**2012 |** Faculteit Wetenschappen

DOCTORAATSPROEFSCHRIFT

# Statistical Modelling Strategies for Public Health Research

## Applications in the Surveillance of Antimicrobial Consumption, the Diagnosis of Acute Infections and the Diagnosis of Coronary Heart Disease

*Proefschrift voorgelegd tot het behalen van de graad van doctor in de wetenschappen, wiskunde, te verdedigen door:*

**Girma Minalu Ayele**

*Promotor: prof. dr. Marc Aerts*
*Copromotor: prof. dr. Niel Hens*

D/2012/2451/34

Maastricht University

universiteit hasselt
KNOWLEDGE IN ACTION

# Jury Members

| | | |
|---|---|---|
| **Prof. dr. Marc Aerts** | - | Promotor: I-BioStat, University Hasselt, Belgium |
| **Prof. dr. Niel Hens** | - | Copromotor: I-BioStat, University Hasselt & Chermid, University of Antwerp, Belgium |
| **Prof. dr. Noel Veraverbeke** | - | Chairman: I-BioStat, emeritus, University Hasselt, Belgium |
| **Prof. dr. Geert Molenberghs** | - | I-BioStat, University Hasselt & KULeuven, Belgium |
| **Prof. dr. Herman Goossens** | - | Laboratory of Medical Microbiology, University of Antwerp, Belgium |
| **Prof. dr. Samuel Coenen** | - | Laboratory of Medical Microbiology & Center for General Practice, University of Antwerp, Belgium |
| **Prof. dr. Norbert Donner-Banzhoff** | - | Department of Primary Care, University Marburg, Germany |
| **Prof. dr. Christel Faes** | - | I-BioStat, University Hasselt, Belgium |
| **Prof. dr. Frank Buntinx** | - | Academic Center for General Practice, KULeuven, Belgium |

# Acknowledgements

It is a pleasure to thank many people who made this thesis possible. This work would not have been possible without the support and help of them.

First and foremost I would like to express my heartfelt gratitude to my PhD supervisor Prof. dr. Marc Aerts for his enthusiasm, his inspiration, his patience, and his guidance throughout these four years. I would also like to thank my co-supervisor, Prof. dr. Niel Hens, for his great ideas and suggestions. Thank you so much for both of you. I am grateful for the feedback and suggestions I received from the jury members to an earlier version of this thesis.

I am indebted to the METHUSALEM-consortium and the INTERCHEST-project for funding my PhD. I am grateful for the opportunity to conduct research within an interdisciplinary context together with the Vaccine & Infectious Disease Institute (VAXINFECTIO) at the University of Antwerp, and with an international working group on chest pain in primary care (INTERCHEST), mostly in collaboration with the Department of Primary Care at the University Marburg, Germany. I would like to thank Samuel Coenen (VAXINFECTIO, University of Antwerp), Katherine Loens (VAXINFECTIO, University of Antwerp) and Jörg Haasenritter (Department of Primary Care, University of Marburg) who have been extremely supportive. My gratitude also goes to all my co-authors.

I am indebted to my many colleagues at CENSTAT for providing a stimulating and fun environment in which to learn and grow. I am especially grateful to my office mates in D4: Ariel, Saskia, Amin, Birhanu and Kim. It has been a pleasure to share the office with you. I would like to thank my friends in Belgium (especially Taddesse, Sintayehu, Selam, Berhane, Liesbeth, Senait, Luc, Melkie and Patrick) for helping me get through the difficult times, and for all the emotional support, entertainment and caring they provided. I am grateful to the secretaries (especially Ms. Martine Machiels) for assisting me in many different ways.

I would also like to thank my entire extended family members for their unconditional love and supports. My beloved brothers (Araya, Abraham and Thewodros), my lovely sisters (Nigist, Haymanot and Mekdes), my brother-in-laws (Belay and Solomon) and sister-in-law (Hani) were particularly supportive.

Finally, and most importantly, I am forever indebted to my parents, Minalu Ayele and Aster Bezabeh, who have always encouraged me to pursue my studies. To them I dedicate this thesis.

Thank you everyone that I couldn't mention here, who have encouraged and helped me one way or the other.

† *"... the LORD is my strength and my shield; my heart trusts in him, and I am helped. My heart leaps for joy and I will give thanks to him ..." Psalm 28:7*

Girma Minalu Ayele
17 September 2012
Diepenbeek

# Contents

# List of Publications

This dissertation has been based on the following scientific papers:

Adriaenssens, N., Coenen, S., Versporten, A., Muller, A., **Minalu, G.**, Faes, C., Vankerckhoven, V., Aerts, M., Hens, N., Molenberghs, G., Goossens, H. (2011). European Surveillance on Antimicrobial Consumption (ESAC): outpatient antibiotic use in Europe (1997–2009). *Journal of Antimicrobial Chemotherapy*, **66** Suppl 6:vi3–vi12.

Adriaenssens, N., Coenen, S., Versporten, A., Muller, A., **Minalu, G.**, Faes, C., Vankerckhoven, V., Aerts, M., Hens, N., Molenberghs, G., Goossens, H. (2011). European Surveillance on Antimicrobial Consumption (ESAC): outpatient macrolide, lincosamide and streptogramin use in Europe (1997–2009). *Journal of Antimicrobial Chemotherapy*, **66** Suppl 6:vi37–vi45.

Adriaenssens, N., Coenen, S., Versporten, A., Muller, A., **Minalu, G.**, Faes, C., Vankerckhoven, V., Aerts, M., Hens, N., Molenberghs, G., Goossens, H. (2011). European Surveillance on Antimicrobial Consumption (ESAC): outpatient quinolone use in Europe (1997–2009). *Journal of Antimicrobial Chemotherapy*, **66** Suppl 6:vi47–vi56.

Coenen, S., Adriaenssens, N., Versporten, A., Muller, A., **Minalu, G.**, Faes, C., Vankerckhoven, V., Aerts, M., Hens, N., Molenberghs, G., Goossens, H. (2011). European Surveillance of Antimicrobial Consumption (ESAC): outpatient use of tetracyclines, sulphonamides and trimethoprim, and other antibacterials use in Europe (1997–2009). *Journal of Antimicrobial Chemotherapy*, **66** Suppl 6:vi57–vi70.

**Minalu, G.**, Aerts, M., Coenen, S., Versporten, A., Muller, A., Adriaenssens, N., Beutels, P., Molenberghs, G., Goossens, H., Hens, N. (2011). Application

of mixed-effects models to study the country-specific outpatient antibiotic use in Europe: a tutorial on longitudinal data analysis. *Journal of Antimicrobial Chemotherapy*, **66** Suppl 6:vi79–vi87.

Versporten, A., Coenen, S., Adriaenssens, N., Muller, A., **Minalu, G.**, Faes, C., Vankerckhoven, V., Aerts, M., Hens, N., Molenberghs, G., Goossens, H. (2011). European Surveillance on Antimicrobial Consumption (ESAC): outpatient penicillin use in Europe (1997–2009). *Journal of Antimicrobial Chemotherapy*, **66** Suppl 6:vi13–vi23.

Versporten, A., Coenen, S., Adriaenssens, N., Muller, A., **Minalu, G.**, Faes, C., Vankerckhoven, V., Aerts, M., Hens, N., Molenberghs, G., Goossens, H. (2011). European Surveillance on Antimicrobial Consumption (ESAC): outpatient cephalosporin use in Europe (1997–2009). *Journal of Antimicrobial Chemotherapy*, **66** Suppl 6:vi25–vi35.

Haasenritter, J., Aerts, M., Bösner, S., Buntinx, F., Burnand, B., Herzig, L., Knottnerus, J.A., **Minalu, G.**, Nilsson, S., Renier, W., Sox, C., Sox, H., Donner-Banzhoff, N. (2012). Coronary heart disease in primary care: accuracy of medical history and physical findings in patients with chest pains - a study protocol for a systematic review with individual patient data. *BMC Family Practice*, **13**:81.

## Papers in Revision and Working Papers

Haasenritter, J., Nilsson, S., Aerts, M., Bösner, S., Buntinx, F., Burnard, B., Herzig, L., Knottnerus, A., **Minalu, G.**, Renier, W., Sox, C., Sox, H., Donner-Banzhoff, N. (2012). Diagnostic accuracy of signs and symptoms for diagnosing CHD in primary care. *Working paper*.

Loens, K., **Minalu, G.**, Hens, N., Ursi, D., Goossens, H., Aerts, M., and Ieven, M. (2012a). Importance of methodology in detection of *Mycoplasma pneumoniae* infections in patients with lower respiratory tract infections. *Working paper*.

Loens, K., **Minalu, G.**, Hens, N., Ursi, D., Goossens, H., Aerts, M., and Ieven, M. (2012b). Evaluation of 2 amplification methods and 4 serological tests for the detection of *Chlamydophila pneumoniae* in patients with community-acquired pneumonia. *Working paper*.

**Minalu, G.**, Aerts, M., Coenen, S., Versporten, A., Muller, A., Adriaenssens, N., Beutels, P., Molenberghs, G., Goossens, H., Hens, N. (2012). Adaptive Change-point Mixed Models Applied to Data on Outpatient Antibiotic Use in Europe. Hasselt: University of Hasselt, Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BIOSTAT). *In revision*.

**Minalu, G.**, Aerts, M., Bösner, S., Buntinx, F., Burnand, B., Haasenritter, J., Herzig, L., Knottnerus, J.A., Nilsson, S., Renier, W., Sox, C., Sox, H., Donner-Banzhoff, N. (2012). Medical history and physical findings for coronary heart disease in primary care - a meta-analysis with individual patient data. *In revision*.

**Minalu, G.**, Aerts, M., Loens, K., Ursi, D., Goossens, H., Ieven, M., and Hens, N. (2012). Bayesian Latent Class Models with Conditionally Dependent Diagnostic Tests: A Simulation Study. *Working paper*.

## Proceedings

**Minalu, G.**, Hens, N., Aerts, M., Coenen, S., Muller, A., Adriaenssens, N., Beutels, P., Molenberghs, G., Goossens, H. In Bowman, A.W., editor. Time Series and Mixed Models to Study the Country-Specific Outpatient Antibiotic Use in Europe. Proceedings of the $25^{th}$ International Workshop on Statistical Modelling, July 5–9, 2010 Glasgow, Scotland, UK. pp. 53–56.

**Minalu, G.**, Aerts, M., Coenen, S., Adriaenssens, N., Versporten, A., Muller, A., Vankerckhoven, V., Hens, N., Molenberghs, G., Goossens, H. Bayesian Change-point Mixed Models Applied to Data on Outpatient Antibiotic Use in Europe. Proceedings of the $58^{th}$ World Statistics Congress of the International Statistical Institute, August 21–26, 2011, Dublin, Ireland.

# List of Abbreviations

| | |
|---|---|
| AIC | Akaike's Information Criterion |
| AL | ANILab |
| AT | Austria |
| ATC | Anatomical Therapeutic Chemical |
| AUC | The Area Under the ROC Curve |
| BE | Belgium |
| BG | Bulgaria |
| BIC | Bayesian Information Criterion |
| CAP | Community-acquired Pneumonia |
| CHD | Coronary Heart Disease |
| CI | Credible Interval |
| CPR | Clinical Prediction Rule |
| CY | Cyprus |
| CZ | Czech Republic |
| DDD | Defined Daily Dose |
| DE | Germany |
| DIC | Deviance Information Criterion |
| DID | DDD Per 1000 Inhabitants Per Day |
| DK | Denmark |
| EE | Estonia |
| EGS | Expanded Gold Standard |
| EI | EUROIMMUN |
| ELISA | Enzyme-linked Immunosorbent Assay |
| EIA | Enzyme Immunoassay |
| ES | Spain |

| | |
|---|---|
| FI | Finland |
| FR | France |
| GR | Greece |
| HR | Croatia |
| HU | Hungary |
| IE | Ireland |
| IgA | Immunoglobulin A |
| IgG | Immunoglobulin G |
| IgM | Immunoglobulin M |
| IL | Israel |
| IM | ImmunoWELL |
| IPD | Individual Patient Data |
| IS | Iceland |
| IT | Italy |
| LCM | Latent Class Model |
| LR | Likelihood Ratio |
| LRTI | Lower Respiratory Tract Infections |
| LT | Lithuania |
| LU | Luxembourg |
| LV | Latvia |
| ME | Medac |
| MIF | Microimmunofluorescence Test |
| ML | Maximum Likelihood |
| MT | Malta |
| NAAT | Nucleic Acid Amplification Tests |
| NASBA | Nucleic Acid Sequence Based Amplification |
| NL | Netherlands |
| NO | Norway |

| | |
|---|---|
| PCR | Polymerase Chain Reaction |
| PL | Poland |
| PT | Portugal |
| ROC | Receiver Operating Characteristic |
| RU | Russian Federation |
| RX | X-ray |
| SE | Sweden |
| SI | Slovenia |
| SK | Slovakia |
| UK | United Kingdom |

# List of Tables

# List of Figures

# Chapter 1

# Introduction

This doctoral thesis is motivated by the LMM-projects (Laboratory of Medical Microbiology, University of Antwerp) of the METHUSALEM-consortium and by the INTERCHEST-project. The main objective of this research is to propose and study appropriate statistical modelling strategies for public health research, with applications in the surveillance of antimicrobial consumption, the diagnosis of acute infections and the diagnosis of coronary heart disease. The METHUSALEM-consortium is an interuniversity collaboration between the Vaccine & Infectious Disease Institute (VAXINFECTIO), University of Antwerp and the Centre for Statistics (CENSTAT), Hasselt University. Total outpatient antibiotics use in Europe is introduced in Section 1.1. The aim of this study is to describe total outpatient antibiotics use in Europe from 1997 to 2009, to analyse the trend of total antibiotics use, to analyse the seasonal variation and to assess the change in the trend of total antibiotics use. Section 1.2 introduces a research project on the application of different statistical models to decide on the strategy for the diagnosis of *Chlamydophila pneumoniae* and *Mycoplasma pneumoniae* infections. In Section 1.3 we introduce the INTERCHEST-project, funded by the German Federal Ministry of Education and Research (BMBF-grant number FKZ01GK0920). In this study, the aim is to determine the diagnostic value of an optimal combination of signs and symptoms for myocardial ischemia (chest pain) in primary care patients. Finally, Section 1.4 gives a short overview of the contents of the dissertation.

## 1.1 Outpatient Antibiotics Use in Europe

Antibiotics are drugs that inhibit or abolish the growth of bacteria. They are not active against viruses which cause diseases such as flu, common cold and acute bronchitis. These viral infections usually resolve spontaneously and antibiotics treatment would not be helpful. Antibiotics resistance is a major European and global public health problem and international efforts are needed to counteract the emergence of resistance. The increase in resistance rates of many important pathogens to the currently most available antibiotics has now been recognized as a universal health hazard and potentially life-threatening problem. A large number of studies strongly suggest that this increase is directly related to the actual use of antibiotics. Antibiotics use is increasingly recognized as the main driver for resistance and differential selection pressure of antibiotics agents may be responsible for some of the observed differences.

Yearly and quarterly data on total outpatient use of antibiotics aggregated at the level of the active substance were collected by the European Surveillance of Antimicrobial Consumption (ESAC) project for the period 1997–2009 from 31 and 27 European countries, respectively, and expressed in defined daily doses (DDD) per 1000 inhabitants per day (DID) (Adriaenssens *et al.*, 2011a; Coenen *et al.*, 2011; Minalu *et al.*, 2011). The ESAC project is an international network of surveillance systems, coordinated by VAXINFECTIO, University of Antwerp, with the aim of collecting comparable and reliable data on antimicrobial use in Europe.

Specific actions such as campaigns aimed at general practitioners as well as the public appeared essential, because outpatient antibiotics use account for a large part of the overall antibiotics usage, but evaluating their impact is not straightforward (Goossens *et al.*, 2005; Davey *et al.*, 2008; Huttner *et al.*, 2010). Campaigns, directed to the public aim at

(1) informing about antibiotics resistance and to warn about the medical and general health issues related to the inappropriate use of antibiotics and

(2) fostering the patient-physician and patient-pharmacist dialogue about the appropriate use of antibiotics, and to increase the awareness of the public to a more rational use of antibiotics.

In some European countries (for example in Belgium, France, Germany, Greece, Iceland, Italy, Luxembourg, Portugal, Spain, and United Kingdom), campaigns were planned as part of a national strategy to reduce resistance to antimicrobial drugs. These strategies also included measures to promote appropriate use of antimicrobial

drugs in hospitals, long-term care facilities, and the agricultural sector. All campaigns used a multifaceted approach.

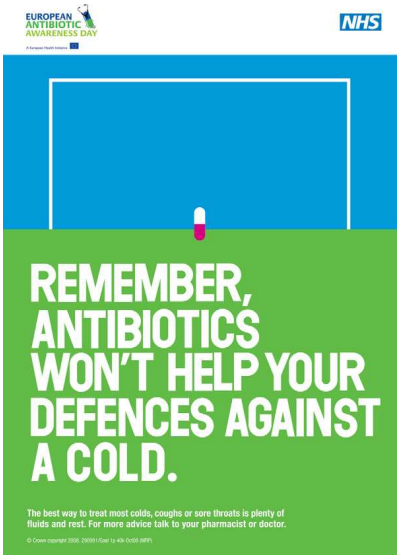



**Figure 1.1:** Poster of the English campaign to promote appropriate use of antimicrobial drugs (Huttner *et al.*, 2010).


**Figure 1.2:** Poster of the "Wise Use of Antibiotics" campaign in New Zealand (Huttner *et al.*, 2010).

The most common intervention, used by all campaigns, was the distribution of patient informational material. The methods for the distribution of printed material varied, but the most common form was direct mailing to physicians and pharmacists for distribution to the patients (pamphlets) and display in waiting rooms or pharmacies (posters) (Huttner *et al.*, 2010). The posters of the English and New Zealand campaigns are shown in Figures 1.1 and 1.2, respectively.

The main objective of the study is to develop an appropriate statistical model to assess the significance of country-specific trends in Europe and to identify possible changes in the trend of antibiotics use, while accounting for country-specific global use as well as seasonal effects.

We proposed a change-point mixed model to assess the use of antibiotics and to assess the change in the trend of outpatient antibiotics use in a Bayesian framework, where the change-points are unknown parameters in the model. The location of the change-points may be related to points in time where public-health strategies aiming at increasing the awareness of the public to a more rational use of antibiotics or targeting to reduce overconsumption of antibiotics were initiated. The application of the model may yield new and important insights in the evolution of outpatient antibiotics use in Europe.

## 1.2 Diagnosis of Acute Infections

Diagnosis of an infectious disease is often based on one or multiple diagnostic tests, none of which is a gold standard. In the assessment of the accuracy of diagnostic tests for infectious diseases, the true disease status of the subject is often unknown due to the lack of a gold standard test. Latent class models (LCMs) with two latent classes, representing diseased and non-diseased subjects, are often used to assess diagnostic tests accuracy when a gold standard assessment of disease is not available.

Diagnostic studies used to diagnosis *Mycoplasma pneumoniae* and *Chlamydophila pneumoniae* respiratory tract infections are described in Sections 1.2.1 and 1.2.2, respectively. For the diagnosis of *C. pneumoniae* and *M. pneumoniae* infections an expanded gold standard (EGS) was defined (Loens *et al.*, 2012a; Loens *et al.*, 2012b).

### 1.2.1 Diagnosis of *Mycoplasma Pneumoniae* Infection

*Mycoplasma pneumoniae* is a very small bacterium that causes the disease called mycoplasma pneumonia, a form of atypical bacterial pneumonia. Serology is often

used to diagnose *Mycoplasma pneumoniae* respiratory infections. Enzyme-linked immunosorbent assays (ELISAs) and enzyme immunoassays (EIAs) were commercially developed; they are relatively simple to perform, are considered to be objective because of photometrically reading of the results, and are easy to standardize as their results are expressed in international units. Hence different serology studies have been applied in different seroepidemiological studies investigating the association between *Mycoplasma pneumoniae* and lower respiratory tract infections (LRTI). The use of different assays would be no problem if the agreement between the tests is high. However, the performances of these tests depend on several factors, including the antigen preparation used (Beersma *et al.*, 2005; Nir-Paz *et al.*, 2006).

