820	4.926	3.904	3.959	3.182	2.786
Inhbition test Yes	5.812	4.920	3.951	3.841	3.071	2.625
Inhibiton test only Negative samples	6.024	5.103	4.203	3.833	3.314	3.069
Not answered	5.816	4.972	3.976	3.680	2.894	2.471

5.6.1.1 Full HBV Model

The full model fitted to the HBV quantitative data includes all possible covariates. The results obtained from the full model for each sample group can be found in Tables C.1 to C.6 in Appendix C. A summary of the results obtained can be found in Table 5.4. The results in bold are significant at the two-sided 5% level.

Sample Group 6 $\log_{10}$ Copies/ml Viral Load

Table 5.4 indicates that the covariates time, subtype, technology, use of an anti-contamination system, experience performing serum tests and use of inhibition tests are significant factors associated with participants' performance.

The estimated mean for the results of participants under baseline conditions is 3.743 $\log_{10}$ copies/ml. In particular, for samples from 2002 to 2004 participants tend to provide higher estimates of viral load than for samples from 2005.

The estimates of viral load for samples of subtype D are more likely to be higher than the estimates for samples of subtype A.

Participants using RTIH and bDNA technologies tend to provide significantly higher estimates of viral load than participants using CC technology.

Participants using an anti-contamination system are more likely to return higher estimates of viral load than participants not using an anti-contamination system.

Table 5.4: *Mean and SD of the parameter estimates from the full QTBM for HBV sample groups.*

Covariates	Mean (SD)					
	6	5	4	3.5	3	2.3
Intercept	3.743 (0.504)	4.554 (0.100)	3.509 (0.125)	3.139 (0.129)	2.769 (0.117)	2.491 (0.148)
Year- baseline "2005"						
Year 2002	**0.590 (0.124)**	**0.314 (0.063)**	**0.181 (0.085)**	**0.614 (0.058)**	-0.015 (0.078)	0.151 (0.082)
Year 2003	**1.664 (0.477)**	**0.155 (0.063)**	0.006 (0.082)		**0.544 (0.051)**	**-0.436 (0.009)**
Year 2004	**1.648 (0.478)**	**0.337 (0.052)**	**0.130 (0.065)**			
Subtype- baseline "A"	**1.289 (0.482)**	-0.045 (0.037)	**-0.138 (0.059)**		0.053 (0.044)	
Technology group- baseline "CC"						
Tech CIH-CC	0.235 (0.215)	-0.178 (0.160)	-0.151 (0.199)	-0.244 (0.186)	-0.030 (0.168)	**0.650 (0.210)**
Tech RTIH-CC	**0.308 (0.147)**	-0.034 (0.093)	0.116 (0.114)	**0.402 (0.115)**	0.125 (0.0100)	**0.679 (0.127)**
Tech RTC-CC	0.104 (0.121)	**-0.120 (0.060)**	**-0.172 (0.075)**	0.288 (0.246)	-0.076 (0.067)	**0.446 (0.111)**
Tech bDNA-CC	**0.518 (0.146)**	**0.353 (0.090)**	**0.246 (0.116)**	0.546 (0.423)	-0.065 (0.116)	-0.244 (0.499)
Tech HC-CC	0.438 (0.227)	**0.372 (0.130)**	0.092 (0.350)	0.444 (0.304)	**1.593 (0.263)**	**2.363 (0.255)**
Anti- baseline "No"	**0.202 (0.095)**	**0.158 (0.049)**	**0.180 (0.060)**	0.025 (0.068)	0.067 (0.057)	0.030 (0.078)
Accred- baseline "No"	0.013 (0.100)	-0.022 (0.052)	-0.016 (0.066)	-0.096 (0.053)	-0.037 (0.062)	-0.069 (0.07)
OthrSpc.- baseline "No"	-0.085 (0.120)	-0.053 (0.075)	-0.119 (0.095)	**0.275 (0.079)**	**-0.180 (0.068)**	-0.055 (0.089)
Analysis method- baseline Singly						
Analysis Duplicated	-0.111 (0.095)	-0.063 (0.052)	-0.111 (0.064)	0.013 (0.071)	-0.035 (0.063)	-0.034 (0.079)
Analysis Other	-0.110 (0.189)	-0.013 (0.108)	-0.020 (0.131)	0.231 (0.125)	-0.061 (0.143)	0.087 (0.133)
Plasma- baseline 0-10						
Group 1: 11-100	-0.192 (0.196)	-0.054 (0.114)	0.120 (0.148)	0.128 (0.104)	-0.108 (0.114)	-0.109 (0.122)
Group 2: 101-1,000	-0.050 (0.122)	0.039 (0.070)	0.179 (0.090)	-0.059 (0.077)	0.037 (0.079)	-0.083 (0.090)
Group 3: 1,001-2,000	0.051 (0.174)	0.138 (0.094)	0.196 (0.131)	0.060 (0.114)	0.063 (0.120)	-0.119 (0.146)
Group 4: 2,001-10,000	0.237 (0.149)	0.116 (0.088)	0.230 (0.111)	0.079 (0.156)	**0.220 (0.086)**	0.035 (0.115)
Group 5: > 10,000	0.150 (0.235)	-0.041 (0.200)	0.145 (0.192)	0.966 (0.750)	0.092 (0.131)	**-0.481 (0.217)**
Serum- baseline 0-10						
Group 1: 11-100	0.160 (0.168)	0.054 (0.088)	0.081 (0.124)	-0.126 (0.104)	0.130 (0.101)	0.135 (0.131)
Group 2: 101-1,000	0.211 (0.127)	-0.028 (0.065)	0.166 (0.091)	0.019 (0.074)	0.065 (0.081)	-0.056 (0.085)
Group 3: 1,001-2,000	0.042 (0.193)	-0.122 (0.107)	0.101 (0.139)	-0.150 (0.101)	0.053 (0.112)	0.059 (0.110)
Group 4: 2,001-10,000	**0.323 (0.153)**	**0.264 (0.080)**	**0.418 (0.106)**	-0.101 (0.305)	**0.278 (0.088)**	-0.155 (0.129)
Group 5: > 10,000	**0.606 (0.282)**	**0.487 (0.156)**	**0.749 (0.198)**	1.147 (1.214)	**0.590 (0.178)**	-0.436 (0.570)
Labtype- baseline Hospital						
Public Health	0.068 (0.155)	0.079 (0.076)	0.042 (0.105)	0.125 (0.080)	-0.031 (0.094)	-0.092 (0.102)
Private	-0.113 (0.141)	-0.083 (0.081)	-0.216 (0.117)	0.091 (0.150)	**-0.265 (0.087)**	-0.075 (0.128)
Reference	-0.422 (0.384)	-0.027 (0.121)	-0.039 (0.156)	0.014 (0121)	0.014 (0.163)	-0.056 (0.203)
Manufacture	0.259 (0.227)	0.125 (0.110)	**0.327 (0.143)**	0.075 (0.106)	**0.363 (0.180)**	-0.051 (0.153)
Research	-0.193 (0.212)	**-0.424 (0.109)**	-0.207 (0.142)	-0.095 (0.238)	**-0.462 (0.099)**	0.030 (0.179)
Inhibition Test-baseline "No"						
Inhbition test Yes	0.118 (0.096)	**0.133 (0.049)**	**0.124 (0.060)**	**0.169 (0.079)**	-0.016 (0.058)	0.000 (0.084)
Inhibiton test only Negative samples	**0.534 (0.198)**	**0.503 (0.100)**	**0.495 (0.126)**	0.277 (0.154)	0.193 (0.132)	**0.361 (0.163)**

Results in bold are significant at the two-sided 5% level.

Experience performing serum tests significantly influences the estimates of viral load. Estimates from participants testing more than 2,000 test annually tend to be higher than estimates from participants performing 0 to 10 tests annually.

Participants who performed inhibition tests only in negative samples are more likely to provide higher estimates of viral load than participants not performing an inhibition test.

No significant differences of the estimates of viral load were found for participants with or without accreditation, plasma test experience, other specimens test experience, laboratory type or method of analysis used.

As an illustrative example, the estimated mean is obtained from the full QTBM model for the results of participants with a particular combination of laboratory practices. Then, the estimated mean is compared with the observed viral load from a participant fulfilling the same conditions.

Consider a participant from 2004 with the following laboratory practice for analysing a sample of 6 $\log_{10}$ copies/ml viral load of subtype A:

- used RTIH technology to analyse the sample,
- used an anti-contamination system,
- was not an accredited laboratory (baseline),
- had no experience performing other specimens test (baseline),
- was using duplicated method of analysis,
- had experience testing between 0 and 10 plasma tests annually (baseline),
- had experience testing between 2,001 and 10,000 serum tests annually,

- was a hospital laboratory (baseline),
- performed an inhibition test.

From the estimated means given in Table 5.4, the estimated mean for participants with such characteristics is derived. Note that the baseline does not add any parameter value since all the information is picked up by the intercept:

$$\hat{\mu} = 3.743 + 1.648 \times (\textit{Year=2004}) + 0.308 \times (\textit{RTIH=yes}) + 0.202 \times (\textit{Antic.=yes})$$

$$-0.111 \times (\textit{Anal.Dupl.=yes}) + 0.323 \times (\textit{Serum4=yes}) + 0.118 \times (\textit{Inhib.=yes}).$$

Since the covariates are indicator variables for non-baseline information, the above expression can be rewritten as

$$\hat{\mu} = 3.743 + 1.648 \times 1 + 0.308 \times 1 + 0.202 \times 1 - 0.111 \times 1 + 0.323 \times 1 + 0.118 \times 1 = 6.231.$$

Thus, results from participants with those characteristics have an estimated mean of 6.231 $\log_{10}$ copies/ml. Since the participants used an in-house technology, the estimated variance is 1.122 (see Table 5.6). Thus, the reported viral load for a sample of 6 $\log_{10}$ copies/ml viral load from a participant with these characteristics is from the normal distribution N(6.231,1.122). If the observed data are checked, a randomly chosen participant in 2004 with the same laboratory practice returned an observed viral load of 6.825 $\log_{10}$. The observed datum has a p-value of 0.322 when assuming to be from the normal distribution N(6.231,1.122), which is in agreement with assuming this distribution for the participant's result.

Now, consider a participant from 2004 who returned a censored observation. Suppose that the participant has the following laboratory practice for analysing a sample of 6 $\log_{10}$ copies/ml viral load subtype A:

- used CC technology to analyse the sample (baseline),

- did not an use anti-contamination system (baseline),
- was an accredited laboratory,
- had experience performing other specimens test,
- was using a single method of analysis (baseline),
- had experience testing between 2,001 and 10,000 plasma tests annually,
- had experience testing between 101 and 1,000 serum test annually,
- was a hospital laboratory (baseline),
- did not perform an inhibition test (baseline).

Then,

$$\hat{\mu} = 3.743 + 1.648 \times (\textit{Year=2004}) + 0.013 \times (\textit{Accred.=yes}) - 0.085 \times (\textit{Other.Spc.=yes})$$

$$+0.237 \times (\textit{Plasma4=yes}) + 0.211 \times (\textit{Serum2=yes}).$$

Thus, the estimated mean is $\hat{\mu} = 5.767$ $\log_{10}$ copies/ml. Since the participant used a commercial technology, the estimated variance is 0.238 (see Table 5.6). The reported viral load in $\log_{10}$ copies/ml from a participant with these characteristics is from the normal distribution N(5.767,0.238).

If the observed data are checked for a participant in 2004 with the same laboratory practice returning a censored observation because its assay provides the upper limit of detection of 5.301 $\log_{10}$ copies/ml for the sample viral load, this censored observation is not in agreement with the estimated distribution for the reported value. If the censored observation is assumed to be from that distribution, the p-value would be 0.017. However, the model takes into account the fact that this observation is censored. In this case the distribution assumed

is truncated to the participant's censored observation, and the estimated mean is different to the one obtained here. The estimated mean for a participant who provided a censored observation with these characteristics is $\hat{\mu} = 7.220$. The estimated value for the censored observation obtained with the model is 7.229, assuming that it is from a normal distribution with mean $\hat{\mu} = 7.220$ and variance 0.238, the p-value was found to be 0.749. Thus, the estimated mean for the result of this participant is in agreement with the estimated value for the censored observation.

Finally, consider a participant from 2004 with the following laboratory practice (the sample to be analysed was a sample of 6 $\log_{10}$ copies/ml viral load of subtype A):

- used bDNA technology to analyse the sample.
- used an anti-contamination system.
- was an accredited laboratory.
- had experience performing other specimens test.
- was using a duplicated method of analysis.
- had experience testing more than 10,000 plasma tests annually.
- had experience testing more than 10,000 serum tests annually.
- was a private laboratory.
- returned a missing value for performing an inhibition test.

From the estimated means given in Table 5.4, the estimated mean for the results of participants with such characteristics as described above is derived. The covariate performing an inhibition test does not provide any information, so the model estimates that the most likely option is not performing any inhibition test.

Note that the baseline does not add any parameter value since all the information is picked up by the intercept:

$$\hat{\mu} = 3.743 + 1.648 * \textit{Year=2004} + 0.518 * \textit{bDNA=yes} + 0.013 * \textit{Accred.=yes}$$

$$-0.111 * \textit{Anal.Dupl.=yes} + 0.150 * \textit{Plasma5=yes} + 0.606 * \textit{Serum5=yes}$$

$$-0.113 * \textit{Labtype=private} = 6.454.$$

Since the participant did not provide information about the covariate, the mean is adjusted to take into account the missing information provided by the participant. The estimated adjusted mean for the results of the participant with unknown information about the covariate level is obtained by the model assuming that the participant did not perform an inhibition test (which is the most likely possibility calculated by the model for this participant). This mean is $\hat{\mu} = 6.462$. The technology used by this participant was a commercial technology, for which the estimated variance is 0.238. The reported value for the sample viral load of a participant with these characteristics is 6.299 $\log_{10}$ copies/ml. Thus, assuming the participant's response is from a N(6.462,0.238), then the p-value associated to the reported value is 0.711, which indicates an agreement between the observed value and the model estimates.

Sample Group 5 $\log_{10}$ Copies/ml Viral Load

The estimated mean for the results of participants under baseline conditions is 4.554 $\log_{10}$ copies/ml for samples of subtype A and 4.509 $\log_{10}$ copies/ml for samples of subtype D. Significant differences were found for the estimated means for samples across years. For samples from years 2002 to 2004 participants tend to provide higher estimates of viral load than for samples from 2005.

Participants using bDNA and HC technologies are more likely to provide higher estimates of viral load than CC technology users. However, participants using RTC technology tend to

provide lower estimates of viral load than CC technology users. No differences were found for the performance of participants using in-house technologies with respect to CC technology users.

Estimates of viral load from participants using an anti-contamination system tend to be higher than estimates from participants not using any anti-contamination system.

Participants with experience performing more than 2,000 serum tests annually are more likely to provide higher estimates of viral load than participants performing very few (0 to 10 tests annually).

Participants from research laboratories tend to return lower estimates of viral load than participants from hospital laboratories.

Participants performing inhibition tests tend to provide higher estimates of viral load than participants who did not perform any inhibition test.

Sample Group 4 $\log_{10}$ Copies/ml Viral Load

Performance of participants are significantly different across time (see Table 5.4). The estimated mean for the results of participants under baseline conditions is 3.509 $\log_{10}$ copies/ml. For samples from 2002 and 2004 participants are more likely to provide higher estimates of viral load than for samples from 2005. The estimates of viral load for samples of subtype D tend to be lower than the estimates for samples of subtype A.

The use of different technologies has an influence on participants' performance. RTC users tend to return lower estimates of viral load than CC users, in contrast to bDNA users who are more likely to provide higher estimates than CC users.

Participants using an anti-contamination system are more likely to return higher estimates of viral load than participants not using an anti-contamination system.

The method of analysis, the accreditation status and experience in performing other specimen tests are not significant covariates when estimating the viral load.

Participants performing over 2,000 serum tests annually tend to provide higher estimates of viral load than participants performing less than 11 tests.

Participating manufacture laboratories are more likely to return higher estimates of viral load than participating hospital laboratories.

The estimates of viral load from participants performing an inhibition test on the samples tend to be higher than the estimates from participants not performing any inhibition test.

Sample Group 3.5 $\log_{10}$ Copies/ml Viral Load

The estimated mean for the results of participants under baseline conditions is 3.139 $\log_{10}$ copies/ml. Significant differences were found between participants' performance samples from 2002 and 2003. For samples from 2002 participants tend to provide higher estimates of viral load than for samples from 2003.

Participants using RTIH technology are more likely to return higher estimates of viral load than participants using CC technology, if all the other covariates are the same.

Those participants with experience in testing other specimens are more likely to provide higher estimates of viral load than participants without experience.

Participants performing an inhibition test tend to return higher estimates of viral load than participants not performing an inhibition test.

Sample Group 3 $\log_{10}$ Copies/ml Viral Load

The estimated mean for the participants' results under baseline conditions is 2.769 $\log_{10}$ copies/ml (see Table 5.4). No differences were found between the performance from 2003 and 2005. However, for samples from 2004 participants tend to provide higher estimates of viral load than for samples from 2005. No significant differences were found for the estimates of viral load when testing samples of different subtypes.

Participants using HC technology are more likely to provide higher estimates of viral load than CC technology users.

The estimated mean for the results of participants with experience testing other specimens is lower than for the results of participants with no experience. Participants with experience testing plasma and serum samples (between 2,000 and 10,000 tests annually) tend to provide higher estimates of viral load than participants performing 0 to 10 tests annually.

Private and research laboratories are more likely to return lower estimates of viral load than hospital laboratories, while manufacture laboratories tend to provide higher estimates than hospital laboratories.

Sample Group 2.3 $\log_{10}$ Copies/ml Viral Load

The estimated mean for the participants' results under baseline conditions is 2.491 $\log_{10}$ copies/ml. For samples from 2003 participants tend to provide lower estimates of viral load than for samples from 2004.

The estimates of viral load from users of CIH, RTIH, RTC and HC tend to be higher than the estimates from CC users.

Participants with extensive experience performing plasma tests (more than 10,000 tests annually) are more likely to provide lower estimates than participants with less experience.

Those participants performing an inhibition test only in negative samples tend to return higher estimates of viral load than participants not performing an inhibition test.

5.6.1.2 Reduced HBV Quantitative Model

As with the qualitative model considered in Chapter 4, model reduction and simplification is carried out. For each sample group those covariates which are not significant at the two-sided 5% level are removed from the full model. However, covariates behaving as confounders are not excluded from the model even if they are not significant.

Interactions between covariates and correlations of the estimated parameters were studied. No interaction was found between covariates. Correlations between the estimated parameters were checked when reducing the full model to the final one. The estimated parameters were not correlated or only low correlations occur (with correlation coefficients lower than 0.25).

The results obtained from the reduced model for each sample group can be found in Tables C.7 to C.12 in Appendix C. Table 5.5 shows a summary of the results obtained from the reduced model. The estimated mean and standard deviation (SD) of the parameter estimates for each sample group are presented. Results in bold are significant at the two-sided 5% level. In this section, only the significant findings are reported.

Table 5.5: *Mean and SD of the parameter estimates from the reduced QTBM for HBV sample groups.*

Covariates	Mean (SD)					
	6	5	4	3.5	3	2.3
Intercept	3.592 (0.486)	4.489 (0.091)	3.444 (0.126)	3.117 (0.114)	2.833 (0.097)	2.359 (0.074
Year- baseline "2005"						
Year 2002	**0.565 (0.115)**	**0.318 (0.062)**	**0.191 (0.084)**	**0.641 (0.053)**	-0.053 (0.074)	**0.176 (0.069)**
Year 2003	**1.71 (0.477)**	**0.147 (0.062)**	-0.001 (0.084)		**0.531 (0.049)**	**-0.404 (0.079)**
Year 2004	**1.745 (0.477)**	**0.333 (0.051)**	0.120 (0.068)			
Subtype- baseline "A"	**1.366 (0.477)**		**-0.138 (0.060)**			
Technology group- baseline "CC"						
Tech CIH-CC	0.206 (0.208)	-0.201 (0.155)	-0.225 (0.188)	-0.228 (0.182)	-0.080 (0.160)	**0.630 (0.198)**
Tech RTIH-CC	0.270 (0.142)	-0.051 (0.091)	0.089 (0.109)	**0.430 (0.115)**	0.119 (0.094)	**0.627 (0.116)**
Tech RTC-CC	0.145 (0.110)	**-0.119 (0.060)**	**-0.194 (0.073)**	0.317 (0.211)	-0.093 (0.060)	**0.367 (0.084)**
Tech bDNA-CC	**0.551 (0.139)**	**0.364 (0.090)**	**0.259 (0.116)**	0.614 (0.491)	-0.130 (0.108)	-0.511 (0.500)
Tech HC-CC	**0.521 (0.211)**	**0.388 (0.130)**	0.074 (0.367)	0.253 (0.252)	**1.553 (0.263)**	**2.379 (0.214)**
Anti- baseline "No"	**0.263 (0.087)**	**0.174 (0.049)**	**0.200 (0.057)**			
OthrSpc.- baseline "No"				**0.252 (0.077)**	**-0.197 (0.071)**	
Analysis method- baseline Singly						
Analysis Duplicated				-0.009 (0.064)		
Analysis Other				**0.269 (0.118)**		
Plasma- baseline 0-10						
Group 1: 11-100		-0.067 (0.110)	0.063 (0.148)	0.107 (0.099)	-0.113 (0.111)	
Group 2: 101-1,000		0.030 (0.065)	0.147 (0.091)	-0.048 (0.067)	0.029 (0.080)	
Group 3: 1,001-2,000		0.123 (0.091)	0.151 (0.126)	0.064 (0.101)	0.084 (0.116)	
Group 4: 2,001-10,000		0.084 (0.070)	0.164 (0.095)	0.095 (0.147)	**0.224 (0.089)**	
Group 5: > 10,000		-0.120 (0.178)	0.012 (0.198)	1.256 (0.661)	0.101 (0.136)	
Serum- baseline 0-10						
Group 1: 11-100	0.138 (0.152)	0.066 (0.087)	0.148 (0.120)	-0.153 (0.104)	0.125 (0.098)	
Group 2: 101-1,000	0.175 (0.120)	-0.028 (0.063)	**0.188 (0.090)**	0.008 (0.070)	0.072 (0.079)	
Group 3: 1,001-2,000	-0.017 (0.186)	-0.140 (0.099)	0.060 (0.147)	-0.139 (0.086)	0.060 (0.114)	
Group 4: 2,001-10,000	**0.251 (0.142)**	**0.245 (0.074)**	**0.388 (0.102)**	-0.087 (0.357)	**0.297 (0.085)**	
Group 5: > 10,000	**0.536 (0.245)**	**0.478 (0.159)**	**0.736 (0.198)**	0.871 (0.1359)	**0.589 (0.178)**	
Labtype- baseline Hospital						
Public Health		0.102 (0.073)	0.113 (0.102)		-0.024 (0.091)	
Private		-0.074 (0.078)	-0.201 (0.128)		**-0.272 (0.085)**	
Reference		-0.043 (0.131)	-0.072 (0.179)		0.018 (0.159)	
Manufacture		0.146 (0.104)	**0.383 (0.140)**		**0.362 (0.162)**	
Research		**-0.394 (0.109)**	-0.103 (0.132)		**-0.466 (0.098)**	
Inhibition Test-baseline "No"						
Inhbition test Yes	0.121 (0.090)	**0.135 (0.049)**	**0.129 (0.062)**	**0.165 (0.070)**	-0.022 (0.059)	-0.003 (0.072)
Inhibiton test only Negative samples	**0.630 (0.179)**	**0.498 (0.098)**	**0.472 (0.121)**	**0.312 (0.149)**	0.212 (0.123)	**0.367 (0.148)**

Results in bold are significant at the two-sided 5% level.

The summary of the results and estimated means provided in the subsequent section are based on the assumption that all covariates, except the one that is commented on at each time, are under baseline conditions.

Sample Group 6 $\log_{10}$ Copies/ml Viral Load

The estimated mean for the results of participants under baseline conditions is 3.592 $\log_{10}$ copies/ml, while the estimated mean for the results of participants for samples from 2004 is $3.592+1.745 = 5.337$ $\log_{10}$ copies/ml. Thus, for samples from 2004 participants, is more likely to obtain an estimate of the viral load closer to the 'target' estimate than for samples from 2005. Estimates of viral load for samples of subtype D tend to be higher than estimates for samples of subtype A. For samples of subtype D the estimated mean is $3.592 + 1.366 = 4.958$ $\log_{10}$ copies/ml.

Participants using CC technology tend to return lower estimates of viral load than participants using other technologies. The highest estimated means are obtained for the results of participants using bDNA and HC technologies, $3.592 + 0.551 = 4.143$ and $3.592 + 0.521 = 4.113$ $\log_{10}$ copies/ml, respectively.

The estimated mean for the results of participants using an anti-contamination system is 0.263 $\log_{10}$ copies/ml, which is higher than the estimated mean for the results of participants not using an anti-contamination system.

The estimated mean for the results of participants with experience testing more than 10,000 serum tests annually is 4.128 $\log_{10}$ copies/ml, which is the highest estimated mean amongst the results of participants with different level of experience performing serum tests annually.

The estimated mean for the results of participants performing inhibition tests only in negative samples is significantly higher than the estimated mean for the results of participants not performing any inhibition test.

Sample Group 5 $\log_{10}$ Copies/ml Viral Load

The estimated mean for the results of participants under baseline conditions is 4.489 $\log_{10}$ copies/ml. For samples from 2002 to 2004 participants are more likely to provide higher estimates of viral load than for samples from 2005. The estimated means were found to be around 4.8 $\log_{10}$ copies/ml for the results of participants for samples from 2002 and 2004.

For the results of participants using in-house technologies and RTC technology the estimated means are under 4.5 $\log_{10}$ copies/ml. In contrast, for the results of participants using bDNA and HC technologies the estimated means are over 4.5 $\log_{10}$ copies/ml.

For the results of participants with experience performing more than 2,000 and less than 10,000 serum tests annually the estimated mean is 4.73 $\log_{10}$ copies/ml. For the results of participants with experience of more than 10,000 serum tests annually the estimated mean is 4.96 $\log_{10}$ copies/ml.

Research laboratories tend to provide significantly lower estimates of viral loads than hospital laboratories. The estimated mean for the results of research laboratories is around 4 $\log_{10}$ copies/ml.

For participants who did not perform any inhibition test the estimated mean is less than 4.5 $\log_{10}$ copies/ml. In turn, the mean for the results of participants who performed inhibition tests, is more than 4.5 $\log_{10}$ copies/ml.

Sample Group 4 $\log_{10}$ Copies/ml Viral Load

For the results of participants for 2005 samples the estimated mean is 3.444 $\log_{10}$ copies/ml, which is lower than the estimated mean for the results of participants for 2002 samples (3.635 $\log_{10}$ copies/ml).

The highest estimated mean was found for the results of bDNA users, which is 3.703 $\log_{10}$ copies/ml (versus 3.444 $\log_{10}$ copies/ml for the results of CC users). In contrast, the lowest estimated mean is obtained for the RTC users, 3.25 $\log_{10}$ copies/ml.

For the results of participants testing more than 10,000 serum samples annually the estimated mean is higher than 4 $\log_{10}$ copies/ml. For the results of manufacture laboratories the estimated mean is close to 4 $\log_{10}$ copies/ml, in contrast to the mean for the results of private laboratories, which is around 3.2 $\log_{10}$ copies/ml.

The estimated mean for the results of participants performing inhibition tests is closer to the target, 4 $\log_{10}$ copies/ml provided by QCMD, than the estimated mean for the results of participants not performing any inhibition test.

Sample Group 3.5 $\log_{10}$ Copies/ml Viral Load

The estimated mean for the results of participants under baseline conditions is 3.117 $\log_{10}$ copies/ml. The estimated mean for the results of participants from 2002, 3.7 $\log_{10}$ copies/ml, is higher than the mean for the results of participants from 2003.

The lowest estimated mean was found for the results of CIH technology users, which is 2.889 $\log_{10}$ copies/ml. In turn, the highest estimated mean is obtained for the results of bDNA technology users, that is 3.731 $\log_{10}$ copies/ml. The closest estimated mean to the target sample viral load was found for the results of RTIH users.

There are differences between estimated means for the results of participants depending on the level of experience performing serum and plasma tests annually; although these difference are not significant, these factors confound other results.

The estimated mean for the results of participants using other methods of analysis, such as triplicate methods, is higher than 3.3 $\log_{10}$ copies/ml.

Participants performing inhibition tests tend to provide higher estimates of viral load than participants who did not perform any inhibition test.

Sample Group 3 $\log_{10}$ Copies/ml Viral Load

The estimated mean for the results of participants under baseline conditions is 2.833 $\log_{10}$ copies/ml. The estimated mean for the results of participants for samples from 2004 is over 3.3 $\log_{10}$ copies/ml.

The closest estimated mean to the target viral load is obtained for the results of RTIH technology users, which is 2.952 $\log_{10}$ copies/ml assuming all other variables are at their baseline levels. For the results of participants using HC technology the estimated mean is 4.386 $\log_{10}$ copies/ml.

The estimated means for the results of participants performing more than 2,000 serum tests annually are higher than 3 $\log_{10}$ copies/ml.

The lowest estimated mean was found for the results of participants from research laboratories, around 2.3 $\log_{10}$ copies/ml, whilst the highest estimated mean is obtained for the results of participants from manufacture laboratories. The closest estimated mean to the target viral load was found for the results of participants from reference laboratories.

Sample Group 2.3 $\log_{10}$ Copies/ml Viral Load

For the results of participants under baseline condition the estimated mean is 2.359 $\log_{10}$ copies/ml. The estimated means for the results of participants for 2002 and 2003 samples are significantly higher and lower, respectively, than the estimated mean for the results of participants for 2004 samples.

The highest estimated mean is obtained for the results of HC technology users, which is 4.738 $\log_{10}$ copies/ml. The closest estimated mean to the target value was found for the results of CC technology users.

The estimated mean for the results of participants performing inhibition tests only in negative samples is higher than 2.7 $\log_{10}$ copies/ml.

5.6.1.3 Estimated Variances for the HBV Quantitative Data

The estimated variances for the distributions of participants' results are obtained by the use of the QTBM. Table 5.6 shows the estimated variances by the model per sample group and technology type. Note that the estimated variance refers to the mean of the posterior distribution of the variance.

The QTBM allows different variances depending on whether the technology used by the participant is commercial or in-house type. It is observed that the estimated variances for the results of in-house technologies users are higher than the estimated variances for the results

of commercial technologies users for all sample groups. These results are in agreement with the classical test performed in Section 5.2, where the variability of the results obtained from participants differs depending on the technology used for some of the sample groups.

Table 5.6: *Summary statistics of the variance estimates from the reduced QTBM for HBV sample groups classified by technology type: estimated mean, standard deviation (SD), confidence interval.*

Estimated Variances per sample group	Mean	SD	95% Confidence Interval		
			2.50%	Median	97.50%
6 Log10 copies/ml					
In-house	1.122	0.190	0.808	1.102	1.556
Commercial	0.238	0.031	0.185	0.236	0.305
5 Log10 copies/ml					
In-house	0.982	0.113	0.789	0.973	1.228
Commercial	0.149	0.012	0.128	0.149	0.174
4 Log10 copies/ml					
In-house	0.883	0.130	0.665	0.873	1.167
Commercial	0.129	0.014	0.103	0.128	0.159
3.5 Log10 copies/ml					
In-house	0.473	0.104	0.311	0.459	0.718
Commercial	0.057	0.009	0.041	0.056	0.078
3 Log10 copies/ml					
In-house	0.598	0.098	0.434	0.589	0.823
Commercial	0.127	0.013	0.104	0.126	0.154
2.3 Log10 copies/ml					
In-house	0.480	0.114	0.306	0.464	0.740
Commercial	0.089	0.016	0.063	0.088	0.127

5.6.1.4 Model Checking

Figure 5.1 represents the density function of the probabilities that the observed standardised residuals are equal or exceed the predicted standardised residuals for sample group 4 $\log_{10}$ copies/ml in 2005. It is observed that the distribution of the proportions is within the interval given by the confidence limits of 0.025 and 0.975, showing no significant differences between the observed and the replicated data. Similar results were found for all other sample groups.

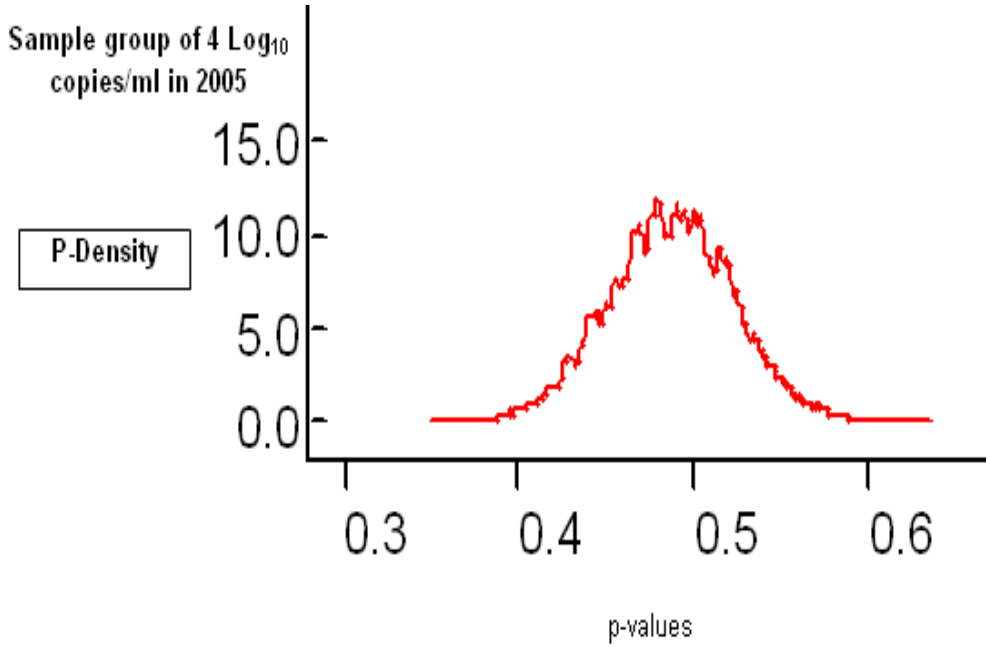


Figure 5.1: *Density function of the Bayesian p-values for sample group 4* $\log_{10}$ *copies/ml in 2005.*

Table 5.7 shows the mean for the distribution of the probabilities of $T(y_{si}^{rep}, \mu_s, \sigma_{s_i}) \geq T(y_{si}, \mu_s, \sigma_{s_i})$, as well as its 95% confidence intervals per sample group and year; in other words, the mean and confidence intervals of the Bayesian 'p-values' that assess the statistical significance of discrepancies between observed and predicted data. Extreme Bayesian 'p-values' indicate conflict between data and aspects of the model. No discrepancies were found for any of the sample groups and years, since the Bayesian 'p-values' are higher than 0.025 and lower than 0.975 for all cases. Therefore, it is concluded that the model fits the data adequately.

As in the previous chapter, note that the definition of function T is chosen depending on the data to be analysed. A function that is sensible for the quantitative responses and can describe the data appropriately has been chosen (Gelman et al., 2004).

Table 5.7: *HBV mean and confidence intervals of Bayesian 'p-values' per year and sample group.*

Test		Bayesian P-Values	95% Confidence Interval	
Year	Sample group		2.50%	97.50%
2002	6	0.499	0.369	0.631
	5	0.472	0.385	0.562
	4	0.485	0.359	0.609
	3.5	0.504	0.415	0.592
	2.3	0.438	0.302	0.571
2003	6	0.492	0.371	0.629
	5	0.488	0.395	0.581
	4	0.473	0.344	0.607
	3.5	0.482	0.344	0.607
	3	0.435	0.312	0.557
	2.3	0.436	0.310	0.569
2004	6	0.482	0.366	0.598
	5	0.499	0.421	0.579
	4	0.505	0.390	0.610
	3	0.487	0.407	0.568
	2.3	0.435	0.329	0.544
2005	6	0.489	0.390	0.590
	5	0.463	0.395	0.530
	4	0.487	0.417	0.558
	3	0.474	0.405	0.542

5.6.1.5 Model Comparison

Figure C.1 (see Appendix C) shows the estimated posterior means of sample viral loads and their 95% confidence intervals from the full and the reduced model. As in the previous chapter, two selected plots are chosen in order to illustrate with more clarity the tendencies of the estimated means. It is shown that the estimated means from the reduced model lie within the confindence intervals indicating that the reduced model is appropriate.

Generally speaking, the reduced model fits the data appropriately since it can be observed that the posterior means from the reduced model lie within the confidence intervals obtained from the full model for all sample groups and years (see Appendix C, Figure C.1).

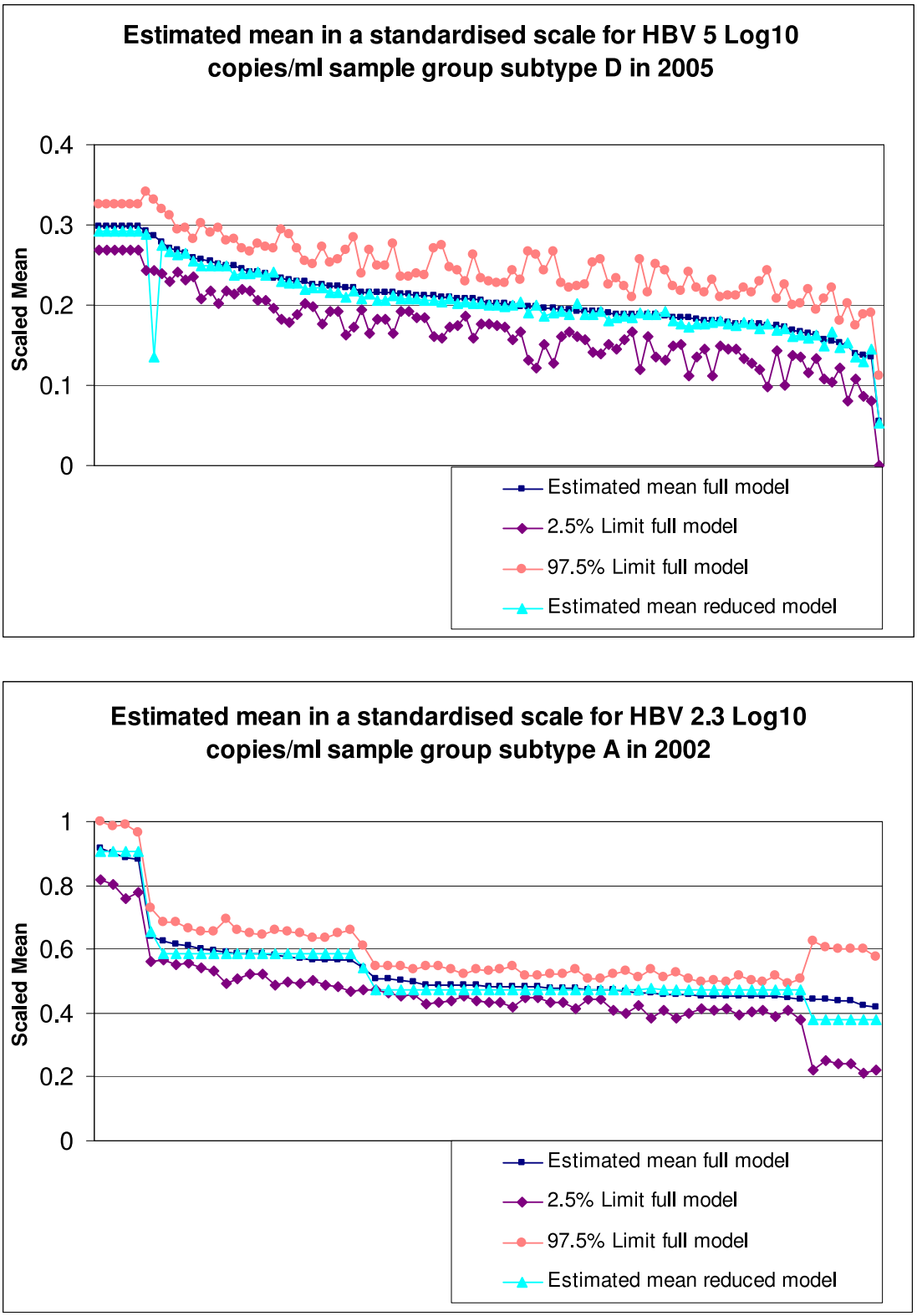


Figure 5.2: *HBV estimated means of sample viral load from the full and reduced QTBM for the sample group 2.3* $\log_{10}$ *copies/ml in 2002 and for the sample group 5* $\log_{10}$ *copies/ml in 2005. Results are ordered by decreasing estimated means. The x-axis represents an identification number of participants' results.*

5.6.1.6 Estimates of Censored Data of HBV

A special feature of the quantitative data considered in this chapter is the censored observations. The QTBM takes into account the censored observations via the use of a probability distribution. Therefore, with the use of the probability distribution an estimate for the sample viral load can be obtained for a participant who provided a censored observation.

The posterior distribution for each censored observation is obtained using the QTBM. The mean of the distribution is used to impute the unknown value of the sample viral load.

Figure 5.3 shows the censored observations and their estimates for sample groups 2.3 and 6 $\log_{10}$ copies/ml. The estimates are ordered for each technology group by the censored observation and then by the estimated values obtained from the model. The estimates for the sample group 2.3 $\log_{10}$ copies/ml are observed to be lower and closer to the target than the censored observation provided by the participant.