According to the European Respiratory Society guidelines (Woodhead *et al.*, 2011), serology for the management of the individual patient with LRTI are not recommended and is considered to be more useful in epidemiologic studies. Application of nucleic acid amplification tests (NAATs) for the detection of atypical pathogens may be considered provided the tests are validated and the results can be obtained sufficiently rapidly to be therapeutically relevant. NAATs applied to respiratory specimens are nowadays widely used for the rapid diagnosis of respiratory tract infections but here again, positive and negative results are not always confirmed by other techniques (Loens *et al.*, 2003; Loens *et al.*, 2010). More and more, a combination of serology and NAAT-detection is recommended for the diagnosis of a *Mycoplasma pneumoniae* infection (Daxboeck *et al.*, 2003; Loens *et al.*, 2010; Woodhead *et al.*, 2011).

Latent class models were used to estimate the diagnostic accuracy, specifically the test sensitivity and specificity, of polymerase chain reaction (PCR), nucleic acid sequence based amplification (NASBA) and 4 different commercially available Immunoglobulin M (IgM) and Immunoglobulin G (IgG) EIA and 2 different commercially available Immunoglobulin A (IgA) assays for the detection of *Mycoplasma pneumoniae* in adult patients with lower respiratory tract infections in order to identify the most appropriate test (Loens *et al.*, 2012a). The results of the latent class models were compared with the pre-defined expanded gold standard.

### 1.2.2 Diagnosis of *Chlamydophila Pneumoniae* Infection

*Chlamydophila pneumoniae* is a bacterial pathogen that infects humans and is a major cause of pneumonia. The diagnosis of an acute *Chlamydophila pneumoniae* infection is usually based on the demonstration of at least a fourfold increase in IgG antibody

levels in paired serum samples, or the presence of IgM antibodies in any serum sample. The microimmunofluorescence test (MIF) is still considered to be the "gold" standard. For the measurement of Chlamydophila species-specific antibodies. The role of IgA antibodies in the diagnosis of acute phase infection has not been definitely established, and these antibodies are not measured in all laboratories. However, the measurement of IgA antibodies has been shown to increase diagnostic findings in some studies.

Studies comparing serology and nucleic acid amplification methods for diagnosis of *Chlamydophila pneumoniae* in community-acquired pneumonia (CAP) are rare. The latent class models were also used to evaluate PCR, NASBA, IgM and IgG MIF and 3 different IgM and IgG EIA assays for the detection of *C. pneumoniae* in patients with CAP as well as 2 different *C. pneumoniae* IgA EIA assays (Loens *et al.*, 2012b). The results of the latent class models were also compared with the expanded gold standard.

## 1.3 Diagnosis of Coronary Heart Disease

Chest pain is a frequent complaint in many health care settings. In primary care 0.7% to 2.7% of patient encounters are due to chest pain (Svavarsdottir *et al.*, 1996; Verdon *et al.*, 2008; Bösner *et al.*, 2009). However, the prevalence of serious cardiac disease in these patients, e.g., chronic stable coronary heart disease (CHD) or acute coronary syndrome, is low. In unselected patients presenting with chest pain in primary care, the overall prevalence of coronary heart disease is between 12.8 and 14.6% (Verdon *et al.*, 2008; Bösner *et al.*, 2009). In the majority of patients, the underlying etiology is musculoskeletal, esophageal, respiratory, psychological, or is unknown.

The primary care physician must reliably identify serious cardiac disease while also protecting patients from unnecessary testing and hospital admissions. The optimal evaluation of possible CHD uses the patient's clinical probability in order to decide on the value of further testing and to interpret test results using probabilistic reasoning (Doust *et al.*, 2009).

However, individual symptoms and signs are not sufficient to reliably diagnose CHD in these patients. Previous meta-analyses using aggregate data do not address this problem (Mant *et al.*, 2004; Chun *et al.*, 2004; Bruyninckx *et al.*, 2008). Meta-analyses investigating the combined diagnostic value of several tests need to be based on individual patient data (Buntinx *et al.*, 2009a).

INTERCHEST collaborators have conducted a systematic review of studies evalu-

ating the diagnostic accuracy of symptoms and signs for diagnosing coronary heart disease in primary care. Medline, Embase and the references of relevant articles were searched to identify eligible studies. Six studies recruiting patients presenting with chest pain in office-based primary care practice were included. In this thesis, individual patient data (IPD) meta-analyses were used to summarize the diagnostic accuracy of signs and symptoms used for the diagnosis of coronary heart disease in primary care patients. The investigators of six studies which aimed to determine the accuracy of symptoms and signs for CHD in primary care constituted an international working group on chest pain in primary care (INTERCHEST). The data from all studies are pooled in order to perform a meta-analysis with individual patient data (Haasenritter *et al.*, 2012a).

This study aims to summarise the available evidence regarding the diagnostic accuracy of symptoms and signs for myocardial ischemia in primary care, and to provide recommendations regarding the design of future diagnostic studies in primary care and the conduct of diagnostic accuracy reviews based on individual patient data. In order to achieve this aim we constructed a clinical prediction rule (CPR) for individual patients, combining all signs and symptoms with other patient characteristics such as age and gender, and with optimal diagnostic accuracy characteristics and applicable to a general population. Such a prediction rule which is "personalized" on the one hand and applicable to individuals in a broad community (covering several countries or regions and thus exceeding the validity of individual studies) on the other hand, needed to be based on an IPD meta-analysis (Minalu *et al.*, 2012b; Haasenritter *et al.*, 2012a).

## 1.4 Outline of the Thesis

The outline of the thesis is as follows:

**Chapter 2** briefly describes the datasets that have been used in this dissertation. We have analyzed the total outpatient antibiotics use datasets in **Chapters 3** and **4**. **Chapter 3** provides the application of the mixed-effects models for the outpatient antibiotics use datasets. The two-stage model and the linear mixed-effects model are applied to the yearly outpatient antibiotics use data. For the quarterly outpatient antibiotics use data, the non-linear mixed model is applied to assess country-specific trends in Europe while accounting for the seasonal effect. In **Chapter 4** the non-linear mixed model was extended by including known and unknown common change-

points and country-specific random change-points to assess the change in the trend of antibiotics use over time.

In **Chapter 5** latent class models are used to evaluate tests used for the diagnosis of *Mycoplasma pneumoniae* and *Chlamydophila pneumoniae* in adult patients with lower respiratory tract infections in order to identify the most appropriate test. In **Chapter 5** we first consider the conditional independence model which assumes all the diagnostic tests are independent conditional on the true disease status. Then, the conditional dependence model which assumes some or all the diagnostic tests are dependent conditional on the true disease status is considered. We compared the results of the latent class models with the expanded gold standard. In **Chapter 6** we have conducted a simulation study to evaluate the impact of misspecifying the conditional dependency of the tests.

In **Chapter 7** Individual patient data meta-analyses are used to explore the combined diagnostic value of all signs and symptoms for diagnosing coronary heart disease in primary care. Based on the data of all studies, we have constructed a new clinical prediction rule for the diagnosis of coronary heart disease in primary care.

Finally, discussions and concluding remarks are given in **Chapter 8**.

# Chapter 2

# Datasets

In this chapter, the datasets that have been used in this thesis are briefly introduced. In Section 2.1, we introduce the yearly and quarterly total outpatient antibiotics use datasets. Diagnostic tests for *Chlamydophila pneumoniae* and *Mycoplasma pneumoniae* respiratory tract infections are introduced in Section 2.2. Section 2.3 introduces the studies evaluating the diagnostic accuracy of signs and symptoms for diagnosing coronary heart disease in primary care.

## 2.1 Outpatient Antibiotics Use Data

Resistance to antibiotics is a major public health problem and antibiotics use is being increasingly recognized as the main selective pressure driving this resistance. Yearly and quarterly data on total outpatient antibiotics use were collected for the period 1997–2009 within ESAC using the Anatomical Therapeutic Chemical (ATC) Classification and the DDD measurement unit. Longitudinal data on antibiotics use from this surveillance system allows a comparison of countries and assessment of trends in Europe. The antibiotics use data were collected at the therapeutic sub-group (ATC-2) level, at the pharmacological sub-group (ATC-3) level, at the chemical sub-group (ATC-4) and at the chemical substances (ATC-5) level. The Anatomical Therapeutic Chemical classification system is shown in Figure 2.1.

Antibiotics use is expressed as the number of defined daily doses (DDD) per 1000 inhabitants per day (World Health Organization, WHO definition). DDD is the assumed average maintenance dose per day for a drug used for its main indication in adults.

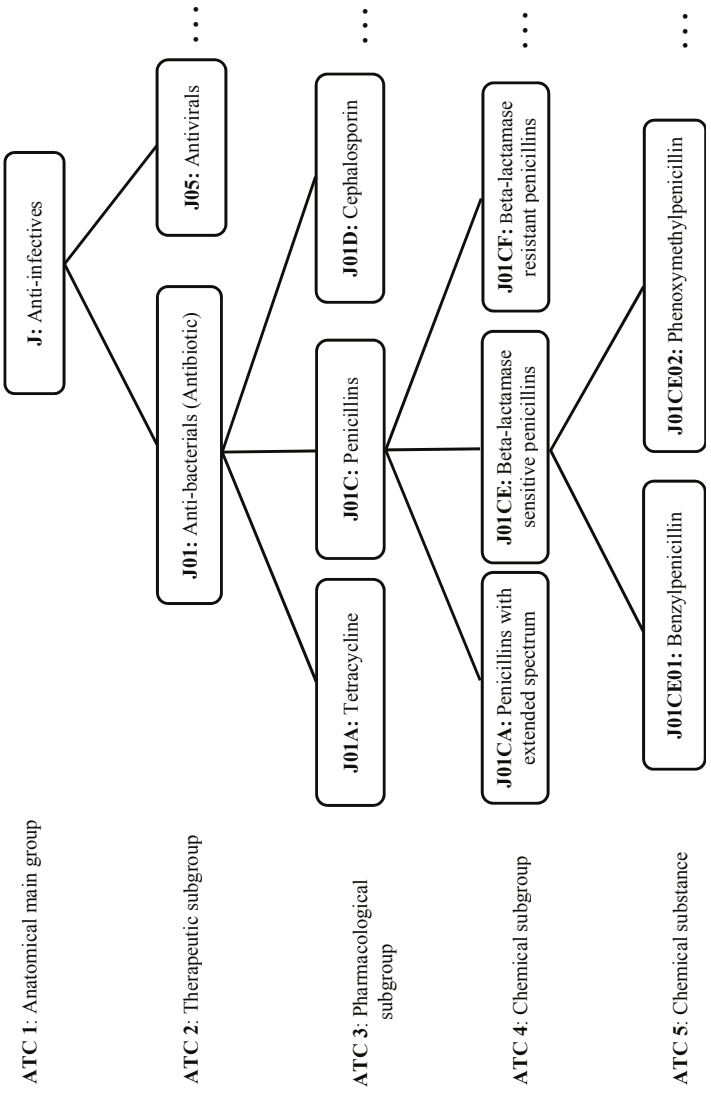



**Figure 2.1:** Anatomical Therapeutic Chemical (ATC) classification system.

Figure 2.2 shows prescription/package of penicillin (at the chemical substance level). The package contains 100 tablets of each 250 milligram. According to World Health Organization (WHO) definition, 2 grams of penicillin is one DDD. Then, the consumption of the prescription/package is 12.5 DDD.


**Figure 2.2:** Prescription/package of penicillin (at the chemical substance (ATC-5) level).

The methods of data collection, validation of the data, and processing for the ESAC project have been described in detail in Adriaenssens *et al.* (2011a), Adriaenssens *et al.* (2011b), Coenen *et al.* (2011) and Versporten *et al.* (2011a). More information on the ESAC project is available on the ESAC web site (www.esac.ua.ac.be). The yearly and quarterly uses of tetracycline in Europe are shown in Sections 2.1.1 and 2.1.2, respectively. Tetracycline is prescribed for use against many bacterial infections. It is commonly used to treat acne.

### 2.1.1 Yearly Antibiotics Use Data

Yearly antibiotics use data on total outpatient antibiotics use from 31 European countries were collected for the period 1997–2009 within ESAC project. The observed country-specific trends for the yearly tetracycline use in DID are shown in Figure 2.3.

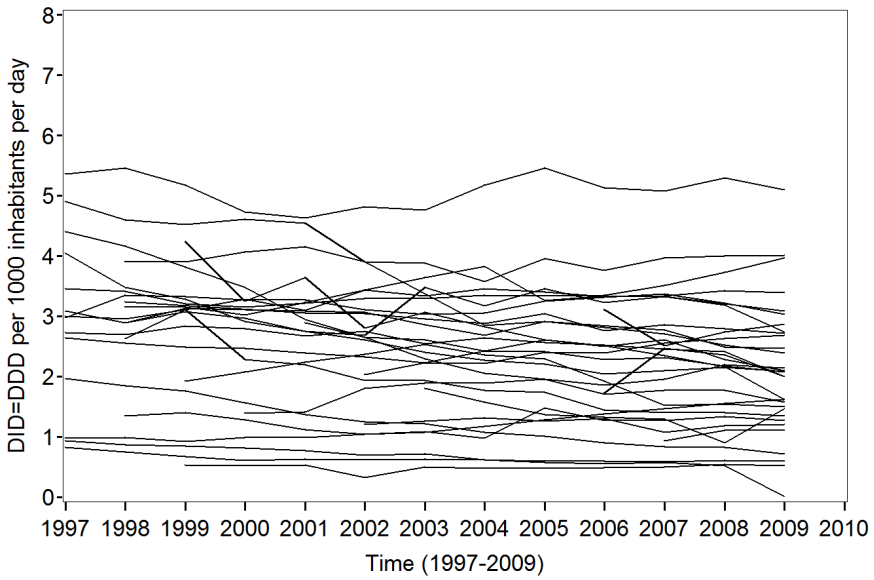



**Figure 2.3:** Observed country-specific evolutions for the yearly use of tetracycline expressed in DID in 31 European countries.

As can be seen in Figure 2.3, there is variability across repeated measurements from the same country (i.e., within-country variability) as well as variability between countries (i.e., between-country variability), which suggests that country-specific intercepts and slopes should be incorporated into the model to account for heterogeneity across countries.

### 2.1.2 Quarterly Antibiotics Use Data

Observed country-specific evolutions for the quarterly use of tetracycline in 27 European countries are shown in Figure 2.4. The longitudinal profiles show clear seasonal variation of total outpatient tetracycline use in all countries, with upward peaks in the winter season. Thus a non-linear model needs to be adopted to take the seasonality into account. Figure 2.4 also shows within-country variability and between-country variability. From the longitudinal profiles it can be clearly seen that countries with higher tetracycline use at baseline (in 1997) have a higher amplitude (higher seasonal variation).

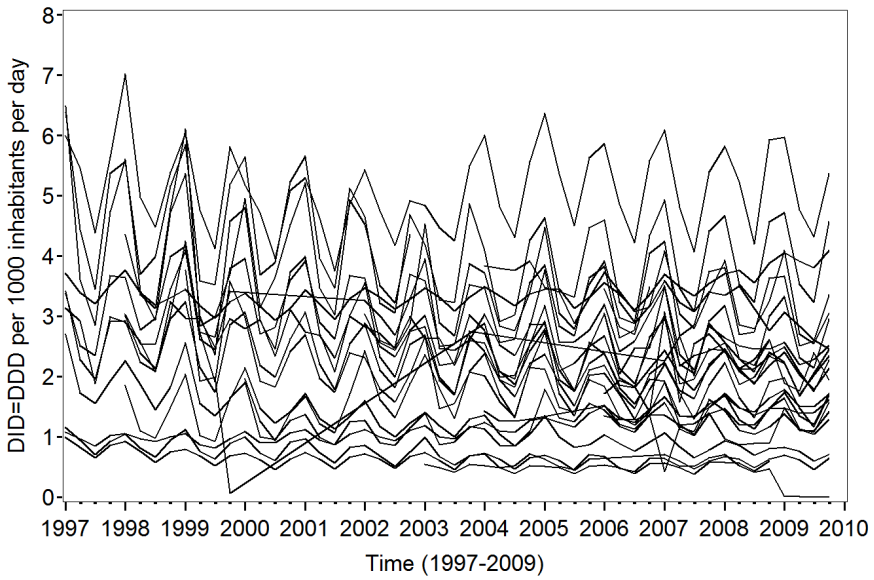



**Figure 2.4:** Observed country-specific evolutions for the quarterly use of tetracycline expressed in DID in 27 European countries.

Figures 2.3 and 2.4 also show that not all longitudinal profiles are complete for all countries. Some profiles start later in time and others show intermediate missing values. As the missingness mechanism is assumed to be missing completely at random (MCAR), all analyses were based on all available cases.

We have conducted the analysis at the therapeutic subgroup (ATC 2), at chemical subgroup (ATC 3) and at the chemical sub-group (ATC-4) (Adriaenssens *et al.*, 2011a; Adriaenssens *et al.*, 2011b; Adriaenssens *et al.*, 2011c; Coenen *et al.*, 2011; Minalu *et al.*, 2011; Versporten *et al.*, 2011a; Versporten *et al.*, 2011b). In this thesis, for illustration of the proposed methods, we mainly focuses on the total outpatient use of tetracycline (at ATC-3 level).

## 2.2 Diagnosis of Acute Infections

In this section, we briefly describe the tests used to diagnose *Chlamydophila pneumoniae* and *Mycoplasma pneumoniae* respiratory tract infections respectively.

### 2.2.1 Diagnosis of *Mycoplasma Pneumoniae*

In this section, we describe the study population, nucleic acid amplification tests and the serology tests used to diagnose *Mycoplasma pneumoniae* respiratory tract

infections.

## Study Population

Two hundred and twelve patients with RX-proven community-acquired pneumonia (CAP) were prospectively included in a hospital-based study across Europe starting from October 2002 until May 2003. The ages of the patients ranged from 19.0 to 84.8 years. Furthermore, 8 *M. pneumoniae* positive CAP-patients enrolled in a Belgian CAP-study between October 2000 and May 2001 as well as 20 *M. pneumoniae* positive patients with a lower respiratory tract infection (LRTI) and enrolled in both above mentioned CAP-studies were included in this study. Throat swabs and paired serum samples were collected and stored locally at -20$^o$C and were regularly shipped on dry ice to the microbiology laboratory of the University Hospital of Antwerp.

## Nucleic Acid Amplification Protocols

Upon arrival in the laboratory, the dry throat swabs were suspended in 1 *ml* of sterile saline, aliquoted in portions of 100$\mu l$ and stored at $-70^o$C until they were batch-wise processed. *M. pneumoniae* DNA was extracted from the throat swabs using the QiaAmp blood mini kit (Qiagen, Hilden, Germany) according to the instructions of the manufacturer. Elution was done in 100$\mu l$ elution buffer. *M. pneumoniae* real-time PCR was done in the Lightcycler as described previously by Loens *et al.* (2002).

For extraction by the NucliSens miniMAG platform, 900$\mu l$ lysis buffer (bioMérieux) was added to a second protease treated aliquot of 100$\mu l$. The samples were mixed vigorously for rapid lysis and stored at $-70^o$C. *M. pneumoniae* RNA was extracted by the NucliSens miniMAG platform with the NucliSens magnetic extraction reagents (bioMérieux, Boxtel, The Netherlands) according to the instructions of the manufacturer. Elution of these nucleic acid extracts was done in 20 $\mu l$.

Nucleic acid extracts from the throat swabs obtained with the NucliSens miniMAG were investigated by real-time NASBA using the NucliSens EasyQ® *M. pneumoniae* assay according to the instructions of the manufacturer. The assay contains an internal control RNA. The amplification process was run in a fluorescent reader, the NucliSens EasyQ Analyzer (bioMérieux). The results obtained with the NucliSens EasyQ assay were calculated with the NucliSens EasyQ software, and were classified as positive, negative or invalid in case the internal control was not detected or the signal was too weak. In negative control reactions, target nucleic acid was replaced by RNase-/DNase-free water. All samples with a positive nucleic acid amplification

result were reanalyzed starting from the extraction (Loens *et al.*, 2012a).

### Serology

Two hundred and eleven paired and 29 acute phase sera from 240 patients with CAP/LRTI were available. The time range between the collection of the acute and convalescent sera was 8 to 63 days. Ten commercially available EIAs were evaluated: *M. pneumoniae*-IgM-ELISA medac (Medac, Wedel, Germany), *M. pneumoniae*-IgG-ELISA medac, *M. pneumoniae*-IgA-ELISA medac; ANILabsystems *M. pneumoniae* IgM, ANILabsystems *M. pneumoniae* IgG (ANILab-systems, Vantaa, Finland distributed by Biomedical Diagnostics, Antwerp, Belgium); and *Anti-Mycoplasma pneumoniae* ELISA (IgM, IgG, and IgA) (EUROIMMUN AG, Lübeck, Germany, distributed by Biognost, Heule, Belgium), ImmunoWELL *M. pneumoniae* IgG and IgM EIA (Genbio, distributed by Biomedical Diagnostics, Bruges, Belgium).