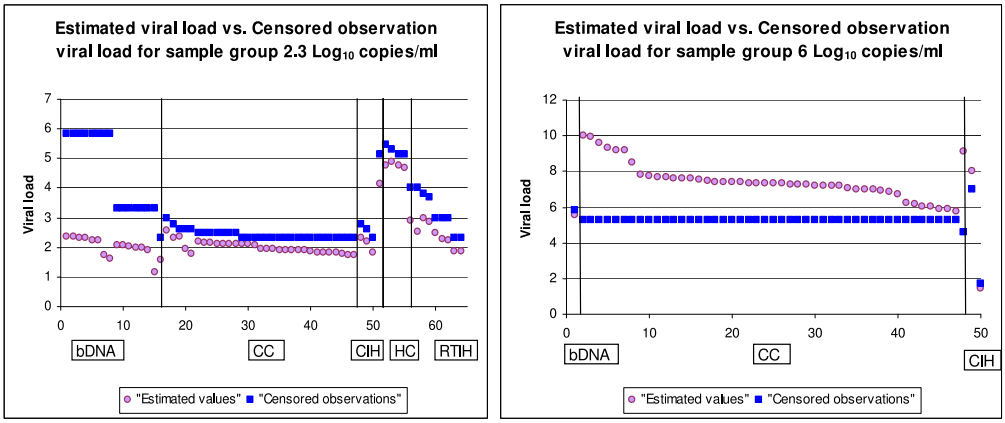


Figure 5.3: *Censored observations and their estimates for the HBV sample groups 2.3 and 6* $\log_{10}$ *copies/ml. Labs' codes are order by technology used as follows: bDNA;17 data for the sample group 2.3 and 1 for the sample group 6* $\log_{10}$ *copies/ml, CC; 31 data for the sample group 2.3 and 47 for the sample group 6* $\log_{10}$ *copies/ml, CIH; 3 data for the sample group 2.3 and 2 for the sample group 6* $\log_{10}$ *copies/ml, HC; 5 data for the sample group 2.3* $\log_{10}$ *copies/ml, RTIH; 9 data for the sample group 2.3* $\log_{10}$ *copies/ml.*

The variability between the estimates and the censored observations for the sample group 6 $\log_{10}$ copies/ml is higher than the variability for the sample group 2.3 $\log_{10}$ copies/ml. Also, some estimates for the strongest sample group are far from the target viral load, in contrast to the results obtained for the weakest viral load sample group. It is observed that for the strongest sample group, estimates from the bDNA and CIH users are close to the observed censored value. This may be due to the fact that those censored observations are left censored (see Figure 2.1 in Chapter 1). The fact that censored responses have estimates further away from the censored observation and from the target viral load may be explained by a different combination of all other covariates.

5.6.2 Modelling HCV Quantitative Data

In order to check if the QTBM fits data from another pathogen, the model is applied to the Hepatitis C Virus data. Since the steps to reduced the full model and the procedure to follow is the same as in the previous subsection, details of the conclusions obtained from the full model are not described. Only the results from the final reduced model along with the conclusions from model comparison and checking are presented.

Hepatitis C virus (HCV) proficiency panels are described in Chapter 2. Information about the percentage of censored data per sample group can also be found in Chapter 2. The variables included in the model are presented in Table 5.8. Table 5.9 shows the mean of the observed responses per sample group and classified by covariate level (with the exception of the censored observations). The mean of the responses from participants with other method of analysis appears fairly low for the sample group 3.5 IU/ml. The reason for this is due to the fact the there is only one observation available for calculating the mean.

Table 5.8: *Covariates for the analysis of HCV quantitative data.*

Covariate	Description	Values
Year	Year when the sample was analysed	Indicator variables for year 2002, 2003 and 2004 compared with year 2005
Technology	Technology used to analyse the sample	Indicator variables for technology groups: CIH, RTC, RTIH, bDNA compare with CC
Anti	Use of anti-contamination system	Indicator variable with 'No use of anti-contamination system' as baseline
Accred	Laboratory accreditation status	Indicator variable with 'No accredited laboratory' as baseline
OtherSp.	Experience on other specimen sample performance such as biopsies, swabs, etc..	Indicator variable with 'No experience performing test of other specimens' as baseline
Serum	Annual number of serum tests performed by the participant	Indicator variables per group of number of plasma tests: '0-10' baseline, '11-100'- group 1, '101-1,000' - group 2, '1,001-2,000'- group 3, '2,001-10,000'- group 4, '$> 10,000$'- group 5
Plasma	Annual number of plasma tests performed by the participant	Indicator variables per group of number of plasma tests: '0-10' baseline, '11-100'- group 1, '101-1,000' - group 2, '1,001-2,000'- group 3, '2,001-10,000'- group 4, '$> 10,000$'- group 5
Analysis	Method of analysis used by the participant	Indicator variables for analysis method. The baseline is 'Simply method of analysis' compared with 'Duplicate' and 'Other methods'
Test Inhibition	Performance of inhibition test by the participant	Indicator variables with 'Non performance of inhibition tests' as baseline compare with 'Performance of inhibition test' and 'Performance of inhibition test only in negative samples'
Labtype	Laboratory type where the sample was analysed	Indicator variables with 'Hospital laboratories' as baseline compare with 'Public Health laboratory', 'Private laboratories', 'Reference laboratories', 'Manufactures laboratories' and 'Research/Others laboratories'
Genotype	Genotype of the sample analysed	Indicator variables with 'Genotype 1' as baseline compare with 'Genotype 3', 'Genotype 4' and 'Genotype 5'

Technologies: CC=Conventional Commercial, RTC=Real Time Commercial, CIH=Conventional Commercial, RTIH=Real Time In-house and bDNA=bDNA.

Table 5.9: *Mean of participants' results per covariate level for HCV programmes.*

Covariates	Mean of paticipants' results					
	5.9	4.9	3.9	3.5	3.2	2.2
Year						
Year 2002	5.254	4.782	3.912	3.182	2.634	2.687
Year 2003	5.921	4.722	3.782	3.332	3.011	1.915
Year 2004	6.039	4.643	3.707	3.771	2.980	2.017
Year 2005	-	4.581	3.690	-	2.924	2.538
Genotype						
Genotype 1	5.748	4.838	3.834	3.271	2.960	2.190
Genotype 3	-	4.580	3.512	-	-	-
Genotype 4	-	4.547	-	-	-	2.692
Genotype 5	-	4.643	-	-	-	-
Technology group						
Tech CC	5.783	4.790	3.878	3.322	3.014	2.540
Tech CIH	4.862	3.981	3.403	-	2.421	2.532
Tech RTIH	5.831	4.343	3.602	3.031	2.842	1.970
Tech RTC	5.624	4.443	3.604	3.454	2.633	2.101
Tech bDNA	5.780	4.487	3.481	3.133	3.024	3.104
Anticontamination						
Yes	5.762	4.720	3.842	3.271	2.991	2.433
No	5.728	4.535	3.385	3.233	2.892	2.194
Not answered	5.485	4.541	3.949	3.282	2.904	2.542
Accreditation						
Yes	5.750	4.722	3.840	3.281	2.990	2.320
No	5.681	4.571	3.709	3.360	2.957	2.391
Not answered	5.809	4.654	3.712	3.233	2.909	2.423
Other Specimens						
Yes	5.701	4.654	3.743	3.412	3.002	2.410
No	5.732	4.657	3.792	3.284	2.961	2.332
Not answered	5.791	4.638	3.714	3.242	2.914	2.432
Analysis method						
Analysis Singly	5.780	4.691	3.780	3.248	2.990	2.440
Analysis Duplicated	5.612	4.462	3.621	3.426	2.843	2.154
Analysis Other	6.643	4.987	4.250	1.202	-	2.143
Not answered	-	-	-	3.272	2.901	2.582
Plasma						
Group 0: 0-10	5.532	4.710	3.830	3.404	3.074	2.363
Group 1: 11-100	5.544	4.684	3.853	3.486	3.018	-
Group 2: 101-1,000	5.861	4.732	3.842	3.279	2.981	2.202
Group 3: 1,001-2,000	5.564	4.754	3.871	2.785	2.870	2.643
Group 4: 2,001-10,000	5.880	4.432	3.604	3.362	2.883	2.354
Group 5: > 10,000	5.921	4.660	3.647	3.431	2.812	2.391
Not answered	5.811	4.641	3.709	3.234	2.914	2.420
Serum						
Group 0: 0-10	5.803	4.623	3.716	3.208	2.922	2.533
Group 1: 11-100	5.912	4.779	3.939	3.410	3.093	2.142
Group 2: 101-1,000	5.837	4.676	3.790	3.284	3.034	2.491
Group 3: 1,001-2,000	5.403	4.611	3.841	3.321	2.841	2.153
Group 4: 2,001-10,000	5.758	4.653	3.725	3.452	3.082	2.504
Group 5: > 10,000	5.039	4.784	3.770	-	-	-
Not answered	5.810	4.642	3.701	3.226	2.908	2.424
Labtype						
Hospital	5.737	4.703	3.804	3.321	3.011	2.353
Public Health	5.631	4.635	3.757	3.230	2.851	2.220
Private	5.847	4.528	3.802	3.323	2.982	2.516
Reference	5.379	4.490	3.453	2.955	-	-
Manufacture	4.760	4.246	3.612	3.362	2.434	-
Research	5.740	4.743	3.714	3.621	2.971	-
Not answered	5.811	4.642	3.711	3.232	2.913	2.427
Inhibition Test						
Inhibition test No	5.563	4.535	3.610	3.150	2.989	2.609
Inhbition test Yes	5.844	4.709	3.826	3.287	2.990	2.257
Inhibiton test only Negative samples	5.342	4.391	3.649	3.427	2.454	2.312
Not answered	5.576	4.473	3.457	3.271	2.887	2.601

The associations of the covariates in Table 5.8 with the estimated sample viral load are studied by adjusting the QTBM to the HCV data. The full model is fitted to the HCV data, and the parameter corresponding to each variable is tested at the two-sided 5% significance level. Those covariates that are not significant, do not interact with others and/or do not behave as a confounder in the model are discarded. Thus, the full model fitted to HCV data is simplified to obtain a reduced model by using a backwards elimination. Here, the results and tables summarising the reduced model fitted to HCV data are presented (see Appendix C for a summary of the results for the full model).

5.6.2.1 Reduced HCV Quantitative Model

The reduced model includes those covariates, per sample group, that are significant at the two-sided 5% level in the full model. Confounding covariates, interactions between them and correlations of the estimated parameters were checked. The confounding covariates were included in the reduced model, interactions were not significant and correlations between the estimated parameters were not found.

Tables C.14 to C.19, in the Appendix C, show the results obtained from the reduced model for each sample group. The Table 5.10 shows the mean and standard deviation for the parameter estimates from the reduced model for each sample group. The results in bold are the significant parameters at the two-sided 5% level.

In the next subsection, a description of the results is obtained assuming that the covariates are under baseline levels except for the commented variable at each time.

Table 5.10: *Mean and SD of the parameter estimates from the reduced QTBM for HCV sample groups.*

Covariates	Mean (SD)					
	5.9	4.9	3.9	3.5	3.2	2.2
Intercept	5.897 (0.074)	5.074 (0.076)	3.870 (0.066)	3.276 (0.094)	2.335 (0.316)	2.332 (0.092)
Year- baseline "2005"						
Year 2002	**-0.806 (0.065)**				0.235 (0.323)	-0.075 (0.122)
Year 2003	-0.077 (0.061)				**0.730 (0.317)**	**-0.446 (0.132)**
Year 2004					**0.726 (0.317)**	**-0.368 (0.112)**
Genotype- baseline "1"						
Genotype 3		**-0.228 (0.065)**	**-0.242 (0.061)**		**0.561 (0.316)**	
Genotype 4		**-0.229 (0.093)**				
Genotype 5		-0.130 (0.080)			**0.687 (0.315)**	
Technology group- baseline "CC"						
Tech CIH-CC		**-0.744 (0.276)**	**-0.467 (0.204)**	-0.435 (0.326)	-0.405 (0.211)	0.291 (0.200)
Tech RTIH-CC		-0.428 (0.124)	**-0.364 (0.141)**	-0.357 (0.211)	**-0.208 (0.088)**	-0.092 (0.127)
Tech RTC-CC		**-0.136 (0.089)**	**-0.170 (0.083)**	-0.224 (0.123)	-0.004 (0.088)	-0.072 (0.138)
Tech bDNA-CC		**-0.286 (0.079)**	**-0.374 (0.075)**	**-0.405 (0.128)**	-0.072 (0.097)	**0.282 (0.132)**
Anti- baseline "No"				0.083 (0.099)		
Accred- baseline "No"			**0.140 (0.06)**			
Analysis method- baseline Singly						
Analysis Duplicated			-0.025 (0.078)			
Analysis Other			**0.719 (0.275)**			
Plasma- baseline 0-10						
Group 1: 11-100		-0.101 (0.127)	-0.044 (0.134)			
Group 2: 101-1,000		-0.061 (0.077)	-0.007 (0.075)			
Group 3: 1,001-2,000		-0.057 (0.107)	-0.019 (0.098)			
Group 4: 2,001-10,000		**-0.372 (0.090)**	**-0.194 (0.086)**			
Group 5: > 10,000		-0.006 (0.144)	-0.045 (0.155)			
Labtype- baseline Hospital						
Public Health	0.010 (0.103)	0.037 (0.100)			-0.029 (0.100)	
Private	0.115 (0.098)	-0.056 (0.092)			0.01 (0.081)	
Reference	0.034 (0.202)	0.052 (0.179)			**-1.251 (0.164)**	
Manufacture	**-2.946 (0.335)**	**-1.391 (0.170)**			-0.117 (0.177)	
Research	-0.119 (0.193)	-0.066 (0.148)			0.050 (0.148)	
Inhibition Test-baseline "No"						
Inhbition test Yes	**0.265 (0.070)**				-0.077 (0.078)	
Inhibiton test only Negative samples	0.149 (0.157)				**-0.382 (0.130)**	

Results in bold are significant at two-sided 5% level.

Sample Group 5.9 $\log_{10}$ **IU/ml Viral Load**

The estimated mean for the results of participants under baseline conditions is 5.897 $\log_{10}$ IU/ml. There are significant differences on performance over time. The estimated mean for the results from 2002, 5.091 $\log_{10}$ IU/ml, is lower than the mean for the results from 2004, .

Participants perform similar independently of the technology used, their experience testing other type of samples, their accreditation status and the use of an anti-contamination system.

Manufacture laboratories tend to provide significantly lower estimates of viral load than hospital laboratories, resulting in an estimated mean for their responses of 2.951 $\log_{10}$ IU/ml.

The estimated means for the results of participants performing inhibition tests are higher than for the results of participants not performing any inhibition test; for the results of participants performing inhibition tests only in negative samples the estimated mean is 6.162 $\log_{10}$ IU/ml.

Sample Group 4.9 $\log_{10}$ **IU/ml Viral Load**

The estimated mean for the results of participants under baseline conditions is 5.074 $\log_{10}$ IU/ml. There are significant differences on performance for samples with different genotypes: samples of genotypes 3, 4 or 5 are estimated with lower viral load than samples of genotype 1. The estimated means for samples of genotype 3 and 4 are around 4.84 $\log_{10}$ IU/ml.

Differences between the estimated means for the results of participants using different technologies were found. In particular, CC users tend to provide higher estimates of viral load than users of all other technologies.

RTIH and CIH users are more likely to return lower estimates than the target viral load of 4.9 $\log_{10}$ IU/ml; their estimated means are 4.6 and 4.3 $\log_{10}$ IU/ml, respectively. The estimated mean for the results of bDNA users is 4.78 $\log_{10}$ IU/ml.

Participants with experience performing between 2,001 and 10,000 plasma tests annually tend to return lower estimates of viral loads than participants with less or more experience. The estimated mean is the furthest away from the target viral load.

The lowest estimated mean was found for the results of manufacture laboratories, having a value of 3.68 $\log_{10}$ IU/ml.

Sample Group 3.9 $\log_{10}$ **IU/ml Viral Load**

The estimated mean for the results of participants under baseline conditions is 3.87 $\log_{10}$ IU/ml, being almost the target viral load. Lower estimates of viral load were found for samples of genotype 3 than for samples of genotype 1.

CIH, RTIH, RTC and bDNA users tend to provide significantly lower estimates than CC users. The estimated means are 3.4, 3.5, 3.7, 3.5 $\log_{10}$ IU/ml, respectively.

Accredited participants are more likely to return higher estimates than non-accredited laboratories, with an estimated mean of 4 $\log_{10}$ IU/ml.

Participants using other method of analysis tend to return higher estimates of viral load than participants using a single method of analysis (with an estimated mean of 4.58 $\log_{10}$ IU/ml).

Sample Group 3.5 $\log_{10}$ IU/ml Viral Load

The estimated mean for the results of participants under baseline conditions is 3.276 $\log_{10}$ IU/ml. bDNA users tend to return lower estimates of viral load than other commercial technologies. The closest estimated mean to the target viral load was found for the results of CC users.

Sample Group 3.2 $\log_{10}$ IU/ml Viral Load

Significant differences were found for the estimates of viral loads for samples of genotypes 3 and 5 with respect to samples of genotype 1. The estimated mean for samples of genotype 1 is 2.335 $\log_{10}$ IU/ml, whilst for samples of genotypes 3 and 5 the estimated means are 2.89 and 3.02 $\log_{10}$ IU/ml, respectively. For samples from 2003 and 2004 participants, the model tends to provide significantly higher estimates, which are closer to the target viral load, than for samples from 2005. The estimated means are around 3 $\log_{10}$ IU/ml for the results of participants from 2003 and 2004, and for the results of participants from 2005 the estimated mean is 2.335 $\log_{10}$ IU/ml).

Users of technologies different from the CC technology tend to provide lower estimates of viral loads than CC technology users.

Participants from reference laboratories and manufacture laboratories are more likely to return lower estimates of viral load, which are further away from the target viral load, than hospital laboratories.

Participants performing inhibition tests tend to provide lower estimates of viral load than participants not performing any inhibition test.

Sample Group 2.2 $\log_{10}$ IU/ml Viral Load

Estimates of viral load for samples from 2003 and 2004 were found to be lower than estimates for samples from 2005. The estimated mean for the results of participants from 2005 is 2.332 $\log_{10}$ IU/ml, whilst the estimated means for the results of participants from 2003 and 2004 are around 1.9 $\log_{10}$ IU/ml.

Participants using real time technologies tend to provide closer estimates of viral load to the target than users of all other technologies, under baseline conditions. The estimated mean for their results is close to 2.2 $\log_{10}$ IU/ml. The estimated means for the results of bDNA and CIH technologies users are higher than 2.6 $\log_{10}$ IU/ml.

5.6.2.2 Estimated Variances for the HCV Quantitative Data

The estimated variances for the distributions of the participants' results are obtained using the QTBM. Table 5.11 shows the estimated variances from the model per sample group and technology type. The QTBM takes into account different variances depending on the technology used by the participant, if it is a commercial or an in-house technology.

The estimated variances for the results of in-house technology users are higher than the estimated variances for the results of commercial technology users for almost all sample groups. However, for the weakest viral load the results of commercial technology users are more variable than results of in-house technology users. For the sample group of 3.2 $\log_{10}$ IU/ml both estimated variances are quite similar suggesting small differences for the variability of the results.

Table 5.11: *Summary statistics of the variance estimates from the reduced QTBM for HCV sample groups classified by technology type: estimated mean, standard deviation (SD), confidence interval.*

Estimated Variances per sample group	Mean	SD	95% Confidence Interval		
			2.50%	Median	97.50%
5.9 Log10 IU/ml					
In-house	0.373	0.182	0.128	0.337	0.830
Commercial	0.099	0.015	0.073	0.098	0.133
4.9 Log10 IU/ml					
In-house	0.622	0.136	0.407	0.605	0.934
Commercial	0.286	0.022	0.245	0.285	0.333
3.9 Log10 IU/ml					
In-house	0.561	0.135	0.355	0.542	0.875
Commercial	0.224	0.019	0.190	0.223	0.264
3.5 Log10 IU/ml					
In-house	0.567	0.229	0.263	0.520	1.145
Commercial	0.148	0.020	0.114	0.147	0.193
3.2 Log10 IU/ml					
In-house	0.136	0.049	0.064	0.129	0.256
Commercial	0.168	0.020	0.133	0.167	0.209
2.2 Log10 IU/ml					
In-house	0.187	0.099	0.063	0.164	0.445
Commercial	0.238	0.049	0.157	0.232	0.349

5.6.2.3 Model Checking

As in the previous section 5.5.5, the posterior predictive results are assessed for model consistency. With the use of a measure of discrepancy between the observed and replicated data, the goodness of fit of the model is studied. Formula 5.1 is used to obtain the standardised residuals as a measure of the discrepancy between replicated and observed data.

Figure 5.4 represents the density function of the probability that the observed standardised residuals are equal or exceed the predicted standardised residuals for the results of sample group 3.2 $\log_{10}$ IU/ml in 2004. It is observed that the distribution of the proportions are within the confidence interval of 0.025 and 0.975. Therefore, no discrepancy between the observed and replicated data was found.

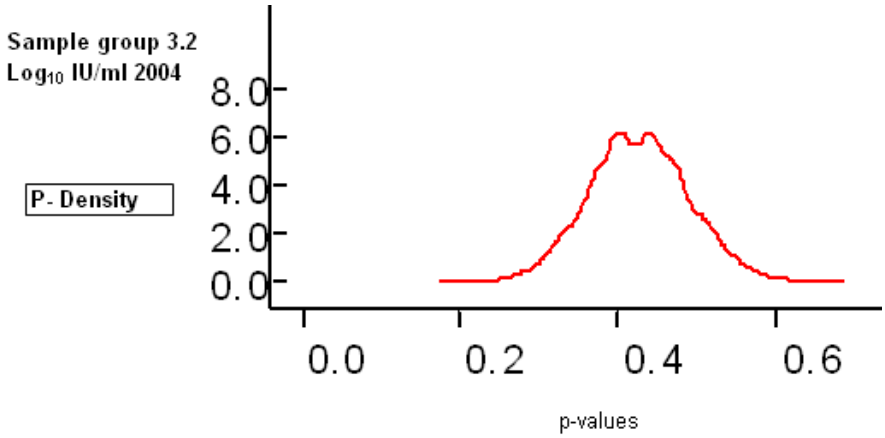


Figure 5.4: *Density function of the Bayesian p-values for sample group 3.2* $\log_{10}$ *IU/ml in 2004 from the reduced QTBM.*

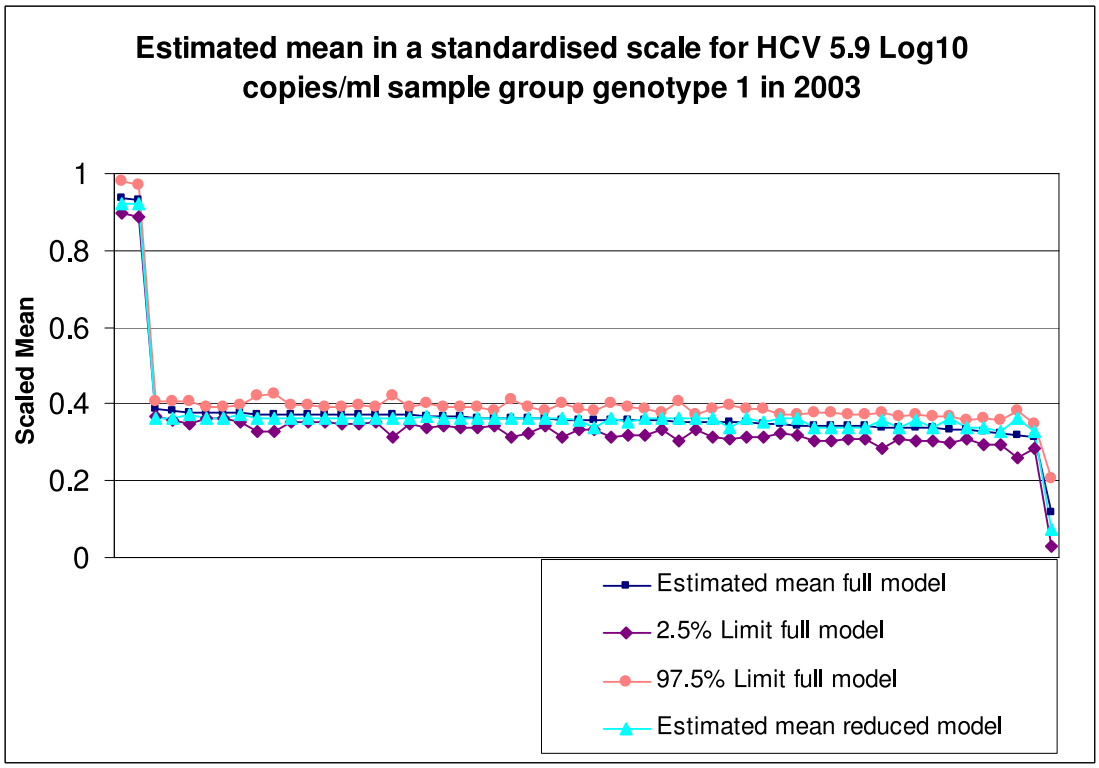
Table 5.12: *HCV mean and confidence intervals of Bayesian 'p-values' from the reduced QTBM per year and sample group.*

Test		Bayesian P-Values	95% Confidence Interval	
Year	Sample group		2.50%	97.50%
2002	5.9	0.485	0.340	0.640
	4.9	0.407	0.280	0.540
	3.9	0.423	0.330	0.510
	3.5	0.519	0.380	0.660
	3.2	0.405	0.267	0.533
	2.2	0.411	0.277	0.553
2003	5.9	0.568	0.446	0.679
	4.9	0.472	0.375	0.571
	3.9	0.465	0.339	0.589
	3.5	0.371	0.250	0.500
	3.2	0.426	0.278	0.574
	2.2	0.487	0.325	0.625
2004	5.9	0.512	0.400	0.629
	4.9	0.486	0.406	0.565
	3.9	0.491	0.386	0.600
	3.5	0.418	0.304	0.536
	3.2	0.427	0.304	0.551
	2.2	0.482	0.352	0.611
2005	4.9	0.485	0.407	0.565
	3.9	0.454	0.375	0.534
	3.2	0.401	0.325	0.482
	2.2	0.404	0.293	0.520

Table 5.12 shows the mean of the distribution for the probability of $T(y_{si}^{rep}, \mu_s, \sigma_{s_i}) \geq T(y_{si}, \mu_s, \sigma_{s_i})$, as well as the 95% confidence intervals per sample group and year. There is no extreme Bayesian 'p-values' indicating any disagreement between data and some aspects of the model for any of the sample groups and years. Therefore, it is concluded that the model fits the data appropriately.

5.6.2.4 Model Comparison

A model comparisons between nested models is carried out in the same way as described previously in Section 5.5.6. Figure 5.5 shows the plots of two randomly chosen samples. They show the estimated posterior scaled means for participants' results from the full model and their 95% confidence intervals together with the estimated posterior means from reduced model. The graph shows that the posterior means from the reduced model lie within the confidence intervals from the full model indicating that the reduced model fits the data appropriately. (For other plots see Figure C.2).



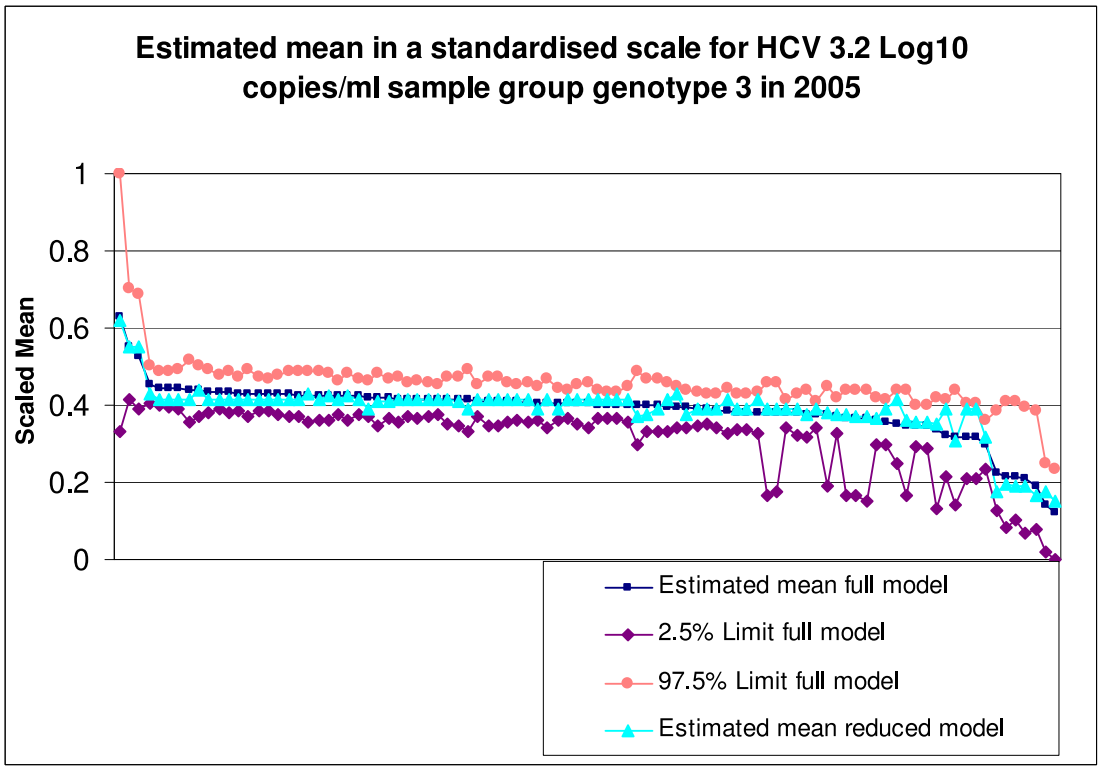


Figure 5.5: *HCV estimated means of sample viral load from the full and reduced QTBM for the sample group 3.2* $\log_{10}$ *IU/ml in 2005 and for the sample group 5.9* $\log_{10}$ *IU/ml in 2003. Results are ordered by decreasing estimated means. The x-axis represents an identification number of participants' results.*

5.6.2.5 Estimates of Censored Data of HCV

As in the HBV data analysis, the estimated posterior distribution for each censored observation is checked and used to obtain the mean of the distribution, which is inputed for the unknown censored observation.

Figure 5.6 shows the censored observations and their estimates obtained from the QTBM for sample groups 2.2 and 5.9 $\log_{10}$ IU/ml. It is observed that the variability of the estimated posterior means in comparison with the censored observations is higher for the weakest sample group. Their estimates are closer to the target viral load than the observed censored values. For the strongest sample group the estimates of the censored observations are close to their corresponding censored value and to the target viral load.

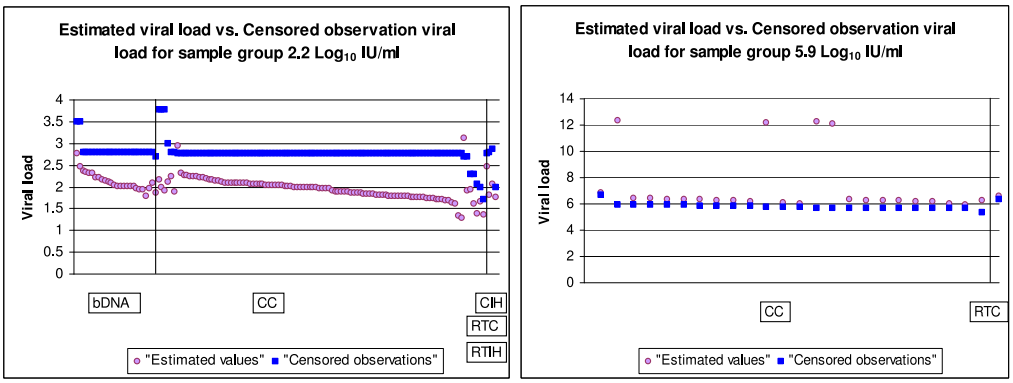


Figure 5.6: *Censored observations and their estimates for the HCV sample groups 2.2 and 5.9* $\log_{10}$ *IU/ml. Labs' codes are order by technology used as follows: bDNA; 27 data for the sample group 2.2* $\log_{10}$ *IU/ml, CC; 104 data for the sample group 2.2 and 24 for the sample group 5.9* $\log_{10}$ *IU/ml, CIH; 1 data for the sample group 2.2* $\log_{10}$ *IU/ml, RTC; 1 data for the sample group 2.2 and 1 for the sample group 5.9* $\log_{10}$ *IU/ml, RTIH; 2 data for the sample group 2.2 and 1 for the sample group 5.9* $\log_{10}$ *IU/ml.*

5.7 Model Sensitivity Analysis

As described and explained in Chapter 4, depending on the prior distributions defined for the model, other reasonable models may also provide a good fit to the data, but they may also lead to different conclusions. Therefore, in order to study the robustness of the model, a sensitivity analysis is performed for a range of alternative prior distributions.

As in Chapter 4, the following models with the alternative priors are used:

- Model 1 with Prior 1: 'Non-informative' or flat prior distributions, Beta(1,1), for the hyperparameters of the covariates to be estimated.
- Model 2 with Prior 2: A hierarchical model with hyperparameters from a flat uniform and inverse-gamma distribution for some of the primary parameters of interest.
- Model 3 with Prior 3: Full hierarchical model with non-informative distributions for the hyperparameters of the regression parameters to be estimated, and flat prior distributions for the hyperparameters of the covariates to be estimated.

5.7.1 Sensitivity Analysis of QTBM Applied to HBV Data

The three alternative models, described above, are applied to the HBV data and the results are compared to the original model.

No differences were found between the results obtained from the new models and the results from the QTBM with the prior distributions defined in Section 4.5.

As an illustrating example, Figure 5.7 shows the difference between the posterior estimated means from the three models and the estimated means from the QTBM for the sample group 4 $\log_{10}$ copies/ml in 2002. The differences are close to zero for each participant and may be attributed to the random nature of the data. Therefore, no incongruent results or inappropriate model fit is expected.

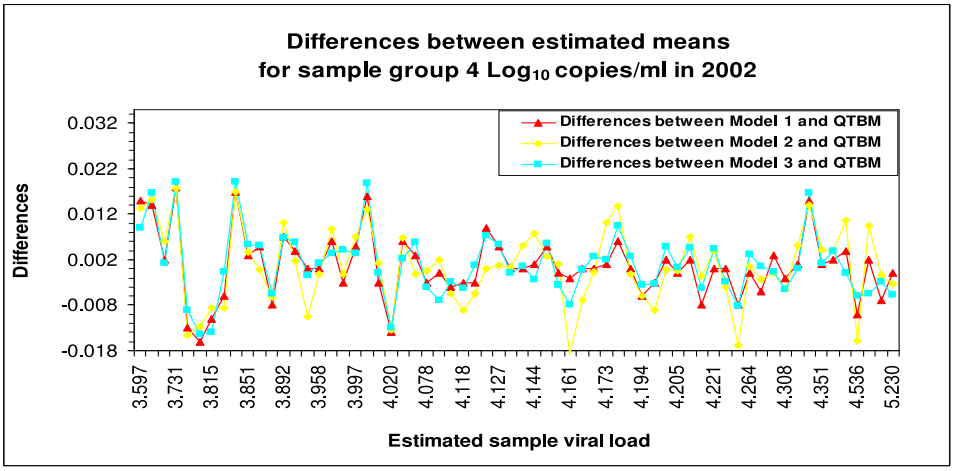


Figure 5.7: *Differences between HBV estimated means of viral load from the QTBM. The x-axis represents an identification number of participants' results.*

5.7.2 Sensitivity Analysis of QTBM Applied to HCV Data

Here, the robustness of the QTBM applied to the HCV data by a sensitivity analysis is checked. No differences between the estimated posterior means from the three alternative models and the estimated means from the QTBM were found. Figure 5.8 shows the differences between the estimated means from the alternative models and the estimated mean from the QTBM, for the sample group 3.9 $\log_{10}$ IU/ml in 2002.

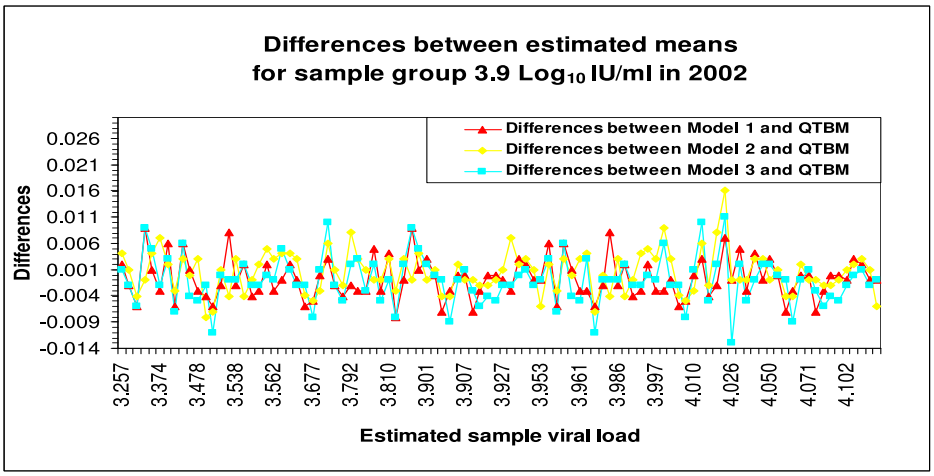


Figure 5.8: *Differences between HCV estimated means of viral load from the QTBM. The x-axis represents an identification number of participants' results.*

5.8 Summary and Conclusions

The proposed QTBM allows to identify factors that are associated with participant's performance of individual samples from a single pathogen across time. It also allows the inclusion of missing information and censored observations from participating laboratories. Thus, there is no need to discard information as in classical models. Furthermore, problems related to asymptotic theory that occur in classical models are avoided.

The QTBM developed and reported by García-Fernández et al. (2007) (see Appendix F) have been reviewed and adjusted to provide more appropriate results that explain the difference of performance in a more exhaustive way. The corresponding applications and results have been presented in this chapter.

The full QTBM applied to HBV and HCV provides a better overview of how the covariates influence the performance. It may help laboratories to improve their performance for future analysis by predicting the best procedure to take when analysing the samples, such as technology to be used, use of anti-contamination system, experience needed to improve their performance, etc. Furthermore, the reduced model provides estimates of the real impact of the significant covariates on the results and can be used for making future predictions. It also provides a measurement for the influence of the laboratory practices over its performance.

As a general overview, from the application of the QTBM to the HBV and HCV data, performance is shown to vary depending on the virus to be analysed and the sample load.

The best technologies used to test the samples from HCV data are CC and RTC, but for the HBV data the best technology varies depending on the sample viral load. Users of bDNA and HC perform poorly for lower strength sample groups, but their performance is better on samples with increased viral load.

The use of an anti-contamination system provides closer estimates of viral loads to the target for stronger samples of HBV, assuming other covariates are at their baseline levels; however it does not change the performance when HCV samples are tested. Differences of performance between laboratories were found when analysing some sample groups of HBV and HCV. Manufacture labs perform poorly for HCV samples, while for HBV samples these labs perform better.

Differences were found for participants performing inhibition tests for HBV sample groups, who in most cases provided better estimates than those participants not performing inhibition tests. However, for the HCV data, those differences appear only for the strongest and the 3.2 $\log_{10}$ IU/ml sample groups. In both cases the results from participants performing inhibition tests are further away from the target than the results for participants not performing them.

Experience of testing plasma and serum tests annually influences the performance on some sample groups for both HBV and HCV datasets. Given that these covariates behave as confounders in some situations, the interpretation about participants' performance depending on the experience of the laboratories should be done carefully.

The estimated variances for the responses of participants using different technologies (commercial or in-house) indicate that the variability is higher for the results of participants using in-house methods for all sample groups of HBV data. This is consistent with the classical analysis of variances that has been performed in this chapter. For the HCV data similar conclusions are obtained for sample groups of viral loads higher than 3.2 $\log_{10}$ IU/ml. However, for weaker sample groups, more variability was found in the results of commercial technology users than in the results of in-house technology users. Nevertheless, these differences were not substantial.

The estimates of the censored observations provide an exact value that can be used to obtain the participants' score for its performance (scoring system described in Chapter 3). The estimates provided from the QTBM take into account the laboratory practices. In some cases these estimates are close to the target viral load, but in other cases the estimates are further away from the target viral load. Participants that so far were excluded can now be scored.

Both model applications were checked for goodness of fit, and the results showed that the QTBM fitted the data appropriately. The prior probability distribution used for the data is a well known distribution chosen according to the characteristics and the definition of the data to be analysed. The sensitivity analysis proved the robustness of the QTBM under changes to prior knowledge.

The main advantage of the proposed QTBM is its capability to deal with censored observations and missing covariates. Since the developed model is derived from a Bayesian perspective, there is no need to have extra considerations about issues with multiple parameter testing and asymptotic theory when estimating parameters of the model. As in the previous chapter no continuous covariates were included in the model as none was collected in the QCMD questionnaire. However, the QTBM is able to be used with continuous covariates after some adjustment (García-Fernández et al., 2007).

As for the QLBM model, one assumption in the QTBM model is that the responses over time are independent, which reduces model complexity and computational time. However, participants can repeat programmes, and those results may be treated as repeated measure if the sample is the same over time. This aspect has not been taken into account, although the QTBM fits the data appropriately and replicated data are consistent with the observed one. Suggestions on how to approach this aspect can be found in Chapter 8. Furthermore, in one year a participant may return several results, and even if the technology used to analyse the sample is different, the laboratory practice and the technician may be the same. In this case, the responses may be correlated, which may lead to biased results. Suggestions about how to approach this can be found in Chapter 8.

One feature of the data is the censored observations provided by some participants. The QTBM updates this partially missing information by the use of probability distributions and provides a distribution of the best possible value to estimate the censored observation. Thus, the model has a very large number of parameters to estimate. On the other hand, the likelihood considers an estimation of the censored observation as an exact observation. So, the likelihood does not distinguish between exact results and censored results; once these are estimated, all of them contribute to the likelihood in the same way.

It is assumed for the QTBM that the variances are different across sample groups and type of technologies used (commercial or in-house). It has been shown that in some cases these differences may not be significant (classical analysis). Therefore, it may be more convenient to reduce the number of estimations within the model and consider a likelihood with different contributions for participants who provided censored observations than for participants who provided exact results.

Chapter 6

Improved Bayesian Model for Quantitative Responses and Pilot Study for its Application

The QTBM, described in Chapter 5, allows the inclusion of missing values and censored observations into the model. However, the QTBM has some limitations. The values of the censored observations are used to estimate exact responses which are used to replace those censored observations in the model. These estimates of the censored observations are then used to estimate the parameters of interest of the model. As a consequence of the contribution of censored information to the model via their estimates, the number of nuisance parameters increases the model's complexity and makes posterior predictive checks, model validation and model comparison difficult. Furthermore, the QTBM assumes that the censoring data mechanism can be ignored, but this is not always appropriate.

In this chapter an improved Bayesian model (Censored Bayesian Model-CBM) is proposed and developed where the contribution of the censored information to the CBM is included via its real observed censored value.

6.1 Introduction

Modelling censored data problems from a Bayesian perspective generally implies the inclusion of a probability distribution into the model that relates the censored information to the variables of interest (Chapter 5 for the QTBM). Thus, censored observations are estimated according to their posterior distributions conditional on the data and the current values for the unknown parameters.

In the QTBM, the censored observations are estimated as additional unknown parameters, which are simulated from the specified likelihood distribution given the current values for all relevant unknown parameters and the information provided by the censored observations. Then, assuming that the response variable follows a normal distribution, the censored observations are simulated from the normal distribution given the current value of the unknown parameters and based on the fact that the real unobserved responses are higher or lower than the observed censored responses (Gelman et al., 2004).

Usually, the prime objective is not to replace the censored observation by an estimate, but to study the relation of the complete data set to the covariates of interest. For this case the QTBM simulates an unnecessary number of parameters (censored observations).

A Censored Bayesian Model (CBM) that takes into account the information provided by the censored observations, and where the conditional posterior distributions of the parameters of interest do not depend on simulated values for the censored observations, is proposed in order to reduce the number of unnecessary estimations and implicitly, to reduce model complexity. The model is developed and coded in the statistical software R (the R code can be found on the CD attached to this thesis).

The inferences on the parameters from the CBM are based on a realistic likelihood function with the observed censored and non-censored information. Furthermore, the CBM has the capacity to incorporate a function that explains the censoring mechanism when this cannot be ignored. Additionally, a simulated value for the censored observation based on the current values of the estimated parameters and the exact observed data can be obtained from the CBM if required. In other words, the CBM can be used to obtain an estimate of the censored value, but this is not necessary to carry out model parameter estimations. This is in contrast with the QTBM that needs the estimate of the censored observation in order to carry out model estimation.

While the CBM is more appropriate and efficient, the new structure of the likelihood introduces a further difficulty since existing techniques for model comparison cannot be applied to the CBM. Thus, in this chapter some modifications of a well known model comparison technique are introduced and applied to the CBM.

The CBM is tested on a subgroup of datasets from QCMD's HBV programmes and on simulated data. Firstly, the results of the modified comparison tool are presented. Then, the results of the CBM application to the complete data set of HBV programmes are described, the chapter concluding with a summary of results and benefits gained from using the proposed model.

6.2 Censored $\log_{10}$-normal Model from a Bayesian Perspective

In this section the general theory of the CBM for the analysis of $\log_{10}$ transformed normal data is presented, allowing for censored responses in a Bayesian framework.

The CBM is a linear regression model allowing right and left censored responses, where the number of covariates may change depending on the observations. First, the notation used in this chapter is introduced. Then, the model derivation is shown and finally the estimation procedure is described.

6.2.1 Notation

The notation used throughout this chapter is as follows:

- Y_{si} is used to denote a quantitative response for the i^{th} observation, $i=1,...,n_s$, within the s^{th} sample group, $s=1,..,l$ with l being the total number of sample groups.

- υ_{si} is an indicator variable for the group of right censored values, with

$$\upsilon_{si} = \begin{cases} 0 & \text{if } y_{si} \text{ is right censored,} \\ 1 & \text{otherwise.} \end{cases}$$

- ρ_{si} is an indicator variable for the group of left censored values, with

$$\rho_{si} = \begin{cases} 0 & \text{if } y_{si} \text{ is left censored,} \\ 1 & \text{otherwise.} \end{cases}$$

- $\psi_{si} = \upsilon_{si}\rho_{si}$ is an indicator variable for the group of censored values, with

$$\psi_{si} = \begin{cases} 0 & \text{if } y_{si} \text{ is censored,} \\ 1 & \text{if } y_{si} \text{ is not censored.} \end{cases}$$

- $\vec{x}_{si} = (x_{si1}, .., x_{sip})$ denotes the p-dimensional vector of covariates for the i^{th} observation in sample group s with $x_{si1} = 1$.

- $\vec{\beta}_s$ is the p-dimensional vector of regression coefficients,
$\vec{\beta}_s^T = (\beta_{s1}, ..., \beta_{sp})$.

6.2.2 Likelihood Function

It is assumed that the $\log_{10}$ of the response variable, y_{si}, follows a normal distribution with mean μ_{si} and variance $\sigma^2_{s_i}$ with $s_i = 1, ..., q$ where q is the total number of different variances within the sample group s. As outlined in Chapter 5, models with different variances (e.g. accounting for different technology clusters) may be necessary. For HBV data, $s_i = 1$ refers to the group of participants' responses using in-house technologies and $s_i = 2$ refers to the group of participants' responses using commercial technologies for sample group s. In this case, two variances per sample group (target viral load) are considered, so $q = 2$ for each sample group s.

- Let Y_{si} follow a $\log_{10}$-normal distribution with parameters μ_{si} and $\sigma^2_{s_i}$. Then, the probability density function is given by

$$f(y_{si}|\mu_{si}, \sigma^2_{s_i}) = (2\pi)^{-1/2}(y_{si}\ln(10)\sigma_{s_i})^{-1}\exp\left\{-\frac{1}{2\sigma^2_{s_i}}(\log_{10}(y_{si}) - \mu_{si})^2\right\}.$$

Assuming independent censoring, a likelihood that appropriately takes into account the censored observation is proposed (formula (6.1)). The censored observation contributes to the likelihood through censored and probability distribution functions as follows

$$Likelihood = \prod_{si} f(y_{si})^{\psi_{si}}(G(y_{si})H(y_{si}))^{\psi_{si}}\left(S(y_{si})^{1-\upsilon_{si}}g(y_{si})^{1-\upsilon_{si}}F(y_{si})^{1-\rho_{si}}h(y_{si})^{1-\rho_{si}}\right)^{1-\psi_{si}}, \tag{6.1}$$

where $G(.)$ and $H(.)$ are the censoring distribution functions, while $g(.)$ and $h(.)$ are their corresponding censoring densities, respectively. $S(.) = 1 - F(.)$, where $F(.)$ is the cumulative distribution function for the response variable with density $f(.)$.

Since the censoring mechanism is assumed to be of no real interest in this application, only the partial likelihood is considered. Thus, the partial likelihood function of the full set of parameters (μ, σ^2) can be written as

$$L(\mu, \vec{\sigma}^2|y) = \prod_{si} f(y_{si}|\mu_{si}, \sigma^2_{s_i})^{\psi_{si}} \left(S(y_{si}|\mu_{si}, \sigma^2_{s_i})^{1-\upsilon_{si}} F(y_{si}|\mu_{si}, \sigma^2_{s_i})^{1-\rho_{si}}\right)^{1-\psi_{si}}. \tag{6.2}$$

6.2.3 Regression Model

In order to fit a linear regression model, $\mu_{si} = \vec{x}_{si}\vec{\beta}_s$, the model is rewritten as a function of $(\beta, \vec{\sigma}^2)$ with $\beta = (\vec{\beta}_1, ..., \vec{\beta}_l)$, where l is the total number of sample groups and $\vec{\sigma}^2 = (\sigma^2_1, ..., \sigma^2_m)$ with m the total number of variances in the model, i.e across all viral loads.

$$L(\beta, \vec{\sigma}^2|y) = \prod_{si} f(y_{si}|\vec{\beta}_s, \sigma^2_{s_i})^{\psi_{si}} \left((1 - F(y_{si}|\vec{\beta}_s, \sigma^2_{s_i}))^{1-\upsilon_{si}} F(y_{si}|\vec{\beta}_s, \sigma^2_{s_i})^{1-\rho_{si}}\right)^{1-\psi_{si}}$$

$$L(\beta, \vec{\sigma}^2|y) = \left((2\pi)^{-k/2} \left(\prod_{si} (\sigma^2_{s_i})^{-1/2}\right)\right)^{\psi_{si}} \exp\left\{-\frac{1}{2}\sum_{si}\left(\frac{\psi_{si}(\log_{10}(y_{si}) - \vec{x}_{si}\vec{\beta}_s)^2}{\sigma^2_{s_i}}\right)\right\}$$

$$\times \prod_{si}\left((y_{si}\ln(10))^{-\psi_{si}}\right)\left(\left(1 - \phi\left(\frac{\log_{10}(y_{si}) - \vec{x}_{si}\vec{\beta}_s}{\sigma_{s_i}}\right)\right)^{1-\upsilon_{si}}\left(\phi\left(\frac{\log_{10}(y_{si}) - \vec{x}_{si}\vec{\beta}_s}{\sigma_{s_i}}\right)\right)^{1-\rho_{si}}\right)^{1-\psi_{si}}. \tag{6.3}$$

In formula (6.3), k is the total number of non-censored observations and ϕ is the cumulative distribution function of the standard normal distribution for the $\log_{10}$ of the responses.

6.2.4 Bayesian Framework

The prior distributions for each $\vec{\beta}_s$ and missing covariate x_{sij} are considered the same as for the previous models QLBM and QTBM. However, the prior distribution for $\sigma^2_{s_i}$ and its hyperparmeters are defined as follows

- Prior distribution for $\sigma^2_{s_i}$ with $s_i = 1, ..., q$, where

$$\sigma^2_{s_i} | c_{s_i}, d_{si} \sim InvGamma(c_{s_i}, d_{si}).$$

The hyperparameters c_{s_i} and d_{s_i}, in the subsequent application, are set up as 3 and 1 in order to provide the variances $\sigma^2_{s_i}$ with a proper inverse-gamma prior distribution with expected variability of responses of 0.5 and existing variances (this facilitates the starting values of the MCMC chain and improves convergence without assuming strong prior knowledge). The hyperparameters for each $\vec{\beta}_s$ and missing covariate x_{sij} are defined as in previous models.

6.2.4.1 Posterior Distributions

Considering the partial likelihood (formula 6.3) and the prior distributions from the previous section, the posterior distribution for the parameters given the data is:

$$\pi(\beta, \vec{\sigma}^2 | y) \propto likelihood \times prior \propto L(\beta, \vec{\sigma}^2) \pi(\beta | \beta_0, V_0) \pi(\vec{\sigma}^2 | c, d).$$

Since $\vec{\beta}_s$, σ_{s_i} and x_{sij}, with $s = 1, .., l$, $s_i = 1, .., q$ and $j = 2, .., r$, are independent, then, the posterior distribution of each parameter can be expressed in terms of its conditional posterior distribution, which is described below:

- Conditional posterior distribution for $\vec{\beta}_s$ with $s = 1, .., l$:

$$\pi(\vec{\beta}_s | y, \vec{\sigma}^2, \vec{\beta}_{-s}) \propto L(\vec{\beta}_s, \sigma^2_{s_i}) \pi(\vec{\beta}_s | \vec{\beta}_{s0}, \mathbf{V}_{s0})$$

$$\pi(\vec{\beta}_s | y, \vec{\sigma}^2, \vec{\beta}_{-s}) \propto \exp\left\{ -\frac{1}{2} \sum_{i_{(s)}} \left(\frac{\psi_{si} (\log_{10}(y_{si}) - \vec{x}_{si} \vec{\beta}_s)^2}{\sigma^2_{s_i}} \right) \right\}$$

$$\times \prod_{i_{(s)}} \left(\left(1 - \phi \left(\frac{\log_{10}(y_{si}) - \vec{x}_{si} \vec{\beta}_s}{\sigma_{s_i}} \right) \right)^{1 - \upsilon_{si}} \left(\phi \left(\frac{\log_{10}(y_{si}) - \vec{x}_{si} \vec{\beta}_s}{\sigma_{s_i}} \right) \right)^{1 - \rho_{si}} \right)^{1 - \psi_{si}}$$

$$\times \exp\left\{-\frac{1}{2}(\vec{\beta}_s - \vec{\beta}_{s0})^T \mathbf{V}_{s0}^{-1}(\vec{\beta}_s - \vec{\beta}_{s0})\right\},$$

where $i_{(s)}$ is the subgroup of responses from participants of sample group s.

- Conditional posterior distribution of $\sigma^2_{s_i}$ with $s_i = 1, .., q$:

$$\pi(\sigma^2_{s_i}|y, \beta, \sigma^2_{-s_i}) \propto L(\beta, \sigma^2_{s_i})\pi(\sigma^2_{s_i}|c_{s_i}, d_{s_i})$$

$$\pi(\sigma^2_{s_i}|y, \beta) \propto (2\pi\sigma^2)^{\frac{-k_{(\sigma^2_{s_i})}}{2}} \exp\left\{-\frac{1}{2\sigma^2_{s_i}} \sum_{i_{(\sigma^2_{s_i})}} (\psi_{si}(log_{10}(y_{si}) - \vec{x}_{si}\vec{\beta}_s)^2)\right\}$$

$$\times \prod_{i_{(\sigma^2_{s_i})}} \left(\left(1 - \phi\left(\frac{\log_{10}(y_{si}) - \vec{x}_{si}\vec{\beta}_s}{\sigma_{s_i}}\right)\right)^{1-\upsilon_{si}} \left(\phi\left(\frac{\log_{10}(y_{si}) - \vec{x}_{si}\vec{\beta}_s}{\sigma_{s_i}}\right)\right)^{1-\rho_{si}}\right)^{1-\psi_{si}}$$

$$\times (\sigma^2_{s_i})^{-c_{s_i}-1} exp\left\{-\frac{d_{s_i}}{\sigma^2_{s_i}}\right\}$$

$$\pi(\sigma^2_{s_i}|y, \beta) \propto \prod_{i_{(\sigma^2_{s_i})}} \left(\left(1 - \phi\left(\frac{\log_{10}(y_{si}) - \vec{x}_{si}\vec{\beta}_s}{\sigma_{s_i}}\right)\right)^{1-\upsilon_{si}} \left(\phi\left(\frac{\log_{10}(y_{si}) - \vec{x}_{si}\vec{\beta}_s}{\sigma_{s_i}}\right)\right)^{1-\rho_{si}}\right)^{1-\psi_{si}}$$

$$\times \exp\left\{-\frac{1}{\sigma^2_{s_i}}\left(d_{s_i} + \frac{1}{2}\sum_{si_{(\sigma_{s_i})}} (\psi_{si}(\log_{10}(y_{si}) - \vec{x}_{si}\vec{\beta}_s)^2)\right)\right\} \times (\sigma^2_{s_i})^{-1-\left(\frac{k_{(\sigma^2_{s_i})}}{2}+c_{s_i}\right)},$$

where $i_{(\sigma^2_{s_i})}$ is the group of observations having the same variance $\sigma^2_{s_i}$ and $k_{(\sigma^2_{s_i})}$ is the number of non-censored observations with variance $\sigma^2_{s_i}$.

- Conditional posterior distribution for x_{sij} with $j = 2, ..., r$

  - For missing binary variables

$$\pi(x_{sij}|y_{si}, x_{-sij}, \beta_{sj}, \sigma^2_{s_i}) \propto L(\beta_{sj}, x_{sij}, \sigma^2_{s_i})\pi(x_{sij}|b_j)$$

$$\pi(x_{sij}|y_{si}, x_{-sij}, \beta_{sj}, \sigma^2_{s_i}) \propto \exp\left\{-\frac{1}{2}\left(\frac{\psi_{si}(\log_{10}(y_{si}) - x_{sij}\beta_{sj})^2}{\sigma^2_{s_i}}\right)\right\}$$
$$\times\left(\left(1-\phi\left(\frac{\log_{10}(y_{si}) - x_{sij}\beta_{sj}}{\sigma_{s_i}}\right)\right)^{1-\upsilon_{si}}\left(\phi\left(\frac{\log_{10}(y_{si}) - x_{sij}\beta_{sj}}{\sigma_{s_i}}\right)\right)^{1-\rho_{si}}\right)^{1-\psi_{si}}$$
$$\times(b_j)^{x_{sij}}(1-b_j)^{(1-x_{sij})}.$$

  - For missing categorical variables

$$\pi(x_{sij}|y_{si}, x_{-sij}, \beta_{sj}, \sigma^2_{s_i}) \propto L(\beta_{sj}, x_{sij}, \sigma^2_{s_i})\pi(x_{sij}|g_j[])$$

$$\pi(x_{sij}|y_{si}, x_{-sij}, \beta_{sj}, \sigma^2_{s_i}) \propto \exp\left\{-\frac{1}{2}\left(\frac{\psi_{si}(\log_{10}(y_{si}) - x_{sij}\beta_{sj})^2}{\sigma^2_{s_i}}\right)\right\}$$
$$\times\left(\left(1-\phi\left(\frac{\log_{10}(y_{si}) - x_{sij}\beta_{sj}}{\sigma_{s_i}}\right)\right)^{1-\upsilon_{si}}\left(\phi\left(\frac{\log_{10}(y_{si}) - x_{sij}\beta_{sj}}{\sigma_{s_i}}\right)\right)^{1-\rho_{si}}\right)^{1-\psi_{si}}\times g_j[x_{sij}],$$

    where x_{-sij} are all the covariate values except the observed j^{th} covariate value for the i^{th} observation in sample group s and $\vec{g}_j[]$ is the vector of assigned probabilities to each category of the variable (see Chapter 5 in Section 5.4 for more details).

6.2.5 MCMC and Simulation of Posterior Distributions

In this subsection, the well known algorithm used to fit the CBM is described. In order to obtain samples from the multivariate posterior distribution of the joint parameters β and $\vec{\sigma}^2$, samples from their conditional posterior distributions are simulated separately.

The MCMC algorithm for the censored $\log_{10}$-normal model applied to HBV data makes use of the Metropolis-Hastings (M-H) algorithm with multivariate random walk proposals. The Metropolis-Hastings algorithm is a technique to simulate iteratively a sample chain where at each iteration step a new 'candidate' value is proposed and this value is accepted with an appropriate probability (as a rule of thumb: minimum acceptance probability 24% (Gamerman and Lopes, 2006)).

The M-H algorithm is an iterative procedure given by the following steps:

1. Initialise the chain with some starting values $(\vec{\beta}_1^0, .., \vec{\beta}_l^0, (\sigma_1^2)^0, .., (\sigma_m^2)^0)$.

2. Update the parameters as follows:

   Draw a candidate $\vec{\beta}_1^{can}$ from a multivariate normal distribution with the mean being the current value of the chain and a covariance matrix given by the inverse of the curvature of the log of the normal distribution at its mode multiplied by a predefined tuning factor. The covariance estimate is pre-calculated, at the beginning of the iteration process, based on the non-censored observations.
   The next value in the chain, $\vec{\beta}_1^1$, is actualised by $\vec{\beta}_1^0$ with probability $1-p$ or by $\vec{\beta}_1^{can}$ with probability p, where

$$p = \min\left\{1, \frac{\pi(\vec{\beta}_1^{can}|\vec{\beta}_2^0, ..\vec{\beta}_l^0, (\sigma^2)^0, y)}{\pi(\vec{\beta}_1^0|\vec{\beta}_2^0, ..\vec{\beta}_l^0, (\sigma^2)^0, y)} \frac{\xi(\vec{\beta}_1^0|\vec{\beta}_1^{can}, \vec{\beta}_2^0, ..\vec{\beta}_p^0, (\sigma^2)^0, y)}{\xi(\vec{\beta}_1^{can}|\vec{\beta}_1^0, \vec{\beta}_2^0, ..\vec{\beta}_p^0, (\sigma^2)^0, y)}\right\}$$

   and $\xi(.)$ being the candidate generator density, in this case having a multivariate normal density with mean equal to the current value of the chain. Thus, the probability p given in the previous formula is reduced to:

$$p = \min\left\{1, \frac{\pi(\vec{\beta}_1^{can}|\vec{\beta}_2^0, ..\vec{\beta}_p^0, (\sigma^2)^0, y)}{\pi(\vec{\beta}_1^0|\vec{\beta}_2^0, ..\vec{\beta}_p^0, (\sigma^2)^0, y)}\right\}.$$

Repeating the above process for $\vec{\beta}_s^1$ and $\vec{\beta}_s^{can}$ for each s, with $s = 2, .., l$, at the current values of the parameters, a simulated vector of regression parameters for each category s is obtained.

3. Update the parameter $(\sigma_1^2)^1$ using a similar procedure as for the regression parameters in Step 2. Here, the proposal distribution is a normal distribution. The proposed candidate is accepted with probability p if it is bigger than zero (truncated normal distribution from zero since the variances can only take on positive values). Repeat the procedure to update $(\sigma_{s_i}^2)^1$ with $s_i = 2, .., q$.

4. Iterate the updating procedure from Step 2.

The simulation of $\vec{\beta}_s$-vectors per block of sample group breaks the correlation between the parameters of sample blocks. This improves convergence, since in high-dimensional models simulating individual parameters can slow down the convergence (Gamerman and Lopes, 2006).

6.3 Pilot Study of the CBM

The main aim of the pilot study, where the CBM has been applied to the HBV data with a subgroup of covariates, is to check the results from the proposed model and study the variability assumption of the data. This study was performed via a representative reduced dataset from the HBV full data in order to save time and computational efforts (in Chapter 7 the application of the CBM to the full HBV data is shown).

For practicality, selected covariates without missing values (year, technology group and viral subtype) are chosen to test the model in this pilot study. Models are fitted to all sample groups using censored and non-censored observations from participants' viral load estimates over time. An exploratory analysis of the differences in technology cluster variances using a classical statistical test has been carried out in Chapter 5, and significant differences between

variances have been found for some sample groups. Therefore, two models are fitted to the selected subgroup of HBV data; one with two variances for all six target viral loads and the second allowing for two variances within each viral target load. Equal variances per year and within technology group are assumed.

The covariates used in these models are:

- Sample subtype: this binary variable takes on the value 0 if the data are from a sample of subtype A and value 1 if the data are from a sample of subtype D. The reason for the inclusion of this variable into the models is to study significant differences in the performance of samples with different subtypes.
- Sample year: for convenience, to simplify the pilot study and to test the model with non-categorical covariates, year is considered as a continuous variable. Year takes values from 1 to 4 corresponding to the years 2002 to 2005. A linear relationship over time is assumed to check for a trend in the performance over time.
- Technology group: this variable identifies the technology used for analysing the sample. It consists of five indicator variables comparing bDNA, RTC, HC, CIH, RTIH using CC as a baseline.

The $\vec{\beta}_s$ is a 8-dimensional vector of regression coefficients for those sample groups that contain samples with both subtypes A and D. However, for those sample groups containing only one kind of subtype, a 7-dimensional vector of regression coefficients is used. The sample groups are defined as in Chapter 2.

6.3.1 Results from the Model with two Variances

The CBM with two variances is considered in this section. First, convergence is checked, then autocorrelations are studied and finally the results are discussed.

6.3.1.1 Convergence and Autocorrelations

Chains of parameter estimations are run until convergence is achieved. Only the parameter estimations, the convergence of which is achieved, are used for the analysis. The convergence of the results is achieved to ensure that the draws obtained can be used as a sample from the posterior distribution. Convergence is assessed using graphical diagnostic tools, by plotting the cumulative posterior means for the parameters against the number of iteration of the chain.

Figure 6.1 illustrates the convergence for some selected parameters in the model. The graph in the upper-left corner is showing the chain of the cumulative posterior means of the parameter estimates for the intercept. The corresponding graph appears to stabilize around 2,000 iterations. Therefore, the first 2,000 iterations are considered as a burn-in period and the rest of iterations are used as samples from the posterior distribution.

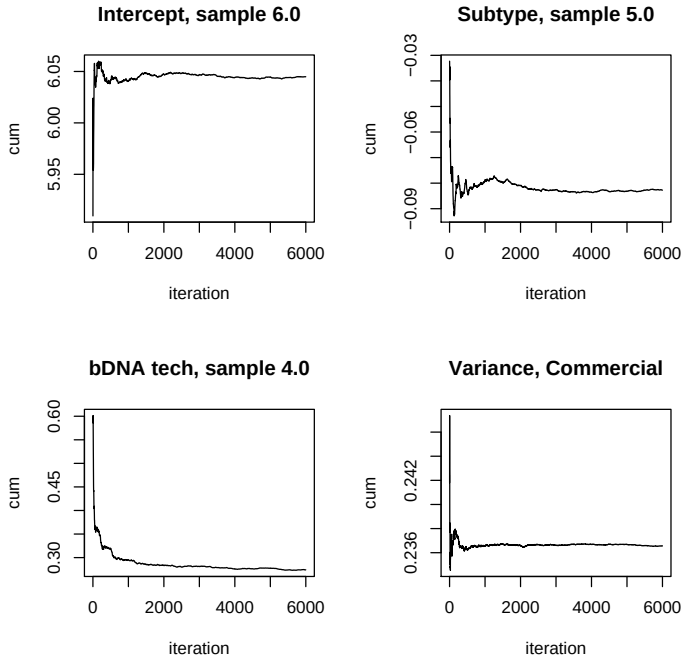


Figure 6.1: *Plots of posterior means of parameters against the number of iterations. The vertical axis represents the cumulative posterior means of the parameter (cum).*

Figure 6.2 shows the autocorrelation of the unfiltered chain for some selected parameters. The time lag of the autocorrelation function is decreasing indicating a good mixing of the chain. The autocorrelations are approximately zero after a time lag of 10. Since the simulated values of the chain should be independent realisations from the posterior distributions, a thinning period to break the autocorrelation is needed (see Section 1.4). The figure suggests that a thinning period of 10 iterations is appropriate for breaking autocorrelations between consecutive simulated values. That is, every 10^{th} observation is taken from the simulated chain and used for drawing conclusions from the posterior distribution.

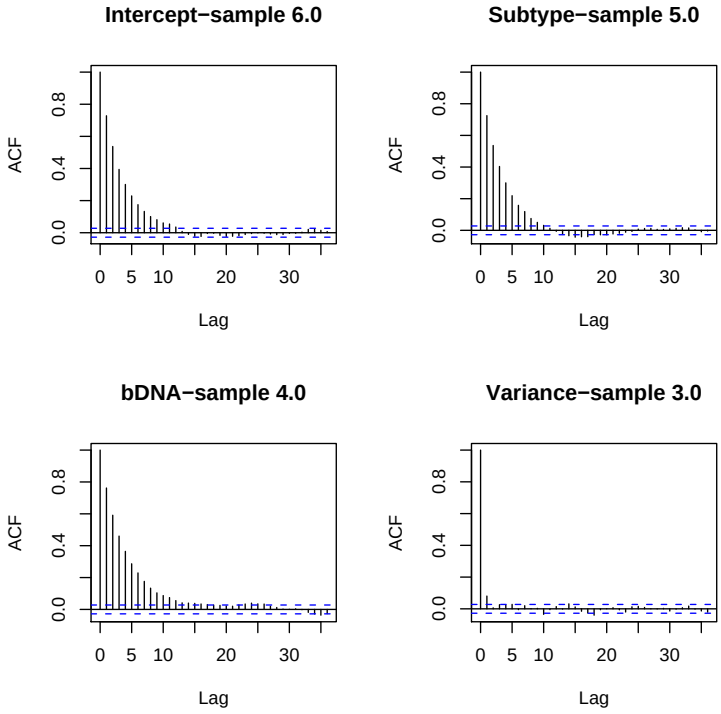


Figure 6.2: *Autocorrelations plots of the chain for four selected parameters.*

Figure 6.3 illustrates the correlation between some of the selected parameters. Correlations for two parameters would be present if the pairs follows a particular pattern, as for example an increasing or decreasing line. Since the plots show a random scattering of parameter pair-values, there is no clear evidence of presence of correlation between pairs of parameters.

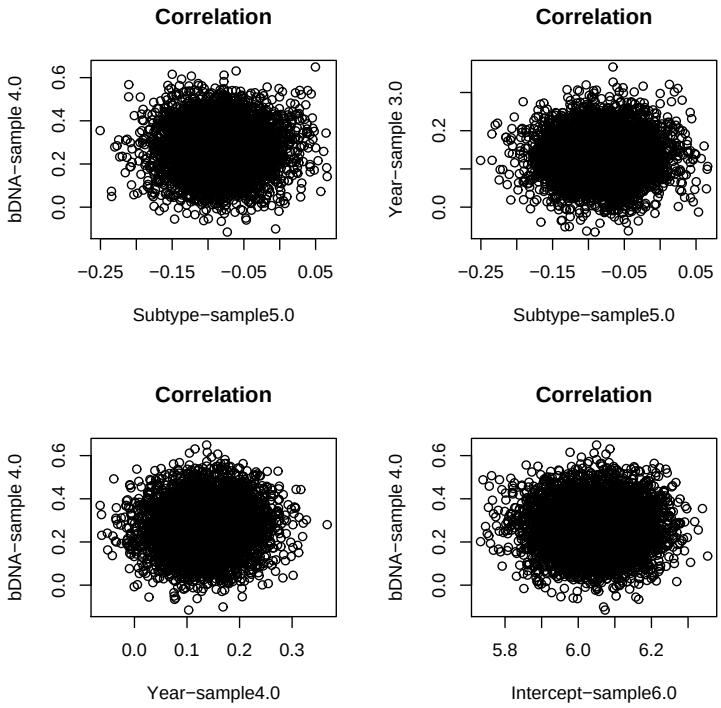


Figure 6.3: *Scatter plots to show correlation between two selected parameters.*

The chain is run for 60,000 iterations with a thinning period of 10 iterations to break autocorrelation between simulated values. Therefore, 6,000 simulated values are kept. Convergence is achieved after approximately 2,000 iterations. A total of 4,000 simulated values are used to estimate the posterior distribution function.

6.3.1.2 Numerical Results

Table 6.1 to 6.6 illustrate the estimated posterior means, standard deviations and probabilities for the regression coefficients for each sample group. The posterior probability is calculated, first, by summing all of the simulation values obtained from the conditional posterior distribution which are higher than zero. Then, the sum is divided by the total number of iterations. If this posterior probability is close to zero (smaller than 0.025), zero is in the upper tail of the distribution. In this situation, it is concluded that parameter is significant (at the one-sided 2.5% significant level). If the posterior probability is close to 1 (higher than 0.975), zero is on the lower tail of the distribution, and it is concluded that parameter is significant (at the one-sided 2.5% significance level).

Numerical results obtained for the posterior probabilities can be summarised for each sample group as follows:

- Sample group 6 $\log_{10}$ copies/ml:

  The variable year is significant at the one-sided 2.5% level. In other words, the estimated means for the participants' results decrease significantly over time. The results from the row data in Table 5.3 show the observed means which decrease from 6.032 to 5.532 for samples from 2002 to 2005. Both bDNA and RTIH users are significantly more likely to provide higher estimates of viral loads than CC users.

Table 6.1: *Table with posterior means, standard deviations and probabilities for the regression parameters from the CBM corresponding to sample group 6* $\log_{10}$ *copies/ml.*

Posterior	Intercept	Year	Subtype	CIH	RTIH	RTC	bDNA	HC
Mean	6.045	-0.095	-0.062	-0.004	0.239	0.253	0.377	0.015
SD	0.091	0.030	0.069	0.178	0.103	0.165	0.113	0.095
Prob	1.000	0.001	0.187	0.487	0.988	0.937	0.999	0.566

- Sample group 5 $\log_{10}$ copies/ml:

  There is a significant negative linear trend over time. Users of bDNA technology are significantly more likely to return higher estimates of viral loads than CC users.

However, CIH users are more likely to return lower estimates of viral loads than CC users. Estimates of viral loads for samples of subtype A tend to be significantly lower than estimates for samples of subtype D, at the one-sided 3% level.

Table 6.2: *Table with posterior means, standard deviations and probabilities for the regression parameters from the CBM corresponding to sample group 5* $\log_{10}$ *copies/ml.*

Posterior	Intercept	Year	Subtype	CIH	RTIH	RTC	bDNA	HC
Mean	5.185	-0.067	-0.083	-0.301	0.007	-0.092	0.303	0.157
SD	0.065	0.021	0.043	0.126	0.071	0.063	0.084	0.132
Prob	1.000	0.001	0.029	0.009	0.539	0.070	0.999	0.880

- Sample group 4 $\log_{10}$ copies/ml:

  The estimates of viral loads for samples of subtype D tend to be lower than the estimates for samples of subtype A. Estimates of viral loads from participants using bDNA and CIH tend to be lower than estimates from CC users.

Table 6.3: *Table with posterior means, standard deviations and probabilities for the regression parameters from the CBM corresponding to sample group 4* $\log_{10}$ *copies/ml.*

Posterior	Intercept	Year	Subtype	CIH	RTIH	RTC	bDNA	HC
Mean	4.147	-0.04	-0.146	-0.406	0.063	-0.077	0.269	-0.129
SD	0.089	0.031	0.072	0.161	0.088	0.070	0.108	0.354
Prob	1.000	0.096	0.023	0.006	0.765	0.144	0.995	0.364

- Sample group 3.5 $\log_{10}$ copies/ml:

  Estimates of viral loads are significantly different over time. The estimates decrease over time. RTIH and bDNA users tend to provide significantly higher estimates of viral loads than CC users.

Table 6.4: *Table with posterior means, standard deviations and probabilities for the regression parameters from the CBM corresponding to sample group 3.5* $\log_{10}$ *copies/ml.*

Posterior	Intercept	Year	CIH	RTIH	RTC	bDNA	HC
Mean	4.480	-0.587	-0.201	0.521	0.189	1.314	0.174
SD	0.124	0.090	0.193	0.125	0.251	0.205	0.369
Prob	1.000	0.000	0.149	1.000	0.780	1.000	0.695

- Sample group 3 $\log_{10}$ copies/ml:

  Estimates of viral loads for samples of subtype D tend to be higher than the estimates for samples of subtype A. RTC users are more likely to provide significantly lower estimates of viral loads than CC users at the one-sided 4% level. Also, a significant difference is found between the estimates of viral loads from HC and CC technology users.

Table 6.5: *Table with posterior means, standard deviations and probabilities for the regression parameters from the CBM corresponding to sample group 3* $\log_{10}$ *copies/ml.*

Posterior	Intercept	Year	Subtype	CIH	RTIH	RTC	bDNA	HC
Mean	3.324	-0.092	0.133	-0.085	0.137	-0.125	0.059	1.104
SD	0.135	0.042	0.057	0.167	0.089	0.069	0.106	0.332
Prob	1.000	0.013	0.989	0.301	0.938	0.036	0.719	0.999

- Sample group 2.3 $\log_{10}$ copies/ml:

  RTIH and HC technologies users tend to return significantly higher estimates of viral loads than CC users. The estimated mean for the results of bDNA users has the highest estimated standard deviation amongst the results from other technology and sample groups. In Chapter 2, it has been shown that all the data provided by bDNA technology users for this sample group are censored, so this may be the reason why the estimated mean has a high standard deviation.

Table 6.6: *Table with posterior means, standard deviations and probabilities for the regression parameters from the CBM corresponding to sample group 2.3* $\log_{10}$ *copies/ml.*

Posterior	Intercept	Year	CIH	RTIH	RTC	bDNA	HC
Mean	2.423	-0.043	0.283	0.542	0.228	-0.811	2.082
SD	0.115	0.052	0.217	0.127	0.138	0.626	0.323
Prob	1.000	0.205	0.904	1.000	0.950	0.088	1.000

Table 6.7 shows the means, standard deviations and probabilities for the estimated variances. The estimated variance for the results given by in-house technologies is approximately three times the estimated variance for the results from users of commercial technologies. This indicates a difference of the variability of results from the two groups of participants.

However, it is not clear if the additional variability is due to change of participants' results or change of different sample groups. Thus, the second CBM fitted to the data, in the next section of this chapter, takes into account differences in the variability not only depending on the technology type, but also on the sample group.

Table 6.7: *Posterior means, standard deviations and probabilities for the variances.*

Variance	Mean	SD	Prob
In-house	0.604	0.035	1
Commercial	0.236	0.009	1

To conclude this section, Figure 6.4 illustrates some simulations from the conditional posterior distribution of two selected regression parameters and their histograms. The sampling paths in the upper row of the Figure 6.4 show that the chain mixes well, while the corresponding histograms indicate a normal approximation for the posterior distribution of the parameters.

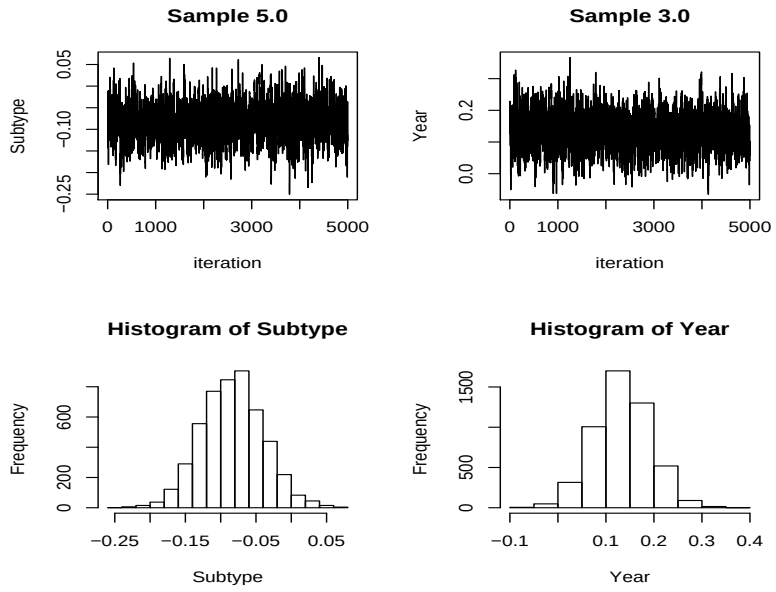


Figure 6.4: *Plots of the simulations and histograms from the conditional posterior distribution for two selected parameters.*

6.3.1.3 Table of Significance

Table 6.8 summarises the results described previously. It shows the two-sided significance levels for each covariate and sample group obtained from the application of the CBM to the subgroup of HBV data.

Table 6.8: *Table of significances of the parameter estimates from the CBM with two variances.*

Sample group	Intercept	Year	Subtype	bDNA	RTC	HC	CIH	RTIH
Sample 6.0	***	***	-	***	-	-	-	***
Sample 5.0	***	***	*	***	-	-	***	-
Sample 4.0	***	-	**	***	-	-	***	-
Sample 3.5	***	***		***	-	-	-	***
Sample 3.0	***	***	***	-	**	***	-	-
Sample 2.3	***	-		-	*	***	-	***

*** 2.5% level, ** 5% level, * 10% level and - not significant at 10% two-sided level.
Blank cells: no parameter to estimate.

6.3.2 Results from the Model with Twelve Variances

Following the findings in the exploratory analysis, an extended model with twelve variances is fitted allowing for different variances for in-house and commercial technology users within each target viral load.

The estimated results obtained from this model are similar to those from the previous model. Convergence is achieved after 1,000 iterations. The chain is run for 60,000 iterations with a thinning period of 10 and no presence of correlations between parameters is detected.

Regression parameters are interpreted as in the previous section for the model with two variances. There are some slight differences with respect to the previous model when testing at the two-sided 5% level. Table 6.9 shows a summary of significances for the estimated parameters from the model with twelve variances.

Table 6.9: *Table of significances of the parameter estimates from the CBM with twelve variances.*

Sample group	Intercept	Year	Subtype	bDNA	RTC	HC	CIH	RTIH
Sample 6.0	***	***	-	***	-	-	-	**
Sample 5.0	***	***	*	***	-	*	***	-
Sample 4.0	***	-	**	***	-	-	***	-
Sample 3.5	***	***		***	-	-	-	***
Sample 3.0	***	***	***	-	*	***	-	*
Sample 2.3	***	-		-	*	***	*	***

*** 2.5% level, ** 5% level, * 10% level and - not significant at 10% two-sided level.
Blank cells: no parameter to estimate.

Table 6.10 shows the estimated posterior means of variances for the distributions of participants' results for each sample group. This illustrates the differences in the variations between the participants' performance for these two major technology clusters. As expected, the variability of results from commercial technology users is lower than the variability of responses from in-house technology users for each sample group.

Table 6.10: *Estimated posterior means of variances for each sample group.*

	Sample group $\log_{10}$ copies/ml					
Variances	6	5	4	3.5	3	2.3
In-house	0.676	0.766	0.606	0.425	0.458	0.380
Commercial	0.407	0.209	0.216	0.255	0.129	0.232

6.3.3 Model Diagnostics

In this section the goodness of fit of the model with two variances is checked. The standardised residuals are used as a model diagnostic tool. In a Bayesian framework the residual for each observation is a random variable since the parameters have probability distributions. A sample from the residual distribution for each data value is calculated based on the posterior parameter samples obtain from the MCMC.

Table 6.11 shows the 5% - lower and 95% - upper limit of the empirical 2.5% and the 97.5 quantile of the standardised residuals for each sample group. For each observation, the empirical 2.5% and 97.5% quantile are obtained from the distribution of the standardised residuals. Then, from the range of all empirical 2.5% quantiles among the observations of a sample group, the cut off point for the 5% quantile is calculated. From the range of all empirical 97.5% quantiles among the observations of a sample group, the cut off point for the 95% quantile is obtained. Thus, there is a lower and upper value cut off for the residuals of the 95% of the data for each sample group. Note that it is not assumed that the distributions of the quantiles follow normal distributions. However, those quantiles should be within the range given by (-3, 3) for a well fitting model.

For sample group 2.3 $\log_{10}$ copies/ml the residuals are far outside of the acceptance limits. This sample group corresponds to the one with the highest proportion of censored data. Therefore, the variation in the standardised residuals may be due to the effect of censored data. For the other sample groups the quantiles of the residual values lie within the upper and lower limits (-3, 3) indicating a good fit of the model. Similar results are obtained with the use of the CBM with twelve variances.

Table 6.11: *Summary of standardised residual.*

Standardised residuals per sample group $\log_{10}$ copies/ml						
Limits	6	5	4	3.5	3	2.3
5% low limit	-1.841	-1.855	-1.782	-1.991	-1.438	-1.317
95% upper limit	1.944	1.692	2.039	1.703	2.435	3.618

6.4 Conclusions

In order to improve the analysis of quantitative data from EQA programmes carried out with the QTBM, a new model has been proposed which has several advantages over the QTBM. The new CBM uses the real censored observation along the model estimation procedure, instead of an estimated value for the censored observation. This reduces the number of nuisance parameters estimated by the QTBM model, saving time and reducing complexity. Since the estimation of the parameters of interest from the model is based on the observed censored value, the results from the CBM are more reliable and robust than the results from the QTBM (as in the QTBM the censored values are estimates of the observed ones). Therefore, predictions and conclusions obtained from the CBM are more consistent with the real data from EQA programmes.

The proposed CBM provides a more appropriate, objective and accurate model to identify factors significantly associated with participants' quantitative performance than the QTBM developed in Chapter 5. The CBM is a step forward to improve the QTBM, as the former takes into account the censored information by using a censored function without having to estimate a value for each censored observation. Therefore, the CBM treats the data information in a more objective, exact and accurate way. Although the censoring mechanism in this application has been assumed to be ignorable and not of real interest, the CBM can be adjusted to any censoring mechanism. Thus, the CBM is a model fulfilling the requirement of flexibility needed to apply it to future EQA data. Furthermore, the CBM can be applied to data from other areas of research with censored observations such as economy or finance.

The pilot study carried out on the HBV data with a subgroup of covariates confirms that the CBM is applicable providing sensible results.

In order to compare competitive models and chose the one that fits the data better, a further study about model selection procedure was carried out. Since the CBM is based on a partial likelihood due to the censored observations, there was a need to adapt and derive a model comparison tool that could be used with CBMs. The next chapter presents the study of a model comparison tool designed for the application of the CBM, its application to the pilot study presented here and the results of the CBM applied to the full dataset.

Chapter 7

Model Comparison Tool and Simulation Study

The aim of this Chapter is to compare the two models proposed in the Chapter 6: the CBM with two variances and the CBM with twelve variances and apply the better of them to a full quantitative dataset. The two model comparison tools, BIC and DIC, provide information about the model fit (see Section 1.3.5). However, there is no agreement whether the BIC is an appropriate measure for model comparison when there is uncertainty that one and only one of the competing models is the "true" one (Bernardo and Smith, 1994). In other words, when the aim of the study is to choose the correct model out of several competing models from where only one of them is the "true" model and the rest are wrong, then the BIC is an appropriate measure for model comparison. Since it is not assumed here that one of the two models, only one of which is the true one, it is preferable to use another measure of comparison, the DIC, which is based on the distances of the data to each of the approximate models.