The plates of the *M. pneumoniae* Medac assays are coated with a highly purified *M. pneumoniae*-specific antigen preparation. The ANILabsystems EIA is a microtiter EIA for *M. pneumoniae* specific antibodies. The antigen used is enriched for cytadhesin protein P1. The ImmunoWELL EIAs use a purified glycolipid extract of *M. pneumoniae* strain FH (ATCC 15531). The test is approved for clinical use by the FDA. The *Anti-Mycoplasma pneumoniae* ELISA from EUROIMMUN uses an ether extract of the *M. pneumoniae* strain "FN".

Serum samples were tested in a single run for each particular IgM, IgG or IgA assay on the same day. Acute and convalescent sera from the same patient were analysed within the same run. Duplicate testing was not performed except for the EUROIMMUN assays. The assays and calculations were performed according to the manufacturer's instructions. IgG was removed if indicated by the manufacturer. A significant rise in IgG titer in paired serum samples was defined as either seroconversion or a fourfold increase in the IgG titer, unless stated otherwise by the manufacturer. Distinct kit vials always shared the same lot or batch number (Loens *et al.*, 2012a).

## 2.2.2 Diagnosis of *Chlamydophila Pneumoniae*

In this section, we describe the study population, nucleic acid amplification tests and the serology tests used to diagnose *Chlamydophila pneumoniae* respiratory tract infections.

## Study Population

Two hundred and twelve patients with RX-proven community-acquired pneumonia were prospectively included in a hospital-based study across Europe starting from October 2002 until May 2003. The ages of the patients ranged from 19.0 to 84.8 years. Collected throat swabs and paired serum samples were stored locally at -20$^o$C and were regularly shipped on dry ice to the microbiology laboratory of the University Hospital of Antwerp.

## Molecular Amplification Protocols

Upon arrival in the laboratory, the dry throat swabs were suspended in 1 $ml$ of sterile saline, aliquoted in portions of 100$\mu l$ and stored at -70$^o$C until they were batchwise processed. *C. Pneumoniae* DNA was extracted from the 212 throat swabs using the QiaAmp blood mini kit (Qiagen, Hilden, Germany) according to the instructions of the manufacturer. Elution was done in 100$\mu l$ elution buffer. *C. pneumoniae* real-time PCR was done in theLightcycler as described previously (Hoymans *et al.*, 2003).

For extraction by the NucliSens miniMAG platform, 900$\mu l$ lysis buffer (bioMérieux) was added to the protease treated aliquots of 100$\mu l$ (Loens *et al.*, 2002). The samples were mixed vigorously for rapid lysis and stored at -70$^o$C. From a second aliquot of the 213 suspended throat swabs, *C. pneumoniae* RNA was extracted by the NucliSens miniMAG platform with the NucliSens magnetic extraction reagents (bioMérieux, Boxtel, The Netherlands) according to the instructions of the manufacturer. Elution of these nucleic acid extracts was done in 20$\mu l$. Nucleic acid extracts from the 213 throat swabs obtained with the NucliSens miniMAG were investigated by real-time NASBA using the NucliSens EasyQ® *C. pneumoniae* assay according to the instructions of the manufacturer. The assay contains an internal control RNA. The amplification process was run in a fluorescent reader, the NucliSens EasyQ Analyzer (bioMérieux). The results obtained with the NucliSens EasyQ assay were calculated with the NucliSens EasyQ software, and were classified as positive, negative or invalid in case the internal control was not detected or the signal was too weak. In negative control reactions, target nucleic acid was replaced by RNase-/DNase-free water. All samples with a positive NAAT result were reanalysed starting from the extraction (Loens *et al.*, 2012b).

### Serology

One hundred and ninety five paired and 17 acute phase sera from 212 patients with community acquired pneumonia (CAP) were available. The range of time between the collection of the acute and convalescent sera was 8 to 63 days. Seventy six acute serum samples and 59 convalescent serum samples had a restricted volume and were omitted only for the Medac IgM test. Four commercial immunoassays were evaluated: Focus MIF IgM and IgG and 3 microtitre EIAs: *C. pneumoniae*-IgM-sELISA medac, *C. pneumoniae*-IgM-sELISA medac (Medac, Wedel, Germany); AniLabsystems *C. pneumoniae* IgM, AniLabsystems *C. pneumoniae* IgG and AniLabsystems *C. pneumoniae* IgA EIA (Biomedical Diagnostics, Antwerp, Belgium); and Anti-*C. pneumoniae* ELISA IgM, Anti-*C. pneumoniae* IgG, and Anti-*C. pneumoniae* IgA (Euroimmun AG, Lbeck, Germany, distributed by Biognost, Heule, Belgium).

The Focus MIF uses purified formalin-fixed elementary bodies of *C. pneumoniae*, C. trachomatis and C. psittaci diluted in yolk sac as antigens and is genus specific. The Anti-*C. pneumoniae* ELISA from Euroimmun uses elementary bodies purified from cell lysates and treated with sodium dodecylsulphate and contains all relevant antigens localized in the outer membrane. The assay is genus specific. The Anilab systems EIA is *C. pneumoniae* specific and utilizes stabilized *C. pneumoniae* elementary body as an antigen. The assays measure only antibodies to surface-exposed proteins and not antibodies targeted to the genus-specific lipopolysaccharide of *C. pneumoniae*. The *C. pneumoniae* SELISA Medac employs a highly purified and specific antigen.

Serum samples were tested in a single run for each particular IgM, IgG or IgA assay on the same day. Acute and convalescent sera from the same patient were analysed within the same run. Repeat testing was not performed except for the Euroimmun assays. The assays and calculations were performed according to the manufacturers instructions. IgG was removed if indicated by the manufacturer.

A significant rise in IgG titer in paired serum samples was defined as either seroconversion or a fourfold increase in the IgG titer, unless stated otherwise by the manufacturer. Distinct kit vials always shared the same lot or batch number. For Focus MIF analysis, sera were initially screened at a dilution of 1/16 as recommended by the manufacturer. An IgG titer of $\geq$1/16 was defined as positive. Evaluation of slides was done blindly (Loens *et al.*, 2012b).

## 2.3 Diagnosis of Coronary Heart Disease

In this study, in order to give GPs evidence-based recommendations that fit the conditions of their clinical setting, we conduct IPD meta-analyses of studies investigating the diagnostic value of signs and symptoms for diagnosing coronary heart disease in primary care.

### 2.3.1 Search Strategy

A comprehensive searches in MEDLINE (National Library of Medicine), and EMBASE (Excerpta Medica) were conducted. Terms identifying chest pain were used along with terms to identify studies conducted in primary care. Search strategies included subject headings (MeSH, Embtree) as well as free-text terms. In order to identify unpublished studies, a hand search in the online published abstracts of the annual meetings of the North American Primary Care Research Group and the European General Practice Research Network were performed. Additionally, the reference lists of all relevant articles were checked. Authors of relevant articles were also asked if they were aware of studies which are unpublished, ongoing, or which the INTERCHEST collaborators have not identified.

Two reviewers independently screened title and abstracts. They retrieved and screened all full texts articles of potential relevant studies. Additionally, a third reviewer reassessed the selected articles. The reviewers resolved disagreements by discussing their findings.

All studies which had prospectively obtained data on signs and symptoms in a consecutive series of adult patients presenting with chest pain in primary care are included. Studies were not eligible if the patients were recruited by paramedics, in emergency departments of hospitals, or if the patients were pre-selected by health professionals based on the likelihood of an underlying CHD. Since the likelihood of CHD is low in the majority of patients presenting in this setting, the INTERCHEST collaborators considered a delayed type reference standard to be the best possible reference standard.

### 2.3.2 Data Acquisition and Study Quality Assessment

The INTERCHEST collaborators extracted information on methodological characteristics of the studies, such as inclusion criteria, patient recruitment, data collection, and reference standard, and the methodological quality from publications or requested it

from the original investigators. Since the original questionnaires or case report forms had been written in different languages, the INTERCHEST collaborators translated the variables in the individual studies into English and created a synopsis showing the names, definitions and categories of all variables used in the respective studies. Using this synopsis the INTERCHEST collaborators recoded the individual data sets and merged the individual study data sets into one pooled data set. The INTERCHEST collaborators included any symptom and sign that was collected in at least two studies in the merged data set. The INTERCHEST collaborators then checked the merged data set whether values seem plausible, realistic, consistent, and at least similar to the results published by the researchers. Two reviewers, who had not been involved in the conduct of the included studies, independently assessed the methodological quality of each study using established criteria (Whiting *et al.*, 2003; Whiting *et al.*, 2006).

Six relevant studies of about 4000 patients in five different countries (USA, Belgium, Sweden, Switzerland and Germany) have been identified (Verdon *et al.*, 2008; Sox *et al.*, 1990; Buntinx *et al.*, 1992; Nilsson *et al.*, 2003; Bösner *et al.*, 2010a; Haasenritter *et al.*, 2012b). All six studies investigated prospectively the accuracy of signs and symptoms for CHD in consecutive series of patients with chest paint. The number of patients ranged from 299 to 1238. Each study used a delayed-type reference standard to establish the reference diagnosis. This standard includes the follow up of the clinical course during an appropriate predefined period. After follow up all relevant data are assessed and a final diagnosis is made (Knottnerus *et al.*, 2009). Study characteristics in these studies are summarized in Table 2.1.

The investigators of the six studies constituted an international working group on chest pain in primary care (INTERCHEST) and agreed to pool the data in a meta-analysis with individual patient data. Minor issues regarding the analyses and interpretation are discussed and decided by the whole group.

However, one study (Haasenritter *et al.*, 2012b) was ongoing at the time we started our analysis. Moreover, this study was similar in regard to many aspects to a study the INTERCHEST collaborators had already included (Bösner *et al.*, 2010a). The main difference was that only a small number of the primary predictors were investigated in the second Germany study. So, the INTERCHEST collaborators decided to exclude this study from the main analysis and to set it aside for external validation of the diagnostic model if possible. This resulted in five studies including about 3100 patients and conducted in five different countries. The INTERCHEST collaborators

considered 61 items of the medical history and physical examination which were available in at least 2 studies as potential predictors. The set of predictors investigated in the respective studies varied strongly. Only the predictors 'sex' and 'age' were available in all studies. Table A.1 lists the predictors which have been included in the merged data set based on the studies that the INTERCHEST collaborators have identified up to date. The merged data set is then checked for internal and external validity. Details of the search strategies and the characteristics of these studies are briefly described in Haasenritter *et al.* (2012a).

**Table 2.1:** Characteristics of the six studies the INTERCHEST collaborators identified to date (Sox *et al.*, 1990; Buntinx *et al.*, 1992; Nilsson *et al.*, 2003; Verdon *et al.*, 2008; Bösner *et al.*, 2010; Haasenritter *et al.*, 2012b) in five different countries.

| Characteristic | Sox *et al.*, 1990 | Buntinx *et al.*, 1992 | Nilsson *et al.*, 2003 | Verdon *et al.*, 2008 | Bösner *et al.*, 2010 | Haasenritter *et al.*, 2012b |
|---|---|---|---|---|---|---|
| Data collection | 1982 | 1988 | 1998–2000 | 2001 | 2004–2005 | 2009–2010 |
| Country | USA | Belgium | Sweden | Switzerland | Germany | Germany |
| Setting | 1 Drop-in clinic | 25 GPs | 3 health care centres each served by 4 GPs | 58 GPs in private practice | 74 GPs in private practice | 56 GPs in private practice |
| Number of patients | 395 | 299 | 523 | 644 | 1238 | 856 |
| Inclusion criteria | Chest pain as presenting complaint. No age limitation (ages were 17 to 81 years; average 41 years) | New episode of chest pain, discomfort or tightness as main or ancillary complaint. No age limitation (ages were 1 to 88 years; average 45 years) | New episode of chest pain, discomfort or tightness as presenting complaint; aged 20-79 years; patients were excluded: if acute MI or coronary revascularization during the previous year | Chest pain as main or ancillary complaint; age$\geq$ 16 years | Chest pain as main or ancillary complaint; age$\geq$ 35 years; excluded: chest pain$\geq$ 1 one month, or had already been investigated | Chest pain as main or ancillary complaint; age$\geq$ 35 years; excluded: chest pain $\geq$ 1 one month, or had already been investigated |
| Reference standard | Delayed-type reference standard | Delayed-type reference standard | Delayed-type reference standard | Delayed-type reference standard | Delayed-type reference standard | Delayed-type reference standard |
| Duration of follow-up | At least 1 year | 2 weeks to 2 months | 3 months | 12 months | 6 months | 6 months |
| RD established by | 2 internist-investigators independently assigned diagnosis. | Treating physicians | Treating physicians | Treating physicians | Independent expert panel (1 GP, 1 cardiologist, 1 research fellow) | Independent expert panel (1 GP, 1 research fellow) |
| Prevalence of CHD as cause of chest pain | 7.2% | 9.6% | 11.2% | 12.6% | 14.4% | 10.6% |

RD: reference diagnosis; MI: myocardial infarction

# Chapter 3

# Modeling Yearly and Quarterly Outpatient Antibiotics Use

Mixed-effects models provide a very flexible approach for analyzing longitudinal data. The linear mixed-effects model is often used to analyze continuous longitudinal data and the generalized linear mixed model is the most frequently used random-effects model for discrete repeated measurements.

Given that repeated measures were taken for each country, intra-country correlation has to be taken into account when analyzing the data. A two-stage model and a linear mixed-effects model are used for the yearly tetracycline use data to assess country-specific trends in Europe. A non-linear mixed model is developed for the quarterly tetracycline use data to assess country-specific trends in Europe, while accounting for country-specific seasonal effects.

## 3.1 A Two-stage Model

In this section the two-stage model for the yearly tetracycline use data is considered in order to assess whether there is a decrease or an increase in tetracycline use in Europe. The model was fitted in two stages. First, a linear regression model was fitted separately for each country. Afterwards, regression methods were used to model the variability of country-specific regression coefficients.

### 3.1.1 Stage 1: Country-specific regression models

In the first stage, a linear regression model (3.1) is used to summarize the observations of country $i$ ($i = 1, 2, \ldots, N$) by their regression parameters:

$$Y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + \varepsilon_{ij}, \tag{3.1}$$

where $Y_{ij}$ is the tetracycline use in DID for country $i$ at time points $t_{ij}$ ($j = 1, 2, ..., n_i$), $n_i$ is the number of observations from the $i$th country, time=1 corresponds to the start of the study (year 1997), $\beta_{0i}$ and $\beta_{1i}$ are unknown country-specific regression coefficients, and $\boldsymbol{\varepsilon_i}$ is an $n_i$-dimensional vector of unexplained error terms $\varepsilon_{ij}$. $\boldsymbol{\varepsilon_i}$ is assumed to follow normally distributed with mean vector zero and $n_i \times n_i$ covariance matrix $\boldsymbol{\Sigma_i}$. To account for the serial correlation of the error terms, a first-order autoregressive structure was used for the variance structure for the error terms. The (co)variance of the errors at time points $j$ and $j$' for country $i$ equals

$$(\sigma_{jj'})_i = (\sigma_i)^2 \rho_i^{|j-j'|}, \quad j, j' = 1, 2, \ldots, 13, \tag{3.2}$$

where $\sigma_i^2$ is the error variance for country $i$ and $\rho_i$ is the AR(1) parameter (correlation parameter) for country $i$.

### 3.1.2 Stage 2: Modelling the variability of country-specific regression coefficients

In the second stage, a multivariate model of the form

$$(\beta_{0i}, \beta_{1i}) = (\beta_0, \beta_1) + (b_{0i}, b_{1i}), \tag{3.3}$$

is used to explain the observed variability between the countries, in terms of their country-specific regression coefficients ($\beta_{0i}$,$\beta_{1i}$). $\beta_0$ and $\beta_1$ are unknown regression parameters, and $b_{0i}$ and $b_{1i}$ are country-specific random effects, where $b_{0i}$ expresses how much the intercept of country $i$ deviates from the global intercept $\beta_0$ and $b_{1i}$ expresses how much the slope of country $i$ deviates from the global slope $\beta_1$. The random effects are assumed to follow a normal distribution with mean vector zero and general covariance matrix $\boldsymbol{D}$, with elements $d_{ij} = d_{ji}$:

$$\boldsymbol{D} = \begin{pmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{pmatrix} \quad \text{and} \quad \rho_{12} = corr(b_{0i}, b_{1i}) = \frac{d_{12}}{\sqrt{d_{11}}\sqrt{d_{22}}}, \tag{3.4}$$

where $d_{11}$ is the variance of the random intercept $b_{0i}$, $d_{22}$ is the variance of the random slope $b_{1i}$, $d_{12}$ and $\rho_{12}$ are the covariance and correlation of the random intercept and the random slope, respectively. In the two-stage analysis, information is lost in summarizing the observed measurements for the $i$th country by the country-specific regression coefficients, the number of observations per country is not taken into account when analysing the estimated regression coefficients (in the second stage), and random variability is introduced by replacing the $\beta_{0i}$ and $\beta_{1i}$ by their estimates. These drawbacks can be avoided by combining the two stages into one model, the so-called linear mixed (-effects) model (see e.g. Verbeke and Molenberghs, 2000).

## 3.2 A Linear Mixed-effects Model

Linear mixed models provide a very flexible environment for modelling data with many types of repeated measurements, whether repeated in time, space, or both. Correlations among measurements made on the same subject or experimental unit can be modelled using random effects and through the additional specification of a covariance structure. Observations from different countries are assumed to be independent and observations within countries are expected to be correlated. The model is defined as:

$$Y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_{ij} + \varepsilon_{ij}, \tag{3.5}$$

where $\beta_0$ is the global intercept (average outpatient tetracycline use in Europe in 1997), $\beta_1$ is the global slope describing the marginal linear time trend of the time series (average change in outpatient tetracycline use in DID per year in Europe from 1997 to 2009), $\boldsymbol{b_i} = (b_{1i}, b_{0i})$ is a vector of country-specific random effects (for intercept and slope) and we assume $\boldsymbol{b_i} \sim (0, \boldsymbol{D})$. The matrix $\boldsymbol{D}$ is an unstructured covariance matrix defined as in (3.4). The unstructured covariance matrix allows all of the parameters of the variance-covariance matrix to be different. A likelihood ratio test was used to compare the unstructured covariance matrix with various more parsimonious covariance matrices as e.g. a compound symmetry structure. The likelihood ratio test result supports the use of the unstructured covariance structure.

$\boldsymbol{\varepsilon}_i$ is a vector of unexplained error terms $\varepsilon_{ij}$. It is usually assumed that all are independent and normally distributed with mean vector zero and covariance matrix $\boldsymbol{\Sigma_i}$. Often, $\boldsymbol{\Sigma_i}$ is assumed equal to $\sigma^2 I_{ni}$, where $I_{n_i}$ is the $n_i$-dimensional identity matrix. This structure is often referred to as the simple covariance structure. Many possible covariance structures are available for the covariance matrix for the error components. A likelihood ratio test was used to contrast the simple covariance structure with the first-order autoregressive structure. The likelihood ratio test result supports the use of the first-order autoregressive structure (AR-1). In this case, the (co)variance of the errors at time points $j$ and $j$' equals

$$(\sigma_{jj'})_i = \sigma^2 \rho^{|j-j'|}, \quad j, j' = 1, 2, ..., 13, \tag{3.6}$$

where $\sigma^2$ is the error variance and $\rho$ is the correlation parameter also known as the autocorrelation coefficient which reflects the degree to which the errors are auto correlated. The correlation decreases exponentially across the lags of the time points. We tried to extend the assumption of constant within-country variability across all countries but the model did not converge likely due to parameter redundancy as was the case for several other more general covariance matrices.

The need for the inclusion of the random effects was tested using a likelihood ratio test, by comparing the log-likelihoods of models with and without the appropriate random effect. The asymptotic null distribution of the likelihood ratio test statistic for testing the significance of random-effects in a linear mixed model is a mixture of chi-squared distributions, the mixing proportions of which depend on the number of random effects present in the model, as well as on their variance-covariance structure (Morrell, 1998; Verbeke and Molenberghs, 2000).