The CBM, described in Chapter 6, makes use of the partial likelihood (6.2). The analysis of censored observations based on the partial likelihood provides parameter estimates for the complete model, since under the assumption of ignorable censoring mechanism both models are equivalent. However, this raises some important questions in the context of model comparison, which have not been addressed yet in any published researched. If the partial likelihood is used to study model comparison with the DIC, can we still rely on the results as a comparable measurement between models? The behaviour of the Deviance Information Criterion based on the partial likelihood, DIC_p, has not been researched yet. Since there is no information about the complete likelihood, how does the DIC_p behave with respect to the Deviance Information Criterion based on the complete likelihood, DIC_M? Are DIC_p and DIC_M equivalent? How do DIC_p and DIC_M vary within data sets of similar characteristics? How do DIC_p and DIC_M behave when the amount of censored data increases or decreases?

In this Chapter the relation between the DIC_p and DIC_M is investigated from a theoretical point of view. Then, simulated data are used to study the variation of the DIC_p within data sets of similar characteristics and different proportions of censored data within the data sets. The behaviour of the DIC_M in relation with the behaviour of the DIC_p and the proportions of censored data are also studied. The DIC_p obtained from the two models, used in the pilot study in the previous chapter, are compared in order to determine if it is necessary to take into account different variances across sample groups and technology types. Finally, the better model, selected by using the DIC_p, will be applied to the full HBV dataset.

7.1 Relation between DIC_p and DIC_M

Using equation (1.5), the Deviance Information Criterion (Spiegelhalter et al., 2002), based on the complete likelihood in equation (6.1), DIC_M can be defined as follows:

$$DIC_M = E_{\theta|y}[-2\log(L_M(\theta|y))] + \{E_{\theta|y}[-2\log(L_M(\theta|y))] - (-2\log(L_M(\bar{\theta}|y))\}, \tag{7.1}$$

where $L_M(\theta|y)$ is the complete likelihood, $\bar{\theta} = E(\theta|y)$ is the estimated mean of the parameters and the mean $Mean[-2\log(L_M(\theta|y))]$ is replaced by $E_{\theta|y}[-2\log((L_M(\theta|y))]$ in equation (1.5).

The complete likelihood given parameter θ is defined in equation (6.1) and can be expressed as follows:

$$L_M(\theta|y) = \prod_{si} f(y_{si}|\theta)^{\psi_{si}} (G(y_{si})H(y_{si}))^{\psi_{si}} \left(S(y_{si}|\theta)^{1-\upsilon_{si}} g(y_{si})^{1-\upsilon_{si}} F(y_{si}|\theta)^{1-\rho_{si}} h(y_{si})^{1-\rho_{si}}\right)^{1-\psi_{si}}, \tag{7.2}$$

which can be rewritten in vector form as:

$$L_M(\theta|y) = f(Y_{non}|\theta)(G(Y_{non})H(Y_{non}))$$

$$\times \left(S(Y_{cen}|\theta)^{1-\upsilon_{cen}} g(Y_{cen})^{1-\upsilon_{cen}} F(Y_{cen}|\theta)^{1-\rho_{cen}} h(Y_{cen})^{1-\rho_{cen}}\right)$$

$$L_M(\theta|y) = f(Y_{non}|\theta) \left(S(Y_{cen}|\theta)^{1-\upsilon_{cen}} F(Y_{cen}|\theta)^{1-\rho_{cen}}\right)$$

$$\times((G(Y_{non}))(H(Y_{non})) \left(g(Y_{cen})^{1-\upsilon_{cen}} h(Y_{cen})^{1-\rho_{cen}}\right),$$

where υ_{cen} is 0 for the group of left censored observations, ρ_{cen} is 0 for the group of right censored observations, otherwise both take the value 1; Y_{cen} is the vector of censored observations and Y_{non} is the vector of non-censored observations.

Then, the complete likelihood can be expressed as:

$$L_M(\theta|y) = L(\theta|y) \times Z(Y_{cen}, Y_{non}), \tag{7.3}$$

where $Z(Y_{cen}, Y_{non}) = G(Y_{non})H(Y_{non})\,(g(Y_{cen})^{1-\upsilon_{cen}} h(Y_{cen})^{1-\rho_{cen}})$. Combining equation (7.1) and (7.3), the DIC_M can be expressed as:

$$DIC_M = E_{\theta|y}[-2\log(L(\theta|y)Z(Y_{cen}, Y_{non}))] + \{E_{\theta|y}[-2\log(L(\theta|y)Z(Y_{cen}, Y_{non}))]$$

$$+2\log(L(\bar{\theta}|y)Z(Y_{cen}, Y_{non}))\}.$$

Next, by using the properties of the logarithm and the expectation function, it leads to

$$DIC_M = E_{\theta|y}[-2(\theta|y)] + \{E_{\theta|y}[-2\log(L(\theta|y))] - (-2\log(L(\bar{\theta}|y)))\}$$

$$+E_{\theta|y}[-2\log(Z(Y_{cen}, Y_{non}))] + \{E_{\theta|y}[-2\log(Z(Y_{cen}, Y_{non}))] - (-2\log(Z(Y_{cen}, Y_{non})))\}.$$

Based on the above formula, the Deviance Information Criterion based on the partial likelihood can be formulated as:

$$DIC_p = E_{\theta|y}[-2(\theta|y)] + \{E_{\theta|y}[-2\log(L(\theta|y))] - (-2\log(L(\bar{\theta}|y)))\}$$

$$+E_{\theta|y}[-2\log(Z(Y_{cen}, Y_{non}))] + \{E_{\theta|y}[-2\log(Z(Y_{cen}, Y_{non}))]\}$$

Consequently, the DIC_M can be further summarised as:

$$DIC_M = DIC_p - 2logZ(Y_{cen}, Y_{non}),$$

where the function Z depends only on the observations Y.

This result shows that the full model DIC can be expressed as the partial DIC and an additional additive term, which depends only on the observations.

7.2 Model Comparison via DIC_p

In order to compare two models, nested or non-nested, and decide which model fits the data, the DIC_M for model 1 and model 2 are used with the following notation:

- Let DIC_M^1 be the complete Deviance Information Criterion for model 1.
- Let DIC_M^2 be the complete Deviance Information Criterion for model 2.
- Let $D = DIC_M^1 - DIC_M^2$ be the difference of the Deviance Information Criterion between model 1 and 2.

According to the Deviance Information Criterion, the model with the lower value is the preferred choice. Then, by studying the difference of the Deviance Information Criterion, D, it can be deduced which model fits the data better. However, the difference can be written as $D = DIC_p^1 - DIC_p^2$ since the Z function does only depend on the data set. Thus, models can be compared based on the Partial Deviance Information Criterion, DIC_p, concluding that the model with lower DIC_p fits the data better.

7.3 Simulation Study

Simulated data are used to study the variation of the DIC_p within datasets of similar characteristics to datasets from the QC programmes. The simulated datasets will have a similar data structure as the QC datasets, but different proportions of censored data within.

First, the description of how the data are simulated is presented. Then, the results of the DIC_p from the simulation study are shown. The behaviour of the DIC_M in relation to the DIC_p and in relation to the proportion of censored data are also studied (see Section 7.4).

Data are simulated in two different ways with the aim of checking consistency and robustness of the conclusions obtained. The first way of simulating data is based on the estimations obtained after applying the CBM, and the second way of simulating data is based on replicated data from the observed HBV data per sample group.

7.3.1 Simulation Study with Data generated from the CBM

For the first type of simulation study, 20 data sets are simulated, each consisting of 10,000 observations, which have similar characteristics as the HBV data studied in the previous chapter. The observations are simulated from a $\log_{10}$-normal distribution with mean, $\bar{\beta}$, and variances, $\bar{\sigma}^2$, based on the posterior mean of the parameters obtained from the CBM applied to the subgroup of HBV data defined previously. The 20 simulated data sets are classified into four different groups of data (five data sets per group) according to the following data features:

Group 1: Data sets with no censored observations.

Group 2: Data sets with the same proportions of censored observations as the HBV data set for different viral loads (see descriptive analysis in Chapter 2).

Group 3: Data sets with twice the amount of censored data than in group 2.

Group 4: Data sets with three times the amount of censored data than in group 2.

In order to retain the proportions of observations per sample group, values are simulated according to the empirical proportions of the observed HBV data per sample group.

Therefore, a total of 10,000 observations is obtained, where the number of observations per sample group is based on those proportions calculated previously. An approximate proportion of censored observations per each sample group is obtained from the HBV data.

Then, for each of the groups of data, censoring indicator variables are calculated as follows:

Group 1: Since this group does not have any censored observations, no censored indicator variable is simulated.

Group 2: Random variables are simulated for each sample group from a Bernoulli distribution. For each sample group the probability for a censored observation is equal to the estimated proportion of censored observations within the HBV data.

Group 3: Random variables are simulated for each sample group from a Bernoulli distribution. The probability assigned to each sample group is twice the proportion of censored observations within the HBV data.

Group 4: Random variables are simulated for each sample group from a Bernoulli distribution. The probability assigned to each sample group is three times the proportion of censored observations within the HBV data.

To obtain the covariate matrix, $\mathbf{X}_{sim}$, for the simulation study, the variables year, technology group and subtype are generated from a discrete distribution. The probabilities of the discrete distributions are based on the empirical proportions of categories per sample group. The simulated response variable Y_{sim} is generated from a $\log_{10}$-normal distribution with mean $\bar{\beta}\mathbf{X_{sim}}$ and variance $\bar{\sigma}^2$. $\bar{\beta}$ and $\bar{\sigma}^2$ are the estimated posterior means of the parameters obtained from the CBM applied to the HBV data.

Finally, the CBM is fitted to the simulated datasets by applying the MCMC techniques described in previous chapter. This allows for the new parameters β and σ to be estimated and to be used to calculate the DIC_p for each of the datasets.

Results of the Simulation Study

It is of interest to study the behaviour of the Partial Deviance Information Criterion, DIC_p, depending on the amount of censored observations in the data set and the variation of the results from the same classified group. Before considering the results obtained from this simulation study, the notation and definition of the simulated data sets Y are introduced:

- Let $Y^j_{non} \sim N(\bar{\beta} * Xsim, \bar{\sigma}^2)$ be the simulated sample Y for data group 1.
- Let $Y^j_{1cen} \sim N(\bar{\beta} * Xsim, \bar{\sigma}^2)$ be the simulated sample Y for data group 2.
- Let $Y^j_{2cen} \sim N(\bar{\beta} * Xsim, \bar{\sigma}^2)$ be the simulated sample Y for data group 3.
- Let $Y^j_{3cen} \sim N(\bar{\beta} * Xsim, \bar{\sigma}^2)$ be the simulated sample Y for data group 4.

In the above notation, $j = 1, ..., 5$ is the number of data sets replicated for each group. Five replicates are taken as a sensible compromise between illustrating basic features and computational effort.

The CBM, allowing for twelve variances depending on technology and sample group, is applied to each of the data sets and the posterior distributions of the parameters are obtained. Then, the DIC_p is calculated for each data set.

Table 7.1 shows the $Mean[-2\log(L(\theta|y))]$, $-2\log(L(\bar{\theta}|y))$, DIC_p and the 'effective number of parameters', p_D, for each of the data sets. The CBM with twelve variances has 58 parameters. For all data sets the value of the effective number of parameters, p_D, is approximately 58.

Figure 7.1 shows the tendency of the DIC_p depending on the proportion of censored observations. There is a decreasing relationship that appears to be approximately linear with increasing number of censored observations.

Table 7.1: *Partial Deviance Information Criterion (DIC_p): Simulation study based on the CBM.*

	$Mean[-2\log(L(\theta\|y))]$	$-2\log(L(\bar{\theta}\|y))$	p_D	DIC_p
Model Y^1_{non}	230318.2	230260.0	58.12	230376.3
Model Y^2_{non}	229873.2	229814.9	58.25	229931.4
Model Y^3_{non}	230007.1	230065.1	57.96	230123.1
Model Y^4_{non}	231033.9	230976.3	57.64	231091.5
Model Y^5_{non}	230173.5	230115.5	58.03	230231.5
Model Y^1_{1cen}	209555.8	209498.0	57.75	209613.5
Model Y^2_{1cen}	210049.1	209991.9	57.27	210106.4
Model Y^3_{1cen}	210134.9	210077.2	57.73	210192.7
Model Y^4_{1cen}	209978.6	209920.5	58.01	210036.6
Model Y^5_{1cen}	210129.0	210071.0	57.99	210187.0
Model Y^1_{2cen}	191154.2	191096.6	57.64	191211.9
Model Y^2_{2cen}	190647.9	190589.9	58.01	190705.9
Model Y^3_{2cen}	190865.7	190807.9	57.87	190923.6
Model Y^4_{2cen}	190615.5	190557.6	57.87	190673.4
Model Y^5_{2cen}	190384.9	190327.6	57.24	190442.1
Model Y^1_{3cen}	171244.1	171186.4	57.68	171301.8
Model Y^2_{3cen}	171098.1	171041.3	56.76	171154.8
Model Y^3_{3cen}	171378.2	171320.8	57.34	171435.5
Model Y^4_{3cen}	171304.4	171246.3	58.11	171362.5
Model Y^5_{3cen}	171495.6	171437.8	57.83	171553.4

The DIC_p varies slightly within replicated data for each group and decreases linearly between groups of simulated data with different proportions of censored observations. In order to study if the same behaviour occurs when the data are simulated in a different way, new simulated data sets are obtained based on replications of each real observation. Then, the CBM is applied, and the DIC_p is calculated, and checked for consistency.

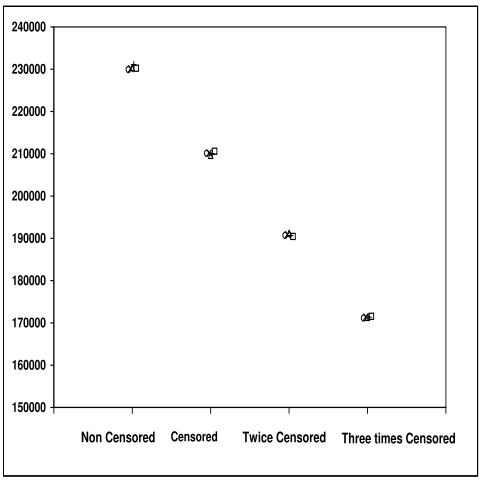


Figure 7.1: *Partial DIC for simulated data sets based on the CBM with different proportions of censored observations; no censored observations on the data sets, same proportion of censored observations as HBV data, twice the proportion of censored observations of HBV data, three times the proportion of censored observations of HBV data.*

7.3.2 Simulation Study using Replicated Data based on the HBV Dataset

In this simulation study the aim is to reproduce random data of approximately 10,000 observations. The procedure is set up in a way that preserves the structure and proportions of the original HBV data. Four different groups of data are simulated each of which differs in the number of censored observations. For each group five data sets are replicated in order to study variation within group.

Using the observations from the HBV programmes, data are replicated in the following way: data from a normal distribution are simulated for each sample group, and the same covariate values of the observations for the replicated data are used. For each sample group s, the corresponding proportion of data from the total number of HBV data are obtained, $p_s = n_s/N$; here the n_s is the total number of HBV observations for the sample group s and N is the total number of HBV data among sample groups. Then, $10000 * p_s$ data are simulated for each sample viral load group s. The way of simulating data from a normal distribution varies as follows:

- Group 1: $|\frac{10000*p_s}{n_s}|$ data are simulated from a $N(\log_{10}(y_{si}), 0.5)$ for each observation y_{si}. Thus, the simulated values of the $\log_{10}$ sample viral load are obtained for sample group s and participant i. Note that the variation of the $\log_{10}(y_{si})$ is 0.5, which is a standard assumption in molecular diagnostics (see Chapter 2). All observations are considered non-censored. The covariates associated with each simulated datum from $N(\log_{10}(y_{si}), 0.5)$ are the covariates associated with the observation y_{si}.

- Group 2: As in group 1, for each observation y_{si}, $|\frac{10000*p_s}{n_s}|$ data are simulated from a $N(\log_{10}(y_{si}), 0.5)$. The associated covariates for each simulated datum are the covariates associated with participant i in sample group s. If the observation y_{si} is censored then the simulated data are considered censored observations in the same direction as observed. Thus, the proportion of censored information of the data set is the same as in the HBV data.

- Group 3: In order to obtain a higher proportion of censored information for the simulated data, twice the amount of data for those censored observations y_{si} is simulated. Then, $2 * |\frac{10000*p_s}{n_s}|$ simulated data are obtained from the normal distribution for each censored observation y_{si}, and $|\frac{10000*p_s}{n_s}|$ simulated data are obtained for each non-censored observation. Then, $|\frac{10000*p_s}{n_s}| * n_{c_s}$ non-censored values are selected randomly and removed from all the simulated data for sample s, where n_{c_s} is the total number of censored observations for sample group s. Thus, the proportion of censored observations is double as in the HBV data.

- Group 4: Triple of the amount of data for those censored observations y_{si} are simulated. The set up is similar to Group 3, i.e. $3 * |\frac{10000*p_s}{n_s}|$ simulated data are obtained from the normal distribution for each censored observation y_{si}, and $|\frac{10000*p_s}{n_s}|$ simulated data are obtained for each non-censored observation. Then, $2 * |\frac{10000*p_s}{n_s}| * n_{c_s}$ non-censored values are selected randomly and removed from all the simulated data for sample group s, where n_{c_s} is the total number of censored observations for sample group s.

Table 7.2 shows the values of $Mean[-2\log(L(\theta|y))]$, $-2\log(L(\bar{\theta}|y))$, DIC_p and the p_D for each of the data sets. Figure 7.2 shows a linear decreasing trend on the DIC_p as the proportion of censored observations increases. However, a slight decrease in the p_D values is observed as the proportion of censored observations increases.

Table 7.2: *Partial Deviance Information Criterion (DIC_p): Simulation study based on HBV replicated data.*

	$Mean[-2\log(L(\theta\|y))]$	$-2\log(L(\bar{\theta}\|y))$	p_D	DIC_p
Model Y^1_{non}	241840.9	241783.2	57.77	241898.7
Model Y^2_{non}	241804.5	241746.6	57.91	241862.5
Model Y^3_{non}	241828.3	241773.3	57.71	241886.0
Model Y^4_{non}	241831.0	241773.3	57.69	241888.7
Model Y^5_{non}	241831.0	241772.9	58.11	241889.1
Model Y^1_{1cen}	218083.8	218026.3	57.51	218141.3
Model Y^2_{1cen}	218123.5	218066.5	56.97	218180.5
Model Y^3_{1cen}	218077.0	218019.9	57.08	218134.0
Model Y^4_{1cen}	218091.4	218034.4	57.01	218148.5
Model Y^5_{1cen}	218053.3	217996.2	57.11	218110.4
Model Y^1_{2cen}	196575.7	196518.7	56.97	196632.6
Model Y^2_{2cen}	196702.4	196645.4	56.94	196759.3
Model Y^3_{2cen}	196307.4	196250.8	56.57	196364.0
Model Y^4_{2cen}	196349.4	196292.2	57.23	196406.6
Model Y^5_{2cen}	196598.0	196541.4	56.58	196654.5
Model Y^1_{3cen}	174568.2	174512.1	56.06	174624.2
Model Y^2_{3cen}	174510.4	174454.6	55.81	174566.2
Model Y^3_{3cen}	174488.2	174432.5	55.74	174544.0
Model Y^4_{3cen}	174488.5	174432.0	56.47	174545.0
Model Y^5_{3cen}	174374.5	174319.1	55.40	174429.9

From both simulation studies it is concluded that the DIC_p for the CBM decreases when the censored information of the data to be analysed increases. The decrease of the DIC_p values appears to be linear with respect to the proportion of censored information contained in the data. Therefore, the DIC_p can be approximated by a linear function depending on the proportion of censored information contained on the data. In the next section the theoretical relation between DIC_M and DIC_p will be derived based on an approximation of the DIC_p by a linear function.

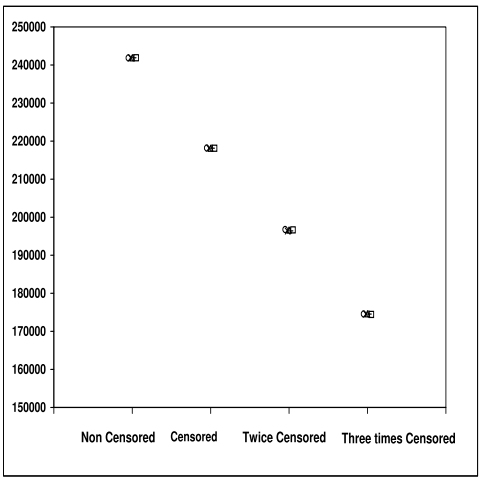


Figure 7.2: *Partial DIC for simulated data sets based on replicated observations from the HBV programmes with different proportions of censored observations; no censored observations on the data sets, same proportion of censored observations as HBV data, twice the proportion of censored observations of HBV data, three times the proportion of censored observations of HBV data.*

7.4 Relation of the DIC_M with Respect to the DIC_p

In the previous section the simulation study showed that the DIC_p follows approximately a negative linear relationship with respect to the proportion of censored observations. Thus, the DIC_p can be approximated to a linear function $a + b * \frac{n_c}{N}$ where n_c is the number of censored observations within the data set and N the total number of data.

On the other hand, $\log(Z(Y_{cen}, Y_{non}))$ can be written as

$$\log(Z(Y_{cen}, Y_{non})) = \log\left(\prod_{i \in I_{N-n_c}} G(y_i)H(y_i)\right) + \log\left(\prod_{i \in I_{n_c}} g(y_i)^{1-\upsilon_i} h(y_i)^{1-\rho_i}\right)$$

$$= \sum_{i \in I_{N-n_c}} \log\left(G(y_i)H(y_i)\right) + \sum_{i \in I_{n_c}} \log\left(g(y_i)^{1-\upsilon_i} h(y_i)^{1-\rho_i}\right),$$

where I_{n_c} is the group of censored observations and I_{N-n_c} is the group of non-censored observations.

If $(N - n_c)$ and n_c are sufficiently large then $\log(Z(Y_{cen}, Y_{non}))$ can be approximated by

$$\log(Z(Y_{cen}, Y_{non})) \approx (N - n_c)(E_{y_{non}}[\log((G(y)H(y)))]) + (n_c)\left(E_{y_{cen}}[\log(g(y)^{1-\upsilon_{cen}}h(y)^{1-\rho_{cen}})]\right)$$

$$\approx n_c(E_{y_{cen}}[\log(g(y)^{1-\upsilon_{cen}}h(y)^{1-\rho_{cen}})] - E_{y_{non}}[\log(G(y)H(y))]) + N\left(E_{y_{non}}[\log(G(y)H(y))]\right).$$

Since the last term is constant with respect to n_c, the $\log(Z(Y_{cen}, Y_{non}))$ can be expressed as:

$$\log(Z(Y_{cen}, Y_{non})) \approx n_c(E_{y_{cen}}[\log(g(y)^{1-\upsilon_{cen}}h(y)^{1-\rho_{cen}})] - E_{y_{non}}[\log(G(y)H(y))]) + d,$$

where $d = N\left(E_{y_{non}}[\log(G(y)H(y))]\right)$.

Finally, the DIC_M can be approximated by:

$$DIC_M \approx a - 2d + \frac{n_c}{N}(b - 2N(E_{y_{cen}}[\log(g(y)^{1-\upsilon_{cen}}h(y)^{1-\rho_{cen}})] - E_{y_{non}}[\log(G(y)H(y))])).$$

7.5 DIC from the Pilot Study

In Chapter 5, it has been assumed that the variability of the data depends on the sample group and the technology classification. However, the results and the classical analysis of variances performed in Chapter 5 showed that this assumption is not valid for all sample groups. The differences in the variability of the response depending on technology type are only significant for some sample groups.

In order to study a reduced model in which the variability in the data is assumed to be the same across all groups, the CBM was applied to a subgroup of HBV data in a pilot study. Two applications of the CBM were carried out: a CBM assuming two different variances for the observations and a CBM assuming twelve different variances for the data. To check which of the two models fits the data better, a variant of the DIC, in the form of DIC_p was proposed. In previous section, its theoretical derivation and application via a simulation study was studied, showing that the proposed DIC can be used to test which of the two models fits the data better.

In this section the proposed DIC_p is applied to the two models from the pilot study. The results from the two models are shown in Table 7.3. Assuming twelve variances the CBM has a better average fit to the data, $Mean[-2\log(L(\theta|y))]$, a better fitting point estimate $-2\log(L(\bar{\theta}|y))$ and a lower estimated predictive error DIC_p, compared to the two variances model.

Table 7.3: *Partial Deviance Information Criterion (DIC_p): Pilot study.*

	$Mean[-2\log(L(\theta\|y))]$	$-2\log(L(\bar{\theta}\|y))$	p_D	DIC_p	BIC
Model 2 variances	43707.35	43661.29	46.06	43753.41	44023.52
Model 12 variances	43640.33	43585.20	55.12	43695.45	44022.84

The BIC is quite similar for both models, although a bit smaller when 12 variances are included in the model. Both information criteria suggest that the CBM assuming twelve variances fits the data better.

7.6 CBM Applied to HBV Data

So far the CBM has been proposed and developed as a more appropriate model to fit the data in this application. The effectiveness and robustness of a variant of the DIC, the DIC_p, as a measure of goodness of fit of the model has also been considered. The CBM assuming twelve variances is concluded to fit the data better than the model based on two variances, when applying it to the HBV data with a subgroup of covariates and to the simulated data.

In this section the CBM assuming twelve variances is applied to the complete HBV data set. The results obtained from the application of the full CBM to the HBV data are presented. If it is not specifically stated otherwise, the results are given with the remaining covariates under baseline conditions.

Sample Group 6 $\log_{10}$ Copies/ml Viral Load

The estimated mean for the results of participants under baseline conditions is roughly 4 $\log_{10}$ copies/ml (see Table 7.4). The estimated mean viral load for samples from 2005 is significantly lower than the means for samples from 2002 to 2004, indicating that participants from those years tend to provide closer estimates of viral load to the target viral load of 6 $\log_{10}$ copies/ml than participants from 2005. Although it is not significant, the estimated mean viral load for samples of subtype D is approximately 1.2 $\log_{10}$ copies/ml higher than the estimated mean for samples of subtype A.

bDNA users are more likely to provide significantly closer estimates of viral load to the target than CC users. Participants using an anti-contamination system tend to provide significantly higher estimates of viral load than participants not using an anti-contamination system. Participants performing an inhibition test in negative samples tend to return higher estimates of viral load than participants not performing any inhibition test.

Table 7.4: *Summary statistics of the parameter estimates from the CBM for HBV sample group 6* $\log_{10}$ *copies/ml viral load: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	3.920	0.629	2.636	3.929	5.129		
Year- baseline "2005"							
Year 2002	0.411	0.146	0.117	0.411	0.687	+	Yes
Year 2003	1.423	0.623	0.208	1.424	2.652	+	Yes
Year 2004	1.507	0.612	0.313	1.507	2.719	+	Yes
Subtype- baseline "A"	1.179	0.618	-0.036	1.182	2.398	+	No
Technology group- baseline "CC"							
Tech CIH-CC	0.013	0.291	-0.554	0.018	0.579	+	No
Tech RTIH-CC	0.253	0.176	-0.081	0.249	0.615	+	No
Tech RTC-CC	-0.030	0.141	-0.318	-0.022	0.233	-	No
Tech bDNA-CC	0.507	0.176	0.146	0.512	0.838	+	Yes
Tech HC-CC	0.523	0.280	-0.037	0.523	1.064	+	No
Anti- baseline "No"	0.251	0.108	0.048	0.251	0.469	+	Yes
Accred- baseline "No"	0.040	0.078	-0.115	0.039	0.197	+	No
OtherSpc. baseline "No"	-0.013	0.085	-0.189	-0.009	0.149	-	No
Analysis method- baseline Singly							
Analysis Duplicated	-0.070	0.114	-0.283	-0.073	0.155	-	No
Analysis Other	-0.080	0.243	-0.570	-0.078	0.399	-	No
Plasma- baseline 0-10							
Group 1: 11-100	-0.136	0.172	-0.497	-0.127	0.168	-	No
Group 2: 101-1,000	-0.006	0.114	-0.236	-0.004	0.216	-	No
Group 3: 1,001-2,000	0.161	0.149	-0.137	0.163	0.449	+	No
Group 4: 2,001-10,000	0.090	0.104	-0.129	0.093	0.293	+	No
Group 5: > 10,000	-0.092	0.140	-0.405	-0.078	0.147	-	No
Serum- baseline 0-10							
Group 1: 11-100	0.116	0.122	-0.127	0.117	0.358	+	No
Group 2: 101-1,000	0.191	0.109	-0.013	0.189	0.409	+	No
Group 3: 1,001-2,000	0.034	0.149	-0.280	0.041	0.311	+	No
Group 4: 2,001-10,000	0.145	0.099	-0.038	0.141	0.347	+	No
Group 5: > 10,000	-0.086	0.140	-0.402	-0.075	0.156	-	No
Labtype- baseline Hospital							
Public Health	0.075	0.121	-0.176	0.076	0.316	+	No
Private	0.027	0.124	-0.232	0.031	0.260	+	No
Reference	-0.138	0.256	-0.676	-0.122	0.345	-	No
Manufacture	0.056	0.132	-0.239	0.066	0.301	+	No
Research	-0.200	0.159	-0.543	-0.189	0.076	-	No
Inhibition Test-baseline "No"							
Inhbition test Yes	0.127	0.099	-0.062	0.125	0.325	+	No
Inhibiton test only Negative samples	0.383	0.167	0.047	0.382	0.708	+	Yes

In the next section a summary of the results is presented (more detailed results can be found in Appendix D).

Sample Group 5 $\log_{10}$ Copies/ml Viral Load

For samples from 2002 and 2004 participants are more likely to provide significantly higher estimates of viral load that are closer to the target than for samples from 2005 (see Table D.1 in Appendix D).

The performance of bDNA and HC technology users is significantly different from performance of CC technology users. Users of bDNA and HC technologies tend to return higher estimates, which are closer to the target viral load than CC users.

Participants using an anti-contamination system are more likely to provide higher estimates of viral load that are closer to the target than participants not using anti-contamination system.

Participants performing between 2,001 to 10,000 serum tests annually tend to return significantly higher estimates of viral load than participants testing less than 11 serum samples annually.

The estimated mean for the results of research laboratories is significantly lower and further away from the target viral load than the estimated mean for the results of hospital laboratories.

Participants that performed an inhibition test for negative samples are more likely to report significantly higher estimates of viral load than participants who did not perform any inhibition test.

Sample Group 4 $\log_{10}$ Copies/ml Viral Load

The estimated mean for the results of participants under baseline conditions is 3.5 $\log_{10}$ copies/ml (see Table D.2 in Appendix D). For samples from 2002 participants are more likely to provide significantly higher estimates of viral load than for samples from 2005.

Users of bDNA technology tend to return higher estimates of viral load that are closer to the target than CC technology users.

Users of an anti-contamination system are more likely to report higher estimates of viral load than participants not using an anti-contamination system.

Experience performing between 2,000 and 10,000 serum tests annually tends to improve participants' results. The estimated mean for the results of participants with such experience is higher and closer to the target viral load than the estimated mean for the results of participants performing fewer than 11 serum tests annually.

Participants performing inhibition tests only in negative samples tend to provide higher estimates of viral load than those not performing any inhibition test.

Sample Group 3.5 $\log_{10}$ Copies/ml Viral Load

The estimated mean for the results of participants under baseline conditions is approximately 3 $\log_{10}$ copies/ml (see Table D.3 in Appendix D). For samples from 2002, participants tend to provide significantly higher estimates of viral load than for samples from 2003.

bDNA and RTIH users are more likely to report higher estimates of viral load than CC users. However, bDNA users tend to provide estimates of viral load further away from the target than CC users.

Participants performing other methods of analysis, such as triplicate, are more likely to provide higher estimates of viral load that are closer to the target than participants using single analysis method.

Participants performing an inhibition test tend to return closer estimates of viral load to the target than participants not performing any inhibition test.

Sample Group 3 $\log_{10}$ **Copies/ml Viral Load**

The estimated mean for the results of participants under baseline conditions is 2.6 $\log_{10}$ copies/ml (see Table D.4 in Appendix D). Participants in 2004 tend to provide estimates of viral load that are significantly closer to the target than participants in 2005.

HC users are more likely to return estimates of viral load significantly further away from the target than CC users.

Participants performing between 2,000 and 10,000 serum tests annually are more likely to obtain estimates of viral load that are closer to the target than participants performing fewer than 11 serum tests annually.

Research laboratories are more likely to provide significantly lower estimates of viral load than hospital laboratories. Those estimates are further away from the target viral load.

Participants performing an inhibition test only in negative samples tend to report significantly higher estimates of viral load and closer to the target than participants not performing any inhibition test.

Sample Group 2.3 $\log_{10}$ **Copies/ml Viral Load**

For samples from 2004 participants tend to provide closer estimates of viral load to the target than for samples from 2003 and 2002; the estimated mean viral load for samples from 2004 is 2.39 $\log_{10}$ copies/ml (Table D.6 in Appendix D).

The closest estimated mean to the target is found for the results of CC users. RTC, RTIH and CIH users tend to provide estimates of viral load significantly further away from the target than CC users. The estimated means for the results of bDNA and HC users are not significantly different from the estimated mean for the results of CC users, although participants using bDNA are more likely to report lower estimates of viral load than participants using CC technology.

Participants performing inhibition tests only in negative samples tend to provide higher estimates of viral load that are further away from the target than participants not performing any inhibition test.

As in Chapter 5, model reduction can be carried out using the same selection procedure. Since the aim of this chapter is to provide a better model from a statistical point of view, it is considered unnecessary to present the results from the reduced model here, which of course it would be more interesting from a practical point of view than from a statistical point of view. The table with the estimated variances for the full model can be found in Appendix D. As in the pilot study described in Section 6.3, here the residuals of the model have been studied to check the goodness of fit of the CBM applied to HBV data. It has been observed that the quantiles of the residuals lie between -3 and 3 indicating that the model fits the data well.

7.6.1 Technical Report of the CBM Applied to HBV Data

The CBM is run for 100,000 iterations. To break up the autocorrelations of the estimates, every tenth observation is recorded and the rest are discarded. Thus, a sample of 10,000 values is obtained for each parameter in order to describe the posterior distribution of the parameter. Convergence is studied and achieved after a burn in period of 3,000 iterations.

No correlation between the estimated parameters is found. The results for each parameter are obtained by a summary of the sample after the burn in period, that is the last 7,000 values from the sample. The acceptance probabilities for each block of parameters (one block of estimated parameters per sample group) is checked. Table 7.5 shows the acceptance probability rates obtained. All the acceptance rates are over 30% indicating that the chain is mixing well (Gamerman and Lopes, 2006).

Table 7.5: *Acceptance probability rates of the parameters estimated from the CBM for HBV sample groups.*

Parameters	Acceptance probability rate					
	Sample 2.3	**Sample 3**	**Sample 3.5**	**Sample 4**	**Sample 5**	**Sample 6**
$\vec{\beta}_s$	32.40	40.69	39.25	39.80	42.43	44.08
$\vec{\sigma}_1^2$ **Commercial**	57.77	35.60	37.59	37.19	30.77	57.41
$\vec{\sigma}_2^2$ **In-house**	68.82	62.02	62.89	71.79	70.85	80.93

The primary aim of this chapter is to develop a better model than the proposed QTBM in the previous chapter, and developed an appropriate measure of fit to discriminate between models (nested and non-nested models), the DIC_p. The results of the DIC_p from a simulation study were used to select between the two variances and twelve variances model applied to HBV data with a reduced group of covariates (chosen as more relevant from a practical point of view). Based on these results, the CBM was applied to the HBV data with the full set of covariates to confirm that the model is working well in a large data setting. This model can be used to provide participants with a full and objective feedback on their laboratory practice when analysing samples of HBV.

7.6.2 Summary

A further application of the CBM to the HBV with a complete set of covariates was carried out and it was concluded that for samples from 2004, participants were more likely to return closer estimates of viral load to the target than for samples from 2005. No differences were found between estimates of viral load for samples of different subtypes. However, for the strongest viral load of 6 $\log_{10}$ copies/ml, the difference between the estimated means was approximately 1.2 $\log_{10}$ copies/ml higher for samples of subtype D than for samples of subtype A (note that the standard deviation for this estimate is 0.6 $\log_{10}$ copies/ml). This result indicates that participants are more likely to provide closer estimates of viral load to the target for samples of subtype D than for samples of subtype A.

For sample groups of viral load 6, 5 and 4 $\log_{10}$ copies/ml, participants using bDNA technology tend to provide closer estimates of viral load to the targets than participants using CC technology. However, for the lowest sample group of 2.3 $\log_{10}$ copies/ml participants using bDNA technology are more likely to return estimates further away from the target viral load than CC users.

The use of an anti-contamination system tends to improve the estimates of viral load, providing higher estimates that are closer to the target viral load for sample groups of 6, 5 and 4 $\log_{10}$ copies/ml, but no differences are found when the sample groups have a lower viral load. Estimates of viral load from participants with different accreditation status are not significantly different for any sample group. For almost all sample groups, participants performing an inhibition test only in negative samples tend to provide closer estimates to the target viral load than participants not performing any inhibition test.

7.7 Conclusions

The measure of the expected predicted error, DIC, developed by Spiegelhalter et al. (2002) fails to work well when censored data are estimated in the models, as it is the case of the QTBM. On the other hand, a better proposed model (the CBM) for quantitative data was developed and described in Chapter 6 which did not use estimates of censored data to provide parameters estimates. However, the CBM makes used of the partial likelihood instead of full likelihood and therefore the existing tool for model comparison, DIC, could not be applied directly to the CBM. In order to be able to perform a model comparison for the developed CBM, a variant of the Bayesian comparison tool DIC was developed in this thesis and results of its applications to real and simulated data were provided.

The partial Deviance Information Criterion, DIC_p provides a model comparison tool that is accurate and appropriate to discriminate between models. The simulation study carried out shows the behaviour of the DIC_p and its robustness. The DIC_p measure obtained from the CBM applied to the HBV data with a subgroup of covariates has made possible to discriminate between a model with two variances and a model with twelve variances, leading to the conclusion that the CBM with twelve variances fits the HBV data better.

It is concluded that the application of the CBM to the HBV data provides more objective and accurate results, which are going to provide better feedback to participants based on their laboratory practices. This is the first time that a derivation of the DIC based on partial likelihood is developed, proving a robust tool for model comparison, which allows to discriminate between nested and non-nested models.

There are however, few concerns with the newly proposed selection criterion. In particular, the assumption that the censoring mechanism is ignorable, which may not always be true.

On the other hand, in common with the previous chapters independence between responses was also assumed. However, some participants return several results per year or repeat the EQA programme leading to non-independent or repeated samples. Both types of dependences may be addressed using methods that count for repeated measures.

The areas of concerns of the data analysis and statistical methodology presented in Chapters 3 to 7 and their evidence of originality will be discussed in Chapter 8. Other areas of research where these models can be applied will be proposed, and further research that needs to be carried out in this field of work will also be suggested.

Chapter 8

Discussion and Further Work

An important way of monitoring laboratory quality by an independent body is to considerer the entire laboratory practice and procedure as part of the EQA programme. It allows a laboratory to benchmark its performance against other laboratories and provides feedback to identify and investigate potential areas of concern. However, there are no uniformly accepted criteria for molecular diagnostic kit EQAs for assessing laboratory performance. The data collected by QCMD is one of the world's largest and oldest molecular diagnostic EQA provider, nevertheless, there is no record of previous research undertaken to interrogate this reservoir of data, in order to study the risk factors associated with laboratories' performance, to provide better feedback to participants about their laboratory practice and to improve the design of EQA programmes.

In this thesis a measure for scoring participants' quantitative performance, which is an improvement on previous measures, has been proposed. Risk factors associated with qualitative and quantitative responses over time and EQA programmes with different pathogens have also been studied. In addition, a statistical model has been developed that can be applied to other research areas and can take into account missing and censored information.

A novel approach has been chosen by investigating the data from a non-classical perspective. Bayesian techniques are widely used in other areas of research (Gelman et al., 2004). However, Bayesian models for qualitative data within the molecular diagnostic field have not been investigated previously. Although in other research fields Bayesian methods have been used to deal with missing data (Clogg et al., 1991), these have not previously been applied to molecular diagnostic performance and missing data as in the QLBM model. Under the classical approach censored observations are currently discarded when scoring participants. In this study, a model that uses censored observations and treats the data in a more natural way is developed for the first time using a Bayesian approach. In particular, the CBM presented in Chapter 6 and the model comparison tool for quantitative data (presented in Chapter 7) is a clear theoretical advancement in the area of Bayesian statistics, which have not been investigated previously in any other area of research.

The methods proposed performed very well when tested on the large reservoir of QCMD data. Despite that, the model proposed are based on a series of assumptions, which sometimes are not met by the data under consideration. This represents a limitation that need to be taken into account when reporting the results. Here some of the limitations of the proposed methodology are presented and further work to be done for improving this research is outlined.

8.1 Proposed Scoring System for Quantitative Participants' Performance

In order to provide appropriate and independent performance indicators to measure the quality of laboratories' performance, a Bayesian approach to estimate the assigned value was suggested in this thesis. This Bayesian estimate is the target value suggested by the EQA provider (from internal investigations and previous panels) amended by estimates from

high quality reference laboratories. In addition, this estimate may be calculated taking into account values outside the limits of detection for the assay used from reference laboratories. However, the estimate depends on the chosen reference laboratories and their performance. Therefore, special care needs to be taken when choosing the reference laboratories. It should be noted that good laboratories score low marks for the proposed score, but high marks for many existing measures, which may be due to the way of choosing reference laboratories. Instead of using the estimates from reference laboratories, an alternative way to estimate the Bayesian value is to amend the target value suggested by the EQA provider by estimates from participants in previous pathogen EQA programmes.

When calculating the Bayesian mean estimate, it has been assumed that the observations from the reference laboratories follow a normal distribution. If this is not the case Bayesian techniques can still be used to obtain the mean. However, the proposed Bayesian mean should be modified and adjusted for each EQA programme.

Although the Bayesian mean estimate takes into account the censored observations from the reference laboratories, the current scoring system cannot provide a score for those participants reporting values outside the limit of detection. Thus, an adjusted scoring indicator should be developed in order to provide participants containing censored observations with a score of their performance.

A proposed area for further work to be done is to obtain, via the use of Bayesian techniques, an estimate of the censored observations from participants to be imputed and then scored. The estimated value to be imputed may be calculated based on the laboratory practice and the censored observation reported by the participant.

8.2 Modelling Qualitative Performance of Participants in QCMD Quality Control Programmes over Time

Given that classical methods failed to model the data adequately, a Bayesian model, the QLBM, is proposed in this thesis in order to analyse and model qualitative performance of participants in QCMD programmes over time. The QLBM was developed and coded using the statistical software WinBUGs. The QLBM is a GLM from a Bayesian perspective that takes into account missing information. The model was applied to data of two different pathogen programmes over time and goodness of fit for the model was checked using posterior predictive techniques for model checking, while model robustness was studied via a sensitivity analysis. It was concluded that the QLBM fits appropriately the qualitative data from EQA programmes. Furthermore, since Bayesian techniques were used to develop the model, there is no need to take care of the asymptotic theory and the assumptions underlying classical GLM theory such as sample size or multiple testing.

The prior specification of the QLBM was built using a hierarchical structure and non-informative priors. The hierarchical structure of the model allows to include enough parameters in the model, avoiding problems of overfitting because of the use of probability distributions to structure the dependency of some parameters. Since there is no prior information available about the primary parameters, the use of non-informative priors influence the results with 'objective' information. Therefore, the large number of parameters fitted and subjective concerns are not drawbacks from the QLBM.

Although, there is not a concern about sample size as in classical analysis, the reliability of the QLBM increases with the increase of the sample size. In this case, the prior information will not have a high impact on the results. However, if the sample size is small the prior distributions specified for the QLBM has a high influence on the estimates. The prior information about the parameters is updated via the likelihood. If the sample size is small then

the updated process is carried out adding little information. Then, the posterior information of the parameter is highly influenced by the prior information. Therefore, in applications with a small number of observations, special care needs to be taken when choosing priors (Gelman et al., 2004).

The priors specified in this thesis for the missing covariates are specific for categorical covariates. If instead continuous variables with missing values are considered, the prior distributions for those need to be changed to take into account the continuous nature of the covariate.

An important area of concern in the QLBM is that it uses the independence of responses as a working assumption. This is done to keep the complexity of the model structure and computational burden within acceptable limits. However, this assumption might not be fulfilled for some responses. Although the model validation shows that replicated responses from the QLBM were in concordance with the observed data (the model fitted appropriately), some participants returned several responses for the same sample in one year. In this case, the responses may be correlated. One possibility to address possible correlations between responses is to include a covariance matrix in the model accounting for these correlations. On the other hand, it is possible that participants repeat the EQA programme, so repeated responses from participants over time for the same sample can be otained (assuming that the sample is the same over time). In this case, the model to analyse the data should incorporate techniques that can deal with repeated measures. One approach in this direction is to use random effects for each participant with a specific correlation distribution or to develop a GEE model for the responses from a Bayesian perspective (Dugard et al., 2010; Hand and Crowder, 1996; Twisk, 2003).

The model developed in this thesis were tested on EQA programmes from 2002 to 2005. In further analysis, it would be of interest to perform an external validation of the model using cross-validation methods, where new data obtained on EQA programmes from 2006 to 2010 are compared with the estimates from the QLBM.

From another perspective, it would also be of interest to compare the different countries where the samples were analysed by mapping the results of participants (Banerjee et al., 2004; Waller and Gotway, 2004). Thus, performance depending on the country and its laboratory practice in relation to the level of development of the country and its resources can be studied.

Alternatively, the model can include a covariate depending on whether the participant has previous experience with QCMD programmes or a variable quantifying experience itself. This may answer questions whether taking part in a EQA really improves performance.

Finally, the proposed QLBM model can be applied in other areas of research such as medicine. An illustrative example in this sense is the application of QLBM in oncology to study risk factors associated with positive sentinel lymph node. A sentinel lymph node is found to be positive after surgery is taken place. If previous to surgery it is known which patients with lymph nodes are more likely to have a positive sentinel, doctors could avoid surgery on certain types of patients saving cost and time for health providers; this may save the patient from having an unnecessary surgery. Although research approaching this topic from a classical perspective was carried out in the past, information about patients containing missing data has had to be discarded (Katz, et al., 2008; van la Parra et al., 2009; Tyler and Balch, 2005). This is only one example of research area where, by applying the proposed QLBM, complete datasets can be analysed and more appropriate conclusions can be obtained from a Bayesian perspective.

8.3 Modelling Quantitative Performance of Participants in QCMD Quality Control Programmes over Time

Given that the classical model failed to model the data appropriately, a Bayesian method has been proposed to analyse and model the quantitative performance of participants in QCMD Quality Control programmes over time. As for the qualitative model, the QTBM is not a standard model that can be found implemented in any statistical software. Although WinBUGs is a Bayesian software which uses implemented MCMC to obtain estimates for basic Bayesian models, the user has to code up non-standard models such as the QTBM. A considerable advantage of the QTBM is that it takes into account the missing information and censored observations.

In order to check the robustness of the model when applied to different datasets, the proposed QTBM was applied to data of two different pathogen programmes over time. Once the results from its application were obtained, the goodness of fit of the model was checked by using posterior predictive techniques for model checking, while the robustness of the model was studied by using a range of different priors via a sensitivity analysis. Based on these results, it was concluded that the QTBM fits appropriately the quantitative data from EQA programmes. Since Bayesian techniques were used, there was no need to take into account extra considerations about issues that may be problematic when applied classical theory, such as discarding data due to missing information.

There are some drawbacks of the approach taken to model quantitative data. This was the case with the QLBM in Chapter 3, where independence of responses was assumed. However, a model validation was carried out and it was concluded that the QTBM fitted the data appropriately. Nevertheless, this might be improved by incorporating the dependence struc-

ture which is present within the data. One possibility is to include a correlation matrix into the model, which might reflect dependence of responses for those participants who provided more than one result for a sample within the same year. On the other hand, since some of the samples are repeated over time, another possibility is to extend the model by using techniques from longitudinal data models to account for the correlation structure arising from those participants who repeat a programme over time (Hand and Crowder, 1996; Gelman and Hill, 2007).

A classical approach to analyse censored data is the Tobit regression model (Tobin, 1958). However, this model still cannot deal with missing covariates and has restriction about the number of parameter estimates because it is based on asymptotic theory. This model was developed from a Bayesian perspective and implemented in statistical software (Goodrich and Lu, 2007), but the Bayesian Tobit model does not take into account missing covariates. Therefore, it would be of interest to develop the Tobit model from the Bayesian perspective in such a way that it can deal with missing covariates, and to compare the results of that model with those of the QTBM. Furthermore, new research to introduce a model that takes into account the country where the sample was analysed would be of high interest. This will determine which country has better performance.

Finally, the QTBM can also be applied to other areas of research such as the estimation of concentrations of compounds in biological samples. Such an example is the cockroach allergen concentration in the homes of asthma sufferers. In this application the analysis of data from a single plate is undertaken by a serial dilution assay. This can result in the concentration being recorded as 'below detection limit'. Standard computer programmes for analysing these data give no estimate at all. This analysis was approached with a Bayesian model developed by Gelman et al. (2004). It is expected the results to be improved by using the QTBM.

8.4 Improved Bayesian Model for Quantitative Responses and Model Comparison Tool

The Bayesian model for quantitative responses provided in Chapter 6 improves the way of treating the censored data in comparison with the methodology presented in Chapter 5. The new methodology is based on describing the likelihood as a function of the observed and censored exact values without having to obtain an estimate for the censored observation. Thus, the complete likelihood can incorporate the censoring mechanism of the data, which may lead to substantial improvements if the censoring mechanism is known. A further advantage is the considerable reduction in the number of nuisance parameters in the CBM in comparison with the QTBM. This makes the CBM less complex for subsequent analysis and interpretation.

Under the working assumption of ignorable censoring mechanism, the CBM model estimates and its application are simplified by the use of a partial likelihood. For this particular situation, no previous research was carried out and published about model comparison tools that can be used to discriminate between nested or non nested models. Therefore, a variant of the existing model comparisons tool Deviance Information Criterion (DIC), the partial DIC (DIC_p) was proposed in this thesis. The behaviour of this proposed measure for model comparison via a simulation study and general conclusions was also investigated by applying the DIC_p to the quantitative data and help decide between two models.

However, there are some areas of concern: the CBM assumes independence for the responses which may not be a reasonable assumption when participants provided repeated measures. The working assumption of ignorable censoring mechanism may be inappropriate in some cases, and the variant of the DIC does not take into account the amount of nuisance parameters due to missing covariates.

Incorporating a correlation matrix structure to the CBM would take into account the dependency of responses. Also, approaching the analysis via other techniques for repeated measurements can be studied from a Bayesian perspective, as explained in previous sections.

The working assumption, which ignores the censoring mechanism can be relaxed by studying the censoring mechanism and proposing a function to be incorporated in the complete likelihood.

The partial DIC proposed in Chapter 7 can be adjusted to obtain a better method for model comparison when the covariates have missing information. An adjusted DIC was studied by Celeux et al. (2006), and the approach presented there can be taken into account to develop an adjusted DIC_p.

The CBM can be applied to other areas of research. An example may be found in industrial applications, where experiments to assess the failure times of springs at cycles of repeated loading under a given stress, in units of 10^3 cycles of loading, are undertaken and resulted data need to be analysed. When, by the end of the experiment, the recorded value is a lower bound for the number of cycles to failure (Cox and Oakes, 1984), then the failure time may be assumed to be log_{10} normally distributed with censored observations. Therefore, the CBM could very well be applied to find out the association of the stress with the failure time.

8.5 General Remarks

It should be mentioned that some of the conclusions about participants' performance reported in Chapters 5 and 6 are obtained with respect to a target viral load estimate provided by QCMD, assuming that this target value is correct. However, if this assumption does not hold true then reported conclusions may be misleading. This situation may be improved if conclusions are obtained with respect to the Bayesian estimated target (as described in Chapter 3).

For the quantitative analysis, another fruitful approach would be to use these models to investigate a measure of performance based on the difference of participants' responses from a 'true' value, instead of participants' estimates of viral loads. In this situation the conclusions obtained from the model would directly reflect the quality of their performance. However, this would depend on the correctness of the 'true' value used to measure these differences. The models proposed for quantitative data are not only applicable to data including one sided censored responses, but they can be easily modified to allow interval censored observations.

Note that extending the models in a way which include two or more pathogens simultaneously is in theory possible, but in practice may not provide a great advantage. Let us consider for example two different approaches to extend the model to two pathogens. A first approach would be including an indicator variable to distinguish between pathogens. This is only an appropriate choice if the remaining covariates included in the model behave the same or similar across pathogens. However, in the analysis it is found that, for example, the association of the variable technology with the responses behave differently across pathogens.

The second approach is to solve the problem presented in the first approach by including additional parameters that take into account the differences in the association of covariates depending on the pathogen analysed. However, as a consequence of the increasing number of parameters, the model complexity and computational cost would increase without gaining further insight into the structure of the data.

A test statistic T has been used in Chapters 4 and 5 for checking models consistency. However, since there is no unique function T to be generally applied to all problems, a function was defined in a way to provide a sensible description of the particular type of data. The test statistic based on this particular choice of function shows that the model provides a good fit to the data. However, there might be other sensible choices of functions which could be used, and it would be of interest to check if they lead to similar conclusions.

The models suggested in this thesis were applied to the data to answer a particular set of questions. However, the models are not limited to just answer the research questions addressed in this thesis. For example, from the point of view of a laboratory it may be of interest to check how good participants' performance are using a certain technology when analysing samples with different viral loads. In this case, the laboratories' results from a panel should be standardised (for having the same scale of values across samples) and introduced in the model as responses using as a covariate an indicator variable for the sample.

It should be mentioned at this point that the conclusions obtained from the model application for a particular pathogen can not be generalised to other pathogens. Furthermore, the laboratory practices have a different effect on the performance depending on the pathogen to be analysed. Although the models suggested in this thesis were only applied to performance of viral samples, they can also be used to analyse the performance of samples of other pathogens or biomarkers such as KRAS.

The conclusions from applying models to QCMD data in this thesis can be considered to reflect well the performance of participants in the geographical area of Europe, as QCMD participants are placed mainly around western Europe and the sample size of the data analysed are large. However, as participants are not randomly selected, the conclusions from the model application cannot be extrapolated for other EQA participants or molecular diagnostic users. Nevertheless, the models proposed in this thesis can be applied to other datasets from EQA providers around the world.

Going back to the motivating scenario, which it was introduced in this thesis (see introduction in Chapter 1), it can be concluded that the quality of the information used by the doctor may depend on the practice within the laboratory which analysed the sample. Thus, a doctor who sends a patient's sample for being analysed with respect to a specific pathogen should be aware of the laboratory practice. Furthermore, the doctor who sends samples to be analysed in order to monitor a treatment should be informed about changes on laboratory practices since this may modified the result provided.

Finally, the research undertaken in this thesis provides a series of statistical methods that can be implemented by EQA providers in order to improve the analysis of the EQA data and provide better recommendations. As a consequence, EQA programmes will gain more participants due to the improvement of the feedbacks received by EQA providers. This will lead to an increase of quality of the laboratories' advice to clinicians, which in turn will lead to better patient diagnoses and care.

Appendix A

Probability Distributions

Here we present the standard notation and probability density functions, means and standard deviations for the probability distributions used in this thesis.

- Normal

$$\theta \sim N(\mu, \sigma^2)$$

$$f(\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(\theta - \mu)^2\right).$$

The $E(\theta) = \mu$ and $Var(\theta) = \sigma^2$

- Gamma

$$\theta \sim Gamma(\alpha, \beta)$$

where the shape parameter $\alpha > 0$ and the inverse scale $\beta > 0$ for $\theta > 0$.

$$f(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{(\alpha-1)} \exp^{-\beta\theta}.$$

The $E(\theta) = \frac{\alpha}{\beta}$ and the $Var(\theta) = \frac{\alpha}{\beta^2}$.

- Inverse Gamma

$$\theta \sim InvGamma(\alpha, \beta)$$

where the shape parameter is $\alpha > 0$ and the scale parameter is $\beta > 0$ for $\theta > 0$.

$$f(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{-(\alpha+1)} \exp^{-\beta/\theta}.$$

The $E(\theta) = \frac{\beta}{\alpha-1}$ for $\alpha > 1$ and the $Var(\theta) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$ for $\alpha > 2$.

- Beta

$$\theta \sim Beta(\alpha, \beta)$$

where $\alpha > 0$ and $\beta > 0$ are 'prior sample size' and $\theta \in [0, 1]$.

$$f(\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}.$$

The $E(\theta) = \frac{\alpha}{\alpha+\beta}$ and $Var(\theta) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

- Dirichlet

$$\theta \sim Dirichlet(\alpha_1, .., \alpha_k)$$

where the 'prior sample sizes' are $\alpha_j > 0$; $\sum_{j=1}^{k} \alpha_j \equiv \alpha_0$.

$$f(\theta) = \frac{\Gamma(\alpha_1 + ... + \alpha_k)}{\Gamma(\alpha_1)...\Gamma(\alpha_k)} \theta_1^{\alpha_1}...(\theta_k)^{\alpha_k}$$

where $\theta_1, .., \theta_k \geq 0$; $\sum_{j=1}^{k} \theta_j = 1$. The $E(\theta_j) = \frac{\alpha_j}{\alpha_0}$ and $Var(\theta) = \frac{\alpha_j(\alpha_0-\alpha_j)}{\alpha_0^2(\alpha_0+1)}$.

- Bernoulli

$$\theta \sim Ber(p)$$

where p is the probability of success, $p \in [0, 1]$ and $\theta = 0, 1$.

$$f(\theta) = p^{\theta}(1-p)^{1-\theta}.$$

The $E(\theta) = p$ and $Var(\theta) = p(1-p)$.

- Categorical

$$\theta \sim Cat(p[])$$

where $\theta = 1, .., dim(p)$; $\sum_{i=1}^{dim(p)} p[i] = 1$ and $f(\theta) = p[\theta]$.

Appendix B

Tables of Results from the QLBM Applied to EV and HBV Data

B.1 Results from the Full QLBM Applied to EV Data

Sample Group Dilution Series 1×10^{-4}

Table B.1: *Summary statistics of the parameter estimates from the full QLBM for EV sample group dilution series* 1×10^{-4}*: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	1.669	0.524	0.658	1.660	2.738		
Year- baseline "2005"							
Year 2004	-0.467	0.347	-1.150	-0.462	0.184	-	No
Technology group- baseline "CIH"							
Tech CC-CIH	0.951	0.789	-0.546	0.939	2.543	+	No
Tech RTIH-CIH	-0.642	0.390	-1.418	-0.644	0.112	-	No
Tech NASBA-CIH	0.242	0.929	-1.553	0.237	2.073	+	No
Tech RTC-CIH	-0.860	0.604	-2.027	-0.867	0.342	-	No
Anti- baseline "No"	0.919	0.472	0.029	0.905	1.869	+	Yes
Accred- baseline "No"	-0.438	0.406	-1.215	-0.446	0.384	-	No
CSF- baseline "No"	0.871	0.558	-0.226	0.874	1.964	+	No
Serum- baseline "No"	-0.472	0.504	-1.428	-0.482	0.551	-	No
Swab- baseline "No"	-0.127	0.530	-1.146	-0.130	0.929	-	No
Biopsies- baseline "No"	-0.131	0.511	-1.133	-0.135	0.869	-	No
Analysis method- baseline Singly							
Analysis Duplicated	0.818	0.377	0.099	0.808	1.580	+	Yes
Analysis Other	0.041	0.993	-1.967	0.050	1.974	+	No
Plasma- baseline 0-10							
Group 1: 11-100	-0.530	0.458	-1.433	-0.530	0.366	-	No
Group 2: 101-1,000	-0.734	0.541	-1.784	-0.737	0.344	-	No
Group 3: 1,001-2,000	1.133	0.698	-0.211	1.116	2.540	+	No
Group 4: 2,001-10,000	-0.068	0.745	-1.513	-0.081	1.429	-	No
Group 5: > 10,000	0.632	0.772	-0.852	0.628	2.228	+	No
Labtype- baseline Hospital							
Public Health	-0.016	0.512	-0.979	-0.032	1.038	-	No
Private	-1.734	0.661	-3.041	-1.728	-0.474	-	Yes
Reference	0.130	0.783	-1.355	0.112	1.749	+	No
Manufacture	0.794	0.832	-0.813	0.791	2.450	+	No
Research	0.522	0.765	-0.971	0.515	2.046	+	No
Inhibition Test-baseline "No"							
Inhbition test Yes	-0.543	0.401	-1.331	-0.538	0.245	-	No
Inhibiton test only Negative samples	-0.853	0.709	-2.228	-0.865	0.566	-	No

Sample Group Dilution Series 1×10^{-5}

Table B.2: *Summary statistics of the parameter estimates from the full QLBM for EV sample group dilution series* 1×10^{-5}*: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	1.053	0.408	0.283	1.042	1.890		
Year- baseline "2005"							
Year 2002	0.579	0.395	-0.195	0.571	1.363	+	No
Year 2003	0.941	0.359	0.259	0.938	1.659	+	Yes
Year 2004	-0.023	0.266	-0.554	-0.021	0.502	-	No
Technology group- baseline "CIH"							
Tech CC-CIH	0.788	0.647	-0.395	0.767	2.123	+	No
Tech RTIH-CIH	-0.519	0.260	-1.036	-0.514	-0.019	-	No
Tech NASBA-CIH	-0.064	0.783	-1.550	-0.088	1.538	-	No
Tech RTC-CIH	-1.418	0.448	-2.306	-1.424	-0.549	-	No
Anti- baseline "No"	0.584	0.312	-0.022	0.579	1.206	+	No
Accred- baseline "No"	0.173	0.276	-0.350	0.169	0.720	+	No
CSF- baseline "No"	0.743	0.367	-0.001	0.746	1.444	+	No
Serum- baseline "No"	-0.205	0.323	-0.826	-0.207	0.439	-	No
Swab- baseline "No"	0.061	0.381	-0.689	0.058	0.799	+	No
Biopsies- baseline "No"	0.637	0.349	-0.065	0.637	1.323	+	No
Analysis method- baseline Singly							
Analysis Duplicated	0.091	0.236	-0.375	0.093	0.551	+	No
Analysis Other	0.915	0.749	-0.479	0.883	2.468	+	No
Plasma- baseline 0-10							
Group 1: 11-100	-0.259	0.319	-0.877	-0.260	0.385	-	No
Group 2: 101-1,000	0.067	0.369	-0.652	0.070	0.804	+	No
Group 3: 1,001-2,000	0.878	0.582	-0.241	0.865	2.065	+	No
Group 4: 2,001-10,000	0.551	0.625	-0.634	0.535	1.828	+	No
Group 5: > 10,000	0.210	0.679	-1.081	0.191	1.565	+	No
Labtype- baseline Hospital							
Public Health	1.010	0.413	0.236	0.996	1.844	+	Yes
Private	-0.368	0.472	-1.253	-0.380	0.579	-	No
Reference	0.457	0.621	-0.709	0.430	1.727	+	No
Manufacture	1.443	0.710	0.099	1.423	2.883	+	Yes
Research	1.015	0.595	-0.101	1.003	2.255	+	No
Inhibition Test-baseline "No"							
Inhbition test Yes	-0.137	0.256	-0.636	-0.134	0.360	-	No
Inhibiton test only Negative samples	0.023	0.525	-0.976	0.011	1.090	+	No

Sample Group Dilution Series 1×10^{-6}

Table B.3: *Summary statistics of the parameter estimates from the full QLBM for EV sample group dilution series* 1×10^{-6}*: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	1.056	0.342	0.376	1.050	1.740		
Year- baseline "2005"							
Year 2002	-0.602	0.269	-1.140	-0.601	-0.077	-	Yes
Year 2003	-0.078	0.265	-0.597	-0.079	0.450	-	No
Year 2004	-0.374	0.219	-0.811	-0.374	0.049	-	No
Technology group- baseline "CIH"							
Tech CC-CIH	0.580	0.403	-0.167	0.564	1.404	+	No
Tech RTIH-CIH	-0.064	0.177	-0.408	-0.065	0.285	-	No
Tech NASBA-CIH	-0.728	0.469	-1.667	-0.728	0.193	-	No
Tech RTC-CIH	-0.728	0.375	-1.455	-0.732	0.010	-	No
Anti- baseline "No"	0.012	0.195	-0.365	0.015	0.395	+	No
Accred- baseline "No"	-0.156	0.194	-0.533	-0.158	0.220	-	No
CSF- baseline "No"	0.591	0.298	-0.034	0.601	1.152	+	No
Serum- baseline "No"	-0.216	0.227	-0.663	-0.217	0.241	-	No
Swab- baseline "No"	0.201	0.289	-0.362	0.200	0.778	+	No
Biopsies- baseline "No"	0.277	0.270	-0.257	0.276	0.798	+	No
Analysis method- baseline Singly							
Analysis Duplicated	0.060	0.156	-0.246	0.058	0.374	+	No
Analysis Other	1.291	0.512	0.337	1.272	2.341	+	Yes
Plasma- baseline 0-10							
Group 1: 11-100	-0.399	0.273	-0.929	-0.401	0.135	-	No
Group 2: 101-1,000	0.145	0.292	-0.401	0.139	0.719	+	No
Group 3: 1,001-2,000	-0.321	0.358	-1.011	-0.328	0.384	-	No
Group 4: 2,001-10,000	0.094	0.447	-0.766	0.088	0.974	+	No
Group 5: > 10,000	0.071	0.614	-1.087	0.056	1.327	+	No
Labtype- baseline Hospital							
Public Health	0.306	0.228	-0.135	0.303	0.762	+	No
Private	0.085	0.387	-0.668	0.080	0.855	+	No
Reference	0.777	0.462	-0.089	0.763	1.732	+	No
Manufacture	2.098	0.568	1.058	2.075	3.272	+	Yes
Research	0.781	0.443	-0.064	0.767	1.676	+	No
Inhibition Test-baseline "No"							
Inhbition test Yes	-0.301	0.173	-0.635	-0.300	0.032	-	No
Inhibiton test only Negative samples	-0.108	0.326	-0.746	-0.112	0.541	-	No

Sample Group Dilution Series 1×10^{-7}

Table B.4: *Summary statistics of the parameter estimates from the full QLBM for EV sample group dilution series* 1×10^{-7}*: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	0.008	0.412	-0.836	0.011	0.822		
Year- baseline "2005"							
Year 2002	-0.846	0.282	-1.400	-0.842	-0.302	-	Yes
Year 2003	1.323	0.336	0.663	1.321	1.972	+	Yes
Year 2004	0.739	0.269	0.215	0.735	1.269	+	Yes
Technology group- baseline "CIH"							
Tech CC-CIH	-0.839	0.440	-1.707	-0.840	0.041	-	No
Tech RTIH-CIH	-0.095	0.223	-0.536	-0.097	0.345	-	No
Tech NASBA-CIH	-0.411	0.550	-1.491	-0.420	0.677	-	No
Tech RTC-CIH	-1.368	0.474	-2.312	-1.362	-0.453	-	Yes
Anti- baseline "No"	-0.102	0.244	-0.586	-0.103	0.379	-	No
Accred- baseline "No"	0.314	0.230	-0.126	0.310	0.770	+	No
CSF- baseline "No"	0.354	0.371	-0.419	0.363	1.079	+	No
Serum- baseline "No"	-0.067	0.278	-0.614	-0.063	0.477	-	No
Swab- baseline "No"	0.360	0.341	-0.287	0.359	1.038	+	No
Biopsies- baseline "No"	0.308	0.333	-0.360	0.313	0.954	+	No
Analysis method- baseline Singly							
Analysis Duplicated	0.033	0.195	-0.348	0.031	0.416	+	No
Analysis Other	1.043	0.555	-0.023	1.027	2.164	+	No
Plasma- baseline 0-10							
Group 1: 11-100	-0.640	0.327	-1.286	-0.640	-0.018	-	No
Group 2: 101-1,000	-0.063	0.356	-0.743	-0.064	0.643	-	No
Group 3: 1,001-2,000	-0.562	0.449	-1.451	-0.561	0.320	-	No
Group 4: 2,001-10,000	-0.635	0.506	-1.617	-0.634	0.371	-	No
Group 5: > 10,000	-0.679	0.667	-1.984	-0.697	0.672	-	No
Labtype- baseline Hospital							
Public Health	0.069	0.304	-0.514	0.070	0.665	+	No
Private	-0.434	0.458	-1.344	-0.436	0.453	-	No
Reference	0.178	0.498	-0.785	0.168	1.173	+	No
Manufacture	1.468	0.535	0.417	1.464	2.529	+	Yes
Research	1.002	0.555	-0.074	0.994	2.113	+	No
Inhibition Test-baseline "No"							
Inhbition test Yes	-0.280	0.217	-0.701	-0.280	0.148	-	No
Inhibiton test only Negative samples	0.712	0.445	-0.147	0.705	1.599	+	No

Sample Group Dilution Series 1×10^{-8}

Table B.5: *Summary statistics of the parameter estimates from the full QLBM for EV sample group dilution series* 1×10^{-8}*: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	-0.948	0.468	-1.848	-0.954	-0.020		
Year- baseline "2005"							
Year 2002	-0.069	0.377	-0.814	-0.070	0.664	-	No
Year 2003	0.810	0.348	0.127	0.805	1.501	+	Yes
Year 2004	0.273	0.307	-0.332	0.275	0.874	+	No
Technology group- baseline "CIH"							
Tech CC-CIH	-0.540	0.561	-1.691	-0.528	0.551	-	No
Tech RTIH-CIH	0.192	0.275	-0.351	0.195	0.728	+	No
Tech NASBA-CIH	-0.568	0.737	-2.071	-0.565	0.848	-	No
Tech RTC-CIH	-0.062	0.571	-1.229	-0.049	1.001	-	No
Anti- baseline "No"	0.026	0.308	-0.579	0.027	0.627	+	No
Accred- baseline "No"	-0.280	0.293	-0.848	-0.282	0.295	-	No
CSF- baseline "No"	-0.263	0.389	-1.004	-0.271	0.515	-	No
Serum- baseline "No"	0.526	0.344	-0.134	0.525	1.223	+	No
Swab- baseline "No"	0.023	0.406	-0.771	0.029	0.809	+	No
Biopsies- baseline "No"	-0.174	0.410	-0.973	-0.171	0.619	-	No
Analysis method- baseline Singly							
Analysis Duplicated	-0.334	0.248	-0.821	-0.333	0.138	-	No
Analysis Other	0.068	0.620	-1.187	0.076	1.255	+	No
Plasma- baseline 0-10							
Group 1: 11-100	-0.640	0.327	-1.286	-0.640	-0.018	-	No
Group 2: 101-1,000	-0.063	0.356	-0.743	-0.064	0.643	-	No
Group 3: 1,001-2,000	-0.562	0.449	-1.451	-0.561	0.320	-	No
Group 4: 2,001-10,000	-0.635	0.506	-1.617	-0.634	0.371	-	No
Group 5: > 10,000	-0.679	0.667	-1.984	-0.697	0.672	-	No
Labtype- baseline Hospital							
Public Health	0.067	0.372	-0.689	0.073	0.775	+	No
Private	0.753	0.503	-0.238	0.758	1.739	+	No
Reference	0.004	0.596	-1.164	0.011	1.157	+	No
Manufacture	0.369	0.686	-1.028	0.385	1.663	+	No
Research	0.465	0.582	-0.705	0.480	1.591	+	No
Inhibition Test-baseline "No"							
Inhbition test Yes	-0.251	0.273	-0.789	-0.249	0.282	-	No
Inhibiton test only Negative samples	-0.512	0.500	-1.538	-0.498	0.432	-	No

Sample Group Negative

Table B.6: *Summary statistics of the parameter estimates from the full QLBM for EV negative sample group: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	1.895	0.556	0.822	1.893	2.986		
Year- baseline "2005"							
Year 2002	-0.137	0.487	-1.090	-0.137	0.822	-	No
Year 2003	-0.167	0.456	-1.050	-0.167	0.728	-	No
Year 2004	0.157	0.431	-0.686	0.152	0.998	+	No
Technology group- baseline "CIH"							
Tech CC-CIH	-0.153	0.660	-1.398	-0.171	1.193	-	No
Tech RTIH-CIH	0.307	0.398	-0.457	0.306	1.112	+	No
Tech NASBA-CIH	0.657	0.839	-0.915	0.642	2.331	+	No
Tech RTC-CIH	0.644	0.807	-0.824	0.612	2.316	+	No
Anti- baseline "No"	0.036	0.445	-0.789	0.019	0.940	+	No
Accred- baseline "No"	0.164	0.388	-0.583	0.165	0.936	+	No
CSF- baseline "No"	0.773	0.470	-0.153	0.770	1.679	+	No
Serum- baseline "No"	-0.389	0.403	-1.183	-0.392	0.408	-	No
Swab- baseline "No"	0.307	0.460	-0.593	0.308	1.214	+	No
Biopsies- baseline "No"	0.179	0.467	-0.722	0.175	1.101	-	No
Analysis method- baseline Singly							
Analysis Duplicated	0.239	0.346	-0.435	0.236	0.933	+	No
Analysis Other	1.091	0.763	-0.344	1.077	2.639	+	No
Plasma- baseline 0-10							
Group 1: 11-100	-0.027	0.464	-0.926	-0.028	0.889	-	No
Group 2: 101-1,000	0.544	0.479	-0.380	0.552	1.471	+	No
Group 3: 1,001-2,000	-0.173	0.561	-1.234	-0.193	0.939	-	No
Group 4: 2,001-10,000	-0.441	0.664	-1.707	-0.464	0.870	-	No
Group 5: > 10,000	0.415	0.901	-1.297	0.402	2.222	+	No
Labtype- baseline Hospital							
Public Health	-0.392	0.425	-1.227	-0.402	0.449	-	No
Private	0.228	0.728	-1.142	0.208	1.748	+	No
Reference	-0.191	0.694	-1.485	-0.219	1.196	-	No
Manufacture	0.824	0.787	-0.638	0.814	2.443	+	No
Research	-0.085	0.631	-1.278	-0.100	1.183	-	No
Inhibition Test-baseline "No"							
Inhbition test Yes	0.675	0.406	-0.109	0.668	1.487	+	No
Inhibiton test only Negative samples	-1.095	0.484	-2.016	-1.104	-0.115	-	No

Sample Group Non-EV

Table B.7: *Summary statistics of the parameter estimates from the full QLBM for Non-EV sample group: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	0.888	0.446	0.035	0.882	1.777		
Year- baseline "2005"							
Year 2003	-0.325	0.336	-0.987	-0.326	0.336	-	No
Year 2004	0.095	0.353	-0.583	0.090	0.805	+	No
Technology group- baseline "CIH"							
Tech CC-CIH	-0.434	0.500	-1.393	-0.440	0.595	-	No
Tech RTIH-CIH	0.890	0.329	0.259	0.888	1.539	+	Yes
Tech NASBA-CIH	0.559	0.863	-1.075	0.536	2.294	+	No
Tech RTC-CIH	0.772	0.662	-0.470	0.754	2.135	+	No
Anti- baseline "No"	0.035	0.364	-0.667	0.029	0.757	+	No
Accred- baseline "No"	0.094	0.337	-0.569	0.093	0.759	+	No
CSF- baseline "No"	0.116	0.417	-0.722	0.123	0.921	+	No
Serum- baseline "No"	0.305	0.379	-0.440	0.304	1.043	+	No
Swab- baseline "No"	-0.243	0.430	-1.061	-0.249	0.609	-	No
Biopsies- baseline "No"	0.612	0.447	-0.263	0.609	1.490	+	No
Analysis method- baseline Singly							
Analysis Duplicated	-0.130	0.289	-0.690	-0.132	0.436	-	No
Analysis Other	1.114	0.753	-0.292	1.092	2.615	+	No
Plasma- baseline 0-10							
Group 1: 11-100	0.422	0.433	-0.411	0.412	1.303	+	No
Group 2: 101-1,000	0.244	0.412	-0.556	0.236	1.075	+	No
Group 3: 1,001-2,000	0.178	0.529	-0.859	0.174	1.221	+	No
Group 4: 2,001-10,000	0.385	0.725	-1.017	0.368	1.846	+	No
Group 5: > 10,000	-0.349	0.713	-1.721	-0.361	1.105	-	No
Labtype- baseline Hospital							
Public Health	0.429	0.428	-0.379	0.418	1.308	+	No
Private	-0.057	0.579	-1.134	-0.078	1.112	-	No
Reference	0.733	0.788	-0.748	0.710	2.335	+	No
Manufacture	0.595	0.808	-0.885	0.564	2.287	+	No
Research	0.059	0.657	-1.166	0.039	1.434	+	No
Inhibition Test-baseline "No"							
Inhbition test Yes	0.752	0.324	0.135	0.742	1.411	+	Yes
Inhibiton test only							
Negative samples	-0.494	0.494	-1.444	-0.505	0.505	-	No

B.2 Results from the Reduced QLBM Applied to EV Data

Sample Group Dilution Series 1×10^{-3}

Table B.8: *Summary statistics of the parameter estimates from the reduced QLBM for EV sample group dilution series* 1×10^{-3}*: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	1.861	0.518	0.901	1.845	2.908		
CSF- baseline "No"	1.591	0.608	0.355	1.595	2.743	+	Yes

Sample Group Dilution Series 1×10^{-4}

Table B.9: *Summary statistics of the parameter estimates from the reduced QLBM for EV sample group dilution series* 1×10^{-4}*: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	1.156	0.494	0.184	1.153	2.155		
Technology group- baseline "CIH"							
Tech CC-CIH	0.843	0.799	-0.642	0.816	2.482	+	No
Tech RTIH-CIH	-0.604	0.384	-1.363	-0.602	0.153	-	No
Tech NASBA-CIH	0.244	0.926	-1.531	0.236	2.064	+	No
Tech RTC-CIH	-0.897	0.596	-2.055	-0.897	0.298	-	No
Anti- baseline "No"	0.739	0.455	-0.142	0.737	1.652	+	No
CSF- baseline "No"	0.649	0.544	-0.468	0.660	1.712	+	No
Serum- baseline "No"	-0.484	0.498	-1.442	-0.491	0.501	-	No
Swab- baseline "No"	0.054	0.545	-0.995	0.051	1.124	+	No
Biopsies- baseline "No"	-0.208	0.520	-1.263	-0.199	0.799	-	No
Analysis method- baseline Singly							
Analysis Duplicated	0.737	0.366	0.037	0.735	1.465	+	Yes
Analysis Other	0.020	0.994	-1.913	0.000	1.986	+	No
Plasma- baseline 0-10							
Group 1: 11-100	-0.474	0.452	-1.369	-0.481	0.412	-	No
Group 2: 101-1,000	-0.738	0.532	-1.782	-0.739	0.308	-	No
Group 3: 1,001-2,000	1.112	0.711	-0.226	1.083	2.568	+	No
Group 4: 2,001-10,000	-0.060	0.740	-1.477	-0.080	1.408	-	No
Group 5: > 10,000	0.610	0.769	-0.880	0.597	2.173	+	No
Labtype- baseline Hospital							
Public Health	-0.053	0.500	-1.004	-0.061	0.951	-	No
Private	-1.764	0.640	-3.019	-1.757	-0.541	-	Yes
Reference	0.168	0.762	-1.303	0.158	1.680	+	No
Manufacture	0.728	0.849	-0.883	0.704	2.448	+	No
Research	0.361	0.760	-1.090	0.342	1.912	+	No

Sample Group Dilution Series 1×10^{-5}

Table B.10: *Summary statistics of the parameter estimates from the reduced QLBM for EV sample group dilution series* 1×10^{-5}*: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	1.068	0.372	0.320	1.063	1.792		
Year- baseline "2005"							
Year 2002	0.616	0.395	-0.146	0.611	1.397	+	No
Year 2003	0.959	0.359	0.281	0.953	1.679	+	Yes
Year 2004	-0.023	0.266	-0.552	-0.023	0.500	-	No
Technology group- baseline "CIH"							
Tech CC-CIH	0.760	0.644	-0.416	0.731	2.097	+	No
Tech RTIH-CIH	-0.562	0.250	-1.046	-0.564	-0.067	-	Yes
Tech NASBA-CIH	-0.094	0.765	-1.521	-0.126	1.487	-	No
Tech RTC-CIH	-1.437	0.419	-2.233	-1.448	-0.595	-	Yes
Anti- baseline "No"	0.561	0.302	-0.004	0.558	1.173	+	No
CSF- baseline "No"	0.740	0.356	0.039	0.744	1.415	+	Yes
Serum- baseline "No"	-0.179	0.302	-0.765	-0.183	0.416	-	No
Swab- baseline "No"	0.128	0.377	-0.604	0.130	0.862	+	No
Biopsies- baseline "No"	0.592	0.369	-0.125	0.587	1.344	+	No
Analysis method- baseline Singly							
Analysis Duplicated	0.077	0.229	-0.361	0.077	0.527	+	No
Analysis Other	0.909	0.769	-0.506	0.877	2.463	+	No
Plasma- baseline 0-10							
Group 1: 11-100	-0.269	0.311	-0.872	-0.268	0.340	-	No
Group 2: 101-1,000	0.028	0.370	-0.700	0.023	0.751	+	No
Group 3: 1,001-2,000	0.891	0.593	-0.208	0.863	2.131	+	No
Group 4: 2,001-10,000	0.540	0.624	-0.634	0.521	1.806	+	No
Group 5: > 10,000	0.237	0.672	-1.048	0.217	1.605	+	No
Labtype- baseline Hospital							
Public Health	1.033	0.403	0.279	1.015	1.864	+	Yes
Private	-0.312	0.461	-1.168	-0.330	0.644	-	No
Reference	0.497	0.611	-0.647	0.478	1.733	+	No
Manufacture	1.397	0.704	0.091	1.367	2.852	+	Yes
Research	1.027	0.574	-0.035	1.008	2.214	+	No

Sample Group Dilution Series 1×10^{-6}

Table B.11: *Summary statistics of the parameter estimates from the reduced QLBM for EV sample group dilution series* 1×10^{-6}*: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	0.852	0.324	0.251	0.840	1.512		
Year- baseline "2005"							
Year 2002	-0.558	0.251	-1.051	-0.559	-0.049	-	Yes
Year 2003	-0.051	0.250	-0.544	-0.052	0.440	-	No
Year 2004	-0.361	0.214	-0.792	-0.358	0.051	-	No
Technology group- baseline "CIH"							
Tech CC-CIH	0.439	0.398	-0.303	0.426	1.270	+	No
Tech RTIH-CIH	-0.116	0.170	-0.453	-0.117	0.213	-	No
Tech NASBA-CIH	-0.891	0.460	-1.784	-0.890	0.022	-	No
Tech RTC-CIH	-0.850	0.369	-1.565	-0.856	-0.115	-	Yes
CSF- baseline "No"	0.604	0.293	-0.009	0.613	1.138	+	No
Serum- baseline "No"	-0.210	0.217	-0.629	-0.212	0.222	-	No
Swab- baseline "No"	0.268	0.285	-0.310	0.275	0.810	+	No
Biopsies- baseline "No"	0.229	0.284	-0.339	0.226	0.786	+	No
Analysis method- baseline Singly							
Analysis Duplicated	0.052	0.156	-0.253	0.052	0.355	+	No
Analysis Other	1.279	0.511	0.345	1.256	2.339	+	Yes
Plasma- baseline 0-10							
Group 1: 11-100	-0.384	0.264	-0.897	-0.384	0.133	-	No
Group 2: 101-1,000	0.115	0.300	-0.480	0.119	0.686	+	No
Group 3: 1,001-2,000	-0.294	0.367	-1.015	-0.290	0.426	-	No
Group 4: 2,001-10,000	0.060	0.458	-0.835	0.055	0.957	+	No
Group 5: > 10,000	0.046	0.627	-1.155	0.041	1.307	+	No
Labtype- baseline Hospital							
Public Health	0.305	0.227	-0.144	0.306	0.746	+	No
Private	0.082	0.383	-0.640	0.077	0.839	+	No
Reference	0.823	0.457	-0.048	0.817	1.763	+	No
Manufacture	2.086	0.561	1.034	2.066	3.255	+	Yes
Research	0.823	0.441	-0.018	0.818	1.714	+	No