## 3.3 A Non-linear Mixed-effects Model

A non-linear mixed model with a sinusoidal component over time, to account for the seasonal variation, is considered to model the quarterly tetracycline use data. This seasonal-trend model is defined as

$$\begin{aligned} Y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_{ij} \\ +\{(\beta_0^S + b_{0i}^S) + (\beta_1^S + b_{1i}^S)t_{ij}\}\sin(\omega t_{ij} + \delta) + \varepsilon_{ij}, \end{aligned} \tag{3.7}$$

where $Y_{ij}$ is the total outpatient tetracycline use in DID for country $i$ ($i = 1, 2, \ldots, N$)

at time points $t_{ij}$ $(j = 1, 2, \ldots, n_i)$, $n_i$ is the number of observations from the $i$th country, time=1 corresponds to the start of the study (first quarter of 1997), $\beta_0$ is the intercept, $\beta_1$ is the regression coefficient describing the marginal linear time trend $(t)$, $\beta_0^S$ is the fixed amplitude, $\beta_1^S$ is the amplitude varying over time, $\omega$ (in radians) is the frequency which is a known constant $(=2\pi/T)$ where $T(=4)$ is the period for the sine curve, $\delta$ (in radians) is the phase shift or phase angle which is an unknown parameter, $\boldsymbol{b_i} = (b_{0i}, b_{1i}, b_{0i}^S, b_{1i}^S)$ is the country-specific vector of random effects where $b_{0i}$ is the country-specific random intercept, $b_{1i}$ is the country-specific random slope for time and $b_{0i}^S$ is the country-specific random slope for amplitude, $b_{1i}^S$ is the country-specific damping effect on the seasonal variation and we assume $\boldsymbol{b_i} \sim N(0, \boldsymbol{D})$. The matrix $\boldsymbol{D}$ is a general covariance matrix with elements $d_{ij} = d_{ji}$. $\boldsymbol{\varepsilon}_i$ is an $n_i$-dimensional vector of unexplained error terms $\varepsilon_{ij}$. It is usually assumed that all $\boldsymbol{\varepsilon}_i$ are independent and normally distributed with mean vector zero and covariance matrix $\boldsymbol{\Sigma}_i$. Often, $\boldsymbol{\Sigma}_i$ is assumed equal to $\sigma_\varepsilon^2 \boldsymbol{I}_{n_i}$, where $\boldsymbol{I}_{n_i}$ is the $n_i$-dimensional identity matrix.

No convergence was obtained when we fitted the models with an unstructured covariance matrix for the random effects. To obtain convergence, we simplified the covariance structure by setting the covariance between $b_{1i}^S$ and $b_{0i}$, $b_{1i}$ and $b_{0i}^S$ equal to 0.

## 3.4 Results

In this section, we presented the results of the models used to analyze the yearly and the quarterly tetracycline use datasets. The results of the two-stage model and the linear mixed model applied for the yearly data are give respectively in Section 3.4.1 and Section 3.4.2, while in Section 3.4.3 the results of the non-linear mixed model applied for the quarterly data are presented.

### 3.4.1 The Two-stage Model

The scatter plot of the estimated country-specific regression coefficients is presented in Figure 3.1, which shows variation across countries (as expected from the observed country-specific evolutions, shown in Figure 2.4), indicating that the majority of countries have a negative slope, indicating decreasing outpatient tetracycline use from 1997 to 2009.

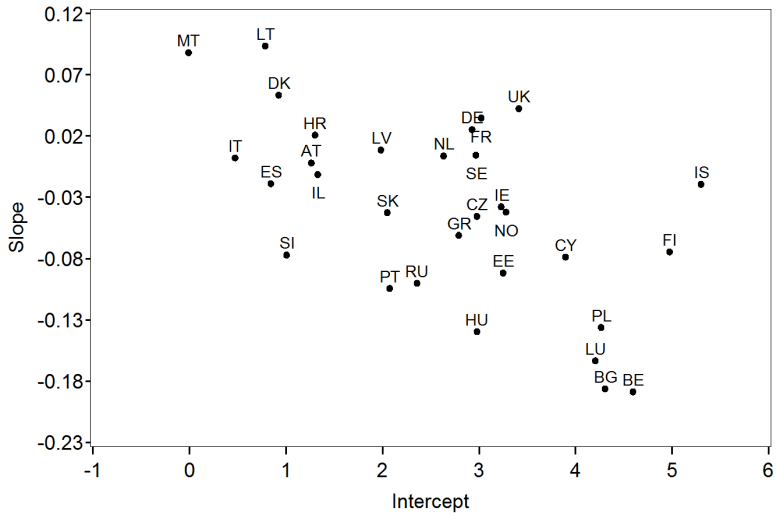



**Figure 3.1:** Scatter plot of country-specific slopes ($\beta_{1i}$) and country-specific intercepts ($\beta_{0i}$) obtained by fitting the two-stage model.

From the scatter plot (Figure 3.1), it can also be clearly seen that there is a negative relationship between the country-specific slopes and country-specific intercepts (so countries with a high level of tetracycline use in 1997 tend to have the largest decrease in use over time).

Table 3.1 shows the parameter estimates and standard errors for the global intercept $\beta_0$ and the global slope $\beta_1$. The parameter $\beta_0$ can be interpreted as the average response at the baseline or the average outpatient tetracycline use in DID in 1997, whereas the parameter $\beta_1$ represents the average linear time effect or the average change in outpatient tetracycline use in DID per year from 1997 to 2009.

**Table 3.1:** Parameter estimates and standard errors for the parameters obtained by fitting the two-stage model.

| Parameter | Estimate (standard error) | $p$-value |
|---|---|---|
| $\beta_0$ | 2.6258 (0.2512) | 0.0001 |
| $\beta_1$ | -0.0401 (0.0134) | 0.0056 |

The results in Table 3.1 indicate that there is an overall significant decrease in the trend of outpatient tetracycline use (slope -0.0401). The estimate for the general covariance matrix $\boldsymbol{D}$ is

$$\boldsymbol{D} = \begin{pmatrix} 1.9570 & -0.0623 \\ -0.0623 & 0.0056 \end{pmatrix}.$$

The estimated variance of the random intercept is 1.957, the estimated variance of the random slope is 0.0056 and the correlation between the random effects is -0.5951 (negative, as expected from Figure 3.1).

### 3.4.2 The Linear Mixed-effects Model

Table 3.2 shows the parameter estimates and standard errors for the fixed-effects parameters using the random-effects model with random intercept and random slope. The results show there is an overall significant decrease in the trend of outpatient tetracycline use in DID.

**Table 3.2:** Parameter estimates and standard errors for the fixed-effects parameters obtained by fitting the linear mixed model.

| Parameter | Estimate (standard error) | $p$-value |
|---|---|---|
| $\beta_0$ | 2.7282 (0.2420) | <0.0001 |
| $\beta_1$ | -0.0481 (0.0137) | 0.0015 |

The parameter estimate for the marginal linear time trend $\beta_1$ in the two-stage model (Table 3.1) is slightly higher than the estimate in the linear mixed model (Table 3.2). The estimates for the variance, covariance and correlation components are

$$\boldsymbol{D} = \begin{pmatrix} 1.2034 & -0.0309 \\ -0.0309 & 0.0022 \end{pmatrix}, \ \rho{=}0.9511 \ \text{ and } \ \sigma^2{=}0.4485.$$

The estimated variance of the random intercept is 1.2034 and the estimated variance of the random slope is 0.0022. The correlation coefficient between the random intercept and the random slope is -0.6005, and indicates that countries with higher tetracycline use in DID at the baseline (in 1997) have a lower slope (decreasing use over time), while countries with lower tetracycline use at baseline have a higher slope (increasing use over time). The within-country variability in DID, which is assumed to be constant across all countries, is estimated to be 0.4485 ($\sigma^2$) and the correlation parameter is 0.9511 ($\rho$). The value of error variance is small, indicating that much of the total variability is captured by the between-country variability. The scatter plot of slopes for time (fixed effect + random effects) and intercepts (fixed effect + random effects) obtained by fitting the linear mixed model is given in Figure 3.2.

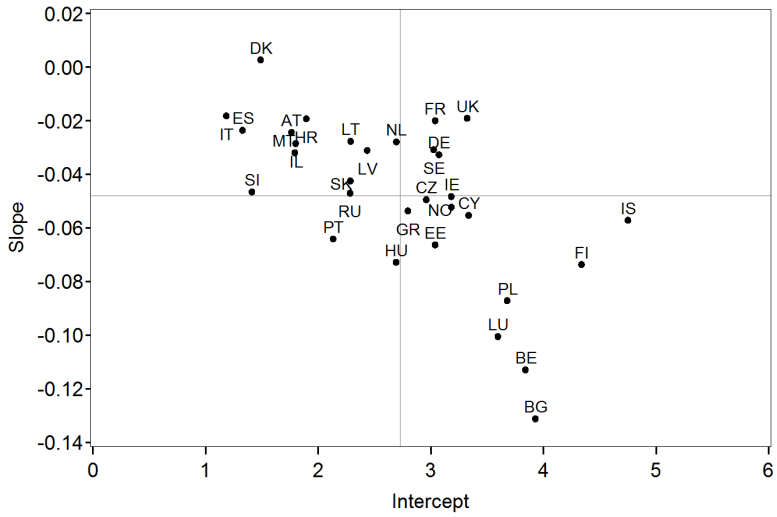



**Figure 3.2:** Scatter plot of slopes for time (fixed effect + random effects) and intercepts (fixed effect + random effects) obtained by fitting the linear mixed model.

Figure 3.2 shows again the strong negative relationship between the intercepts and slopes. Denmark has the highest slope (slope is positive), while Bulgaria has the smallest slope (slope is negative). Figure 3.2 also shows the negative relationship between the random intercept and the random slope from the very similar scatter plot of country-specific slopes and country-specific intercepts (Figure 3.1). The vertical line (=2.7282) is the estimate for the global intercept ($\beta_0$; average outpatient tetracycline use in Europe in 1997) and the horizontal line (=-0.0481) is the estimate for the global linear time effect ($\beta_1$; average change in outpatient tetracycline use in DID per year in Europe from 1997 to 2009).

Residuals and influential diagnostic measures were used to examine model assumptions and to detect outliers and potentially influential observations. From the influential measures, observations of Belgium and Iceland have a small effect on the estimates of the fixed-effect parameters, but are not considered influential.

### 3.4.3 The Non-linear Mixed-effects Model

The parameter estimates and standard errors for the fixed-effects parameters in the non-linear mixed model (3.7) for the quarterly tetracycline use data are given in Table 3.3.

**Table 3.3:** Parameter estimates and standard errors for the fixed-effects parameters obtained by fitting the non-linear mixed model.

| Parameter | Estimate (standard error) | $p$-value |
|---|---|---|
| $\beta_0$ | 2.6041 (0.2510) | <0.0001 |
| $\beta_1$ | -0.0091 (0.0033) | 0.0111 |
| $\beta_0^S$ | 0.6225 (0.0717) | <0.0001 |
| $\beta_1^S$ | -0.0064 (0.0015) | <0.0003 |
| $\delta$ | 0.4947 (0.0235) | <0.0001 |

The results given in Table 3.3 suggest again that there is an overall significant decrease in the use of tetracycline over time. There is also a significant seasonal variation.

The estimates for the variance components are

$$\boldsymbol{D} = \begin{pmatrix} 1.6350 & -0.0093 & 0.3411 & 0 \\ -0.0093 & 0.0003 & -0.0030 & 0 \\ 0.3411 & -0.0030 & 0.1037 & 0 \\ 0 & 0 & 0 & 0.00003 \end{pmatrix} \text{ and } \sigma_\varepsilon^2 = 0.0757.$$

The correlation coefficient between the random effects was estimated to be -0.4199 (random intercept and random slope for time; negative, as expected from Figures 3.1 and 3.2), 0.8284 (random intercept and random slope for amplitude) and -0.5379 (random slope for time and random slope for amplitude), respectively. The high correlation coefficient between the random intercept and random slope for amplitude indicates that countries with higher tetracycline use at the baseline have higher seasonal variation. A similar relationship was also observed between the random intercept and the random slope for amplitude from the observed country-specific profiles (Figure 2.4).

The scatter plot of slopes for time (fixed effect + random effects) and intercepts (fixed effect + random effects) obtained by fitting the non-linear mixed model is given in Figure 3.3. From the scatter plot of slopes and intercepts (Figure 3.3) we can see by how much the country-specific estimates deviate from the overall estimates for the intercept $\beta_0$ and the linear time effect $\beta_1$. After correcting for seasonal variation, the vertical line (=2.6041) is the estimate for the intercept $\beta_0$ and the horizontal line (=-0.0091) is the estimate for the linear time effect $\beta_1$.

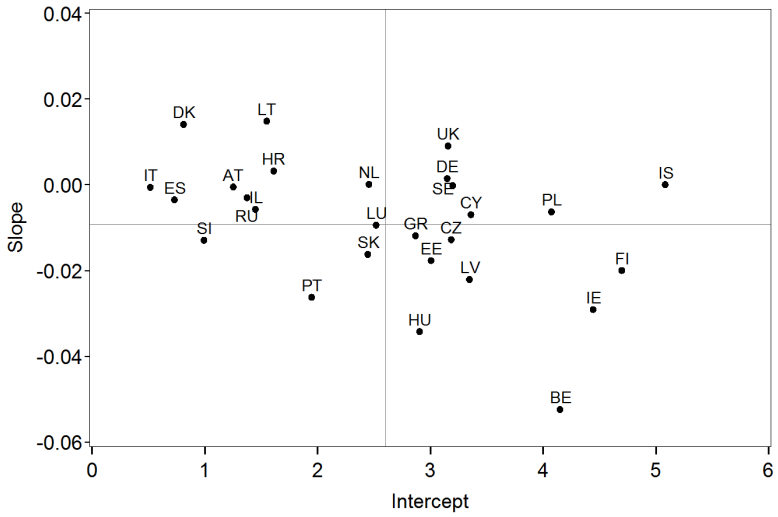



**Figure 3.3:** Scatter plot of slopes for time (fixed effect + random effects) and intercepts (fixed effect + random effects) obtained by fitting the non-linear mixed model.

From the scatter plot, we again observe that there is a strong negative linear relationship between the random intercepts and random slopes for time.

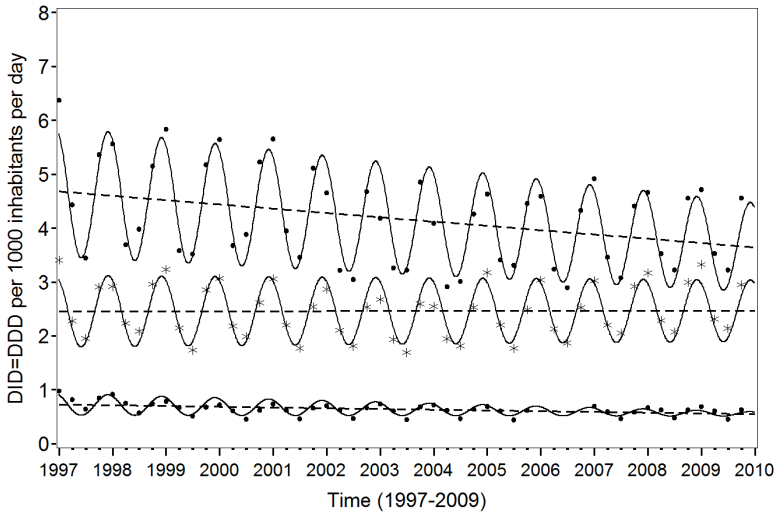



**Figure 3.4:** The predicted country-specific profiles (continuous lines), country-specific predicted linear trends (broken lines) and observed country-specific DID (dots and stars) for three selected countries (Finland, the Netherlands and Spain from top to bottom).

As can be seen from Figure 3.4, which shows the observed country-specific profiles and the predicted country-specific profiles for three selected countries (Finland, the Netherlands and Spain), the predicted country-specific profiles are quite close to the

observed country-specific DID.

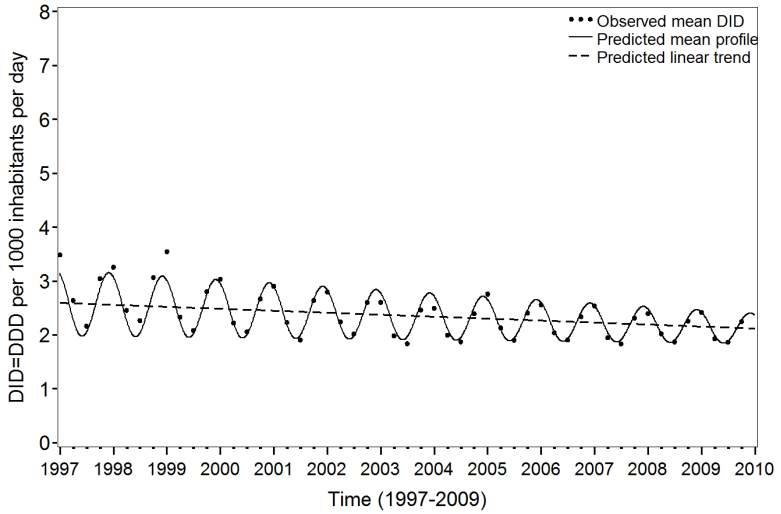



**Figure 3.5:** The predicted mean profile (continuous line), predicted trend (broken line) and observed mean (dots) DID.

The estimated linear trend (broken line), the estimated seasonal-trend model (continuous line) and the observed average DID for Europe are shown in Figure 3.5, again indicating that the model describes the data very well.

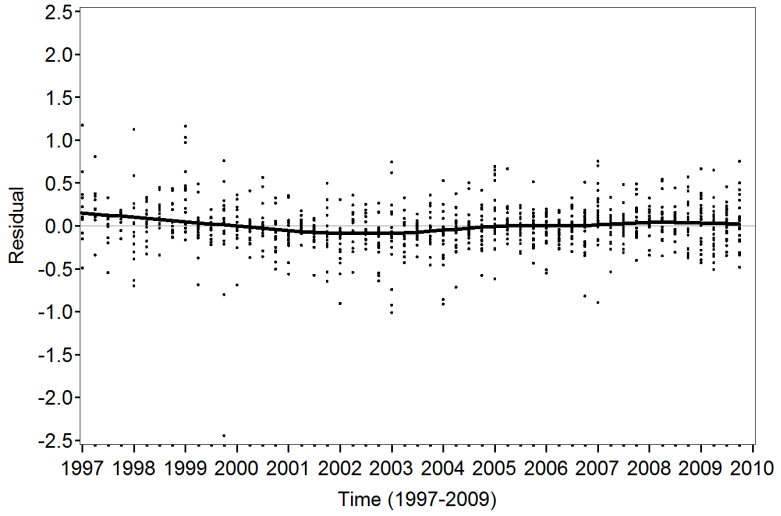



**Figure 3.6:** The scatter plot of residuals (dots) obtained from fitting the non-linear mixed model and smoothed average trend of residuals (solid line).

To get insight into the residual serial correlation, we plotted the residuals versus time. The plot (i.e. Figure 3.6) shows there is essentially no systematic structure in the residual profiles. This supports the assumption that the time dependency and correlation are accounted for by the random effects and the sinusoidal component. Furthermore, the heterogeneity of the structure of the error component across the countries (such as heteroscedasticity) is accounted for by the country-specific seasonal variation (amplitude). In the final model all assignable sources of heterogeneity and variability have been formulated in the mean structure of the model, which allows the model to be used for accurate predictions. From Figure 3.6, one observation seems an outlier. The model was fitted again after removing the outlying observation resulting in no difference in the parameter estimates.

## 3.5 Discussion

This chapter describes a two-stage model and a linear mixed-effects model for the yearly tetracycline use data to assess country-specific trends in Europe. For the quarterly tetracycline use data, a non-linear mixed model was used to assess country-specific trends, while accounting for country-specific seasonal effects. This analysis can be performed at several levels within the Anatomical Therapeutic Chemical (ATC) classification. In this chapter, the analysis was conducted at the pharmacological subgroup (ATC-3) level, but the analysis could be performed at the therapeutic sub-group (ATC-2) level or the chemical sub-group (ATC-4) or substances (ATC-5) level as well.

The non-linear mixed model was also applied to the total outpatient use of antibiotic, penicillin, cephalosporin, macrolide, lincosamide and streptogramin, quinolone, sulphonamide and trimethoprim, and other antibacterials (Adriaenssens *et al.*, 2011a; Adriaenssens *et al.*, 2011b; Adriaenssens *et al.*, 2011c; Coenen *et al.*, 2011; Minalu *et al.*, 2011; Versporten *et al.*, 2011b; Versporten *et al.*, 2011a). An extension of the non-linear mixed model with multiple unknown common change-points and country-specific random change-points have been applied to investigate changes in slope (Chapter 4). Similar techniques could be adopted to assess the impact of public campaigns, like the ones organized in Belgium and France, or the European Antibiotics Awareness Day organized by the European Centre for Disease Prevention and Control (ECDC).

From the results, there is an overall significant decrease in the trend of tetracycline use, and there is a strong positive relationship between the random intercept and

random slope for amplitude, indicating that countries with a higher tetracycline use at the baseline are observed to have a high seasonal variation. We also identified significant difference in total outpatient tetracycline use between countries in Europe.

# Chapter 4

# Adaptive Change-point Modeling

In common regression analysis the response variable is modeled as a linear function of the explanatory variables. Sometimes it may happen that the relationship between the response and some explanatory variables is non-linear, showing a few values where the effect on the response changes abruptly. These values are called break-points, change-points, transition-points or switch-points (Muggeo *et al.*, 2003). To estimate the change-points, Bayesian (Smith *et al.*, 1975; Carlin *et al.*, 1992; Kiuchi *et al.*, 1995; Lange *et al.*, 1992; Ghosh *et al.*, 2007; Dominicus *et al.*, 2008) or likelihood (Hall *et al.* , 2000; Hall *et al.* , 2003; Hens *et al.*, 2010) methods may be used.

Random change-point models have previously been used in several applications to model longitudinal data. These include studies of progression of HIV infection using CD4 T-cell numbers (Lange *et al.*, 1992; Kiuchi *et al.*, 1995; Ghosh *et al.*, 2007) and development of prostate-specific antigen levels as a marker for prostate cancer (Slate *et al.*, 2007). Random change-point models have also been applied to assess the variability in repeated measures of cognitive function (Dominicus *et al.*, 2008). Hall *et al.* (2003) compared the Bayesian approach with the likelihood approach for modeling cognitive function over time, and pointed out that the Bayesian method has an advantage over the likelihood method in that it does not require all subjects to have the same change-point.

In this chapter, we fitted an adaptive Bayesian linear spline model where the number of knots (change-points) and their location are data-driven and determined by the

deviance information criterion (DIC). The presence and the location of one or more change-points is data-driven and can vary across countries.

We start by introducing the non-linear mixed model (3.7) in Section 3.3. We extend the non-linear mixed model by including change-points to identify possible changes in the trend of tetracycline use in DID. All models are fitted in a fully Bayesian paradigm. The models are implemented in R using the R-package R2WinBUGS (Sturtz *et al.*, 2005). The program used to fit the change-point model with one unknown common change-point, one country-specific random change-point and a country-specific latent indicator for the change-point is included in Appendix D.

## 4.1 A Non-linear Mixed Model

In this section, we fitted the non-linear mixed model (3.7) in a Bayesian approach. An extension with known common change-points, unknown common change-points and country-specific random change-points is then considered in Section 4.2. Recall the non-linear mixed model from Section 3.3:

$$Y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_{ij} + (\beta_0^S + b_{0i}^S + \beta_1^S t_{ij})\sin(\omega t_{ij} + \delta) + \varepsilon_{ij}. \qquad (4.1)$$

No convergence was obtained when we fitted the model (4.1) with an unstructured covariance matrix for the random effects. To obtain convergence, we simplified the covariance structure by setting the covariances between the random effects equal to 0.

The non-linear mixed model (4.1) was extended by including a non-linear trend and secondly an amplitude varying non-linearly over time (expressed as $t_{ij}^{\alpha}$). The non-linear mixed model with a non-linear trend and an amplitude varying non-linearly over time is formulated as:

$$Y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_{ij}^{\alpha} + (\beta_0^S + b_{0i}^S + \beta_1^S t_{ij}^{\alpha})\sin(\omega t_{ij} + \delta) + \varepsilon_{ij}, \qquad (4.2)$$

where $\alpha$ is the non-linear trend and the non-linear amplitude varying over time parameter.

## 4.2 An Adaptive Change-point Model

Since there is no a prior knowledge on the number of change-points, we gradually build up the model by first considering a change-point model with a known common change-point and next extending it by including unknown common and country-specific random change-points.

A general mixed model with country-specific mean can be written as

$$\begin{aligned} Y_{ij} &= \mu_i(t_{ij}) + \varepsilon_{ij}, \quad i = 1, 2, ..., N; j = 1, 2, ..., n_i, \\ \mu_i(t_{ij}) &= \mu_i^T(t_{ij}) + \mu_i^S(t_{ij}), \end{aligned} \tag{4.3}$$

where $Y_{ij}$ is the tetracycline use in DID for country $i$ at time points $t_{ij}$, $\mu_i^T(t_{ij})$ is the trend component, $\mu_i^S(t_{ij})$ is the seasonal component and $\varepsilon_{ij}$ is the measurement error which is assumed to be normally distributed with mean zero and constant variance $\sigma_\varepsilon^2$. The country-specific mean components $\mu_i^T(t_{ij})$ and $\mu_i^S(t_{ij})$ are modelled as

$$\begin{aligned} \mu_i^T(t_{ij}) &= (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_{ij} + \mu_i^{CP}(t_{ij}), \\ \mu_i^S(t_{ij}) &= (\beta_0^S + b_{0i}^S + \beta_1^S t_{ij})\sin(\omega t_{ij} + \delta), \end{aligned} \tag{4.4}$$

where $\mu_i^{CP}(t_{ij})$ is a change-point component given by

$$\mu_i^{CP}(t_{ij}) = \sum_{k=1}^{K} (\beta_{(k+1)} + b_{(k+1)i})(t_{ij} - K_{ki})_+, \tag{4.5}$$

where $x_+ = \max(x, 0)$, $K$ is the number of unknown change-points, $K_{ki} = C_k$ or $K_{ki} = C_k + c_{ki}$ or $K_{ki} = c_{ki}$ where $C_k$ denotes a global change-point and $c_{ki}$ a country-specific random change-point. If $\mu_i^{CP}(t_{ij})$=0 then there are no change-points and the model reduces to model (4.1).

Substituting equation (4.5) and (4.4) in equation (4.3), yields the model

$$\begin{aligned} Y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_{ij} + \sum_{k=1}^{K} (\beta_{(k+1)} + b_{(k+1)i})(t_{ij} - K_{ki})_+ \\ + (\beta_0^S + b_{0i}^S + \beta_1^S t_{ij})\sin(\omega t_{ij} + \delta) + \varepsilon_{ij}, \end{aligned} \tag{4.6}$$

where the fixed effects $\beta_0$, $\beta_1$, $\beta_0^S$, $\beta_1^S$ and $\delta$, and the random effects $b_{0i}$, $b_{1i}$ and $b_{0i}^S$ are defined as before, $K$ is the number of change-points, for $k = 1, 2, \ldots, K$,

$\beta_{(k+1)}$ is the global difference in the linear trend before and after the change-point $C_k$, $b_{(k+1)i}$ is the country-specific difference in the linear trend before and after the change-point and $\varepsilon_{ij}$ is an unexplained error term. Random effects for the global level of use, the trend effects, the amplitude of the seasonal effect and the location of the change-point are used to account for heterogeneity across countries. The number of change-points $K$ and the location of the change-point(s) are data-driven.

In equation (4.6) all countries are assumed to have a change in the trend of tetracycline use in DID, but this might not be true because some countries might not have a change in the trend of tetracycline use. To relax this assumption, we extend (4.6) by including country-specific latent indicators for the change-points,

$$\begin{aligned} Y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_{ij} + \sum_{k=1}^{K} \{(\beta_{(k+1)} + b_{(k+1)i})(t_{ij} - K_{ki})_+\} I_{ki} \\ + (\beta_0^S + b_{0i}^S + \beta_1^S t_{ij})\sin(\omega t_{ij} + \delta) + \varepsilon_{ij}, \end{aligned} \tag{4.7}$$

where $I_{ki}$ is an unknown country-specific indicator for the change in the trend of tetracycline use in DID for country $i$ for the $k$th change-point, $k = 1, 2, \ldots, K$ where $K$ is the number of change-points. Here,

$$I_{ki} = \begin{cases} 1 & \text{if there is a change at knot } K_{ki} \text{ in country } i \\ 0 & \text{if there is no change in country } i. \end{cases} \tag{4.8}$$

As there is no prior knowledge on the number of change-points in the study, the number of change-points $K$ in equations 4.6 and 4.7 has to be chosen prior to the data fitting, $k = 1, \ldots, K$. We first start from the simplest model where there is only a known common change-point, i.e. $K = 1$. We gradually extend the model by including a known and an unknown common change-point. And later, we extended the model by including an additional unknown common change-point. Next to the common change-points, country-specific random change-points have also been included in the model

## 4.3 Prior Specification

The following uninformative prior distributions were used for the fixed effects:

$$\begin{aligned} \beta_0, \beta_1, \beta_{(k+1)}, \beta_0^S, \beta_1^S, \delta &\sim \text{Normal}(0,\ 1000), \text{ independently where } k = 1, \ldots, K, \\ C_1 &\sim \text{Uniform}(1,\ 52), \\ C_2 &\sim \text{Uniform}(C_1,\ 52). \end{aligned} \tag{4.9}$$

The normal priors on $\beta_0, \beta_1, \beta_{(k+1)}, \beta_0^S, \beta_1^S$ and $\delta$ have large variances, expressing our lack of knowledge about the regression coefficients. For the random effects, a normal prior distributions was used:

$$\begin{aligned} b_{0i} &\sim \text{Normal}(0,\ \sigma^2_{b_0}), \\ b_{1i} &\sim \text{Normal}(0,\ \sigma^2_{b_1}), \\ b_{(k+1)i} &\sim \text{Normal}(0,\ \sigma^2_{b_{(k+1)}}), \\ b_{0i}^S &\sim \text{Normal}(0,\ \sigma^2_{b_0^S}), \\ c_{ki} &\sim \text{Normal}(C_k,\ \sigma^2_{c_k})I(1, 52). \end{aligned} \tag{4.10}$$

A uniform prior distribution over the total range of time was also assumed for the country-specific random change-point:

$$c_{ki} \sim \text{Uniform}(1,\ 52). \tag{4.11}$$

The country-specific indicator for the $k^{th}$ change-point ($I_{ki}$) is Bernoulli-distributed with probability $P_k$, where the probability $P_k$ is beta-distributed with shape parameters $\alpha_p$ (=1) and $\beta_p$ (=1):

$$\begin{aligned} I_{ki} &\sim \text{dbern}(P_k), \\ P_k &\sim \text{dbeta}(1,1). \end{aligned} \tag{4.12}$$

The hyperparameters in the prior distributions were chosen so that the priors are uninformative. An independent inverse gamma distribution with a shape parameter $\alpha$ (=0.001) and a scale parameter $\beta$ (=0.001) was used for the variance parameters.

$$\sigma^2_{b_0}, \sigma^2_{b_1}, \sigma^2_{b_{(k+1)}}, \sigma^2_{b_0^S}, \sigma^2_{c_k}, \sigma^2_{\varepsilon} \sim \text{IGamma}(0.001,\ 0.001), \text{ independently,} \tag{4.13}$$

where $x \sim \text{IGamma}(\alpha, \beta)$ means that $1/x$ has the Gamma distribution with mean $\alpha/\beta$ and variance $\alpha/\beta^2$.

## 4.4 Model Selection

We use the deviance information criterion (DIC) for model comparison (Spiegelhalter *et al.*, 2002). The deviance information criterion can be represented as:

$$DIC = p_D + \bar{D}. \tag{4.14}$$

DIC is a Bayesian equivalent to Akaike's information criterion (AIC) and consists of two components, a term that measures goodness-of-fit ($\bar{D}$, is defined as the the posterior expectation of the deviance) and a penalty term for model complexity ($p_D$, is defined as the difference between the posterior mean of deviance and the deviances evaluated at the posterior mean $\bar{\theta}$ of the parameters). $p_D = \bar{D} - D(\bar{\theta})$. The smaller the DIC, the better the fit (Dominicus *et al.*, 2008; Gelman *et al.*, 2004; Ghosh *et al.*, 2007; Spiegelhalter *et al.*, 2002).

The quarterly tetracycline use data was analyzed in (Minalu *et al.*, 2011) using the non-linear mixed model. The results of the non-linear mixed models were used as a starting value for the MCMC algorithm. And for the additional change-point parameters, the locations of campaigns or policy changes in antibiotics use in most European countries were used as starting values. To ensure adequate convergence all results were obtained using two chains of 110,000 iterations, of which we discarded the first 10,000 (burn-in) and the chain was then thinned to every 5th sample as there was autocorrelation for some parameters. Trace plots and the potential scale reduction $\hat{R}$ were used to check convergency of the MCMC algorithm (Gelman *et al.*, 2004).

## 4.5 Result

We considered the following models, within the family (4.6):

*Model 1*: Non-linear mixed model without a change-point,
$\mu_i^{CP}(t_{ij})$=0,

*Model 2*: Non-linear mixed model with a known common change-point ($C_1 = 17$),
$\mu_i^{CP}(t_{ij}) = (\beta_2 + b_{2i})(t_{ij} - 17)_+$,

*Model 3*: Non-linear mixed model with a known common change-point ($C_1 = 29$),

$$\mu_i^{CP}(t_{ij}) = (\beta_2 + b_{2i})(t_{ij} - 29)_+,$$

*Model 4*: Non-linear mixed model with one unknown common change-point ($C_1$),

$$\mu_i^{CP}(t_{ij}) = (\beta_2 + b_{2i})(t_{ij} - C_1)_+,$$

*Model 5*: Non-linear mixed model with two unknown common change-points ($C_1$ and $C_2$),

$$\mu_i^{CP}(t_{ij}) = (\beta_2 + b_{2i})(t_{ij} - C_1)_+ + (\beta_3 + b_{3i})(t_{ij} - C_2)_+,$$

where ordering restriction was imposed for the common change-points (i.e. $C_1 < C_2$).

*Model 6*: Non-linear mixed model with one unknown common change-point ($C_1$) and one country-specific random change-point ($c_i$),

$$\mu_i^{CP}(t_{ij}) = (\beta_2 + b_{2i})(t_{ij} - c_i)_+,$$

where the country-specific random change-point is centered around $C_1$ and is restricted to be $\epsilon$ [1,52], $c_i \sim N(C_1, \sigma_c^2)(1, 52)$.

*Model 7*: Non-linear mixed model with one country-specific random change-point ($c_i$),

$$\mu_i^{CP}(t_{ij}) = (\beta_2 + b_{2i})(t_{ij} - c_i)_+,$$

The Model 1 without a change-point is first extended with known common change-points (Models 2 and 3). Because there were public campaigns in some of the European countries during the year 2000–2001 (for example in Belgium, Germany and Greece) and during the year 2004–2005 (for example in Portugal and United Kingdom) (Huttner *et al.*, 2010), we used time=17 (first quarter of 2001) and time=29 (first quarter of 2004) as known common change-points in the trend of tetracycline use in DID, respectively in Model 2 and Model 3. Next, we estimate the change-points by including unknown common and/or country-specific random change-points (Models 4–7). The non-linear mixed model (Model 1) was extended by including a non-linear trend and secondly an amplitude varying non-linearly over time (expressed as $t_{ij}^{\alpha}$). As these extended models did not outperform the change-point models, we only presented the results of the original non-linear and the change-point models (Models 1–7). Various models with three change-points were applied too, but convergency could not be reached for any of these models.

**Table 4.1:** Parameter estimates: posterior means and standard errors, and Model Comparison: $\bar{D}$, $p_D$ and DIC values.

| Parameters | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 | Model 7 |
|---|---|---|---|---|---|---|---|
| $\beta_0$ | 2.6399(0.2630) | 2.7620(0.3121) | 2.5814(0.2941) | 2.6367(0.2505) | 2.7240(0.2770) | 2.5937(0.2986) | 2.6330(0.2815) |
| $\beta_1$ | -0.0087(0.0041) | -0.0253(0.0074) | -0.0118(0.0067) | -0.0115(0.0065) | -0.0212(0.0078) | -0.0133(0.0064) | -0.0146(0.0071) |
| $\beta_2$ | - | 0.0202(0.0076) | 0.0093(0.0089) | 0.0098(0.0088) | 0.0213(0.0119) | 0.0112(0.0102) | 0.0118(0.0109) |
| $\beta_3$ | - | - | - | - | -0.0048(0.0142) | - | - |
| $C_1$ | - | 17* | 29* | 29.3975(1.2912) | 20.2353(3.1875) | 29.4144(2.8802) | - |
| $C_2$ | - | - | - | - | 31.9461(1.4710) | - | - |
| $\beta_0^S$ | 0.6176(0.0629) | 0.6098(0.0622) | 0.6112(0.0614) | 0.6109(0.0618) | 0.6083(0.0616) | 0.6100(0.0612) | 0.6113(0.0618) |
| $\beta_1^S$ | -0.0064(0.0010) | -0.0062(0.0009) | -0.0062(0.0008) | -0.0062(0.0008) | -0.0062(0.0008) | -0.0062(0.0008) | -0.0062(0.0008) |
| $\delta$ | 0.4972(0.0245) | 0.5041(0.0228) | 0.4989(0.0217) | 0.4988(0.0216) | 0.5016(0.0211) | 0.5002(0.0213) | 0.5004(0.0212) |
| $\sigma^2_{b_0}$ | 1.8174(0.5703) | 2.0457(0.6508) | 1.9518(0.6247) | 1.9091(0.5979) | 1.8666(0.5975) | 1.9835(0.6359) | 2.0599(0.6536) |
| $\sigma^2_{b_1}$ | 0.0004(0.0001) | 0.0009(0.0003) | 0.0009(0.0003) | 0.0009(0.0003) | 0.0008(0.0003) | 0.0009(0.0003) | 0.0009(0.0004) |
| $\sigma^2_{b_2}$ | - | 0.0010(0.0004) | 0.0016(0.0006) | 0.0016(0.0006) | 0.0009(0.0004) | 0.0020(0.0008) | 0.0024(0.0010) |
| $\sigma^2_{b_3}$ | - | - | - | - | 0.0018(0.0007) | - | - |
| $\sigma^2_c$ | - | - | - | - | - | 51.3065(32.6960) | - |
| $\sigma^2_{b_0^S}$ | 0.0777(0.0249) | 0.0782(0.0248) | 0.0788(0.0246) | 0.0790(0.0252) | 0.0792(0.0250) | 0.0791(0.0250) | 0.0791(0.0249) |
| $\sigma^2_e$ | 0.0806(0.0038) | 0.0691(0.0033) | 0.0623(0.0030) | 0.0625(0.0030) | 0.0597(0.0029) | 0.0605(0.0029) | 0.0603(0.0030) |
| $\bar{D}$ | 313.8719 | 162.3209 | 60.6070 | 62.9339 | 17.5635 | 30.5115 | 28.6827 |
| $p_D$ | 77.7781 | 91.7648 | 96.3882 | 97.7551 | 102.4649 | 95.3957 | 56.2395 |
| $DIC$ | 391.6500 | 254.0857 | 156.9952 | 160.6891 | 120.0285 | 125.9073 | 84.9222 |

* Because there were public campaigns in some of the European countries during the year 2000–2001 (for example in Belgium, Germany and Greece) and during the year 2004–2005 (for example in Portugal and UK), time=17 and time=29 are used as known common change-points.

For the unknown common change-points in Models 4–6, uniform prior distributions over the total range of time were used. A normal prior distribution with mean $C_1$ (the common change-point) and variance $\sigma^2_{c_k}$ was used for the country-specific random change-point in Model 6, while in Model 7 a uniform prior distribution over the total range of time was assumed for the country-specific random change-point. A summary of the posterior distributions of the model parameters in Models 1–7 is given in Table 4.1.

The results in Table 4.1 clearly indicate the need for one or more change-points. Indeed, Model 1 (no change-points) gets little support with the highest DIC=391.6500. Including a known common change-point reduces the DIC considerably (Models 2 and 3). There is no improvement when the known change-point 29 is replaced by an unknown common change-point (Model 4). There is however a further improvement when two unknown common change-points are included in the model (Model 5). In Models 2–5 all countries are assumed to have the same common change-point, while in Models 6–7 all countries have different change-points. Comparing Model 6 with Model 4 shows a reduction in DIC when including a country-specific random change-point next to the global change-point. A large improvement is achieved when a uniform prior distribution over the total range of time was used for the country-specific random change-point (Model 7). The credible intervals for all parameters in Model 7 are given in Table 4.2. Scatter plots of country-specific estimates, for the change-points in Models 6 and 7 are shown in Figures 4.1 and 4.2, respectively.