Sample Group Dilution Series 1×10^{-7}

Table B.12: *Summary statistics of the parameter estimates from the reduced QLBM for EV sample group dilution series* 1×10^{-7}*: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	0.070	0.378	-0.663	0.072	0.794		
Year- baseline "2005"							
Year 2002	-0.762	0.276	-1.316	-0.759	-0.235	-	Yes
Year 2003	1.322	0.335	0.671	1.314	1.992	+	Yes
Year 2004	0.705	0.263	0.201	0.701	1.229	+	Yes
Technology group- baseline "CIH"							
Tech CC-CIH	-0.880	0.429	-1.716	-0.880	-0.040	-	Yes
Tech RTIH-CIH	-0.231	0.212	-0.635	-0.232	0.180	-	No
Tech NASBA-CIH	-0.507	0.546	-1.588	-0.506	0.556	-	No
Tech RTC-CIH	-1.448	0.477	-2.406	-1.442	-0.517	-	Yes
CSF- baseline "No"	0.381	0.379	-0.371	0.386	1.098	+	No
Serum- baseline "No"	-0.034	0.273	-0.564	-0.035	0.501	-	No
Swab- baseline "No"	0.369	0.356	-0.324	0.369	1.038	+	No
Biopsies- baseline "No"	0.266	0.353	-0.438	0.268	0.959	+	No
Analysis method- baseline Singly							
Analysis Duplicated	-0.023	0.195	-0.403	-0.025	0.362	-	No
Analysis Other	0.984	0.551	-0.067	0.974	2.107	+	No
Plasma- baseline 0-10							
Group 1: 11-100	-0.630	0.315	-1.245	-0.633	-0.010	-	Yes
Group 2: 101-1,000	-0.081	0.348	-0.761	-0.085	0.616	-	No
Group 3: 1,001-2,000	-0.500	0.445	-1.378	-0.499	0.383	-	No
Group 4: 2,001-10,000	-0.639	0.511	-1.633	-0.638	0.371	-	No
Group 5: > 10,000	-0.654	0.650	-1.925	-0.652	0.641	-	No
Labtype- baseline Hospital							
Public Health	0.185	0.294	-0.392	0.180	0.765	+	No
Private	-0.388	0.444	-1.239	-0.391	0.470	-	No
Reference	0.278	0.485	-0.650	0.262	1.266	+	No
Manufacture	1.416	0.516	0.420	1.406	2.438	+	Yes
Research	1.040	0.534	0.011	1.028	2.127	+	Yes

Sample Group Dilution Series 1×10^{-8}

Table B.13: *Summary statistics of the parameter estimates from the reduced QLBM for EV sample group dilution series* 1×10^{-8}*: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	-1.061	0.389	-1.842	-1.052	-0.298		
Year- baseline "2005"							
Year 2002	0.027	0.322	-0.595	0.024	0.669	+	No
Year 2003	0.795	0.307	0.204	0.796	1.391	+	Yes
Year 2004	0.265	0.294	-0.312	0.263	0.856	+	No
Technology group- baseline "CIH"							
Tech CC-CIH	-0.657	0.535	-1.730	-0.638	0.361	-	No
Tech RTIH-CIH	0.114	0.249	-0.371	0.111	0.599	+	No
Tech NASBA-CIH	-0.573	0.703	-2.023	-0.539	0.727	-	No
Tech RTC-CIH	-0.110	0.546	-1.194	-0.102	0.936	-	No
CSF- baseline "No"	-0.437	0.385	-1.214	-0.432	0.303	-	No
Serum- baseline "No"	0.499	0.318	-0.130	0.499	1.121	+	No
Swab- baseline "No"	0.035	0.388	-0.740	0.044	0.782	+	No
Biopsies- baseline "No"	-0.106	0.388	-0.870	-0.113	0.655	-	No

Negative Sample Group

Table B.14: *Summary statistics of the parameter estimates from the reduced QLBM for EV Negative sample group: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	2.004	0.349	1.341	1.991	2.723		
CSF- baseline "No"	0.870	0.388	0.086	0.874	1.628	+	Yes
Inhibition Test-baseline "No"							
Inhbition test Yes	0.662	0.379	-0.058	0.649	1.421	+	No
Inhibiton test only Negative samples	-1.075	0.433	-1.872	-1.092	-0.192	-	Yes

Non-EV Sample Group

Table B.15: *Summary statistics of the parameter estimates from the reduced QLBM for Non-EV sample group: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	1.156	0.375	0.450	1.154	1.934		
Technology group- baseline "CIH"							
Tech CC-CIH	-0.481	0.482	-1.397	-0.496	0.468	-	No
Tech RTIH-CIH	0.827	0.315	0.224	0.822	1.458	+	Yes
Tech NASBA-CIH	0.667	0.827	-0.884	0.637	2.383	+	No
Tech RTC-CIH	0.824	0.653	-0.357	0.787	2.186	+	No
CSF- baseline "No"	-0.042	0.402	-0.857	-0.037	0.723	-	No
Serum- baseline "No"	0.351	0.363	-0.362	0.340	1.069	+	No
Swab- baseline "No"	-0.200	0.415	-1.001	-0.207	0.626	-	No
Biopsies- baseline "No"	0.592	0.416	-0.222	0.590	1.401	+	No
Inhibition Test-baseline "No"							
Inhbition test Yes	0.676	0.306	0.089	0.671	1.300	+	Yes
Inhibiton test only Negative samples	-0.461	0.450	-1.314	-0.480	0.452	-	No

B.3 EV Estimated Probabilities

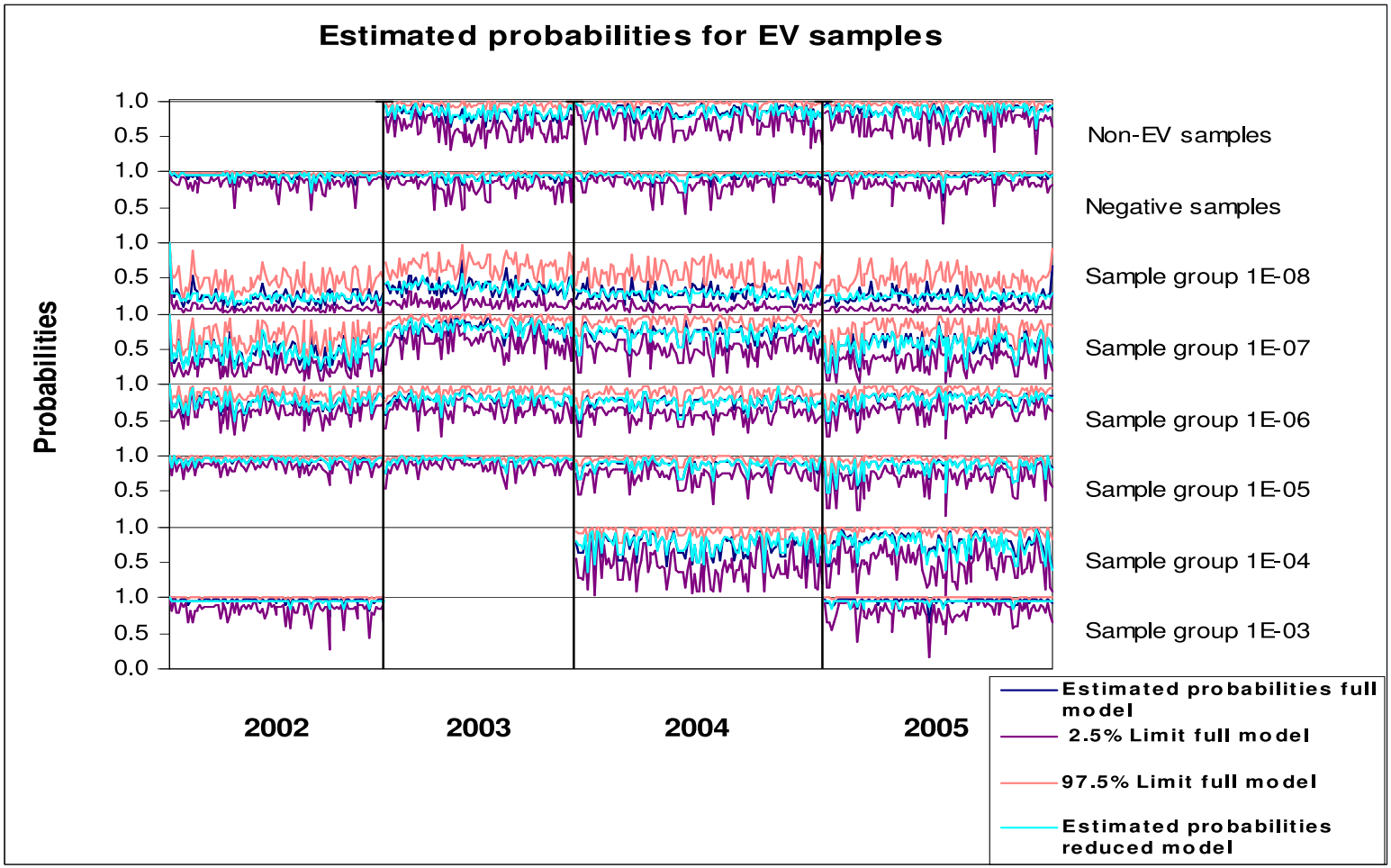


Figure B.1: *EV estimated probabilities from the full and reduced QLBM.*

B.4 Results from the Full QLBM Applied to HBV Data

In this section is presented a summary of the results obtained from the full QLBM applied to HBV data. Table B.16 summaries the mean and standard deviation from the posterior distribution of the parameter estimates per sample group and covariate level. In next section is shown the complete tables of results obtained from the reduced QLBM model.

Table B.16: *Estimated mean and standard deviation of the parameter estimates from the full QLBM applied to the HBV data.*

Covariates	Mean (SD)						
	6	5	4	3.5	3	2.3	Negative
Intercept	1.950 (0.751)	1.885 (0.578)	1.316 (0.566)	1.726 (0.631)	0.505 (0.506)	0.205 (0.483)	2.040 (0.647)
Year- baseline "2005"							
Year 2002	0.260 (0.787)	-0.576 (0.474)	-0.736 (0.514)	**-0.217 (0.427)**		-0.019 (0.357)	0.621 (0.643)
Year 2003	1.093 (0.800)	0.466 (0.556)	**0.049 (0.595)**		0.261 (0.447)	-0.554 (0.355)	0.379 (0.667)
Year 2004	-0.001 (0.725)	0.214 (0.509)	0.219 (0.591)		**1.100 (0.345)**		0.067 (0.600)
Subtype- baseline "A"	0.929 (0.736)	0.041 (0.368)	0.137 (0.556)		-0.211 (0.307)		
Technology group- baseline "CC"							
Tech CIH-CC	0.057 (0.715)	0.352 (0.533)	0.283 (0.556)	-0.062 (0.518)	-0.046 (0.444)	-0.122 (0.388)	0.520 (0.659)
Tech RTIH-CC	0.751 (0.817)	0.901 (0.601)	0.673 (0.579)	1.050 (0.663)	-0.381 (0.408)	0.573 (0.403)	-0.627 (0.611)
Tech RTC-CC	0.707 (0.822)	1.071 (0.659)	1.080 (0.652)	0.046 (0.783)	0.705 (0.455)	0.728 (0.547)	-0.205 (0.652)
Tech bDNA-CC	-0.162 (0.829)	**-2.568 (0.501)**	**-1.639 (0.575)**	**-3.301 (0.601)**	**-1.174 (0.520)**	**-2.621 (0.624)**	-0.171 (0.726)
Tech HC-CC	0.231 (0.940)	**-1.842 (0.621)**	**-3.146 (0.651)**	**-3.225 (0.633)**	**-1.314 (0.715)**	**-1.588 (0.677)**	-0.234 (0.849)
Tech TMA-CC	0.045 (0.985)	0.320 (0.930)	0.228 (0.959)	0.394 (0.937)	-	0.474 (0.910)	0.153 (0.960)
Anti- baseline "No"	0.609 (0.700)	0.755 (0.463)	**1.064 (0.472)**	**1.164 (0.507)**	**1.306 (0.356)**	0.501 (0.335)	**1.128 (0.554)**
Accred- baseline "No"	1.053 (0.727)	**1.662 (0.508)**	**1.495 (0.483)**	0.525 (0.506)	**0.763 (0.364)**	0.394 (0.336)	0.640 (0.615)
OthrSpc.- baseline "No"	0.925 (0.808)	**1.335 (0.614)**	**1.481 (0.586)**	0.503 (0.623)	-0.006 (0.398)	0.282 (0.386)	-0.327 (0.632)
Analysis method- baseline Singly							
Analysis Duplicated	1.017 (0.780)	0.347 (0.448)	-0.031 (0.462)	0.063 (0.476)	0.465 (0.387)	0.112 (0.352)	0.473 (0.607)
Analysis Other	0.234 (0.917)	0.605 (0.830)	0.592 (0.830)	0.499 (0.848)	-0.519 (0.597)	**1.323 (0.721)**	-0.200 (0.826)
Plasma- baseline 0-10							
Group 1: 11-100	0.453 (0.880)	0.688 (0.686)	0.945 (0.682)	0.798 (0.685)	1.097 (0.663)	0.013 (0.475)	-0.151 (0.823)
Group 2: 101-1,000	0.189 (0.764)	0.879 (0.584)	1.055 (0.548)	0.701 (0.572)	0.600 (0.431)	0.034 (0.370)	0.608 (0.715)
Group 3: 1,001-2,000	0.385 (0.892)	0.009 (0.686)	0.555 (0.732)	-0.175 (0.695)	1.345 (0.755)	-0.315 (0.556)	0.773 (0.867)
Group 4: 2,001-10,000	0.299 (0.941)	0.524 (0.741)	0.656 (0.701)	0.085 (0.739)	0.425 (0.545)	0.940 (0.599)	-0.590 (0.750)
Group 5: > 10,000	-0.260 (1.072)	-0.363 (1.154)	0.029 (0.986)	-0.053 (0.985)	-0.195 (0.694)	-0.079 (0.743)	-0.012 (1.006)
Serum- baseline 0-10							
Group 1: 11-100	-0.079 (0.827)	0.532 (0.677)	0.799 (0.693)	0.551 (0.696)	0.901 (0.607)	0.434 (0.476)	0.883 (0.819)
Group 2: 101-1,000	0.809 (0.839)	0.164 (0.562)	0.149 (0.522)	0.329 (0.544)	-0.197 (0.435)	0.369 (0.381)	0.379 (0.683)
Group 3: 1,001-2,000	0.381 (0.886)	0.129 (0.672)	0.696 (0.722)	0.286 (0.677)	0.328 (0.585)	-0.228 (0.493)	0.670 (0.831)
Group 4: 2,001-10,000	-0.068 (0.956)	-0.566 (0.741)	-0.378 (0.661)	-0.934 (0.784)	-0.426 (0.515)	-1.047 (0.551)	-0.055 (0.743)
Group 5: > 10,000	0.109 (0.964)	0.832 (0.865)	0.831 (0.845)	-0.013 (1.013)	-0.627 (0.681)	-0.028 (0.891)	0.273 (0.934)
Labtype- baseline Hospital							
Public Health	0.547 (0.862)	0.682 (0.705)	0.604 (0.693)	0.526 (0.710)	0.294 (0.531)	0.727 (0.512)	0.881 (0.798)
Private	0.373 (0.917)	0.312 (0.661)	-0.228 (0.588)	0.248 (0.688)	-0.487 (0.460)	-0.083 (0.482)	0.888 (0.800)
Reference	-0.493 (1.111)	-1.249 (0.947)	-0.856 (0.925)	-0.179 (0.768)	-1.121 (0.954)	0.210 (0.711)	-0.357 (1.063)
Manufacture	0.051 (1.014)	0.491 (0.960)	0.456 (0.927)	0.576 (0.892)	0.198 (0.994)	0.180 (0.774)	-0.562 (0.923)
Research	0.335 (0.914)	0.752 (0.836)	0.213 (0.783)	0.591 (0.912)	-0.476 (0.567)	-0.186 (0.654)	0.476 (0.880)
Inhibition Test-baseline "No"							
Inhbition test Yes	1.149 (0.709)	-0.009 (0.420)	0.033 (0.440)	-0.317 (0.474)	0.375 (0.351)	0.188 (0.324)	1.071 (0.564)
Inhibiton test only Negative samples	0.418 (0.886)	0.322 (0.749)	0.477 (0.756)	-0.007 (0.736)	0.268 (0.693)	0.697 (0.634)	-0.472 (0.773)

B.5 Results from the Reduced QLBM Applied to HBV Data

Sample Group 6 $\log_{10}$ Copies/ml Viral Load

Table B.17: *Summary statistics of the parameter estimates from the reduced QLBM for HBV sample group 6 Log_{10} copies/ml viral load: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	3.870	0.325	3.268	3.858	4.542		

Sample Group 5 $\log_{10}$ Copies/ml Viral Load

Table B.18: *Summary statistics of the parameter estimates from the reduced QLBM for HBV sample group 5 Log_{10} copies/ml viral load: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	2.528	0.289	1.978	2.524	3.125		
Technology group-baseline "CC"							
Tech CIH-CC	0.180	0.479	-0.723	0.164	1.168	+	No
Tech RTIH-CC	0.992	0.558	-0.022	0.964	2.177	+	No
Tech RTC-CC	1.049	0.616	-0.087	1.018	2.319	+	No
Tech bDNA-CC	-2.512	0.438	-3.360	-2.513	-1.645	-	Yes
Tech HC-CC	-2.064	0.523	-3.077	-2.069	-1.038	-	Yes
Tech TMA-CC	0.296	0.892	-1.379	0.274	2.090	+	No
Accred- baseline "No"	1.505	0.454	0.655	1.494	2.444	+	Yes
OthrSpc. baseline "No"	1.072	0.518	0.096	1.061	2.116	+	Yes

Sample Group 4 $\log_{10}$ Copies/ml Viral Load

Table B.19: *Summary statistics of the parameter estimates from the reduced QLBM for HBV sample group 4 Log_{10} copies/ml viral load: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	1.635	0.389	0.898	1.625	2.388		
Technology group-baseline "CC"							
Tech CIH-CC	0.246	0.505	-0.717	0.226	1.267	+	No
Tech RTIH-CC	0.618	0.537	-0.407	0.611	1.721	+	No
Tech RTC-CC	1.148	0.619	-0.005	1.127	2.454	+	No
Tech bDNA-CC	-1.584	0.534	-2.597	-1.597	-0.530	-	Yes
Tech HC-CC	-3.383	0.599	-4.572	-3.378	-2.193	-	Yes
Tech TMA-CC	0.238	0.911	-1.493	0.229	2.077	+	No
Anti- baseline "No"	0.922	0.417	0.105	0.925	1.742	+	Yes
Accred- baseline "No"	1.482	0.450	0.631	1.463	2.404	+	Yes
OthrSpc. baseline "No"	1.399	0.534	0.402	1.389	2.492	+	Yes

Sample Group 3.5 $\log_{10}$ Copies/ml Viral Load

Table B.20: *Summary statistics of the parameter estimates from the reduced QLBM for HBV sample group 3.5 Log_{10} copies/ml viral load: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	2.189	0.348	1.516	2.185	2.892		
Technology group-baseline "CC"							
Tech CIH-CC	-0.093	0.468	-0.985	-0.099	0.825	-	No
Tech RTIH-CC	0.909	0.642	-0.273	0.881	2.232	+	No
Tech RTC-CC	-0.050	0.767	-1.491	-0.076	1.536	-	No
Tech bDNA-CC	-3.158	0.546	-4.242	-3.145	-2.087	-	Yes
Tech HC-CC	-3.285	0.572	-4.390	-3.277	-2.181	-	Yes
Tech TMA-CC	0.292	0.903	-1.391	0.254	2.144	+	No
Anti- baseline "No"	0.863	0.424	0.031	0.864	1.696	+	Yes

Sample Group 3 $\log_{10}$ Copies/ml Viral Load

Table B.21: *Summary statistics of the parameter estimates from the reduced QLBM for HBV sample group 3 Log_{10} copies/ml viral load: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	0.784	0.353	0.092	0.784	1.478		
Year- baseline "2005"							
Year 2003	0.597	0.375	-0.123	0.586	1.346	+	No
Year 2004	1.011	0.313	0.408	1.008	1.647	+	Yes
Technology group- baseline "CC"							
Tech CIH-CC	-0.022	0.394	-0.776	-0.034	0.762	-	No
Tech RTIH-CC	-0.182	0.362	-0.885	-0.184	0.531	-	No
Tech RTC-CC	0.797	0.413	0.009	0.785	1.638	+	Yes
Tech bDNA-CC	-1.335	0.447	-2.207	-1.339	-0.451	-	Yes
Tech HC-CC	-1.679	0.666	-2.994	-1.675	-0.399	-	Yes
Tech TMA-CC							
Anti- baseline "No"	1.235	0.316	0.609	1.233	1.859	+	Yes
Accred- baseline "No"	0.573	0.316	-0.052	0.569	1.195	+	No

Sample Group 2.3 $\log_{10}$ Copies/ml Viral Load

Table B.22: *Summary statistics of the parameter estimates from the reduced QLBM for HBV sample group 2.3 Log_{10} copies/ml viral load: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	0.978	0.201	0.596	0.974	1.386		
Technology group- baseline "CC"							
Tech CIH-CC	-0.440	0.312	-1.052	-0.439	0.173	-	No
Tech RTIH-CC	0.352	0.361	-0.364	0.347	1.062	+	No
Tech RTC-CC	0.614	0.497	-0.309	0.599	1.620	+	No
Tech bDNA-CC	-2.669	0.584	-3.862	-2.656	-1.585	-	Yes
Tech HC-CC	-1.674	0.632	-2.952	-1.651	-0.454	-	Yes
Tech TMA-CC	0.484	0.864	-1.193	0.475	2.193	+	No
Analysis method- baseline Singly							
Analysis Duplicated	-0.102	0.294	-0.673	-0.099	0.480	-	No
Analysis Other	1.434	0.693	0.173	1.402	2.889	+	Yes

Negative Sample Group

Table B.23: *Summary statistics of the parameter estimates from the reduced QLBM for HBV Negative sample group: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	2.825	0.293	2.272	2.820	3.416		
Anti- baseline "No"	1.036	0.492	0.120	1.031	2.064	+	Yes

B.6 HBV Estimated Probabilities

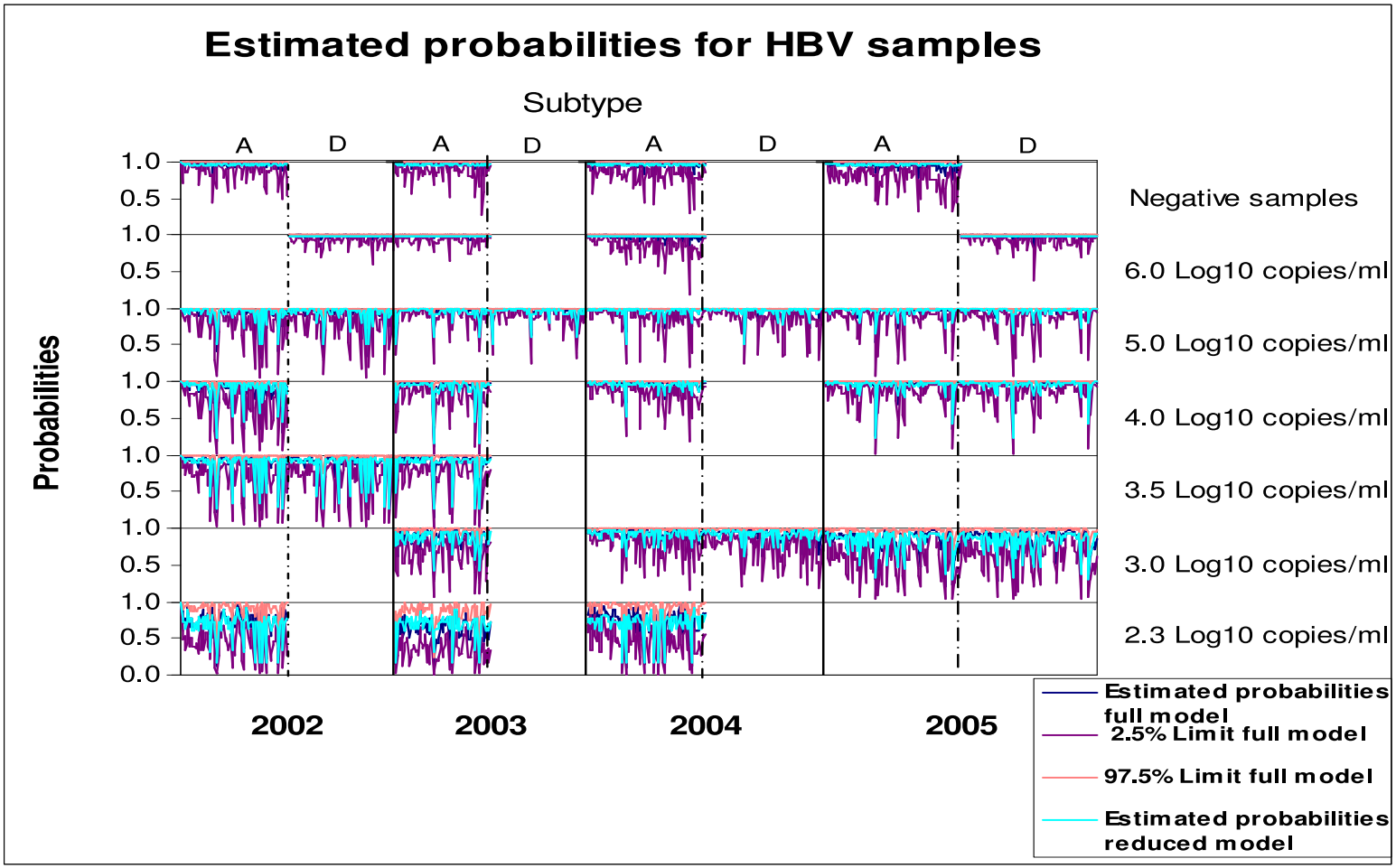


Figure B.2: *HBV estimated probabilities from the full and reduced QLBM.*

Appendix C

Tables of Results from the QTBM Applied to HBV and HCV Data

C.1 Results from the Full QTBM Applied to HBV Data

Sample Group 6 $\log_{10}$ Copies/ml Viral Load

Table C.1: *Summary statistics of the parameter estimates from the full QTBM for HBV sample group 6* $\log_{10}$ *copies/ml viral load: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	3.743	0.504	2.715	3.825	4.573		
Year- baseline "2005"							
Year 2002	0.590	0.124	0.354	0.591	0.834	+	Yes
Year 2003	1.664	0.477	0.931	1.593	2.717	+	Yes
Year 2004	1.648	0.478	0.900	1.582	2.715	+	Yes
Subtype- baseline "A"	1.289	0.482	0.530	1.210	2.358	+	Yes
Technology group- baseline "CC"							
Tech CIH-CC	0.235	0.215	-0.184	0.231	0.670	+	No
Tech RTIH-CC	0.308	0.147	0.021	0.308	0.598	+	Yes
Tech RTC-CC	0.104	0.121	-0.126	0.102	0.345	+	No
Tech bDNA-CC	0.518	0.146	0.236	0.519	0.810	+	Yes
Tech HC-CC	0.438	0.227	-0.014	0.440	0.878	+	No
Anti- baseline "No"	0.202	0.095	0.014	0.204	0.384	+	Yes
Accred- baseline "No"	0.013	0.100	-0.183	0.016	0.207	+	No
OtherSpc. baseline "No"	-0.085	0.120	-0.316	-0.086	0.151	-	No
Analysis method- baseline Singly							
Analysis Duplicated	-0.111	0.095	-0.296	-0.111	0.078	-	No
Analysis Other	-0.110	0.189	-0.478	-0.107	0.262	-	No
Plasma- baseline 0-10							
Group 1: 11-100	-0.192	0.196	-0.569	-0.190	0.194	-	No
Group 2: 101-1,000	-0.050	0.122	-0.286	-0.050	0.191	-	No
Group 3: 1,001-2,000	0.051	0.174	-0.298	0.053	0.393	+	No
Group 4: 2,001-10,000	0.237	0.149	-0.062	0.239	0.524	+	No
Group 5: > 10,000	0.150	0.235	-0.332	0.157	0.595	+	No
Serum- baseline 0-10							
Group 1: 11-100	0.160	0.168	-0.181	0.166	0.472	+	No
Group 2: 101-1,000	0.211	0.127	-0.040	0.213	0.451	+	No
Group 3: 1,001-2,000	0.042	0.193	-0.357	0.049	0.392	+	No
Group 4: 2,001-10,000	0.323	0.153	0.019	0.324	0.630	+	Yes
Group 5: > 10,000	0.606	0.282	0.050	0.606	1.156	+	Yes
Labtype- baseline Hospital							
Public Health	0.068	0.155	-0.237	0.067	0.369	+	No
Private	-0.113	0.141	-0.399	-0.114	0.162	-	No
Reference	-0.422	0.384	-1.082	-0.444	0.316	-	No
Manufacture	0.259	0.227	-0.184	0.260	0.705	+	No
Research	-0.193	0.212	-0.633	-0.181	0.192	-	No
Inhibition Test-baseline "No"							
Inhbition test Yes	0.118	0.096	-0.073	0.121	0.306	+	No
Inhibiton test only Negative samples	0.534	0.198	0.147	0.533	0.925	+	Yes

Sample Group 5 $\log_{10}$ Copies/ml Viral Load

Table C.2: *Summary statistics of the parameter estimates from the full QTBM for HBV sample group 5 $\log_{10}$ copies/ml viral load: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	4.554	0.100	4.344	4.558	4.739		
Year- baseline "2005"							
Year 2002	0.314	0.063	0.191	0.314	0.443	+	Yes
Year 2003	0.155	0.063	0.034	0.154	0.279	+	Yes
Year 2004	0.337	0.052	0.237	0.337	0.439	+	Yes
Subtype- baseline "A"	-0.045	0.037	-0.118	-0.046	0.027	-	No
Technology group- baseline "CC"							
Tech CIH-CC	-0.178	0.160	-0.487	-0.178	0.133	-	No
Tech RTIH-CC	-0.034	0.093	-0.217	-0.035	0.148	-	No
Tech RTC-CC	-0.120	0.060	-0.236	-0.120	-0.002	-	Yes
Tech bDNA-CC	0.353	0.090	0.176	0.353	0.525	+	Yes
Tech HC-CC	0.372	0.130	0.111	0.373	0.626	+	Yes
Anti- baseline "No"	0.158	0.049	0.062	0.158	0.258	+	Yes
Accred- baseline "No"	-0.022	0.052	-0.122	-0.023	0.078	-	No
OtherSpc. baseline "No"	-0.053	0.075	-0.197	-0.055	0.096	-	No
Analysis method- baseline Singly							
Analysis Duplicated	-0.063	0.052	-0.164	-0.063	0.041	-	No
Analysis Other	-0.013	0.108	-0.221	-0.013	0.202	-	No
Plasma- baseline 0-10							
Group 1: 11-100	-0.054	0.114	-0.286	-0.051	0.166	-	No
Group 2: 101-1,000	0.039	0.070	-0.095	0.038	0.178	+	No
Group 3: 1,001-2,000	0.138	0.094	-0.050	0.137	0.323	+	No
Group 4: 2,001-10,000	0.116	0.088	-0.056	0.116	0.285	+	No
Group 5: > 10,000	-0.041	0.200	-0.488	-0.010	0.285	-	No
Serum- baseline 0-10							
Group 1: 11-100	0.054	0.088	-0.124	0.056	0.223	+	No
Group 2: 101-1,000	-0.028	0.065	-0.152	-0.030	0.108	-	No
Group 3: 1,001-2,000	-0.122	0.107	-0.325	-0.124	0.090	-	No
Group 4: 2,001-10,000	0.264	0.080	0.108	0.264	0.420	+	Yes
Group 5: > 10,000	0.487	0.156	0.183	0.486	0.790	+	Yes
Labtype- baseline Hospital							
Public Health	0.079	0.076	-0.070	0.079	0.225	+	No
Private	-0.083	0.081	-0.239	-0.081	0.075	-	No
Reference	-0.027	0.121	-0.271	-0.024	0.201	-	No
Manufacture	0.125	0.110	-0.094	0.126	0.336	+	No
Research	-0.424	0.109	-0.622	-0.429	-0.190	-	Yes
Inhibition Test-baseline "No"							
Inhbition test Yes	0.133	0.049	0.037	0.133	0.228	+	Yes
Inhibiton test only Negative samples	0.503	0.100	0.301	0.505	0.690	+	Yes

Sample Group 4 $\log_{10}$ Copies/ml Viral Load

Table C.3: *Summary statistics of the parameter estimates from the full QTBM for HBV sample group 4 $\log_{10}$ copies/ml viral load: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	3.509	0.125	3.266	3.511	3.758		
Year- baseline "2005"							
Year 2002	0.181	0.085	0.014	0.180	0.351	+	Yes
Year 2003	0.006	0.082	-0.153	0.005	0.168	+	No
Year 2004	0.130	0.065	0.005	0.129	0.258	+	Yes
Subtype- baseline "A"	-0.138	0.059	-0.256	-0.138	-0.021	-	Yes
Technology group- baseline "CC"							
Tech CIH-CC	-0.151	0.199	-0.535	-0.148	0.240	-	No
Tech RTIH-CC	0.116	0.114	-0.108	0.116	0.339	+	No
Tech RTC-CC	-0.172	0.075	-0.316	-0.173	-0.022	-	Yes
Tech bDNA-CC	0.246	0.116	0.025	0.245	0.478	+	Yes
Tech HC-CC	0.092	0.350	-0.605	0.098	0.733	+	No
Anti- baseline "No"	0.180	0.060	0.064	0.179	0.297	+	Yes
Accred- baseline "No"	-0.016	0.066	-0.144	-0.017	0.117	-	No
OtherSpc. baseline "No"	-0.119	0.095	-0.289	-0.126	0.088	-	No
Analysis method- baseline Singly							
Analysis Duplicated	-0.111	0.064	-0.237	-0.111	0.015	-	No
Analysis Other	-0.020	0.131	-0.280	-0.021	0.234	-	No
Plasma- baseline 0-10							
Group 1: 11-100	0.120	0.148	-0.202	0.130	0.383	+	No
Group 2: 101-1,000	0.179	0.090	-0.010	0.183	0.339	+	No
Group 3: 1,001-2,000	0.196	0.131	-0.084	0.203	0.437	+	No
Group 4: 2,001-10,000	0.230	0.111	-0.010	0.238	0.428	+	No
Group 5: > 10,000	0.145	0.192	-0.309	0.165	0.472	+	No
Serum- baseline 0-10							
Group 1: 11-100	0.081	0.124	-0.171	0.085	0.318	+	No
Group 2: 101-1,000	0.166	0.091	-0.018	0.169	0.335	+	No
Group 3: 1,001-2,000	0.101	0.139	-0.202	0.111	0.339	+	No
Group 4: 2,001-10,000	0.418	0.106	0.189	0.423	0.615	+	Yes
Group 5: > 10,000	0.749	0.198	0.351	0.755	1.113	+	Yes
Labtype- baseline Hospital							
Public Health	0.042	0.105	-0.170	0.043	0.245	+	No
Private	-0.216	0.117	-0.440	-0.216	0.014	-	No
Reference	-0.039	0.156	-0.374	-0.033	0.240	-	No
Manufacture	0.327	0.143	0.044	0.327	0.603	+	Yes
Research	-0.207	0.142	-0.490	-0.203	0.058	-	No
Inhibition Test-baseline "No"							
Inhbition test Yes	0.124	0.060	0.006	0.122	0.248	+	Yes
Inhibiton test only Negative samples	0.495	0.126	0.249	0.493	0.743	+	Yes

Sample Group 3.5 $\log_{10}$ Copies/ml Viral Load

Table C.4: *Summary statistics of the parameter estimates from the full QTBM for HBV sample group 3.5* $\log_{10}$ *copies/ml viral load: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	3.139	0.129	2.881	3.144	3.379		
Year- baseline "2003"							
Year 2002	0.614	0.058	0.501	0.614	0.726	+	Yes
Technology group- baseline "CC"							
Tech CIH-CC	-0.244	0.186	-0.605	-0.248	0.124	-	No
Tech RTIH-CC	0.402	0.115	0.171	0.402	0.626	+	Yes
Tech RTC-CC	0.288	0.246	-0.206	0.290	0.751	+	No
Tech bDNA-CC	0.546	0.423	-0.190	0.495	1.505	+	No
Tech HC-CC	0.444	0.304	-0.180	0.460	0.998	+	No
Anti- baseline "No"	0.025	0.068	-0.106	0.024	0.160	+	No
Accred- baseline "No"	-0.096	0.053	-0.200	-0.097	0.010	-	No
OtherSpc. baseline "No"	0.275	0.079	0.116	0.277	0.431	+	Yes
Analysis method- baseline Singly							
Analysis Duplicated	0.013	0.071	-0.131	0.015	0.145	+	No
Analysis Other	0.231	0.125	-0.013	0.230	0.479	+	No
Plasma- baseline 0-10							
Group 1: 11-100	0.128	0.104	-0.074	0.126	0.330	+	No
Group 2: 101-1,000	-0.059	0.077	-0.209	-0.060	0.093	-	No
Group 3: 1,001-2,000	0.060	0.114	-0.159	0.059	0.287	+	No
Group 4: 2,001-10,000	0.079	0.156	-0.218	0.074	0.393	+	No
Group 5: > 10,000	0.966	0.750	-0.413	0.977	2.293	+	No
Serum- baseline 0-10							
Group 1: 11-100	-0.126	0.104	-0.328	-0.126	0.079	-	No
Group 2: 101-1,000	0.019	0.074	-0.123	0.018	0.172	+	No
Group 3: 1,001-2,000	-0.150	0.101	-0.352	-0.149	0.049	-	No
Group 4: 2,001-10,000	-0.101	0.305	-0.772	-0.088	0.468	-	No
Group 5: > 10,000	1.147	1.214	-1.570	1.601	2.659	+	No
Labtype- baseline Hospital							
Public Health	0.125	0.080	-0.030	0.124	0.280	+	No
Private	0.091	0.150	-0.207	0.091	0.389	+	No
Reference	0.014	0.121	-0.224	0.016	0.255	+	No
Manufacture	0.075	0.106	-0.135	0.076	0.280	+	No
Research	-0.095	0.238	-0.544	-0.098	0.374	-	No
Inhibition Test-baseline "No"							
Inhbition test Yes	0.169	0.079	0.015	0.167	0.325	+	Yes
Inhibiton test only Negative samples	0.277	0.154	-0.024	0.275	0.579	+	No

Sample Group 3 $\log_{10}$ Copies/ml Viral Load

Table C.5: *Summary statistics of the parameter estimates from the full QTBM for HBV sample group 3* $\log_{10}$ *copies/ml viral load: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	2.769	0.117	2.545	2.766	2.995		
Year- baseline "2005"							
Year 2003	-0.015	0.078	-0.165	-0.016	0.142	-	No
Year 2004	0.544	0.051	0.446	0.544	0.644	+	Yes
Subtype- baseline "A"	0.053	0.044	-0.032	0.054	0.139	+	No
Technology group- baseline "CC"							
Tech CIH-CC	-0.030	0.168	-0.353	-0.031	0.301	-	No
Tech RTIH-CC	0.125	0.100	-0.076	0.126	0.315	+	No
Tech RTC-CC	-0.076	0.067	-0.207	-0.077	0.064	-	No
Tech bDNA-CC	-0.065	0.116	-0.294	-0.064	0.159	-	No
Tech HC-CC	1.593	0.263	1.059	1.593	2.094	+	Yes
Anti- baseline "No"	0.067	0.057	-0.042	0.067	0.178	+	No
Accred- baseline "No"	-0.037	0.062	-0.161	-0.036	0.086	-	No
OtherSpc. baseline "No"	-0.180	0.068	-0.307	-0.182	-0.040	-	Yes
Analysis method- baseline Singly							
Analysis Duplicated	-0.035	0.063	-0.156	-0.035	0.090	-	No
Analysis Other	-0.061	0.143	-0.343	-0.061	0.217	-	No
Plasma- baseline 0-10							
Group 1: 11-100	-0.108	0.114	-0.335	-0.108	0.111	-	No
Group 2: 101-1,000	0.037	0.079	-0.123	0.038	0.187	+	No
Group 3: 1,001-2,000	0.063	0.120	-0.173	0.063	0.298	+	No
Group 4: 2,001-10,000	0.220	0.086	0.045	0.221	0.390	+	Yes
Group 5: > 10,000	0.092	0.131	-0.167	0.094	0.351	+	No
Serum- baseline 0-10							
Group 1: 11-100	0.130	0.101	-0.071	0.130	0.322	+	No
Group 2: 101-1,000	0.065	0.081	-0.097	0.065	0.226	+	No
Group 3: 1,001-2,000	0.053	0.112	-0.171	0.054	0.272	+	No
Group 4: 2,001-10,000	0.278	0.088	0.104	0.278	0.449	+	Yes
Group 5: > 10,000	0.590	0.178	0.232	0.591	0.939	+	Yes
Labtype- baseline Hospital							
Public Health	-0.031	0.094	-0.212	-0.030	0.152	-	No
Private	-0.265	0.087	-0.441	-0.265	-0.097	-	Yes
Reference	0.014	0.163	-0.320	0.016	0.328	+	No
Manufacture	0.363	0.180	0.019	0.359	0.723	+	Yes
Research	-0.462	0.099	-0.652	-0.464	-0.265	-	Yes
Inhibition Test-baseline "No"							
Inhbition test Yes	-0.016	0.058	-0.136	-0.016	0.094	-	No
Inhibiton test only Negative samples	0.193	0.132	-0.068	0.195	0.444	+	No

Sample Group 2.3 $\log_{10}$ Copies/ml Viral Load

Table C.6: *Summary statistics of the parameter estimates from the full QTBM for HBV sample group 2.3* $\log_{10}$ *copies/ml viral load: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	2.491	0.148	2.232	2.479	2.804		
Year- baseline "2004"							
Year 2002	0.151	0.082	-0.006	0.152	0.311	+	No
Year 2003	-0.436	0.090	-0.615	-0.436	-0.261	-	Yes
Technology group- baseline "CC"							
Tech CIH-CC	0.650	0.210	0.225	0.651	1.053	+	Yes
Tech RTIH-CC	0.679	0.127	0.430	0.680	0.920	+	Yes
Tech RTC-CC	0.446	0.111	0.228	0.446	0.668	+	Yes
Tech bDNA-CC	-0.244	0.499	-1.306	-0.178	0.592	-	No
Tech HC-CC	2.363	0.255	1.813	2.376	2.823	+	Yes
Anti- baseline "No"	0.030	0.078	-0.126	0.033	0.179	+	No
Accred- baseline "No"	-0.069	0.070	-0.206	-0.068	0.068	-	No
OtherSpc. baseline "No"	-0.055	0.089	-0.231	-0.057	0.120	-	No
Analysis method- baseline Singly							
Analysis Duplicated	-0.034	0.079	-0.190	-0.034	0.122	-	No
Analysis Other	0.087	0.133	-0.174	0.087	0.351	+	No
Plasma- baseline 0-10							
Group 1: 11-100	-0.109	0.122	-0.348	-0.109	0.129	-	No
Group 2: 101-1,000	-0.083	0.090	-0.262	-0.082	0.092	-	No
Group 3: 1,001-2,000	-0.119	0.146	-0.411	-0.116	0.162	-	No
Group 4: 2,001-10,000	0.035	0.115	-0.185	0.033	0.270	+	No
Group 5: > 10,000	-0.481	0.217	-0.903	-0.479	-0.060	-	Yes
Serum- baseline 0-10							
Group 1: 11-100	0.135	0.131	-0.125	0.136	0.385	+	No
Group 2: 101-1,000	-0.056	0.085	-0.225	-0.053	0.106	-	No
Group 3: 1,001-2,000	0.059	0.110	-0.162	0.059	0.269	+	No
Group 4: 2,001-10,000	-0.155	0.129	-0.415	-0.151	0.088	-	No
Group 5: > 10,000	-0.436	0.570	-1.574	-0.411	0.653	-	No
Labtype- baseline Hospital							
Public Health	-0.092	0.102	-0.297	-0.093	0.112	-	No
Private	-0.075	0.128	-0.332	-0.071	0.171	-	No
Reference	-0.056	0.203	-0.474	-0.048	0.328	-	No
Manufacture	-0.051	0.153	-0.352	-0.052	0.251	-	No
Research	0.030	0.179	-0.331	0.032	0.376	+	No
Inhibition Test-baseline "No"							
Inhbition test Yes	0.000	0.084	-0.162	0.000	0.160	+	No
Inhibiton test only Negative samples	0.361	0.163	0.041	0.360	0.691	+	Yes