The estimate for the unknown common change-point ($C_1$) obtained from fitting Model 4 is 29.3975 (fourth quarter of 2003) which is quite close to the estimates for the common change-point obtained from fitting Model 6 ($C_1$=29.4144). The average for the estimated country-specific random change-points in Model 7 is 28.7451, which is very close to the estimate for the unknown common change-points in Models 4 and 6. From Model 5, the estimate for the first common change-point ($C_1$) is 20.2353 (fourth quarter of 2001) and 31.9461 (fourth quarter of 2004) for the second common change-point ($C_2$).

**Table 4.2:** Parameter estimates: posterior means and standard errors, 95% Quantile-based and HPD credible intervals (CI) obtained from fitting Model 7.

| Parameters | Estimates(Std.Errors) | 95% Quantile-Based CI | 95% HPD CI |
|---|---|---|---|
| $\beta_0$ | 2.6330(0.2815) | (2.0620, 3.1800) | (2.0390, 3.1550) |
| $\beta_1$ | -0.0146(0.0071) | (-0.0283,-0.0008) | (-0.0284,-0.0009) |
| $\beta_2$ | 0.0118(0.0109) | (-0.0113, 0.0323) | (-0.0091, 0.0339) |
| $\beta_0^S$ | 0.6113(0.0618) | (0.4884, 0.7329) | (0.4862, 0.7302) |
| $\beta_1^S$ | -0.0062(0.0008) | (-0.0078,-0.0046) | (-0.0078,-0.0045) |
| $\delta$ | 0.5004(0.0212) | (0.4592, 0.5420) | (0.4594, 0.5420) |
| $\sigma^2_{b_0}$ | 2.0599(0.6536) | (1.1400, 3.6330) | (1.0190, 3.3620) |
| $\sigma^2_{b_1}$ | 0.0009(0.0004) | (0.0005, 0.0018) | (0.0004, 0.0016) |
| $\sigma^2_{b_2}$ | 0.0024(0.0010) | (0.0011, 0.0048) | (0.0009, 0.0043) |
| $\sigma^2_{\alpha}$ | 0.0791(0.0249) | (0.0436, 0.1398) | (0.0385, 0.1289) |
| $\sigma^2_{e}$ | 0.0603(0.0030) | (0.0548, 0.0664) | (0.0548, 0.0664) |

Table 4.2 shows the parameter estimates (and standard errors), 95% Quantile-based and Highest Posterior Density (HPD) credible intervals for all parameters in Model 7. The two 95% credible intervals for $\beta_1$ indicate that there is a significant decrease in the global trend of tetracycline use in DID. Both intervals for $\beta_0^S$ and $\beta_1^S$ do not include zero, indicating respectively that there is a significant overall seasonal variation and a significant overall seasonal variation trend over time.

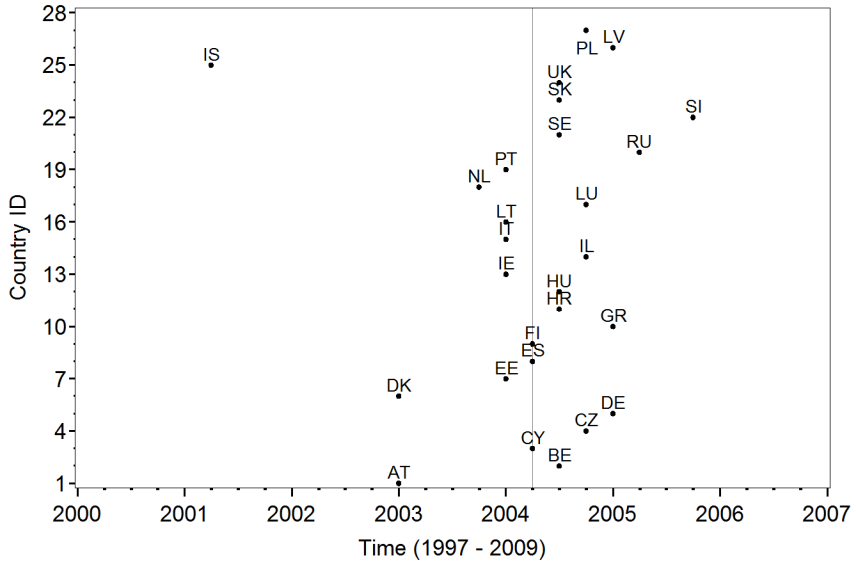



**Figure 4.1:** Scatter plot of estimates for the country-specific change-points obtained from fitting Model 6. The vertical line indicates the estimated global change-point.

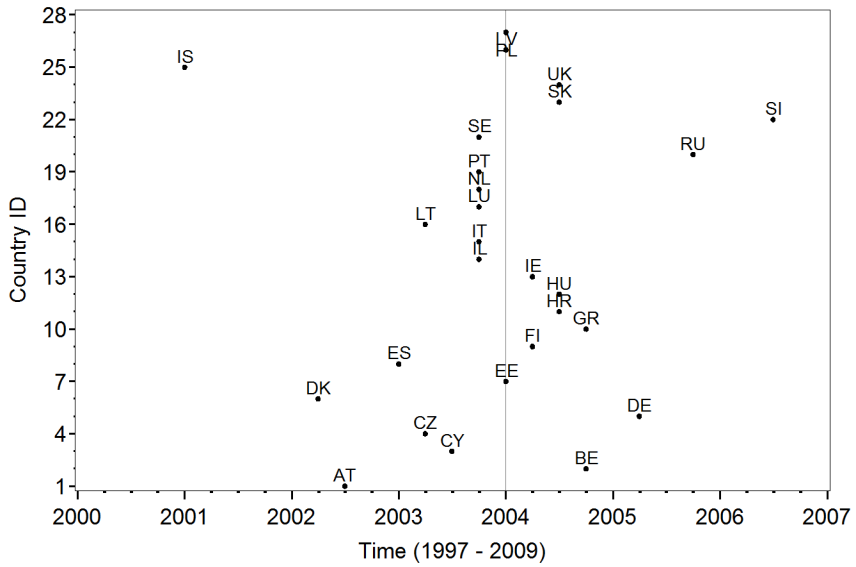



**Figure 4.2:** Scatter plot of estimates for the country-specific change-points obtained from fitting Model 7. The vertical line indicates the average for the estimated country-specific random change-points.

Models 2–7 assume that there are one or more trend changes of tetracycline use in all countries, but for some countries it might be better to have only one or even no change-point. To allow a data-adaptive selection of the number and location of the country-specific change-points, we extend Models 4–7 by including a latent country-specific indicator $I_{ki}$ for the $k$th change-point, $k = 1, 2, \ldots, K$ for country $i$ $(i = 1, 2, \ldots, N)$.

*Model 4* $^*$: Non-linear mixed model with one unknown common change-point $(C_1)$ and a country-specific indicator $I_{1i}$,

$$\mu_i^{CP}(t_{ij}) = \{(\beta_2 + b_{2i})(t_{ij} - C_1)_+\}I_{1i},$$

where $I_{1i}$ is an unknown country-specific indicator for the change in the trend of DID for country $i$. Here, $I_{1i} = 1$ if a change at $C_1$ in the use of tetracycline over time in country $i$ is needed, or $I_{1i} = 0$ if no change in the use of tetracycline over time in country $i$ is needed,

*Model 5* $^*$: Non-linear mixed model with two unknown common change-points $(C_1$ and $C_2)$ and two country-specific indicators $(I_{1i}$ and $I_{2i})$,

$$\mu_i^{CP}(t_{ij}) = \{(\beta_2 + b_{2i})(t_{ij} - C_1)_+\}I_{1i} + \{(\beta_3 + b_{3i})(t_{ij} - C_2)_+\}I_{2i},$$

where ordering restriction was imposed for the common change-points (i.e. $C_1 < C_2$).

*Model 6* *: Non-linear mixed model with one unknown common change-point ($C_1$), one country-specific random change-point ($c_i$) and a country-specific indicator $I_{1i}$,

$$\mu_i^{CP}(t_{ij}) = \{(\beta_2 + b_{2i})(t_{ij} - c_i)_+\}I_{1i},$$

where the country-specific random change-point is centered around $C_1$ and is restricted to be $\epsilon$ [1,52], $c_i \sim N(C_1, \sigma_c^2)(1, 52)$.

*Model 7* *: Non-linear mixed model with a country-specific random change-point ($c_i$) and a country-specific indicator $I_{1i}$,

$$\mu_i^{CP}(t_{ij}) = \{(\beta_2 + b_{2i})(t_{ij} - c_i)_+\}I_{1i}.$$

The parameter estimates for all parameters in Models 4*, 6* and 7* are given in Table 4.3. No convergence was obtained for Model 5*.

**Table 4.3:** Parameter estimates: posterior means and standard errors, and Model Comparison: $\bar{D}$, $p_D$ and DIC values obtained from fitting Models 4*, 6* and 7*.

| Parameters | Model 4* | Model 6* | Model 7* |
|---|---|---|---|
| $\beta_0$ | 2.6488(0.2734) | 2.6618(0.2890) | 2.6322(0.2527) |
| $\beta_1$ | -0.0120(0.0064) | -0.0140(0.0066) | -0.0139(0.0065) |
| $\beta_2$ | 0.0108(0.0105) | 0.0130(0.0111) | 0.0126(0.0126) |
| $C_1$ | 29.4560(1.2975) | 29.1115(2.8179) | - |
| $\beta_0^S$ | 0.6113(0.0630) | 0.6104(0.0613) | 0.6120(0.0615) |
| $\beta_1^S$ | -0.0062(0.0008) | -0.0062(0.0008) | -0.0062(0.0008) |
| $\delta$ | 0.4985(0.0218) | 0.4999(0.0214) | 0.5002(0.0213) |
| $P_1$ | 0.8407(0.1222) | 0.8651(0.1091) | 0.8861(0.1008) |
| $\sigma_{b0}^2$ | 1.9527(0.6229) | 2.0181(0.6382) | 2.0580(0.6472) |
| $\sigma_{b1}^2$ | 0.0009(0.0003) | 0.0009(0.0003) | 0.0009(0.0003) |
| $\sigma_{b2}^2$ | 0.0018(0.0008) | 0.0022(0.0009) | 0.0026(0.0011) |
| $\sigma_c^2$ | - | 46.3954(30.8729) | - |
| $\sigma_{b0^S}^2$ | 0.0791(0.0252) | 0.0788(0.0248) | 0.0790(0.0251) |
| $\sigma_e^2$ | 0.0626(0.0030) | 0.0606(0.0030) | 0.0605(0.0030) |
| $D$ | 65.1738 | 32.7773 | 31.0169 |
| $p_D$ | 87.6091 | 92.3128 | 54.7426 |
| $DIC$ | 152.7831 | 125.0902 | 85.7595 |

* Models 4, 6 and 7 are fitted with a country-specific latent indicator $I_{ki}$.

From the results given in Table 4.3, Model 7* has the lowest DIC value which is quite close to the DIC value of Model 7 (in Table 4.1). The parameter estimates given in Table 4.3 are also close to the corresponding parameter estimates given in Table 4.1. The parameter estimates for the country-specific latent indicators $I_{ki}$ are given in Table 4.4.

**Table 4.4:** Parameter estimates: posterior means and standard errors for the country-specific indicators ($I_{ki}$) obtained from fitting Models 4*, 6* and 7*.

| Country | Parameters | Model 4* | Model 6* | Model 7* |
|---|---|---|---|---|
| Austria | $I_1$ | 0.9516(0.2147) | 0.9693(0.1724) | 0.9629(0.1890) |
| Belgium | $I_2$ | 1.0000(0.0000) | 1.0000(0.0000) | 1.0000(0.0000) |
| Cyprus | $I_3$ | 0.8434(0.3635) | 0.8706(0.3356) | 0.8794(0.3257) |
| Czech Republic | $I_4$ | 0.8026(0.3980) | 0.8746(0.3311) | 0.8963(0.3048) |
| Germany | $I_5$ | 0.9696(0.1718) | 0.9785(0.1450) | 0.9740(0.1592) |
| Denmark | $I_6$ | 0.9054(0.2926) | 0.9194(0.2722) | 0.9241(0.2648) |
| Estonia | $I_7$ | 0.7104(0.4536) | 0.7515(0.4321) | 0.8006(0.3996) |
| Spain | $I_8$ | 0.6616(0.4732) | 0.7266(0.4457) | 0.8132(0.3898) |
| Finland | $I_9$ | 1.0000(0.0000) | 1.0000(0.0000) | 1.0000(0.0000) |
| Greece | $I_{10}$ | 0.8278(0.3776) | 0.8634(0.3434) | 0.8869(0.3167) |
| Croatia | $I_{11}$ | 0.9932(0.0819) | 0.9930(0.0834) | 0.9885(0.1066) |
| Hungary | $I_{12}$ | 0.9576(0.2016) | 0.9656(0.1823) | 0.9557(0.2057) |
| Ireland | $I_{13}$ | 0.8520(0.3551) | 0.8604(0.3466) | 0.8823(0.3222) |
| Israel | $I_{14}$ | 0.7719(0.4196) | 0.8041(0.3969) | 0.8389(0.3677) |
| Italy | $I_{15}$ | 0.8299(0.3757) | 0.8457(0.3613) | 0.8665(0.3401) |
| Lithuania | $I_{16}$ | 0.9705(0.1692) | 0.9773(0.1489) | 0.9756(0.1543) |
| Luxembourg | $I_{17}$ | 0.7520(0.4318) | 0.7994(0.4004) | 0.8286(0.3769) |
| Netherlands | $I_{18}$ | 0.9996(0.0194) | 0.9996(0.0212) | 0.9990(0.0308) |
| Portugal | $I_{19}$ | 0.8623(0.3446) | 0.8908(0.3119) | 0.8903(0.3125) |
| Russian Federation | $I_{20}$ | 0.8067(0.3949) | 0.8569(0.3502) | 0.9025(0.2967) |
| Sweden | $I_{21}$ | 0.5890(0.4920) | 0.6698(0.4703) | 0.8136(0.3894) |
| Slovenia | $I_{22}$ | 0.7065(0.4554) | 0.8022(0.3983) | 0.9013(0.2982) |
| Slovakia | $I_{23}$ | 1.0000(0.0071) | 1.0000(0.0000) | 0.9997(0.0166) |
| United Kingdom | $I_{24}$ | 0.9998(0.0158) | 0.9997(0.0180) | 0.9996(0.0200) |
| Iceland | $I_{25}$ | 0.9768(0.1505) | 0.9999(0.0087) | 1.0000(0.0000) |
| Latvia | $I_{26}$ | 0.7986(0.4011) | 0.8164(0.3872) | 0.8408(0.3659) |
| Poland | $I_{27}$ | 0.8389(0.3677) | 0.8594(0.3476) | 0.8678(0.3387) |

* Models 4, 6 and 7 are fitted with a country-specific latent indicator $I_{ki}$.

The posterior means for the change-point indicator $I_{ki}$ is greater than 0.5 for all countries, which indicates a change in the trend of tetracycline use for all countries.

The estimated linear trend (dashed line), the estimated change-point model (solid line) from Model 7 and the observed average DID for Europe, are shown in Figure 4.3.

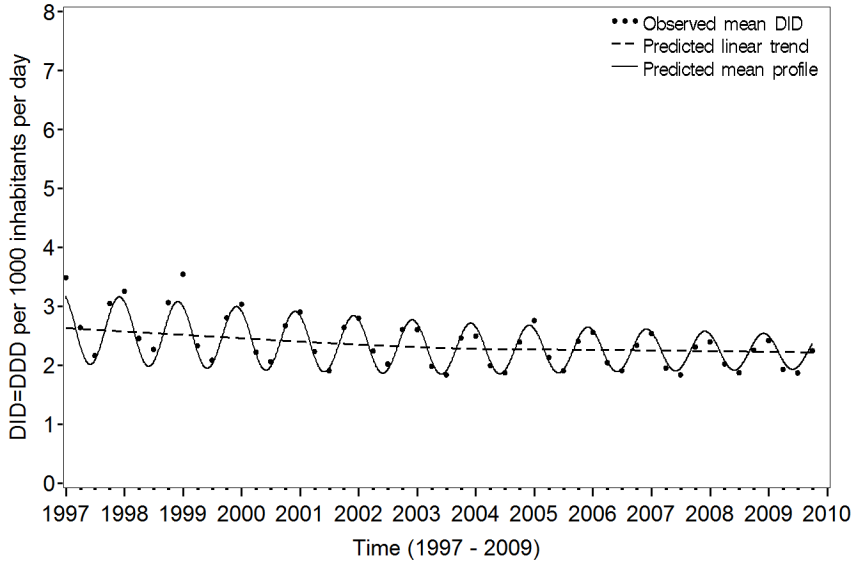



**Figure 4.3:** The observed mean DID (dots), the predicted mean profile (solid line) and the predicted linear trend (dashed line) obtained from fitting Model 7.

The predicted mean is based on the predicted outcomes from the posterior distribution of the country-specific random effects. Figure 4.3 indicates that the model describes the data very well.

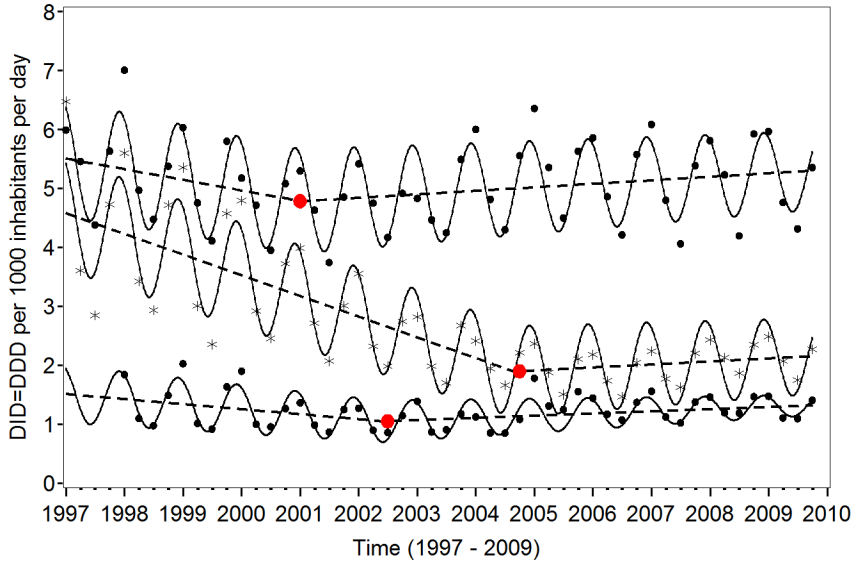



**Figure 4.4:** The observed country-specific DID (dots and stars), the predicted country-specific profiles (solid lines) and the country-specific predicted linear trends (dashed lines) obtained from fitting Model 7 for three selected countries (Iceland, Belgium and Austria from top to bottom).

The observed country-specific profiles and the predicted country-specific profiles from Model 7 for three selected countries (Iceland, Belgium and Austria) are shown in Figure 4.4. As can be seen from Figure 4.4, the predicted country-specific profiles follow closely the observed country-specific DID values. The red dots indicate the estimated country-specific random change-points obtained from fitting Model 7.

A visual inspection of convergence diagnostics graphs for various model parameters showed that the posterior densities are smooth and unimodal shapes. The trace plots indicate that chains appear to have reached a stationary distribution. The chain also has good mixing and is dense.

## 4.6 Discussion

This study was motivated by the need to assess the use of tetracycline in 27 European countries, to assess the change in the trend of tetracycline use over time, and to possibly relate any changes in antibiotics use due to campaigns and policy changes. The data have previously been analyzed based on a non-linear mixed model while taking into account the seasonal effects (Minalu *et al.*, 2011). From the analysis, we have identified significant variation in total outpatient tetracycline use in Europe.

In this chapter, we presented and discussed adaptive change-point Bayesian models to analyse the outpatient tetracycline use from 1997 to 2009. We considered the non-linear mixed model extended with known common change-points, unknown common change-points and country-specific random change-points. The change-point mixed model was also extended by including country-specific indicators for the change-points. A widely used statistic for comparing models in a Bayesian framework, the Deviance Information Criterion (DIC), was used for model comparison. The model with country-specific change-points (Model 7) has the lowest value of DIC. There is some controversy on which criterion to use to compare Bayesian models. Gelman *et al.* (2004) suggested $p_V$=Var(Deviance)/2 as an estimate of the effective number of parameters in the model as an alternative to $p_D$. Note that using $p_V$ as an alternative measure of complexity, the change-point model with two unknown common change-points (Model 5) has the lowest DIC value.

The random change-point models have been applied in many applications. In this Chapter, we extend the existing approaches by a general model building procedure where the number of knots and their location are data-driven. We also extend the previously proposed change-point models by taking into account a country-specific

seasonal variation. The change-point models were also extended by including country-specific latent indicators for the change-points.