C.2 Results from the Reduced QTBM Applied to HBV Data

Sample Group 6 $\log_{10}$ Copies/ml Viral Load

Table C.7: *Summary statistics of the parameter estimates from the reduced QTBM for HBV sample group 6* $\log_{10}$ *copies/ml viral load: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	3.592	0.486	2.711	3.571	4.516		
Year- baseline "2005"							
Year 2002	0.565	0.115	0.336	0.565	0.786	+	Yes
Year 2003	1.710	0.477	0.824	1.721	2.622	+	Yes
Year 2004	1.745	0.477	0.833	1.756	2.664	+	Yes
Subtype- baseline "A"	1.366	0.477	0.471	1.376	2.312	+	Yes
Technology group- baseline "CC"							
Tech CIH-CC	0.206	0.208	-0.198	0.204	0.623	+	No
Tech RTIH-CC	0.270	0.142	-0.005	0.270	0.545	+	No
Tech RTC-CC	0.145	0.110	-0.077	0.145	0.359	+	No
Tech bDNA-CC	0.551	0.139	0.279	0.553	0.821	+	Yes
Tech HC-CC	0.521	0.211	0.102	0.522	0.939	+	Yes
Anti- baseline "No"	0.263	0.087	0.093	0.263	0.432	+	Yes
Serum- baseline 0-10							
Group 1: 11-100	0.138	0.152	-0.161	0.138	0.425	+	No
Group 2: 101-1,000	0.175	0.120	-0.062	0.177	0.406	+	No
Group 3: 1,001-2,000	-0.017	0.186	-0.383	-0.016	0.339	-	No
Group 4: 2,001-10,000	0.251	0.142	-0.026	0.251	0.531	+	No
Group 5: > 10,000	0.536	0.245	0.038	0.538	1.010	+	Yes
Inhibition Test-baseline "No"							
Inhbition test Yes	0.121	0.090	-0.056	0.120	0.301	+	No
Inhibiton test only Negative samples	0.630	0.179	0.283	0.628	0.987	+	Yes

Sample Group 5 $\log_{10}$ Copies/ml Viral Load

Table C.8: *Summary statistics of the parameter estimates from the reduced QTBM for HBV sample group 5 $\log_{10}$ copies/ml viral load: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	4.489	0.091	4.312	4.488	4.671		
Year- baseline "2005"							
Year 2002	0.318	0.062	0.197	0.317	0.442	+	Yes
Year 2003	0.147	0.062	0.025	0.147	0.269	+	Yes
Year 2004	0.333	0.051	0.235	0.332	0.434	+	Yes
Technology group- baseline "CC"							
Tech CIH-CC	-0.201	0.155	-0.516	-0.200	0.101	-	No
Tech RTIH-CC	-0.051	0.091	-0.233	-0.050	0.129	-	No
Tech RTC-CC	-0.119	0.060	-0.238	-0.118	0.000	-	Yes
Tech bDNA-CC	0.364	0.090	0.185	0.364	0.539	+	Yes
Tech HC-CC	0.388	0.130	0.131	0.389	0.642	+	Yes
Anti- baseline "No"	0.174	0.049	0.080	0.174	0.268	+	Yes
Plasma- baseline 0-10							
Group 1: 11-100	-0.067	0.110	-0.291	-0.062	0.136	-	No
Group 2: 101-1,000	0.030	0.065	-0.100	0.030	0.154	+	No
Group 3: 1,001-2,000	0.123	0.091	-0.062	0.125	0.295	+	No
Group 4: 2,001-10,000	0.084	0.070	-0.051	0.084	0.222	+	No
Group 5: > 10,000	-0.120	0.178	-0.491	-0.097	0.180	-	No
Serum- baseline 0-10							
Group 1: 11-100	0.066	0.087	-0.105	0.065	0.236	+	No
Group 2: 101-1,000	-0.028	0.063	-0.148	-0.030	0.100	-	No
Group 3: 1,001-2,000	-0.140	0.099	-0.330	-0.140	0.057	-	No
Group 4: 2,001-10,000	0.245	0.074	0.098	0.245	0.392	+	Yes
Group 5: > 10,000	0.478	0.159	0.165	0.475	0.791	+	Yes
Labtype- baseline Hospital							
Public Health	0.102	0.073	-0.040	0.102	0.247	+	No
Private	-0.074	0.078	-0.238	-0.072	0.073	-	No
Reference	-0.043	0.131	-0.325	-0.037	0.199	-	No
Manufacture	0.146	0.104	-0.059	0.147	0.346	+	No
Research	-0.394	0.109	-0.589	-0.400	-0.152	-	Yes
Inhibition Test-baseline "No"							
Inhbition test Yes	0.135	0.049	0.039	0.134	0.231	+	Yes
Inhibiton test only Negative samples	0.498	0.098	0.302	0.500	0.690	+	Yes

Sample Group 4 $\log_{10}$ Copies/ml Viral Load

Table C.9: *Summary statistics of the parameter estimates from the reduced QTBM for HBV sample group 4 $\log_{10}$ copies/ml viral load: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	3.444	0.126	3.204	3.443	3.697		
Year- baseline "2005"							
Year 2002	0.191	0.084	0.026	0.192	0.355	+	Yes
Year 2003	-0.001	0.084	-0.170	-0.002	0.161	-	No
Year 2004	0.120	0.068	-0.011	0.120	0.251	+	No
Subtype- baseline "A"	-0.138	0.060	-0.255	-0.139	-0.018	-	Yes
Technology group- baseline "CC"							
Tech CIH-CC	-0.225	0.188	-0.597	-0.225	0.132	-	No
Tech RTIH-CC	0.089	0.109	-0.125	0.087	0.304	+	No
Tech RTC-CC	-0.194	0.073	-0.336	-0.193	-0.049	-	Yes
Tech bDNA-CC	0.259	0.116	0.032	0.258	0.483	+	Yes
Tech HC-CC	0.074	0.367	-0.679	0.086	0.756	+	No
Anti- baseline "No"	0.200	0.057	0.086	0.201	0.310	+	Yes
Plasma- baseline 0-10							
Group 1: 11-100	0.063	0.148	-0.249	0.072	0.332	+	No
Group 2: 101-1,000	0.147	0.091	-0.036	0.150	0.318	+	No
Group 3: 1,001-2,000	0.151	0.126	-0.109	0.157	0.389	+	No
Group 4: 2,001-10,000	0.164	0.095	-0.029	0.165	0.345	+	No
Group 5: > 10,000	0.012	0.198	-0.449	0.039	0.336	+	No
Serum- baseline 0-10							
Group 1: 11-100	0.148	0.120	-0.098	0.153	0.373	+	No
Group 2: 101-1,000	0.188	0.090	0.008	0.190	0.358	+	Yes
Group 3: 1,001-2,000	0.060	0.147	-0.245	0.071	0.325	+	No
Group 4: 2,001-10,000	0.388	0.102	0.188	0.390	0.585	+	Yes
Group 5: > 10,000	0.736	0.198	0.341	0.739	1.116	+	Yes
Labtype- baseline Hospital							
Public Health	0.113	0.102	-0.086	0.113	0.315	+	No
Private	-0.201	0.128	-0.444	-0.198	0.042	-	No
Reference	-0.072	0.179	-0.512	-0.053	0.234	-	No
Manufacture	0.383	0.140	0.103	0.382	0.653	+	Yes
Research	-0.103	0.132	-0.382	-0.096	0.141	-	No
Inhibition Test-baseline "No"							
Inhbition test Yes	0.129	0.062	0.011	0.128	0.249	+	Yes
Inhibiton test only Negative samples	0.472	0.121	0.233	0.473	0.703	+	Yes

Sample Group 3.5 $\log_{10}$ Copies/ml Viral Load

Table C.10: *Summary statistics of the parameter estimates from the reduced QTBM for HBV sample group 3.5* $\log_{10}$ *copies/ml viral load: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	3.117	0.114	2.897	3.118	3.355		
Year- baseline "2003"							
Year 2002	0.641	0.053	0.541	0.640	0.747	+	Yes
Technology group- baseline "CC"							
Tech CIH-CC	-0.228	0.182	-0.592	-0.227	0.130	-	No
Tech RTIH-CC	0.430	0.115	0.208	0.432	0.654	+	Yes
Tech RTC-CC	0.317	0.211	-0.110	0.320	0.723	+	No
Tech bDNA-CC	0.614	0.491	-0.132	0.512	1.670	+	No
Tech HC-CC	0.253	0.252	-0.226	0.246	0.769	+	No
OtherSpc. baseline "No"	0.252	0.077	0.096	0.253	0.400	+	Yes
Analysis method- baseline Singly							
Analysis Duplicated	-0.009	0.064	-0.138	-0.009	0.118	-	No
Analysis Other	0.269	0.118	0.043	0.266	0.504	+	Yes
Plasma- baseline 0-10							
Group 1: 11-100	0.107	0.099	-0.086	0.108	0.296	+	No
Group 2: 101-1,000	-0.048	0.067	-0.178	-0.048	0.079	-	No
Group 3: 1,001-2,000	0.064	0.101	-0.137	0.066	0.261	+	No
Group 4: 2,001-10,000	0.095	0.147	-0.193	0.095	0.378	+	No
Group 5: > 10,000	1.256	0.661	-0.160	1.334	2.331	+	No
Serum- baseline 0-10							
Group 1: 11-100	-0.153	0.104	-0.348	-0.153	0.055	-	No
Group 2: 101-1,000	0.008	0.070	-0.133	0.008	0.145	+	No
Group 3: 1,001-2,000	-0.139	0.086	-0.311	-0.137	0.025	-	No
Group 4: 2,001-10,000	-0.087	0.357	-0.987	-0.047	0.514	-	No
Group 5: > 10,000	0.871	1.359	-1.755	1.176	2.705	+	No
Inhibition Test-baseline "No"							
Inhbition test Yes	0.165	0.070	0.023	0.165	0.306	+	Yes
Inhibiton test only Negative samples	0.312	0.149	0.012	0.315	0.594	+	Yes

Sample Group 3 $\log_{10}$ Copies/ml Viral Load

Table C.11: *Summary statistics of the parameter estimates from the reduced QTBM for HBV sample group 3* $\log_{10}$ *copies/ml viral load: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	2.833	0.097	2.640	2.830	3.029		
Year- baseline "2005"							
Year 2003	-0.053	0.074	-0.197	-0.053	0.092	-	No
Year 2004	0.531	0.049	0.434	0.532	0.626	+	Yes
Technology group- baseline "CC"							
Tech CIH-CC	-0.080	0.160	-0.389	-0.083	0.232	-	No
Tech RTIH-CC	0.119	0.094	-0.071	0.119	0.299	+	No
Tech RTC-CC	-0.093	0.060	-0.214	-0.093	0.025	-	No
Tech bDNA-CC	-0.130	0.108	-0.346	-0.128	0.085	-	No
Tech HC-CC	1.553	0.263	1.023	1.559	2.041	+	Yes
OtherSpc. baseline "No"	-0.197	0.071	-0.336	-0.198	-0.058	-	Yes
Plasma- baseline 0-10							
Group 1: 11-100	-0.113	0.111	-0.330	-0.114	0.101	-	No
Group 2: 101-1,000	0.029	0.080	-0.130	0.030	0.184	+	No
Group 3: 1,001-2,000	0.084	0.116	-0.143	0.083	0.314	+	No
Group 4: 2,001-10,000	0.224	0.089	0.047	0.225	0.397	+	Yes
Group 5: > 10,000	0.101	0.136	-0.168	0.104	0.372	+	No
Serum- baseline 0-10							
Group 1: 11-100	0.125	0.098	-0.069	0.124	0.315	+	No
Group 2: 101-1,000	0.072	0.079	-0.084	0.073	0.226	+	No
Group 3: 1,001-2,000	0.060	0.114	-0.162	0.061	0.276	+	No
Group 4: 2,001-10,000	0.297	0.085	0.133	0.297	0.464	+	Yes
Group 5: > 10,000	0.589	0.178	0.232	0.591	0.929	+	Yes
Labtype- baseline Hospital							
Public Health	-0.024	0.091	-0.201	-0.024	0.158	-	No
Private	-0.272	0.085	-0.433	-0.272	-0.106	-	Yes
Reference	0.018	0.159	-0.301	0.021	0.320	+	No
Manufacture	0.362	0.162	0.056	0.358	0.688	+	Yes
Research	-0.466	0.098	-0.652	-0.468	-0.271	-	Yes
Inhibition Test-baseline "No"							
Inhbition test Yes	-0.022	0.059	-0.140	-0.022	0.092	-	No
Inhibiton test only Negative samples	0.212	0.123	-0.026	0.214	0.455	+	No

Sample Group 2.3 $\log_{10}$ Copies/ml Viral Load

Table C.12: *Summary statistics of the parameter estimates from the reduced QTBM for HBV sample group 2.3* $\log_{10}$ *copies/ml viral load: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	2.359	0.074	2.211	2.361	2.502		
Year- baseline "2004"							
Year 2002	0.176	0.069	0.044	0.176	0.316	+	Yes
Year 2003	-0.404	0.079	-0.560	-0.404	-0.251	-	Yes
Technology group- baseline "CC"							
Tech CIH-CC	0.630	0.198	0.238	0.631	1.018	+	Yes
Tech RTIH-CC	0.627	0.116	0.396	0.629	0.854	+	Yes
Tech RTC-CC	0.367	0.084	0.205	0.366	0.530	+	Yes
Tech bDNA-CC	-0.511	0.500	-1.462	-0.487	0.395	-	No
Tech HC-CC	2.379	0.214	1.920	2.392	2.770	+	Yes
Inhibition Test-baseline "No"							
Inhbition test Yes	-0.003	0.072	-0.144	-0.004	0.138	-	No
Inhibiton test only Negative samples	0.367	0.148	0.073	0.368	0.661	+	Yes

C.3 HBV Estimated Means

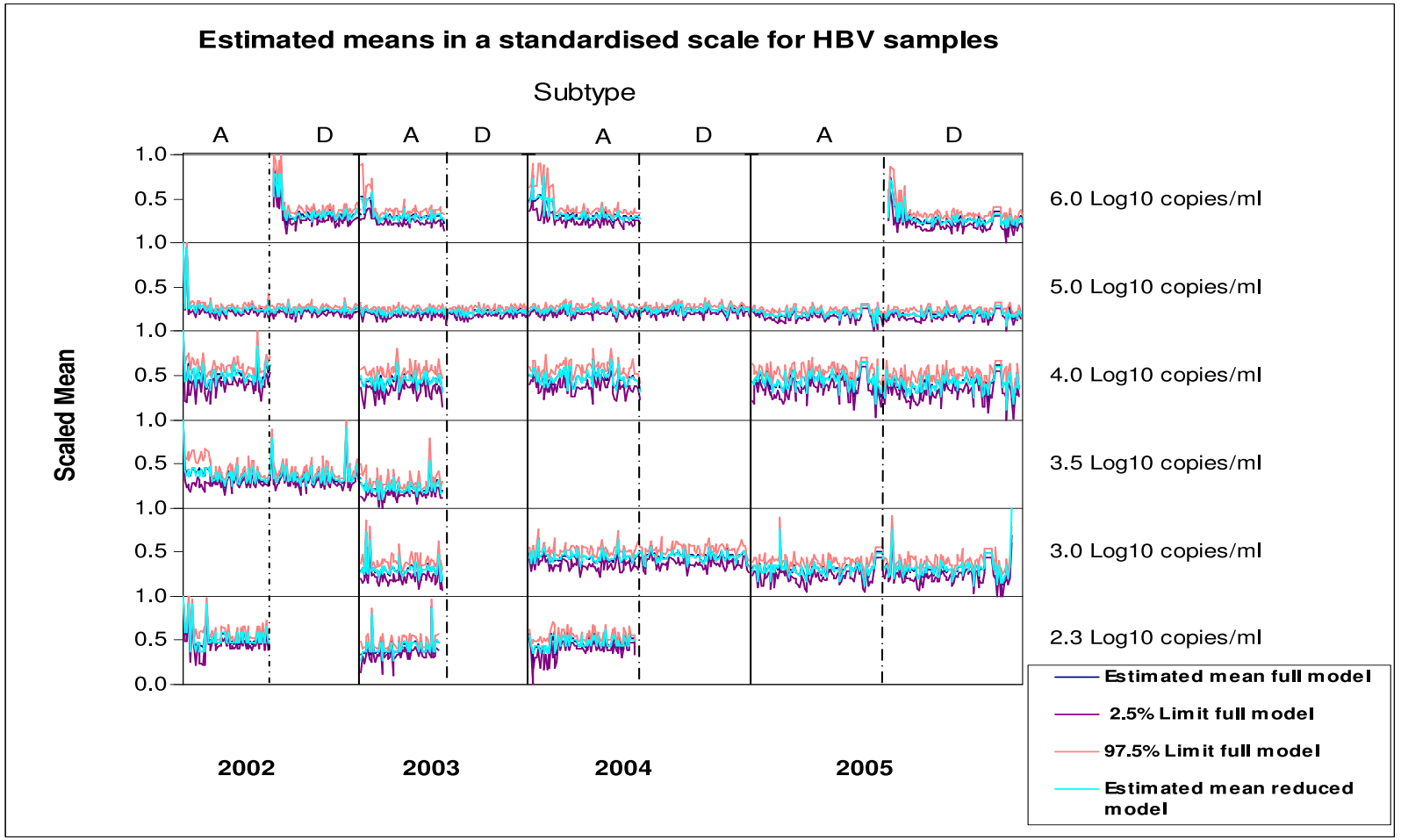


Figure C.1: *HBV estimated means of sample viral load from the full and reduced QTBM.*

C.4 Results from the Full QTBM Applied to HCV Data

In this section is presented a summary of the results obtained from the full QTBM applied to HCV data. Table C.13 summaries the mean and standard deviation from the posterior distribution of the parameter estimates per sample group and covariate level.

Table C.13: *Estimated mean and standard deviation of the parameter estimates from the full QTBM applied to the HCV data.*

Covariates	Mean (SD)					
	5.9	4.9	3.9	3.5	3.2	2.2
Intercept	5.862 (0.173)	5.030 (0.175)	3.727 (0.150)	3.135 (0.184)	2.438 (0.439)	2.548 (0.252)
Year- baseline "2005"						
Year 2002	**-0.843 (0.075)**	0.178 (0.106)	0.015 (0.080)	-0.130 (0.086)	0.187 (0.408)	-0.114 (0.744)
Year 2003	-0.125 (0.067)	0.110 (0.137)	-0.039 (0.091)	0.107 (0.081)	**0.709 (0.404)**	**-0.433 (0.148)**
Year 2004		-0.077 (0.090)	-0.134 (0.082)		**0.694 (0.406)**	**-0.377 (0.129)**
Genotype- baseline "1"						
Genotype 3		**-0.232 (0.092)**	**-0.283 (0.075)**		0.543 (0.405)	
Genotype 4		**-0.320 (0.109)**				0.044 (0.147)
Genotype 5		-0.122 (0.122)			**0.674 (0.405)**	
Technology group- baseline "CC"						
Tech CIH-CC	0.068 (0.276)	**-0.827 (0.277)**	**-0.447 (0.223)**	-0.271 (0.318)	**-0.426 (0.203)**	0.345 (0.257)
Tech RTIH-CC	0.116 (0.179)	**-0.455 (0.142)**	**-0.300 (0.151)**	-0.431 (0.221)	**-0.264 (0.105)**	0.005 (0.195)
Tech RTC-CC	-0.116 (0.127)	-0.140 (0.101)	**-0.138 (0.101)**	-0.211 (0.140)	-0.080 (0.100)	-0.023 (0.168)
Tech bDNA-CC	0.140 (0.145)	**-0.280 (0.128)**	**-0.253 (0.127)**	**-0.347 (0.164)**	-0.170 (0.124)	**0.184 (0.212)**
Anti- baseline "No"	0.101 (0.104)	-0.110 (0.084)	0.078 (0.082)	0.055 (0.111)	-0.107 (0.080)	0.037 (0.137)
Accred- baseline "No"	0.003 (0.070)	0.049 (0.064)	**0.147 (0.059)**	0.046 (0.072)	-0.011 (0.066)	-0.104 (0.108)
OtherSpc- baseline "No"	-0.141 (0.088)	0.007 (0.080)	0.023 (0.080)	-0.018 (0.093)	0.145 (0.079)	0.021 (0.147)
Analysis method- baseline Singly						
Analysis Duplicated	-0.082 (0.085)	-0.035 (0.083)	-0.052 (0.080)	-0.057 (0.095)	-0.080 (0.079)	-0.156 (0.133)
Analysis Other	0.282 (0.540)	0.295 (0.289)	**0.639 (0.281)**	0.729 (0.634)	0.272 (0.209)	0.152 (0.376)
Plasma- baseline 0-10						
Group 1: 11-100	-0.218 (0.174)	-0.122 (0.133)	-0.001 (0.137)	-0.191 (0.174)	-0.123 (0.156)	-0.575 (0.436)
Group 2: 101-1,000	0.068 (0.090)	-0.062 (0.083)	0.042 (0.081)	0.064 (0.096)	-0.008 (0.087)	-0.151 (0.147)
Group 3: 1,001-2,000	0.024 (0.110)	-0.051 (0.110)	0.014 (0.103)	-0.063 (0.124)	0.020 (0.117)	0.109 (0.183)
Group 4: 2,001-10,000	0.166 (0.101)	**-0.413 (0.097)**	**-0.205 (0.092)**	-0.072 (0.108)	-0.077 (0.104)	-0.054 (0.161)
Group 5: > 10,000	-0.054 (0.206)	-0.015 (0.157)	-0.077 (0.179)	-0.192 (0.230)	-0.037 (0.153)	0.133 (0.295)
Serum- baseline 0-10						
Group 1: 11-100	-0.126 (0.134)	0.059 (0.116)	0.145 (0.119)	0.163 (0.152)	-0.038 (0.118)	-0.170 (0.233)
Group 2: 101-1,000	-0.051 (0.089)	-0.031 (0.088)	0.006 (0.095)	0.046 (0.099)	-0.066 (0.093)	-0.099 (0.192)
Group 3: 1,001-2,000	-0.097 (0.104)	-0.116 (0.108)	0.043 (0.101)	0.044 (0.113)	0.042 (0.096)	-0.230 (0.159)
Group 4: 2,001-10,000	-0.069 (0.099)	0.058 (0.092)	0.128 (0.094)	0.119 (0.109)	-0.003 (0.096)	0.063 (0.153)
Group 5: > 10,000	-0.194 (0.432)	-0.128 (0.249)	-0.016 (0.224)	-0.193 (0.383)	**-0.549 (0.302)**	-0.567 (0.420)
Labtype- baseline Hospital						
Public Health	-0.053 (0.110)	0.018 (0.103)	0.048 (0.105)	-0.055 (0.115)	-0.002 (0.111)	-0.018 (0.187)
Private	0.037 (0.102)	-0.042 (0.094)	0.096 (0.089)	0.078 (0.115)	0.066 (0.092)	0.139 (0.160)
Reference	-0.007 (0.207)	0.032 (0.191)	-0.412 (0.273)	-0.281 (0.288)	**-1.062 (0.365)**	-0.160 (0.351)
Manufacture	**-2.496 (0.450)**	**-1.331 (0.198)**	0.177 (0.220)	0.469 (0.589)	-0.194 (0.218)	-0.216 (0.330)
Research	-0.206 (0.194)	-0.100 (0.154)	-0.222 (0.199)	-0.177 (0.223)	0.045 (0.164)	-0.160 (0.377)
Inhibition Test-baseline "No"						
Inhbition test Yes	**0.323 (0.091)**	0.132 (0.085)	0.041 (0.093)	0.137 (0.105)	-0.033 (0.087)	-0.095 (0.166)
Inhibiton test only Negative samples	0.172 (0.170)	0.096 (0.153)	0.015 (0.150)	0.169 (0.190)	**-0.331 (0.141)**	-0.160 (0.244)

C.5 Results from the Reduced QTBM Applied to HCV Data

Sample Group 5.9 $\log_{10}$ IU/ml Viral Load

Table C.14: *Summary statistics of the parameter estimates from the reduced QTBM for HCV sample group 5.9* $\log_{10}$ *IU/ml viral load: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	2.332	0.092	2.143	2.335	2.506		
Year- baseline "2005"							
Year 2002	-0.075	0.122	-0.316	-0.077	0.166	-	No
Year 2003	-0.446	0.132	-0.707	-0.445	-0.189	-	Yes
Year 2004	-0.368	0.112	-0.588	-0.368	-0.150	-	Yes
Technology group- baseline "CC"							
Tech CIH-CC	0.291	0.200	-0.096	0.288	0.683	+	No
Tech RTIH-CC	-0.092	0.127	-0.342	-0.093	0.156	-	No
Tech RTC-CC	-0.072	0.138	-0.330	-0.075	0.211	-	No
Tech bDNA-CC	0.282	0.132	0.015	0.284	0.540	+	Yes

Sample Group 4.9 $\log_{10}$ IU/ml Viral Load

Table C.15: *Summary statistics of the parameter estimates from the reduced QTBM for HCV sample group 4.9* $\log_{10}$ *IU/ml viral load: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	2.335	0.316	1.538	2.373	2.871		
Year- baseline "2005"							
Year 2002	0.235	0.323	-0.326	0.196	1.020	+	No
Year 2003	0.730	0.317	0.197	0.683	1.518	+	Yes
Year 2004	0.726	0.317	0.184	0.681	1.518	+	Yes
Genotype-baseline "1"							
Genotype 3	0.561	0.316	0.025	0.516	1.359	+	Yes
Genotype 5	0.687	0.315	0.144	0.641	1.478	+	Yes
Technology group- baseline "CC"							
Tech CIH-CC	-0.405	0.211	-0.815	-0.409	0.010	-	No
Tech RTIH-CC	-0.208	0.088	-0.380	-0.208	-0.037	-	Yes
Tech RTC-CC	-0.004	0.088	-0.179	-0.005	0.166	-	No
Tech bDNA-CC	-0.072	0.097	-0.262	-0.071	0.118	-	No
Labtype- baseline Hospital							
Public Health	-0.029	0.100	-0.226	-0.028	0.162	-	No
Private	0.010	0.081	-0.150	0.012	0.166	+	No
Reference	-1.251	0.164	-1.583	-1.249	-0.931	-	Yes
Manufacture	-0.117	0.177	-0.471	-0.117	0.224	-	No
Research	0.050	0.148	-0.237	0.050	0.341	+	No
Inhibition Test-baseline "No"							
Inhbition test Yes	-0.077	0.078	-0.230	-0.076	0.077	-	No
Inhibiton test only Negative samples	-0.382	0.130	-0.636	-0.381	-0.120	-	Yes

Sample Group 3.9 $\log_{10}$ IU/ml Viral Load

Table C.16: *Summary statistics of the parameter estimates from the reduced QTBM for HCV sample group 3.9* $\log_{10}$ *IU/ml viral load: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	3.276	0.094	3.092	3.277	3.454		
Technology group- baseline "CC"							
Tech CIH-CC	-0.435	0.326	-1.075	-0.441	0.228	-	No
Tech RTIH-CC	-0.357	0.211	-0.781	-0.355	0.058	-	No
Tech RTC-CC	-0.224	0.123	-0.462	-0.225	0.019	-	No
Tech bDNA-CC	-0.405	0.128	-0.655	-0.405	-0.149	-	Yes
Anti- baseline "No"	0.083	0.099	-0.111	0.081	0.279	+	No

Sample Group 3.5 $\log_{10}$ IU/ml Viral Load

Table C.17: *Summary statistics of the parameter estimates from the reduced QTBM for HCV sample group 3.5* $\log_{10}$ *IU/ml viral load: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	3.870	0.066	3.742	3.870	4.000		
Genotype-baseline "1"							
Genotype 3	-0.242	0.061	-0.362	-0.242	-0.120	-	Yes
Technology group- baseline "CC"							
Tech CIH-CC	-0.467	0.204	-0.862	-0.466	-0.066	-	Yes
Tech RTIH-CC	-0.364	0.141	-0.641	-0.363	-0.087	-	Yes
Tech RTC-CC	-0.170	0.083	-0.336	-0.169	-0.010	-	Yes
Tech bDNA-CC	-0.374	0.075	-0.521	-0.375	-0.228	-	Yes
Accred- baseline "No"	0.140	0.060	0.020	0.141	0.259	+	Yes
Analysis method- baseline Singly							
Analysis Duplicated	-0.025	0.078	-0.180	-0.024	0.127	-	No
Analysis Other	0.719	0.275	0.193	0.716	1.254	+	Yes
Plasma- baseline 0-10							
Group 1: 11-100	-0.044	0.134	-0.322	-0.040	0.206	-	No
Group 2: 101-1,000	-0.007	0.075	-0.154	-0.007	0.140	-	No
Group 3: 1,001-2,000	-0.019	0.098	-0.214	-0.020	0.173	-	No
Group 4: 2,001-10,000	-0.194	0.086	-0.359	-0.194	-0.025	-	Yes
Group 5: > 10,000	-0.045	0.155	-0.377	-0.038	0.245	-	No

Sample Group 3.2 $\log_{10}$ IU/ml Viral Load

Table C.18: *Summary statistics of the parameter estimates from the reduced QTBM for HCV sample group 3.2* $\log_{10}$ *IU/ml viral load: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	5.074	0.076	4.925	5.075	5.224		
Genotype-baseline "1"							
Genotype 3	-0.228	0.065	-0.357	-0.229	-0.102	-	Yes
Genotype 4	-0.229	0.093	-0.415	-0.228	-0.047	-	Yes
Genotype 5	-0.130	0.080	-0.286	-0.130	0.022	-	No
Technology group- baseline "CC"							
Tech CIH-CC	-0.744	0.276	-1.281	-0.747	-0.177	-	Yes
Tech RTIH-CC	-0.428	0.124	-0.670	-0.428	-0.178	-	Yes
Tech RTC-CC	-0.136	0.089	-0.311	-0.135	0.040	-	No
Tech bDNA-CC	-0.286	0.079	-0.439	-0.285	-0.132	-	Yes
Plasma- baseline 0-10							
Group 1: 11-100	-0.101	0.127	-0.347	-0.100	0.142	-	No
Group 2: 101-1,000	-0.061	0.077	-0.212	-0.061	0.089	-	No
Group 3: 1,001-2,000	-0.057	0.107	-0.267	-0.057	0.152	-	No
Group 4: 2,001-10,000	-0.372	0.090	-0.545	-0.371	-0.193	-	Yes
Group 5: > 10,000	-0.006	0.144	-0.287	-0.004	0.275	-	No
Labtype- baseline Hospital							
Public Health	0.037	0.100	-0.161	0.039	0.232	+	No
Private	-0.056	0.092	-0.239	-0.055	0.125	-	No
Reference	0.052	0.179	-0.305	0.053	0.399	+	No
Manufacture	-1.391	0.170	-1.719	-1.397	-1.044	-	Yes
Research	-0.066	0.148	-0.356	-0.064	0.220	-	No

Sample Group 2.2 $\log_{10}$ IU/ml Viral Load

Table C.19: *Summary statistics of the parameter estimates from the reduced QTBM for HCV sample group 2.2* $\log_{10}$ *IU/ml viral load: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	5.897	0.074	5.751	5.898	6.037		
Year- baseline "2004"							
Year 2002	-0.806	0.065	-0.936	-0.806	-0.679	-	Yes
Year 2003	-0.077	0.061	-0.196	-0.078	0.047	-	No
Labtype- baseline Hospital							
Public Health	0.010	0.103	-0.192	0.009	0.210	+	No
Private	0.115	0.098	-0.077	0.115	0.301	+	No
Reference	0.034	0.202	-0.361	0.033	0.430	+	No
Manufacture	-2.946	0.335	-3.581	-2.948	-2.273	-	Yes
Research	-0.119	0.193	-0.493	-0.118	0.254	-	No
Inhibition Test-baseline "No"							
Inhbition test Yes	0.265	0.070	0.130	0.265	0.402	+	Yes
Inhibiton test only Negative samples	0.149	0.157	-0.158	0.146	0.469	+	No

C.6 HCV Estimated Means

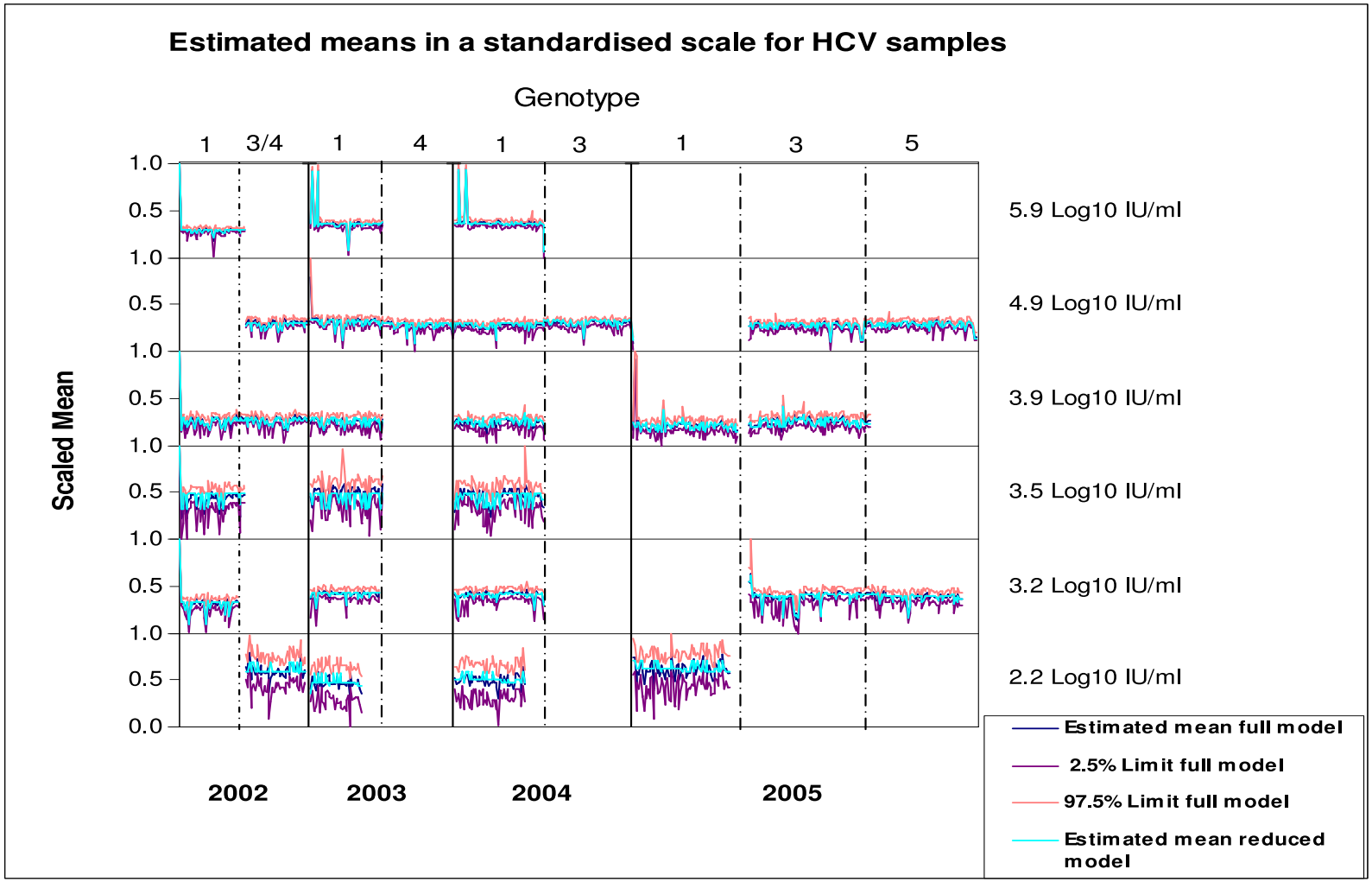


Figure C.2: *HCV estimated means of sample viral load from the full and reduced QTBM.*

Appendix D

Tables of Results from the CBM Applied to HBV Data

Sample Group 5 $\log_{10}$ Copies/ml Viral Load

Table D.1: *Summary statistics of the parameter estimates from the CBM for HBV sample group 5* $\log_{10}$ *copies/ml viral load: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	4.465	0.108	4.247	4.466	4.674		
Year- baseline "2005"							
Year 2002	0.316	0.078	0.155	0.317	0.462	+	Yes
Year 2003	0.114	0.079	-0.042	0.114	0.268	+	No
Year 2004	0.287	0.065	0.155	0.288	0.409	+	Yes
Subtype- baseline "A"	-0.050	0.047	-0.141	-0.051	0.043	-	No
Technology group- baseline "CC"							
Tech CIH-CC	-0.118	0.176	-0.477	-0.117	0.228	-	No
Tech RTIH-CC	0.019	0.113	-0.202	0.019	0.239	+	No
Tech RTC-CC	-0.102	0.076	-0.252	-0.103	0.043	-	No
Tech bDNA-CC	0.427	0.112	0.198	0.431	0.641	+	Yes
Tech HC-CC	0.447	0.168	0.123	0.446	0.772	+	Yes
Anti- baseline "No"	0.226	0.062	0.107	0.225	0.349	+	Yes
Accred- baseline "No"	0.023	0.046	-0.066	-0.022	0.113	+	No
OtherSpc. baseline "No"	0.012	0.046	-0.081	0.013	0.098	+	No
Analysis method- baseline Singly							
Analysis Duplicated	-0.024	0.065	-0.152	-0.024	0.101	-	No
Analysis Other	0.062	0.130	-0.189	0.063	0.313	+	No
Plasma- baseline 0-10							
Group 1: 11-100	-0.059	0.088	-0.247	-0.052	0.092	-	No
Group 2: 101-1,000	0.026	0.059	-0.094	0.026	0.138	+	No
Group 3: 1,001-2,000	0.127	0.073	-0.013	0.126	0.271	+	No
Group 4: 2,001-10,000	0.063	0.062	-0.064	0.065	0.181	+	No
Group 5: > 10,000	-0.066	0.096	-0.277	-0.056	0.093	-	No
Serum- baseline 0-10							
Group 1: 11-100	0.052	0.068	-0.081	0.051	0.185	+	No
Group 2: 101-1,000	-0.008	0.059	-0.125	-0.008	0.106	-	No
Group 3: 1,001-2,000	-0.067	0.090	-0.253	-0.066	0.103	-	No
Group 4: 2,001-10,000	0.128	0.063	0.004	0.127	0.255	+	Yes
Group 5: > 10,000	-0.030	0.081	-0.209	-0.024	0.112	-	No
Labtype- baseline Hospital							
Public Health	0.048	0.063	-0.090	0.051	0.164	+	No
Private	-0.015	0.072	-0.160	-0.011	0.115	-	No
Reference	-0.151	0.129	-0.410	-0.148	0.083	-	No
Manufacture	0.079	0.071	-0.069	0.083	0.211	+	No
Research	-0.179	0.099	-0.386	-0.173	-0.003	-	Yes
Inhibition Test-baseline "No"							
Inhbition test Yes	0.101	0.054	-0.003	0.100	0.206	+	No
Inhibiton test only Negative samples	0.334	0.091	0.156	0.334	0.508	+	Yes

Sample Group 4 $\log_{10}$ Copies/ml Viral Load

Table D.2: *Summary statistics of the parameter estimates from the CBM for HBV sample group 4* $\log_{10}$ *copies/ml viral load: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	3.530	0.128	3.276	3.531	3.775		
Year- baseline "2005"							
Year 2002	0.227	0.110	0.009	0.228	0.437	+	Yes
Year 2003	-0.005	0.108	-0.225	-0.003	0.203	-	No
Year 2004	0.103	0.089	-0.072	0.102	0.281	+	No
Subtype- baseline "A"	-0.140	0.083	-0.306	-0.138	0.015	-	No
Technology group- baseline "CC"							
Tech CIH-CC	-0.107	0.221	-0.542	-0.105	0.327	-	No
Tech RTIH-CC	0.163	0.129	-0.099	0.165	0.404	+	No
Tech RTC-CC	-0.164	0.091	-0.342	-0.165	0.024	-	No
Tech bDNA-CC	0.413	0.138	0.140	0.413	0.687	+	Yes
Tech HC-CC	-0.029	0.489	-1.007	-0.014	0.886	-	No
Anti- baseline "No"	0.255	0.076	0.103	0.257	0.399	+	Yes
Accred- baseline "No"	0.025	0.057	-0.088	0.026	0.135	+	No
OtherSpc. baseline "No"	0.012	0.060	-0.113	0.014	0.125	+	No
Analysis method- baseline Singly							
Analysis Duplicated	-0.112	0.081	-0.268	-0.112	0.049	-	No
Analysis Other	0.076	0.160	-0.238	0.078	0.383	+	No
Plasma- baseline 0-10							
Group 1: 11-100	-0.098	0.115	-0.338	-0.092	0.107	-	No
Group 2: 101-1,000	0.050	0.073	-0.099	0.051	0.191	+	No
Group 3: 1,001-2,000	0.128	0.095	-0.069	0.129	0.312	+	No
Group 4: 2,001-10,000	0.050	0.079	-0.113	0.053	0.196	+	No
Group 5: > 10,000	-0.081	0.112	-0.330	-0.069	0.107	-	No
Serum- baseline 0-10							
Group 1: 11-100	0.086	0.086	-0.090	0.088	0.248	+	No
Group 2: 101-1,000	0.064	0.074	-0.091	0.067	0.203	+	No
Group 3: 1,001-2,000	-0.035	0.116	-0.269	-0.031	0.180	-	No
Group 4: 2,001-10,000	0.159	0.080	0.003	0.159	0.317	+	Yes
Group 5: > 10,000	-0.022	0.097	-0.242	-0.014	0.149	-	No
Labtype- baseline Hospital							
Public Health	0.044	0.095	-0.169	0.051	0.210	+	No
Private	-0.035	0.090	-0.217	-0.031	0.130	-	No
Reference	-0.190	0.169	-0.553	-0.184	0.115	-	No
Manufacture	0.116	0.090	-0.073	0.121	0.280	+	No
Research	-0.108	0.110	-0.343	-0.098	0.084	-	No
Inhibition Test-baseline "No"							
Inhbition test Yes	0.123	0.068	-0.009	0.122	0.258	+	No
Inhibiton test only Negative samples	0.422	0.117	0.197	0.421	0.652	+	Yes