From the results obtained from fitting the change-point model with a country-specific change-point (Model 7), there is a significant decrease in the trend of tetracycline use in DID. There is a significant seasonal variation in the use of tetracycline and also a significant seasonal variation trend over time. The change-point estimates we found from the change-point model correspond to the campaigns initiated in some countries, for instance the one organized in Australia (in 2002), Portugal (in 2004–2007) and United Kingdom (in 2004–2005).

In principle, the adaptive change-point models could be extended with more change-points. But for the tetracycline use data, convergency was not reached when including more than two common change-points or more than one country-specific random change-point. Further research includes performing a simulation study to investigate the performance of the proposed change-point models under different scenarios.

# Chapter 5

# Diagnosis of Acute Infections

Disease diagnosis is often based on information obtained from one or multiple diagnostic tests, none of which is a gold standard. In such a situation, two or more of the diagnostic tests may be conditionally dependent due to a factor other than the disease status, arising from a common biological phenomenon. In order to simplify the modeling and statistical analysis of diagnostic studies, it is often assumed that the results from different diagnostic tests are independent of each other conditional on the true disease status. This assumption may be violated in practice, especially in situations where none of the tests is a perfectly accurate gold standard. However, several authors have demonstrated that it is important to account for this dependence while analyzing the results from diagnostic tests in order to obtain unbiased estimates of prevalence of disease and accuracy of the tests (Dendukuri *et al.*, 2001).

The fixed effects and the random effects latent class models are often used to analyze multiple diagnostic tests results when a perfect reference test is lacking (McCutcheon, 1987; Clogg, 1995; Hagenaars and McCutcheon, 2002; Collins and Lanza, 2010). In these models, the true disease status of a person is an unobserved variable with two mutually exclusive categories, 'diseased' and 'non-diseased'. The dependence structure can be parametrized using the two modeling approaches. Pairwise dependencies between two test can either be modeled by a covariance term in the fixed effects formulation or by a common Gaussian random effect in the random effects formulation. The random effects models capture the conditional dependency between multiple tests by random effects. The sensitivities and specificities of tests are modeled as functions of latent, subject-specific random variables (Dendukuri *et al.*, 2001; Qu *et al.*, 1996; Hadgu *et al.* , 1998). In the fixed effects model formulation, models with four or

more diagnostic tests results in a complicated notation, especially when including higher-order correlations (Menten *et al.*, 2008).

To analyze test results from multiple diagnostic tests, Bayesian or likelihood methods may be used. The frequentist approach requires a minimum of four tests to estimate all parameters. The main advantage of the random effect model over the fixed effect approach is that it provides a solution even in the situation when we have a non-identifiable model, i.e. when the number of tests is less than 4 or 5.

In this chapter, the random effects latent class models are illustrated using diagnostic tests for the diagnosis of *Mycoplasma pneumoniae* and *Chlamydophila pneumoniae* (Loens *et al.*, 2012a; Loens *et al.*, 2012b).

## 5.1 Expanded Gold Standard

Different criteria for the expanded gold standard were used:

(i) An expanded gold standard is positive if the test result is positive by PCR and NASBA **or** positive by one amplification test and at least one serology test (either IgM or a significant rise of IgG antibodies as defined according to the instructions of the manufacturer) **or** a significant rise of IgG antibodies in at least two different EIA's.

(ii) An expanded gold standard is positive if the test result is positive by PCR and NASBA.

(iii) An expanded gold standard is positive if the test result is positive by 1 amplification test and at least one serology test (either IgM or a significant rise of IgG antibodies as defined according to the instructions of the manufacturer).

(iv) An expanded gold standard is positive if the test result is positive by a significant rise of IgG antibodies in at least two different EIA's.

These expanded gold standards are not real gold standards but rather optimal subjective standards based on clinical experience and evidence.

## 5.2 Measures of Diagnostic Accuracy

The aim of diagnostic medicine research is to estimate and compare the accuracy of diagnostic tests to provide reliable information about a patient's disease status

and thereby influencing patient care. When developing screening tools, researchers evaluate the discriminating power of the screening test by using simple measures such as the sensitivity, specificity, the positive and negative predictive values and likelihood ratio of the test.

**Table 5.1:** Contingency table created by comparing the results of the diagnostic test and reference standard.

| | | Reference Standard | |
|---|---|---|---|
| | | 1 | 0 |
| Screen Test Outcome | 1 | True Positive (TP) | False Positive (FP) |
| | 0 | False Negative (FN) | True Negative (TN) |

Sensitivity is the ability of a test to detect the disease when it is truly present, whereas specificity is the probability of a test to exclude the disease status in patients who do not have the disease.

- Sensitivity $= \frac{\text{TP}}{\text{TP + FN}}$
- Specificity $= \frac{\text{TN}}{\text{TN + FP}}$

Clinically, it is always important to know how good the test is at predicting the true disease status given the findings from the proposed test. This is captured by the predictive values. Positive predictive value (PPV) is the probability that a patient has the disease given that the test results are positive, and the negative predictive value (NPV) is the probability that a patient does not have the disease given that the test results are indeed negative.

- Positive predictive value $= \frac{\text{TP}}{\text{TP + FP}}$
- Negative predictive value $= \frac{\text{TN}}{\text{FN + TN}}$

Likelihood ratio is a very useful measure of diagnostic accuracy. It is defined as the ratio of expected test result in subjects with a certain state/disease to the subjects without the disease. Likelihood ratio for positive test results (PLR) tells us how much more likely the positive test result is to occur in subjects with the disease

compared to those without the disease. Likelihood ratio for negative test result (NLR) represents the ratio of the probability that a negative result will occur in subjects with the disease to the probability that the same result will occur in subjects without the disease.

- Positive likelihood ratio = $\frac{\text{Sensitivity}}{\text{1 - Specificity}}$
- Negative likelihood ratio = $\frac{\text{1 - Sensitivity}}{\text{Specificity}}$

## 5.3 The Conditional Independence Model

The conditional independence model assumes that all tests are independent, conditional on the disease status (diseased, $d_i = 1$ or non-diseased, $d_i = 0$). Results for an individual test are Bernoulli distributed with $P(Y_{ij} = 1|D_i = d_i)$, the probability of testing positive on the $j$th test given an individual's true disease status $d_i$. The probability that the $i$th subject has a positive test ($y_{ij} = 1$) on the $j$th test is given by

$$P(Y_{ij} = 1|D_i = d_i) = \eta^{-1}(\alpha_{jd_i}), \tag{5.1}$$

where $y_{ij}$ is the observed binary outcome (0=negative, 1=positive) for the $j$th test $\mathrm{T}_j$ on the $i$th subject with disease status $d_i$, $i = 1, \ldots, N$, $j = 1, \ldots, J$, $\eta^{-1}$ is the logit link function $\eta^{-1}(y) = 1/(1 + e^{-y})$, and $\alpha_{jd_i}$ is an intercept term. The probability of an outcome pattern for individual patients, conditional on their disease status, is given by

$$P(Y_{i1} = y_{i1}, \ldots, Y_{iJ} = y_{iJ}|D_i = d_i) = \Pi_{j=1}^{J} P(Y_{ij} = y_j|D_i = d_i), \tag{5.2}$$

The probability that the $i$th individual will have test results $y_i = (y_{i1}, \ldots, y_{iJ})^T$ is given by

$$P(Y_{i1} = y_{i1}, \ldots, Y_{iJ} = y_{iJ}) = \sum_{d_i=0}^{1} P(D_i = d_i)\Pi_{j=1}^{J} P(Y_{ij} = y_j|D_i = d_i). \tag{5.3}$$

Expression (5.3) can be expressed in terms of sensitivity and specificity as

$$\begin{aligned} &P(Y_{i1} = y_{i1}, \ldots, Y_{iJ} = y_{iJ}) \\ &\quad = \pi\Pi_{j=1}^{J} S_j^{y_j}(1 - S_j)^{(1-y_j)} + (1 - \pi)\Pi_{j=1}^{J} C_j^{(1-y_j)}(1 - C_j)^{y_j}, \end{aligned} \tag{5.4}$$

where $\pi$ $\big(= P(D_i = 1)\big)$ is the prevalence, $S_j$ $\big(= P(Y_{ij} = 1|D_i = 1)\big)$ is the sensitivity and $C_j$ $\big(= P(Y_{ij} = 0|D_i = 0)\big)$ is the specificity of the $j$th test.

Classification of a patient to any of the two latent classes, given his/her outcome patterns, is then based on the individual's set of two probabilities: $P(Y_{i1} = y_{i1}, \ldots, Y_{iJ} = y_{iJ})$, $z = 0, 1$. The probability that a patient is positive given his/her outcome patterns is given by

$$\begin{aligned} &P(D_i = 1|Y_{i1} = y_{i1}, \ldots, Y_{iJ} = y_{iJ}) \\ &\quad = \frac{P(D_i = 1)P(Y_{i1} = y_{i1}, \ldots, Y_{iJ} = y_{iJ}|D_i = 1)}{\sum_{d_i=0}^{1} P(D_i = d_i)P(Y_{i1} = y_{i1}, \ldots, Y_{iJ} = y_{iJ}|D_i = d_i)} \end{aligned} \tag{5.5}$$

## 5.4 The Conditional Dependence Model

The conditional dependence model assumes that some tests are dependent conditional on the disease status. In the conditional dependence model, the probability of testing positive for test $\mathrm{T}_i$ depends not only on the unobserved disease status but also on continuous latent random variables through a regression model. In this model, outcomes for a single test for an individual subject $i$ are Bernoulli distributed with

$$P(Y_{ij} = 1|D_i = d_i, \boldsymbol{Z_i} = \boldsymbol{z_i}) = \eta^{-1}(\alpha_{jd_i} + \boldsymbol{\beta}_{jd_i}^T \boldsymbol{z_i}), \tag{5.6}$$

where $\boldsymbol{z_i}$ is vector of realized values of $K$ random effects. The vector $\boldsymbol{z_i} = (z_{1i}, \ldots, z_{Ki})^T$ of random effects consist of $K$ random variables $Z_{ki} \sim \mathrm{N}(0, 1)$. The coefficient vector $\boldsymbol{\beta}_{jd_i} = (\beta_{j1d_i}, ..., \beta_{jKd_i})^T$ describes the dependency of test $\mathrm{T}_j$ on the $K$ random effects. The random effects can be thought of as unobserved characteristics of the subjects that influence the probability of testing positive for one or more of the diagnostic tests. The dependence structure of the model is defined by the random effects and the coefficient vectors $\boldsymbol{\beta}_{jd_i}$. Tests that share a common random value within each patient will show dependence, conditional on the patient's disease status (Dendukuri *et al.*, 2001; Menten *et al.*, 2008; Qu *et al.*, 1996; Hadgu *et al.* , 1998).

If we assume that tests $\mathrm{T}_1$–$\mathrm{T}_3$ are correlated in the disease subjects only, there would be a single random effect $z_{1i}$ and the coefficient vectors would be scalars with $\beta_{j10} = \cdots = \beta_{j1J} = 0$, $\beta_{111} = \beta_{211} = \beta_{311} = \gamma_{123|D=1}$, $\beta_{411} = \beta_{511} = \beta_{611} = 0$. The

size of $\gamma_{123|D=1}$ indicates the strength of the dependency of $T_1$–$T_3$ on the random effect in disease subjects and consequently is proportional to the strength of the association between $T_1$–$T_3$.

The probability of an outcome pattern for individual patients, conditional on their disease status and random effects value, is given by

$$P(Y_{i1} = y_1, \ldots, Y_{iJ} = y_J | D_i = d_i, \boldsymbol{Z_i} = \boldsymbol{z_i}) = \Pi_{j=1}^{J} P(Y_{ij} = y_j | D_i = d_i, \boldsymbol{Z_i} = \boldsymbol{z_i}). \tag{5.7}$$

The results of the different tests are assumed independent conditional on the disease status and the random effects. To obtain the population-averaged sensitivity and specificity, we average out over the random effects distributions

$$\begin{aligned} S_j &= \int_{-\infty}^{\infty} \eta^{-1}(\alpha_{j1} + \beta_{j1} z_{1i}) f(z_{1i}) dz_{1i}, \\ C_j &= 1 - \int_{-\infty}^{\infty} \eta^{-1}(\alpha_{j0} + \beta_{j0} z_{2i}) f(z_{2i}) dz_{2i}. \end{aligned} \tag{5.8}$$

The probability that the $i$th individual will have test results $y_i = (y_{i1}, \ldots, y_{iJ})^T$ is given by

$$\begin{aligned} &P(Y_{i1} = y_{i1}, \ldots, Y_{iJ} = y_{iJ}) \\ &= \sum_{d_i=0}^{1} P(D_i = d_i) P(Y_{i1} = y_{i1}, \ldots, Y_{iJ} = y_{iJ} | D_i = d_i) \\ &= P(D_i = 0) P(Y_{i1} = y_{i1}, \ldots, Y_{iJ} = y_{iJ} | D_i = 0) \\ &\quad + P(D_i = 1) P(Y_{i1} = y_{i1}, \ldots, Y_{iJ} = y_{iJ} | D_i = 1) \\ &= P(D_i = 0) \int_{-\infty}^{\infty} P(Y_{i1} = y_{i1}, \ldots, Y_{iJ} = y_{iJ} | D_i = 0, Z_i = z_{1i}) f(z_{1i}) dz_{1i} \\ &\quad + P(D_i = 1) \int_{-\infty}^{\infty} P(Y_{i1} = y_{i1}, \ldots, Y_{iJ} = y_{iJ} | D_i = 1, Z_i = z_{2i}) f(z_{2i}) dz_{2i}. \end{aligned} \tag{5.9}$$

Since the probabilities of an outcome for an individual patients, conditional on their disease status and random effect value are independent (equation 5.7), then the probability that the $i$th individual will have test results $y_i = (y_{i1}, \ldots, y_{iJ})^T$ is given by

$$
\begin{aligned}
&P(Y_{i1} = y_{i1}, \ldots, Y_{iJ} = y_{iJ}) \\
&\quad = P(D_i = 0) \int_{-\infty}^{\infty} \Pi_{j=1}^{J} P(Y_{ij} = y_j | D_i = 0, Z_i = z_{1i}) f(z_{1i}) dz_{1i} \\
&\qquad + P(D_i = 1) \int_{-\infty}^{\infty} \Pi_{j=1}^{J} P(Y_{ij} = y_j | D_i = 1, Z_i = z_{2i}) f(z_{2i}) dz_{2i}.
\end{aligned} \tag{5.10}
$$

It is commonly thought that test performance may depend on covariates. In order to know in which tests and to what extent the sensitivity and specificity depends on age and sex, we extend the conditional dependence model (5.6) by including a covariate vector $\boldsymbol{x_i}$

$$
P(Y_{ij} = 1 | D_i = d_i, \boldsymbol{Z_i} = \boldsymbol{z_i}, \boldsymbol{X_i} = \boldsymbol{x_i}) = \eta^{-1}(\alpha_{jd_i} + \boldsymbol{\beta}_{jd_i}^T \boldsymbol{z_i} + \boldsymbol{\gamma}_{jd_i}^T \boldsymbol{x_i}). \tag{5.11}
$$

The conditional dependence model extended by covariates will provide estimates of sensitivities and specificities adjusted for the covariate $\boldsymbol{x_i}$. The probability of an outcome pattern for individual patients, conditional on their disease status, covariates and random effects value, is given by

$$
\begin{aligned}
&P(Y_{i1} = y_1, \ldots, Y_{iJ} = y_J | D_i = d_i, \boldsymbol{Z_i} = \boldsymbol{z_i}, \boldsymbol{X_i} = \boldsymbol{x_i}) \\
&\qquad = \Pi_{j=1}^{J} P(Y_{ij} = y_j | D_i = d_i, \boldsymbol{Z_i} = \boldsymbol{z_i}, \boldsymbol{X_i} = \boldsymbol{x_i}).
\end{aligned} \tag{5.12}
$$

## 5.5 Application to the Datasets

In the next sections we apply the conditional independence model and conditional dependence model to evaluate diagnostic tests used in acute phase serum for the diagnosis of *Mycoplasma pneumoniae* and *Chlamydophila pneumoniae* infections. The programs used to fit the conditional independence and the conditional dependence models are included in Appendix D.

### 5.5.1 Evaluation of Tests Used for the Detection of *Mycoplasma Pneumoniae*

**Seroprevalence for *M. Pneumoniae* IgG in Patients Determined with Different Assays**

All tests could be applied to specimens from 201 patients. When measuring IgG in the acute phase sera, 71.1%, 35.3%, 71.6% and 49.8% of the sera were found positive for the presence of *M. pneumoniae* IgG by ImmunoWELL, Medac, ANILab, and

EUROIMMUN EIAs, respectively (Table 5.2).

**Table 5.2:** Comparison of *M. pneumoniae* seropositivities using different Enzyme Immunoassays.

| Test | No. of positive patients | % of positive patients |
|---|---|---|
| IM-IgM-v0 | 43 | 21.4 |
| ME-IgM-v0 | 7 | 3.5 |
| AL-IgM-v0 | 26 | 12.9 |
| EI-IgM-v0 | 3 | 1.5 |
| IM-IgM-v4 | 45 | 22.4 |
| ME-IgM-v4 | 11 | 5.5 |
| AL-IgM-v4 | 25 | 12.4 |
| EI-IgM-v4 | 7 | 3.5 |
| IM-IgG-v0 | 143 | 71.1 |
| ME-IgG-v0 | 71 | 35.3 |
| AL-IgG-v0 | 144 | 71.6 |
| EI-IgG-v0 | 100 | 49.8 |
| ME-IgA-v0 | 31 | 15.4 |
| EI-IgA-v0 | 27 | 13.4 |
| ME-IgA-v4 | 35 | 17.4 |
| EI-IgA-v4 | 35 | 17.4 |

IM: ImmunoWELL, ME: Medac, AL: ANILab, EI: EUROIMMUN, v0: acute phase serum, v4: convalescent phase serum

### Medac Tests

We considered the conditional independence and dependence models for the analysis of the Medac tests. The two nucleic acid sequence-based amplification techniques (NASBA and PCR) were also included in the model.

(i) Serology Tests

$T_1$ = Me-IgM-v0

$T_2$ = Me-IgG-v0

$T_3$ = Me-IgA-v0

$T_4$ = Me-IgG-sign-rise

(ii) Amplification Techniques

$T_5$ = OS-NASBA-M

$T_6$ = OS-PCR-M

First we consider a conditional independence model where all the six tests are conditionally independent (Model 1). Model 2 is a conditional dependence model where the three Medac tests ($T_1$–$T_3$) are assumed to be dependent in the disease subjects ($d_i = 1$) and in the non-disease subjects ($d_i = 0$). The two amplification techniques ($T_5$ and $T_6$) are assumed to be dependent in the disease subjects and in the non-disease subjects in Model 3. Only the tree Medac tests ($T_1$–$T_3$) and the significant rise test ($T_4$) are assumed to be dependent in the disease and non-disease subjects in Model 4. While in Model 5, the significant rise test ($T_4$) and the two amplification techniques ($T_5$ and $T_6$) are assumed to be dependent in the disease and non-disease subjects. The three Medac tests and the two amplification techniques are assumed dependent in Model 6. In Model 7 all six tests are assumed to be dependent. Model 2 is extended by including sex and age of the patients as a covariate, respectively in Model 8 and Model 9.

**Table 5.3:** Description and model selection criteria for nine models used to analyze the Medac tests for the diagnosis of *M. pneumoniae*.

| Model No. | Correlation in non-diseased subjects | Correlation in diseased subjects | Covariate | Model comparison | | |
|---|---|---|---|---|---|---|
| | | | | *pD* | *DIC* | *p*-value |
| 1 | | | | 8.919 | 136.462 | 0.002 |
| 2 | $T_1$–$T_3$ | $T_1$–$T_3$ | | 8.655 | 105.260 | 0.356 |
| 3 | $T_5$–$T_6$ | $T_5$–$T_6$ | | 8.999 | 136.955 | 0.001 |
| 4 | $T_1$–$T_4$ | $T_1$–$T_4$ | | 9.450 | 114.508 | 0.166 |
| 5 | $T_4$–$T_6$ | $T_4$–$T_6$ | | 8.851 | 137.115 | 0.001 |
| 6 | $T_1$–$T_3$ & $T_5$–$T_6$ | $T_1$–$T_3$ & $T_5$–$T_6$ | | 9.792 | 115.463 | 0.153 |
| 7 | $T_1$–$T_6$ | $T_1$–$T_6$ | | 10.445 | 120.928 | 0.121 |
| 8* | $T_1$–$T_3$ | $T_1$–$T_3$ | Sex | 8.695 | 104.951 | 0.372 |
| 9† | $T_1$–$T_3$ | $T_1$–$T_3$ | Age | 8.913 | 105.722 | 0.351 |

* Gender of the patients is included as a covariate in Model 2.
† Age of the patients (as a categorical variable; =0 if age<45 or =1 if age≥45) is included as a covariate in Model 2.