Sample Group 3.5 $\log_{10}$ Copies/ml Viral Load

Table D.3: *Summary statistics of the parameter estimates from the CBM for HBV sample group 3.5* $\log_{10}$ *copies/ml viral load: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	2.935	0.158	2.624	2.935	3.247		
Year- baseline "2003"							
Year 2002	0.698	0.079	0.545	0.698	0.852	+	Yes
Technology group- baseline "CC"							
Tech CIH-CC	-0.137	0.197	-0.525	-0.136	0.241	-	No
Tech RTIH-CC	0.521	0.124	0.277	0.523	0.766	+	Yes
Tech RTC-CC	0.455	0.309	-0.147	0.456	1.066	+	No
Tech bDNA-CC	1.279	0.199	0.884	1.288	1.646	+	Yes
Tech HC-CC	-0.005	1.095	-2.182	0.013	2.070	-	No
Anti- baseline "No"	0.081	0.091	-0.094	0.080	0.264	+	No
Accred- baseline "No"	-0.053	0.080	-0.215	-0.052	0.100	-	No
OtherSpc. baseline "No"	0.173	0.094	-0.007	0.172	0.361	+	No
Analysis method- baseline Singly							
Analysis Duplicated	0.042	0.096	-0.145	0.040	0.228	+	No
Analysis Other	0.389	0.177	0.046	0.386	0.741	+	Yes
Plasma- baseline 0-10							
Group 1: 11-100	0.050	0.123	-0.213	0.061	0.265	+	No
Group 2: 101-1,000	-0.071	0.092	-0.258	-0.069	0.106	-	No
Group 3: 1,001-2,000	0.046	0.124	-0.214	0.052	0.281	+	No
Group 4: 2,001-10,000	-0.020	0.178	-0.396	-0.007	0.309	-	No
Group 5: > 10,000	-0.293	0.338	-1.121	-0.239	0.209	-	No
Serum- baseline 0-10							
Group 1: 11-100	-0.091	0.145	-0.379	-0.087	0.172	-	No
Group 2: 101-1,000	0.061	0.097	-0.125	0.061	0.256	+	No
Group 3: 1,001-2,000	-0.191	0.124	-0.429	-0.192	0.049	-	No
Group 4: 2,001-10,000	-0.374	0.340	-1.134	-0.353	0.198	-	No
Group 5: > 10,000	-0.691	0.606	-2.063	-0.617	0.193	-	No
Labtype- baseline Hospital							
Public Health	0.116	0.097	-0.077	0.117	0.305	+	No
Private	-0.072	0.172	-0.448	-0.058	0.232	-	No
Reference	-0.098	0.162	-0.433	-0.094	0.202	-	No
Manufacture	0.004	0.129	-0.282	0.013	0.232	+	No
Research	-0.067	0.181	-0.469	-0.045	0.238	-	No
Inhibition Test-baseline "No"							
Inhbition test Yes	0.243	0.106	0.029	0.245	0.447	+	Yes
Inhibiton test only Negative samples	0.165	0.152	-0.159	0.172	0.445	+	No

Sample Group 3 $\log_{10}$ Copies/ml Viral Load

Table D.4: *Summary statistics of the parameter estimates from the CBM for HBV sample group 3* $\log_{10}$ *copies/ml viral load: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	2.619	0.128	2.370	2.619	2.870		
Year- baseline "2005"							
Year 2003	-0.013	0.102	-0.208	-0.012	0.185	-	No
Year 2004	0.502	0.064	0.374	0.502	0.627	+	Yes
Subtype- baseline "A"	0.045	0.063	-0.080	0.045	0.167	+	No
Technology group- baseline "CC"							
Tech CIH-CC	0.031	0.189	-0.340	0.029	0.400	+	No
Tech RTIH-CC	0.182	0.123	-0.072	0.182	0.416	+	No
Tech RTC-CC	-0.078	0.081	-0.236	-0.078	0.079	-	No
Tech bDNA-CC	0.087	0.154	-0.220	0.086	0.394	+	No
Tech HC-CC	1.549	0.368	0.801	1.555	2.257	+	Yes
Anti- baseline "No"	0.150	0.078	-0.011	0.152	0.301	+	No
Accred- baseline "No"	0.032	0.056	-0.078	0.032	0.144	+	No
OtherSpc. baseline "No"	-0.037	0.060	-0.156	-0.035	0.078	-	No
Analysis method- baseline Singly							
Analysis Duplicated	0.017	0.081	-0.139	0.019	0.174	+	No
Analysis Other	0.072	0.170	-0.271	0.076	0.397	+	No
Plasma- baseline 0-10							
Group 1: 11-100	-0.197	0.118	-0.439	-0.194	0.027	-	No
Group 2: 101-1,000	-0.011	0.072	-0.162	-0.008	0.120	-	No
Group 3: 1,001-2,000	0.113	0.094	-0.079	0.116	0.291	+	No
Group 4: 2,001-10,000	0.093	0.071	-0.056	0.097	0.228	+	No
Group 5: > 10,000	-0.078	0.102	-0.299	-0.069	0.100	-	No
Serum- baseline 0-10							
Group 1: 11-100	0.097	0.077	-0.064	0.099	0.246	+	No
Group 2: 101-1,000	0.031	0.072	-0.110	0.032	0.171	+	No
Group 3: 1,001-2,000	-0.016	0.106	-0.232	-0.011	0.177	-	No
Group 4: 2,001-10,000	0.148	0.070	0.014	0.147	0.287	+	Yes
Group 5: > 10,000	-0.062	0.104	-0.289	-0.053	0.116	-	No
Labtype- baseline Hospital							
Public Health	0.024	0.087	-0.157	0.028	0.187	+	No
Private	-0.103	0.099	-0.301	-0.101	0.085	-	No
Reference	-0.086	0.170	-0.437	-0.078	0.215	-	No
Manufacture	0.108	0.088	-0.088	0.113	0.262	+	No
Research	-0.324	0.114	-0.557	-0.321	-0.105	-	Yes
Inhibition Test-baseline "No"							
Inhbition test Yes	0.023	0.071	-0.119	0.024	0.158	+	No
Inhibiton test only Negative samples	0.253	0.120	0.018	0.254	0.486	+	Yes

Sample Group 2.3 $\log_{10}$ Copies/ml Viral Load

Table D.5: *Summary statistics of the parameter estimates from the CBM for HBV sample group 2.3* $\log_{10}$ *copies/ml viral load: estimated mean, standard deviation (SD), confidence interval, tendency describing the sign of the estimated mean and significance at two-sided 5% of the parameter.*

Covariate	Mean	SD	95% Confidence Interval			Tendency	Significance
			2.50%	Median	97.50%		
Intercept	2.392	0.197	2.014	2.388	2.784		
Year- baseline "2004"							
Year 2002	0.226	0.132	-0.028	0.223	0.490	+	No
Year 2003	-0.333	0.135	-0.593	-0.334	-0.067	-	Yes
Technology group- baseline "CC"							
Tech CIH-CC	0.655	0.255	0.131	0.655	1.149	+	Yes
Tech RTIH-CC	0.695	0.147	0.397	0.697	0.973	+	Yes
Tech RTC-CC	0.409	0.171	0.077	0.408	0.745	+	Yes
Tech bDNA-CC	-0.543	0.668	-2.002	-0.489	0.607	-	No
Tech HC-CC	0.011	1.092	-2.045	-0.010	2.168	+	No
Anti- baseline "No"	-0.040	0.120	-0.271	-0.043	0.198	-	No
Accred- baseline "No"	-0.026	0.100	-0.225	-0.024	0.172	-	No
OtherSpc. baseline "No"	-0.058	0.115	-0.304	-0.051	0.149	-	No
Analysis method- baseline Singly							
Analysis Duplicated	-0.039	0.120	-0.269	-0.040	0.206	-	No
Analysis Other	0.164	0.211	-0.244	0.159	0.594	+	No
Plasma- baseline 0-10							
Group 1: 11-100	-0.099	0.168	-0.445	-0.097	0.215	-	No
Group 2: 101-1,000	-0.086	0.132	-0.346	-0.085	0.167	-	No
Group 3: 1,001-2,000	-0.097	0.228	-0.537	-0.097	0.355	-	No
Group 4: 2,001-10,000	0.086	0.146	-0.209	0.092	0.364	+	No
Group 5: > 10,000	-0.366	0.256	-0.916	-0.343	0.072	-	No
Serum- baseline 0-10							
Group 1: 11-100	0.053	0.163	-0.293	0.061	0.361	+	No
Group 2: 101-1,000	-0.021	0.124	-0.258	-0.022	0.219	-	No
Group 3: 1,001-2,000	0.110	0.163	-0.209	0.109	0.427	+	No
Group 4: 2,001-10,000	-0.081	0.182	-0.460	-0.071	0.252	-	No
Group 5: > 10,000	-0.897	0.628	-2.269	-0.814	0.083	-	No
Labtype- baseline Hospital							
Public Health	-0.092	0.155	-0.403	-0.090	0.208	-	No
Private	-0.067	0.180	-0.454	-0.058	0.253	-	No
Reference	-0.335	0.330	-1.001	-0.323	0.262	-	No
Manufacture	-0.102	0.209	-0.542	-0.089	0.274	-	No
Research	-0.012	0.149	-0.334	-0.006	0.263	-	No
Inhibition Test-baseline "No"							
Inhbition test Yes	-0.019	0.120	-0.270	-0.015	0.215	-	No
Inhibiton test only Negative samples	0.434	0.202	0.039	0.431	0.838	+	Yes

Estimates of Variances

Table D.6: *Summary statistics of the variance estimates from the CBM for HBV sample groups classified by technology type: estimated mean, standard deviation (SD), confidence interval.*

Estimated Variances per sample group	Mean	SD	95% Confidence Interval		
			2.50%	Median	97.50%
6 Log_{10} copies/ml					
In-house	1.118	0.183	0.820	1.101	1.521
Commercial	0.313	0.039	0.244	0.310	0.398
5 Log_{10} copies/ml					
In-house	0.955	0.109	0.763	0.947	1.193
Commercial	0.191	0.014	0.165	0.190	0.222
4 Log_{10} copies/ml					
In-house	0.816	0.116	0.615	0.807	1.074
Commercial	0.195	0.019	0.162	0.194	0.236
3.5 Log_{10} copies/ml					
In-house	0.446	0.093	0.296	0.434	0.665
Commercial	0.131	0.021	0.096	0.129	0.178
3 Log_{10} copies/ml					
In-house	0.541	0.086	0.395	0.533	0.732
Commercial	0.187	0.019	0.154	0.186	0.227
2.3 Log_{10} copies/ml					
In-house	0.500	0.113	0.322	0.485	0.765
Commercial	0.237	0.040	0.171	0.234	0.327

Appendix E

Sensitivity Analysis: Classical Tests

We used classical statistical tests to compare the means of the posterior distributions of the estimates in order to prove formally that there is no significant difference between the models for a specific sample group chosen.

E.1 Application to the QLBM

E.1.1 Application to the EV Data Analysis

The ROC curves shows an analysis of the specificity and sensitivity from each of the models proposed for studying the robustness of the QLBM model. Figure E.1 shows the curves for each model. Almost no differences were observed between results obtained from the different models. Furthermore, the differences for the Area Under the Curve (AUC) was tested and Table E.1 shows the results obtained. The AUC is a global measure for comparison between models. It is the probability of a correct classification for a couple of responses of correct and incorrect result. It shows significant difference from 0.5 indicating that the model are better for discriminating between correct and incorrect results than random selection, indicating that all of them are useful at predicting a correct result.

Table E.1: *AUC for the EV data analysis.*

Model	AUC	SD	Confidence Interval 2.5%	97.5%	p-value
Full QTBM	0.697	0.032	0.633	0.760	0.000
Model 1	0.693	0.033	0.629	0.757	0.000
Model 2	0.695	0.032	0.631	0.759	0.000
Model 3	0.695	0.033	0.631	0.759	0.000
Reduced QTBM	0.684	0.033	0.619	0.748	0.000

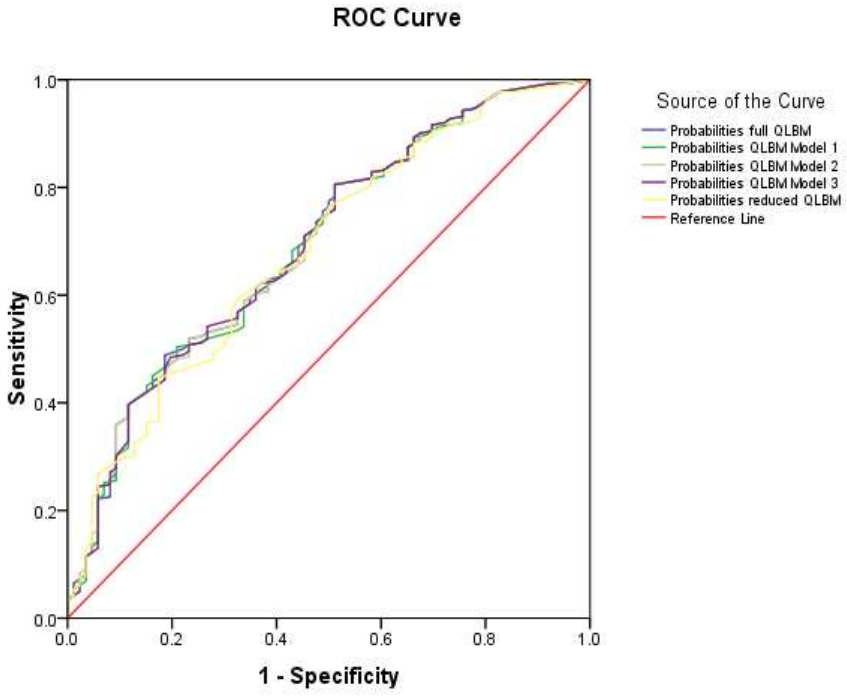


Figure E.1: *ROC curve for the EV data analysis from the proposed models.*

E.1.2 Application to the HBV Data Analysis

As in previous subsection we applied classical tests for differences between the results of the several models proposed in the sensitivity analysis. Figure E.2 and Table E.2 show the curves and the areas under the curves, respectively, from the applied models to the HBV data. Similar conclusions to the previous subsection are obtained.

Table E.2: *AUC test for the HBV data analysis.*

Model	AUC	SD	Confidence 2.5%	Interval 97.5%	p-value
Full QTBM	0.982	0.017	0.950	1.00	0.000
Model 1	0.981	0.016	0.949	1.00	0.000
Model 2	0.981	0.016	0.949	1.00	0.000
Model 3	0.981	0.016	0.949	1.00	0.000
Reduced QTBM	0.924	0.033	0.859	0.988	0.000

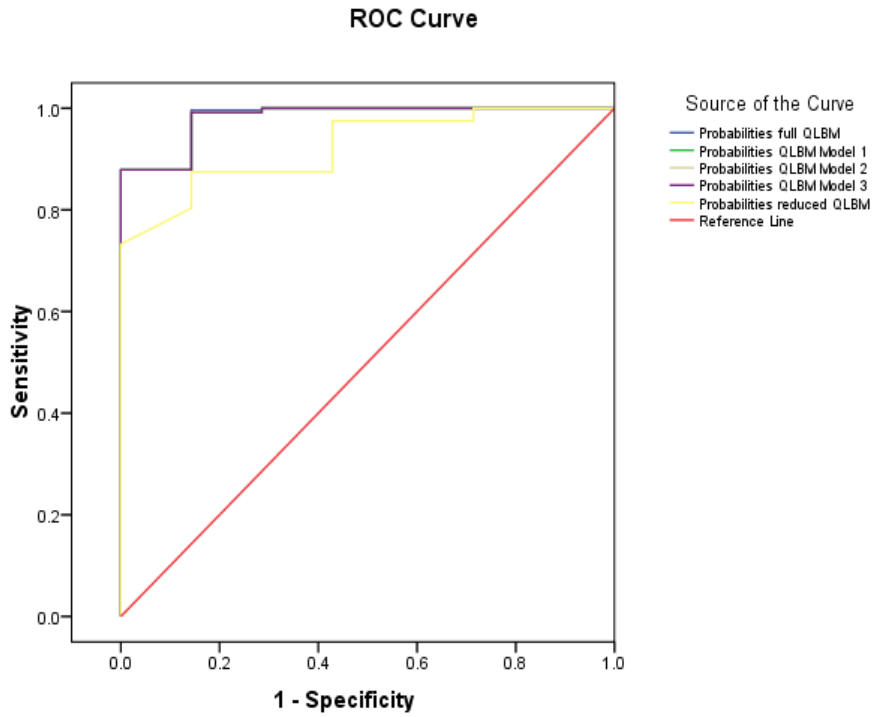


Figure E.2: *ROC curve for the HBV data analysis from the proposed models.*

E.2 Application to the QTBM

To test the differences between the mean estimated from the QTBM and the rest of models proposed we performed classical paired tests for the mean of the posterior distributions for the estimates of viral loads. This test is a general summary for comparing the posterior distributions obtained from the different models.

E.2.1 Application to the HBV Data Analysis

Table E.3 shows the differences on the posterior means, the confidence intervals and the p-values. No significant differences were found between the applied QTBM and the other models proposed in the sensitivity analysis.

E.2.2 Application to the HCV Data Analysis

Table E.4 shows the differences on the posterior means, the confidence intervals and the p-values. It is shown not different between the applied QTBM and the rest of models proposed in the sensitivity analysis.

Table E.3: *Paired tests for differences of means from the HBV analysis.*

Model	Differences	SD	Confidence Interval 2.5%	97.5%	p-value
Full QTBM and Model 1	0.001	0.007	-0.001	0.002	0.517
Full QTBM and Model 2	0.001	0.007	-0.0012	0.002	0.517
Full QTBM and Model 3	0.001	0.007	-0.001	0.002	0.517
Full QTBM and Reduced QTBM	-0.006	0.066	-0.022	0.011	0.496

Table E.4: *Paired tests for differences of means from the HCV analysis.*

Model	Differences	SD	Confidence Interval 2.5%	97.5%	p-value
Full QTBM and Model 1	0.001	0.077	-0.014	0.016	0.890
Full QTBM and Model 2	0.001	0.004	-0.000	0.001	0.202
Full QTBM and Model 3	0.001	0.004	-0.000	0.001	0.065
Full QTBM and Reduced QTBM	-0.003	0.075	-0.018	0.012	0.701

Appendix F

Publications

In this appendix, we present the two posters and corresponding abstract that the author published in relation with the QLBM and QTBM models (García-Fernández, Wallace and Staines, 2007) (García-Fernández, Wallace, Staines and van Loon, 2007).

F.1 Modelling Performances of Quality Control for Molecular Diagnostics Participants in Enterovirus Programmes over Time

F.1.1 Abstract

F.1.1.1 Objectives

To analyse data from Enterovirus (EV) Quality Control (QC) programmes over time to provide a better feedback to participants and improve the design of future QC programmes. These should help improve the performance of molecular diagnostic technologies users.

F.1.1.2 Methods

Homogeneity tests and Generalised Linear Models (GLM) are used to model the positive/negative responses provided in the QC programmes. Homogeneity tests are performed on the ratio of correct results over time for the different category of samples. GLM (logistic regression) is used to find significant factors on the estimated probabilities of correct/incorrect results over time and sample categories.

F.1.1.3 Results

Data from the 1999 to 2005 QCMD EV programmes were analysed. The EV proficiency panel compositions varied by year although they contained series of similar samples (sample category): negative, non-EV and EV samples with different serotype and viral load.

Labs were categorised on whether they had been in previous EV programmes, if they had returned a correct result in that previous programme and whether they were accredited. The technology used was one amongst other factors included as potential explanatory variables.

The difference in the proportions of false positives and false negatives results over time varied depending on sample and lab category. The proportion of false positives for non-EV samples varied on the virus included in the sample. Laboratories that had a correct result in the previously panel are significantly less likely to obtain a false positive that those that are new to the programme. However, no significant differences were found when analysing negative samples. No significance difference was found between the performances of accredited and non-accredited participants. Performances from different technology users varied over time and sample category. In 2004/05 commercial assay users were less likely to detect low dilution samples than in 2002/03. The proportion of correct result over time decreased, as the dilution series are lower.

F.1.1.4 Conclusions

Performance of participants to the EV QC programme depended on the virus. These results suggest that participating in an EV QC programme helps improve the performance of laboratories. However no difference was found between the performance of accredited and non-accredited labs.

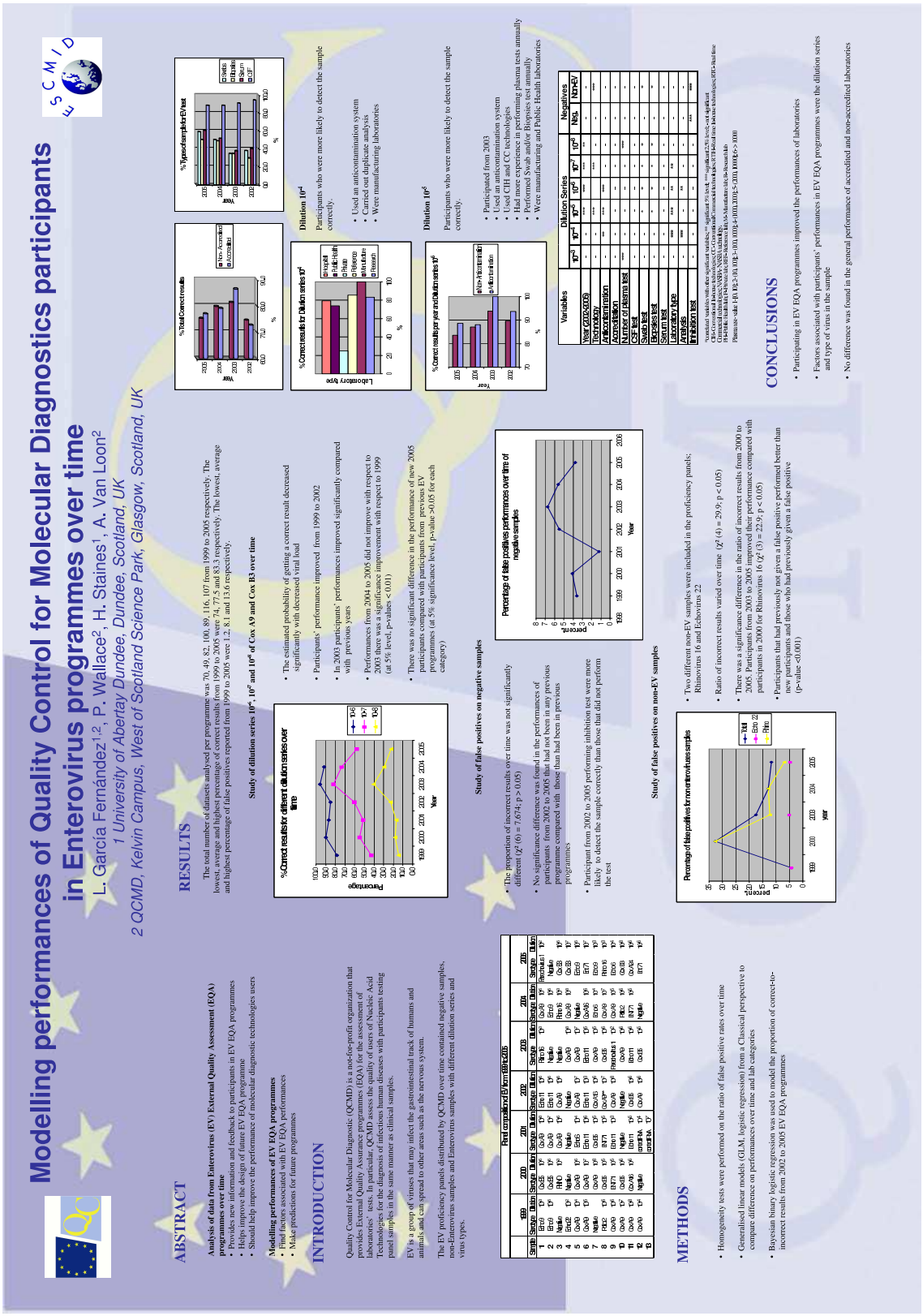


Figure F.1: Modelling performances of Quality Control for Molecular Diagnostics participants in Enterovirus programmes over time

F.2 Statistical Modelling of the Performance of Nucleic Acid Amplification Technologies in Clinical Diagnostic Applied to Quality Control for Molecular Diagnostics Hepatitis B Virus Programmes

F.2.1 Abstract

F.2.1.1 Objectives

Pathogen load estimation provided by Nucleic Acid Technologies (NATs) used to diagnose and manage patients with infectious disease gives more information than positive/negative results available from earlier techniques. Generalised Linear Models (GLM) are currently used to analyse NAT users' performance. However, these ignore the censored quantities lying outside the detection limits of the assays. We introduce an approach that identifies significant factors associated with lab performance including censored values. The model is tested on data from 4 years of QCMD HBV programmes.

F.2.1.2 Methods

We propose a GLM allowing a censored mechanism and Bayesian parameter inferences using Markov Chain Monte Carlo (MCMC) methods. The model assumes that the log10 copies/ml pathogen load estimates are normally distributed.

F.2.1.3 Results

Potential explanatory variables in the model include NAT technology used, year of programme and sample genotype. Lab performance was assessed by the difference between the lab's and the target estimate of the pathogen load.

The proportion of censored data was higher for samples with lower viral load.
Users of Commercial PCR technologies were compared with other technology groups and results depended on viral load. HBV genotype was a significant factor for some sample categories whilst programme year was significant for almost all sample categories.

F.2.1.4 Conclusions

The model deals with multiple parameters and censored values in a simpler way than traditional statistical techniques. Information from the censored values (outside the assay limits of detection) is incorporated in the model and further modelling of the censored values can be made.

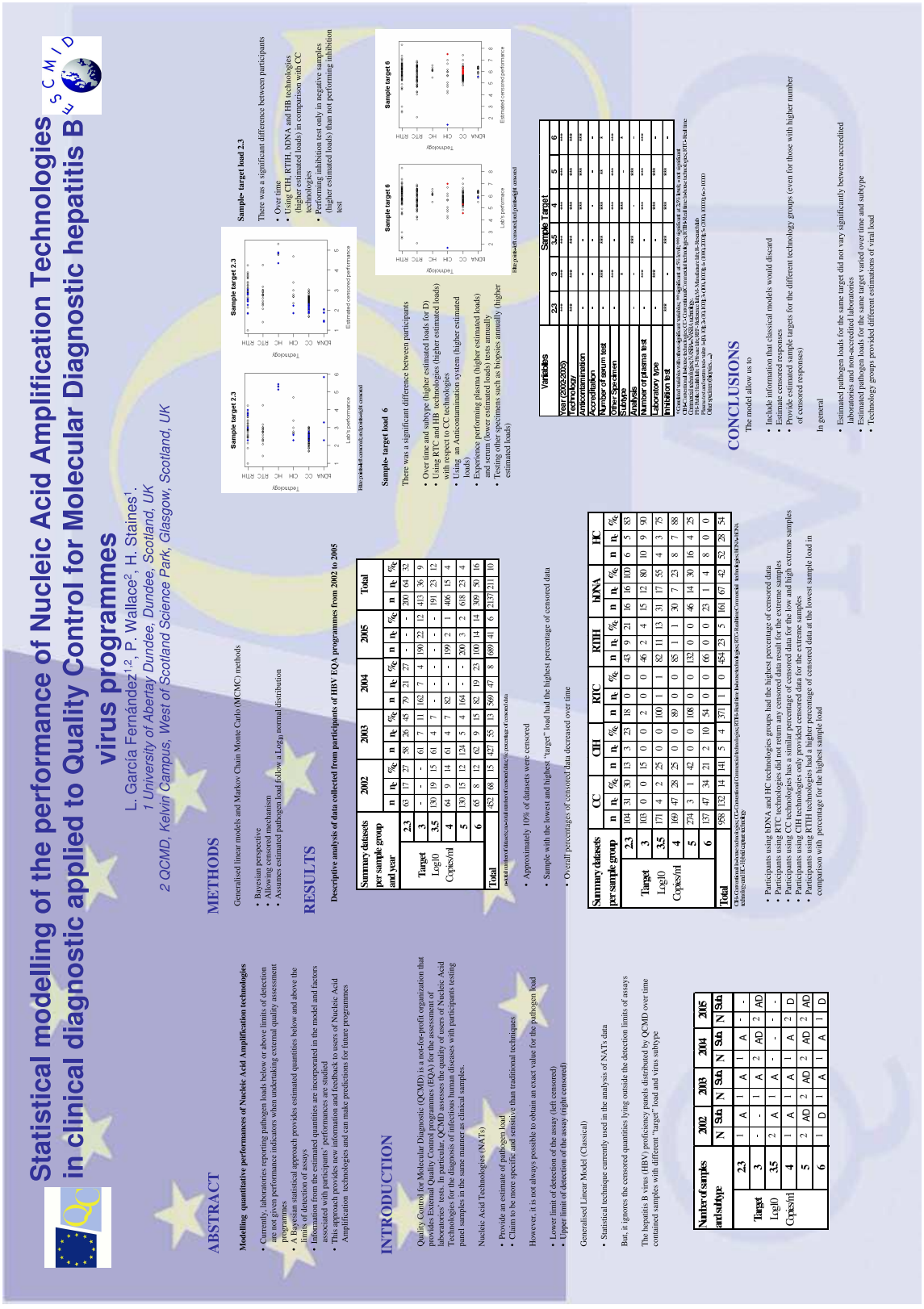


Figure F.2: Statistical modelling of the performance of Nucleic Acid Amplification Technologies in clinical diagnostic applied to Quality Control for Molecular Diagnostics hepatitis B virus programmes

Bibliography

Apfalter, P., Reischl, U. and Hammerschlog, M. R. (2005). In-house Nucleic Acid Amplification Assays in Research: How much quality control is needed before one can rely upon the results?, *Journal of Clinical Microbiology* **43**(12): 5835–5841.

Banerjee, S., Carlin, B. P. and Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis of Spatial Data*, Chapman & Hall, New York.

Barndorff-Nielsen, O. E., Cox, D. R. and Klueppelberg, C. (2001). *Complex Stochastic Systems*, Chapman & Hall, Florida.

Bayes, T. (1763). An eassy towards solving a problem in the doctrine of chances, *Phil. Trans. Roy. Soc.* **53**: 370–418.

Berger, J. (2000). Bayesian Analysis: a look of today and thoughts of tomorrow, *Journal American Statist. Assoc.* **95**: 1269–1276.

Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory*, Wiley, New York.

BioTecniques (2007). *BioTechniques' Protocol Guide 2008*. http://www.biotechniques.com.

Burtis, C. and Ashwood, E. (2007). *Fundamentals of Molecular Diagnostics*, Saunders Elsever, Missouri.

Carlin, B. P. and Louis, T. A. (2001). *Bayes and Empirical Bayes Methods for Data Analysis*, 2^{nd} edn, Chapman & Hall, New York.

Celeux, G., Forbes, F., Robert, C. P. and Titterington, D. M. (2006). Deviance Information Criteria for Missing Data Models, *Bayesian Analysis* **1**(4): 651–674.

CEN TC 140 prEN 14136 (2004). Use of EQA schemes in the assessment of the performance of in vitro diagnostic examination procedures.

Clogg, C., Rubin, D., Schenker, N., Shultz, B. and Wideman, L. (1991). Multiple imputation of industry and occupation codes in Census public use samples using Bayesian logistic regression, *Journal of American Statistical Association* **86**: 68–78.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*, 2^{nd} edn, Lawrence Erlbaum Associates, Inc., New Jersey.

Congdon, P. (2001). *Bayesian Statistical Modelling*, John Wiley & Sons, Ltd., Chichester.

Conraths, F. J. and Schare, G. (2006). Validation of molecular-diagnostic techniques in the parasitological laboratory, *Veterinary Parasitology* **136**: 91–98.

Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data*, Chapman & Hall, London.

David, M. H., Little, R. J. A., Samuhel, M. E. and Triest, R. K. (1986). Alternative methods for CPS income imputation, *Journal of the American Statistical Association* **81**: 29–41.

Davison, A. C. (2003). *Statistical Models*, Cambridge University Press, Cambridge.

Dennis Lo, Y., Chiu, R. and Allenchan, K. (2006). *Clinical Applications of PCR*, 2^{th} edn, Humana Press, Inc., New Jersey.

Dobson, A. (1990). *An Introduction to Generalized Linear Models*, Chapman Hall, U.S.A.

Dugard, P., Todman, J. and Staines, H. (2010). *Approaching Multivariate Analysis. A Practical Introduction*, 2^{nd} edn, Routledge, New York.

Fahrmeir, L. and Tutz, G. (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models*, 2^{nd} edn, Springer, New York.

Fishman, G. S. (1996). *Monte Carlo: Concepts, Algorithms, and Applications*, Springer-Verlag, New York.

Fleige, S. and Pfaffl, M. W. (2006). RNA integrity and RT-PCR performance, *Molecular Aspects Med.* **27**: 126–139.

Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo*, 2^{nd} edn, Chapman & Hall, London.

García-Fernández, L. M., Wallace, P. and Staines, H. (2007). Statistical modelling of the performance of nucleic acid amplification technologies in clinical diagnostic applied to quality control for molecular diagnostic hepatitis B virus programmes, *Clinical Microbiology and Infection* **13. Supplement 1**.

García-Fernández, L. M., Wallace, P., Staines, H. and van Loon, A. (2007). Modelling performance of quality control for molecular diagnostic participants in Enterovirus programmes over time, *Clinical Microbiology and Infection* **13. Supplement 1**.

Gelman, A., Carlin, J. C., Stern, H. and Rubin, D. B. (2004). *Bayesian Data Analysis*, 2^{nd} edn, Chapman & Hall, New York.

Gelman, A., Chew, G. L. and Shnaidman, M. (2004). Bayesian analysis of serial dilution assays, *Biometrics* **60**: 407–417.

Gelman, A., Goegebeur, Y., Tuerlinckx, F. and Van Mechelen, I. (2000). Diagnostic checks for discrete-data regression models using posterior predictive simulations, *Applied Statistics* **49**: 247–268.

Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge University Press, New York.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Trans. Pattn. Anal. Mach. Intel.* **6**: 721–741.

Germann, D. and Telenti, A. (1995). Nucleic Acid amplification methods in diagnostic Virology, *Journal of Microbiological methods* **23**: 31–39.

Geurden, T., Claerebout, E., Vercruysse, J. and Berkvens, D. (2004). Estimation of test characteristics and prevalence of Giardia duondenalis in dairy claves in Belgium using Bayesian approach, *Int. J. Parasitol* **34**: 1121–1127.

Geyer, C. J. (1992). Practical Markov chain Monte Carlo, *Statistical Science* **7**: 473–511.

Gilks, W., Clayton, D., Spiegelhalter, D., Best, N., McNeil, A., Sharples, L. and Kirby, A. (1993). Modelling complexity: Applications of Gibbs sampling in medicine, *J. Royal Statistical Society Series B* **55**: 39–52.

Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1996). *Markov chain Monte Carlo in Practice*, Chapman & Hall, London.

Goodrich, B. and Lu, Y. (2007). *tobit.bayes: Bayesian Linear Regression for a Censored Dependent Variable*. in Kosuke Imai, Gary King, and Olivia Lau, "Zelig: Everyones Statistical Software", http://gking.harvard.edu/zelig.

Hak, E., Verhaij, M., Grobbee, D. E., Nichol, K. L. and Hoes, A. W. (2002). Confounding by indication in non-experimental evaluation of vaccine effectiveness: The example of prevention of influenza complications, *Journal of Epidemiology and Community Health* **56**: 951–955.

Hand, D. J. and Crowder, M. J. (1996). *Practical Longitudinal Data Analysis*, Chapman & Hall, London.

Harrell, J. F. E. (2001). *Regression Modelling Strategies*, Springer Science, New York, USA.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications, *Biometrika* **57**: 97–109.

Heath, S. (1997). Markov chain Monte Carlo methods for radiation hybrid mapping, *J. Computational Biology* **4**: 505–515.

Heid, et al., C. A. (1996). Real time quantitative PCR, *Genome Res.* **6**: 986–994.

ILAC (2010). *International laboratory Accreditation*. http://www.ilac.org.

ILAC-G13 (2000). Guidelines for the requirements for the competence of providers of proficiency testing.

ISO/IEC Guides (1997a). Proficiency testing by inter laboratory comparisons. Part1: Development and operations of proficiency testing scheme.

ISO/IEC Guides (1997b). Proficiency testing by inter laboratory comparisons. Part2: Selection and use of PT schemes by laboratory accreditation bodies.

Katz, et al., A. (2008). Nomogram for the Prediction of Having Four or More Involved Nodes for Sentinel Lymph Node-Positive Breast Cancer, *Journal of Clinical Oncology* **26**: 2093–2098.

Kloek, J. and van Dijk, H. K. (1978). Bayesian estimates of equation system parameters: an application of integration by Monte Carlo, *Econometrica* **46**: 1–19.

Krause, R. D., Anand, V. D., Gruemer, H. D. and Willke, T. A. (1975). The impact of Laboratory Error on the Normal Range: A Bayesian Model, *Clinical Chemistry* **21**: 321–324.

Krzanowski, W. J. (1998). *An Introduction to Statistical Modelling*, Arnold, London.

Lee, P. (2004). *Bayesian Statistics*, London: Arnold.

Lindley, D. V. (1965). *Introduccion to Probability and Statistics from a Bayesian Viewpoint*, Vol. 2, Cambridge University Press, Cambridge.

Little, R. J. A. and Rubin, D. B. (1992). *Statistical Analysis with Missing Data*, Wiley, New York.

Liu, C. (1995). Missing data imputation using the multivariate t distribution, *Journal of Multivariate Analysis* **48**: 198–206.

Lunn, D. J., Thomas, A., Best, N. and Spiegelhalter, D. (2000). WinBUGS–a Bayesian Modelling Framework: Concepts, Structure, and Extensibility, *Statistics and Computing* **10**: 325–337.

McCullagh, P. and Nelder, J. A. (1989). *Generalised Linear Models*, Chapman & Hall, New York.

McCulloch, C. E. and Searle, S. R. (2001). *Generalised, Linear, and Mixed Models*, John Wiley & Sons, Inc., New York.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, N. M., Teller, A. H. and Teller, E. (1953). Equations of state calculations by fast computing machine, *J. Chem. Phys.* **21**: 1087–1091.

Miettinen, O. S. and Cook, E. F. (1981). Confounding: essence and detection, *American Journal of Epidemiology* **114**: 593–603.

Morgan, B. J. (2000). *Applied Stochastic Modelling*, Arnold, London.

NEQAS (2010). *National External Quality Assessment Service*. http://www.ukneqas.org.uk.

Niazi, S. K. (2007). Handbook of bioequivalence testing.

Pogathota, R. (2007). *Evaluating the performance of molecular diagnostic assay users: A statistical analysis of EQA data*, MSc. Thesis. University of Abertay Dundee.

Project, T. B. (1996-2004). *MRC Biostatistics Unit, Cambridge, UK*. http://www.mrc-bsu.cam.ac.uk/bugs.

QCMD (2010). *Quality Control for Molecular Diagnostics*. http://www.qcmd.org.

Rawlings, J. O., Pantula, S. G. and Dickey, D. A. (1998). *Applied Regression Analysis*, 2^{nd} edn, Springer, New York.

Roche, A. F., Wainer, H. and Thissen, D. (1975). *Skeletal Maturity: The Knee Joint as a Biological Indicator*, Plenum, New York.

Spanos, A. (1999). *Probability Theory and Statistical Inference*, Cambridge University Press, Cambridge.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit, *Journal of Royal Statist. Society Series B* **64**: 583–639.

Staines, H. J., García-Fernández, L., Pagothota, R., Wallace, P. S., MacKay, W. G. and Van Loon, A. M. (2009). Monitoring performance of nucleic acid-based diagnostic measurement system users by EQA, *Accred. Qual. Assur.* **14**: 243–252.

Thissen, D. (1986). *MULTILOG Version 5: User's Guide.*, 5^{th} edn, Scientific Software, Inc., Mooresville,IN.

Tobin, J. (1958). Estimation of Relationships for Limited Dependent Variables, *Econometrica* **26**: 24–36.

Twisk, J. W. R. (2003). *Applied Longitudinal Data Analysis for Epidemiology. A Practical Guide*, Cambridge University Press, New York.

Tyler, D. S. and Balch, C. M. (2005). Nomograms and Staging in Melanoma: New Tools for Better Predicting Outcomes, *Annals of Surgical Oncology* **12(4)**: 267–269.

Upton, G. and Cook, I. (2002). *A Dictionary of Statistics*, Oxford University Press, New York.

Valentine-Thon, E., van Loon, A. M., Schirm, J., Reid, J., Klapper, P. E. and Cleator, G. M. (2001). European Proficiency Testing Program for Molecular Detection and Quantification of Hepatitis B Virus DNA, *Journal of Clinical Microbioloy* **39**: 4407–4412.

van la Parra, R. F. D., Ernst, M. F., Bevilacqua, J. L., Mol, S. J., van Zee, K. J., Broekman, J. M. and Boscha, K. (2009). Validation of a Nomogram to Predict the Risk of Nonsentinel Lymph Node Metastases in Breast Cancer Patients with a Positive Sentinel Node Biopsy, *Annals of Surgical Oncology* **16**: 1128–1135.

Volinsky, C. T. and Raftery, A. E. (2000). Bayesian Information Criterion for Censored Survival Models, *Biometrics* **56**: 256–262.

Waller, L. A. and Gotway, C. A. (2004). *Applied Spatial Statistics for Public Health Data*, John Wiley & Sons, Inc., New Jersey.

Westgard, J. O. (2004). Clinical quality vs analytical performance: what are the right targets and target values?, *Accred Qual Assur* **10**: 10–14.

Wong, S. K. (2005). Evaluation of the use of consensus values in proficiency testing programmes, *Accred Qual Assur* **10**: 409–414.

Zar, J. (1999). *Biostatistical Analysis*, 4^{th} edn, Prentice Hall, New Jersey.