The results of the conditional dependence model and the different expanded gold standards are summarized in Tables 5.4 and 5.6. In Tables 5.4 and 5.6, only the results of the simple latent class model (i.e., the conditional independence model) and the best conditional dependence model based on the deviance information criteria (DIC) are included.

**Table 5.4:** Parameter estimates: posterior means and standard errors for the parameters obtained by fitting Model 1 and 2 for the Medac tests used for the diagnosis of *M. pneumoniae*.

| Parameters | Model 1 | Model 2 |
|---|---|---|
| $\alpha_{11}$ | -1.541 (0.664) | -2.241 (1.002) |
| $\alpha_{21}$ | -0.263 (0.517) | -0.229 (0.830) |
| $\alpha_{31}$ | -1.542 (0.656) | -2.269 (1.011) |
| $\alpha_{41}$ | 1.110 (0.598) | 1.108 (0.605) |
| $\alpha_{51}$ | 4.446 (1.699) | 4.437 (1.677) |
| $\alpha_{61}$ | 4.180 (1.731) | 4.180 (1.727) |
| $\alpha_{10}$ | -2.989 (0.343) | -4.049 (0.497) |
| $\alpha_{20}$ | -0.226 (0.147) | -0.343 (0.224) |
| $\alpha_{30}$ | -1.426 (0.185) | -2.077 (0.312) |
| $\alpha_{40}$ | -2.992 (0.342) | -2.993 (0.344) |
| $\alpha_{50}$ | -3.650 (0.499) | -3.644 (0.491) |
| $\alpha_{60}$ | -6.360 (1.432) | -6.355 (1.414) |
| $\beta_{10}$ | | 1.685 (0.311) |
| $\beta_{11}$ | | 2.098 (0.776) |

Using the parameter estimates given in Table 5.4, the estimates for sensitivities and specificities are calculated as in Table 5.5. In order to obtain population averaged sensitivities and specificities, we averaged out over the random effects distributions.

**Table 5.5:** Estimation of sensitivities and specificities from the conditional dependence model (Model 2), where the first three test are assumed to be dependent in the diseased and non-diseased subjects.

| Test | Sensitivity | Specificity |
|---|---|---|
| $T_1$ | $\int_{-\infty}^{\infty} \eta^{-1}(\alpha_{11} + \beta_{11} z_{1i}) f(z_{1i}) dz_{1i}$ | 1 - $\int_{-\infty}^{\infty} \eta^{-1}(\alpha_{10} + \beta_{10} z_{2i}) f(z_{2i}) dz_{2i}$ |
| $T_2$ | $\int_{-\infty}^{\infty} \eta^{-1}(\alpha_{21} + \beta_{11} z_{1i}) f(z_{1i}) dz_{1i}$ | 1 - $\int_{-\infty}^{\infty} \eta^{-1}(\alpha_{20} + \beta_{10} z_{2i}) f(z_{2i}) dz_{2i}$ |
| $T_3$ | $\int_{-\infty}^{\infty} \eta^{-1}(\alpha_{31} + \beta_{11} z_{1i}) f(z_{1i}) dz_{1i}$ | 1 - $\int_{-\infty}^{\infty} \eta^{-1}(\alpha_{30} + \beta_{10} z_{2i}) f(z_{2i}) dz_{2i}$ |
| $T_4$ | $\eta^{-1}(\alpha_{41})$ | 1 - $\eta^{-1}(\alpha_{40})$ |
| $T_5$ | $\eta^{-1}(\alpha_{51})$ | 1 - $\eta^{-1}(\alpha_{50})$ |
| $T_6$ | $\eta^{-1}(\alpha_{61})$ | 1 - $\eta^{-1}(\alpha_{60})$ |

For the conditional independence model (Model 1), the estimates for the sensitivity and specificity for the $j$th test ($j = 1, 2, \ldots, 6$) are $\eta^{-1}(\alpha_{j1})$ and $1 - \eta^{-1}(\alpha_{j0})$, respectively.

**Table 5.6:** Estimates and standard errors for prevalence, sensitivities and specificities for the Medac tests used for the diagnosis of *M. pneumoniae*.

| Parameters | Expanded Gold Standard† | | | | Latent Class Model | |
|---|---|---|---|---|---|---|
| | EGS[1] | EGS[2] | EGS[3] | EGS[4] | Conditional independence model | Conditional dependence model‡ |
| Prevalence | 0.119 | 0.079 | 0.075 | 0.095 | 0.085(0.020) | 0.084(0.020) |
| Sens. Me-IgM-v0 | 0.125 | 0.188 | 0.200 | 0.053 | 0.195(0.095) | 0.219(0.097) |
| Sens. Me-IgG-v0 | 0.500 | 0.438 | 0.333 | 0.421 | 0.439(0.120) | 0.466(0.115) |
| Sens. Me-IgA-v0 | 0.167 | 0.188 | 0.133 | 0.053 | 0.195(0.094) | 0.216(0.097) |
| Sens. Me-IgG-sign-rise | 0.708 | 0.750 | 0.800 | 0.842 | 0.737(0.107) | 0.736(0.108) |
| Sens. NASBA | 0.708 | 1.000 | 1.000 | 0.632 | 0.971(0.038) | 0.971(0.038) |
| Sens. PCR | 0.667 | 1.000 | 0.933 | 0.632 | 0.962(0.051) | 0.962(0.051) |
| Spec. Me-IgM-v0 | 0.949 | 0.952 | 0.952 | 0.939 | 0.950(0.016) | 0.947(0.017) |
| Spec. Me-IgG-v0 | 0.565 | 0.557 | 0.548 | 0.555 | 0.556(0.036) | 0.557(0.037) |
| Spec. Me-IgA-v0 | 0.802 | 0.805 | 0.801 | 0.791 | 0.805(0.029) | 0.804(0.030) |
| Spec. Me-IgG-sign-rise | 0.977 | 0.951 | 0.952 | 0.973 | 0.950(0.016) | 0.950(0.016) |
| Spec. NASBA | 0.977 | 0.973 | 0.968 | 0.951 | 0.972(0.012) | 0.972(0.012) |
| Spec. PCR | 1.000 | 1.000 | 0.989 | 0.978 | 0.996(0.004) | 0.996(0.004) |
| ME\|D=0 | | | | | | 1.685(0.311) |
| ME\|D=1 | | | | | | 2.098(0.776) |
| $pD$ | | | | | 8.919 | 8.655 |
| $DIC$ | | | | | 136.462 | 105.260 |
| Bayesian $p$-value | | | | | 0.002 | 0.356 |

[1] An expanded gold standard is positive if the test result is positive by PCR and NASBA or positive by one amplification test and at least one serology test (either IgM or a significant rise of IgG antibodies as defined according to the instructions of the manufacturer) or a significant rise of IgG antibodies in at least two different EIA's.

[2] An expanded gold standard is positive if the test result is positive by PCR and NASBA.

[3] An expanded gold standard is positive if the test result is positive by 1 amplification test and at least one serology test (either IgM or a significant rise of IgG antibodies as defined according to the instructions of the manufacturer).

[4] An expanded gold standard is positive if the test result is positive by a significant rise of IgG antibodies in at least two different EIA's.

† These expanded gold standards are not real gold standards but rather optimal subjective standards based on clinical experience and evidence.

‡ The three Medac tests (Me-IgM-v0, Me-IgG-v0 and Me-IgA-v0) are assumed to be dependent in the disease subjects and in the non-disease subjects.

Table 5.6 shows the estimated population-averaged sensitivities and specificities for each test under the expanded gold standard and under Models 1 and 2. The conditional dependence model, referred to as Model 2 in Table 5.3, has the smallest DIC and largest Bayesian $p$-value. The sensitivities of the Medac tests (Me-IgM, Me-IgG-v0 and Me-IgA-v0) under Model 1 are slightly lower than those under Model 2.

From the results of the conditional dependence model (Model 2), the estimates for the prevalence is 8.4 percent. Sensitivities of the Medac manufacturers are $<0.5$, but the specificities of the tests are $>0.5$. Sensitivity and specificity of NASBA amplification technique are 97 percent and 97 percent, respectively. For the PCR amplification technique the estimates for the sensitivity and specificity are 96 percent and 99 percent, respectively. The sensitivity of the significant rise of Me-IgG test is close to 74 percent and the specificity of the test is close to 95 percent.

The different criteria for the expanded gold standard were also compared with the tests used for the diagnosis of *M. pneumoniae* and the results of the latent class models. As we can clearly see from Table 5.6, the estimated measures of diagnostic accuracy from the latent class models are quite close to the expanded gold standard using only PCR and NASBA (i.e. $\text{EGS}^2$ in Table 5.6).

The results of the conditional independence, conditional dependence and the expanded gold standard for the ImmunoWELL tests, ANILab tests and EUROIMMUN tests used for the diagnosis of *M. pneumoniae* are given in Tables 5.7–5.9, respectivly. To compare the nucleic acid sequence-based amplification techniques with the serology tests for the detection of *M. pneumoniae*, the two amplification techniques (NASBA and PCR) were also included in the models.

**ImmunoWELL Tests**

**Table 5.7:** Estimates and standard errors for prevalence, sensitivities and specificities for the ImmunoWELL tests used for the diagnosis of *M. pneumoniae*.

| Parameters | Expanded Gold Standard† | | | | Latent Class Model | |
|---|---|---|---|---|---|---|
| | EGS[1] | EGS[2] | EGS[3] | EGS[4] | Conditional independence model | Conditional dependence model‡ |
| Prevalence | 0.119 | 0.079 | 0.075 | 0.095 | 0.086(0.020) | 0.086(0.020) |
| Sens. IM-IgM-v0 | 0.542 | 0.563 | 0.600 | 0.579 | 0.550(0.121) | 0.546(0.120) |
| Sens. IM-IgG-v0 | 0.792 | 0.813 | 0.800 | 0.737 | 0.804(0.094) | 0.797(0.095) |
| Sens. IM-IgG-sign-rise | 0.375 | 0.500 | 0.533 | 0.474 | 0.492(0.120) | 0.490(0.121) |
| Sens. NASBA | 0.708 | 1.000 | 1.000 | 0.632 | 0.971(0.038) | 0.971(0.038) |
| Sens. PCR | 0.667 | 1.000 | 0.933 | 0.632 | 0.954(0.059) | 0.950(0.064) |
| Spec. IM-IgM-v0 | 0.678 | 0.670 | 0.672 | 0.679 | 0.669(0.034) | 0.669(0.035) |
| Spec. IM-IgG-v0 | 0.299 | 0.297 | 0.296 | 0.291 | 0.298(0.034) | 0.299(0.034) |
| Spec. IM-IgG-sign-rise | 0.966 | 0.962 | 0.963 | 0.967 | 0.960(0.014) | 0.960(0.014) |
| Spec. NASBA | 0.977 | 0.973 | 0.968 | 0.951 | 0.973(0.013) | 0.973(0.013) |
| Spec. PCR | 1.000 | 1.000 | 0.989 | 0.978 | 0.996(0.004) | 0.996(0.004) |
| IM\|D=0 | | | | | | 1.434(0.369) |
| IM\|D=1 | | | | | | 0.816(0.621) |
| $pD$ | | | | | 7.029 | 7.658 |
| $DIC$ | | | | | 93.148 | 84.347 |
| Bayesian $p$-value | | | | | 0.017 | 0.118 |

[1] An expanded gold standard is positive if the test result is positive by PCR and NASBA or positive by one amplification test and at least one serology test (either IgM or a significant rise of IgG antibodies as defined according to the instructions of the manufacturer) or a significant rise of IgG antibodies in at least two different EIA's.

[2] An expanded gold standard is positive if the test result is positive by PCR and NASBA.

[3] An expanded gold standard is positive if the test result is positive by 1 amplification test and at least one serology test (either IgM or a significant rise of IgG antibodies as defined according to the instructions of the manufacturer).

[4] An expanded gold standard is positive if the test result is positive by a significant rise of IgG antibodies in at least two different EIA's.

† These expanded gold standards are not real gold standards but rather optimal subjective standards based on clinical experience and evidence.

‡ The two ImmunoWELL tests (IM-IgM-v0 and IM-IgG-v0) are assumed to be dependent in the disease subjects and in the non-disease subjects.

### ANILab Tests

**Table 5.8:** Estimates and standard errors for prevalence, sensitivities and specificities for the ANILab tests used for the diagnosis of *M. pneumoniae*.

| Parameters | Expanded Gold Standard† | | | | Latent Class Model | |
|---|---|---|---|---|---|---|
| | EGS[1] | EGS[2] | EGS[3] | EGS[4] | Conditional independence model | Conditional dependence model‡ |
| Prevalence | 0.119 | 0.079 | 0.075 | 0.095 | 0.087(0.020) | 0.086(0.020) |
| Sens. AL-IgM-v0 | 0.542 | 0.625 | 0.600 | 0.526 | 0.601(0.119) | 0.600(0.116) |
| Sens. AL-IgG-v0 | 0.708 | 0.688 | 0.667 | 0.632 | 0.693(0.110) | 0.682(0.110) |
| Sens. AL-IgG-sign-rise | 0.667 | 0.750 | 0.867 | 0.789 | 0.743(0.105) | 0.743(0.105) |
| Sens. NASBA | 0.708 | 1.000 | 1.000 | 0.632 | 0.971(0.037) | 0.971(0.038) |
| Sens. PCR | 0.667 | 1.000 | 0.933 | 0.632 | 0.943(0.065) | 0.946(0.062) |
| Spec. AL-IgM-v0 | 0.644 | 0.643 | 0.639 | 0.637 | 0.642(0.035) | 0.642(0.036) |
| Spec. AL-IgG-v0 | 0.175 | 0.178 | 0.177 | 0.170 | 0.180(0.028) | 0.180(0.028) |
| Spec. AL-IgG-sign-rise | 0.944 | 0.924 | 0.930 | 0.939 | 0.925(0.019) | 0.925(0.019) |
| Spec. NASBA | 0.977 | 0.973 | 0.968 | 0.951 | 0.974(0.012) | 0.974(0.012) |
| Spec. PCR | 1.000 | 1.000 | 0.989 | 0.978 | 0.996(0.004) | 0.996(0.004) |
| AL\|D=0 | | | | | | 0.703(0.385) |
| AL\|D=1 | | | | | | 0.918(0.657) |
| $pD$ | | | | | 7.296 | 7.897 |
| $DIC$ | | | | | 82.849 | 82.918 |
| Bayesian $p$-value | | | | | 0.265 | 0.259 |

[1] An expanded gold standard is positive if the test result is positive by PCR and NASBA or positive by one amplification test and at least one serology test (either IgM or a significant rise of IgG antibodies as defined according to the instructions of the manufacturer) or a significant rise of IgG antibodies in at least two different EIA's.

[2] An expanded gold standard is positive if the test result is positive by PCR and NASBA.

[3] An expanded gold standard is positive if the test result is positive by 1 amplification test and at least one serology test (either IgM or a significant rise of IgG antibodies as defined according to the instructions of the manufacturer).

[4] An expanded gold standard is positive if the test result is positive by a significant rise of IgG antibodies in at least two different EIA's.

† These expanded gold standards are not real gold standards but rather optimal subjective standards based on clinical experience and evidence.

‡ The two ANILab tests (AL-IgM-v0 and AL-IgG-v0) are assumed to be dependent in the disease subjects and in the non-disease subjects.

**EUROIMMUN Tests**

**Table 5.9:** Estimates and standard errors of prevalence, sensitivities and specificities for the EUROIMMUN tests used for the diagnosis of *M. pneumoniae*.

| Parameters | Expanded Gold Standard† | | | | Latent Class Model | |
|---|---|---|---|---|---|---|
| | EGS[1] | EGS[2] | EGS[3] | EGS[4] | Conditional independence model | Conditional dependence model‡ |
| Prevalence | 0.119 | 0.079 | 0.075 | 0.095 | 0.084(0.020) | 0.084(0.020) |
| Sens. EI-IgM-v0 | 0.125 | 0.188 | 0.200 | 0.053 | 0.196(0.095) | 0.198(0.095) |
| Sens. EI-IgG-v0 | 0.667 | 0.625 | 0.533 | 0.632 | 0.623(0.115) | 0.619(0.115) |
| Sens. EI-IgG-sign-rise | 0.750 | 0.750 | 0.800 | 0.947 | 0.737(0.107) | 0.734(0.106) |
| Sens. EI-IgA-v0 | 0.250 | 0.313 | 0.267 | 0.158 | 0.318(0.111) | 0.317(0.110) |
| Sens. NASBA | 0.708 | 1.000 | 1.000 | 0.632 | 0.971(0.039) | 0.972(0.037) |
| Sens. PCR | 0.667 | 1.000 | 0.933 | 0.632 | 0.962(0.050) | 0.963(0.049) |
| Spec. EI-IgM-v0 | 0.972 | 0.973 | 0.973 | 0.962 | 0.971(0.012) | 0.970(0.013) |
| Spec. EI-IgG-v0 | 0.328 | 0.324 | 0.317 | 0.324 | 0.324(0.035) | 0.325(0.035) |
| Spec. EI-IgG-sign-risev | 0.909 | 0.881 | 0.884 | 0.912 | 0.880(0.024) | 0.879(0.024) |
| Spec. EI-IgA-v0 | 0.712 | 0.719 | 0.715 | 0.703 | 0.718(0.033) | 0.717(0.034) |
| Spec.NASBA | 0.977 | 0.973 | 0.968 | 0.951 | 0.972(0.012) | 0.972(0.012) |
| Spec. PCR | 1.000 | 1.000 | 0.989 | 0.978 | 0.996(0.004) | 0.996(0.004) |
| EI\|D=0 | | | | | | 0.958(0.255) |
| EI\|D=1 | | | | | | 0.581(0.416) |
| $pD$ | | | | | 8.906 | 9.647 |
| $DIC$ | | | | | 119.654 | 112.690 |
| Bayesian $p$-value | | | | | 0.074 | 0.217 |

[1] An expanded gold standard is positive if the test result is positive by PCR and NASBA or positive by one amplification test and at least one serology test (either IgM or a significant rise of IgG antibodies as defined according to the instructions of the manufacturer) or a significant rise of IgG antibodies in at least two different EIA's.

[2] An expanded gold standard is positive if the test result is positive by PCR and NASBA.

[3] An expanded gold standard is positive if the test result is positive by 1 amplification test and at least one serology test (either IgM or a significant rise of IgG antibodies as defined according to the instructions of the manufacturer).

[4] An expanded gold standard is positive if the test result is positive by a significant rise of IgG antibodies in at least two different EIA's.

† These expanded gold standards are not real gold standards but rather optimal subjective standards based on clinical experience and evidence.

‡ The two EUROIMMUN tests (EI-IgM-v0 and EI-IgG-v0) are assumed to be dependent in the disease subjects and in the non-disease subjects.

**Comparison of the Latent Class Model with the Expanded Gold Standard**

The results of the latent class model are compared with the expanded gold standard, for the different manufacturers. The results of the conditional dependence model for the Medac manufacturer are summarized in Table 5.10.

**Table 5.10:** Comparison of the latent class model and the expanded gold standard for tests used for the diagnosis of *M. pneumoniae*.

| | | Latent Class Model | |
|---|---|---|---|
| | | Non-diseased | Diseased |
| Expanded Gold Standard[1] | Non-diseased | 177 | 0 |
| | Diseased | 8 | 16 |
| Expanded Gold Standard[2] | Non-diseased | 185 | 0 |
| | Diseased | 0 | 16 |
| Expanded Gold Standard[3] | Non-diseased | 184 | 2 |
| | Diseased | 1 | 14 |
| Expanded Gold Standard[4] | Non-diseased | 178 | 4 |
| | Diseased | 7 | 12 |

[1] An expanded gold standard is positive if the test result is positive by PCR and NASBA or positive by one amplification test and at least one serology test (either IgM or a significant rise of IgG antibodies as defined according to the instructions of the manufacturer) or a significant rise of IgG antibodies in at least two different EIA's.

[2] An expanded gold standard is positive if the test result is positive by PCR and NASBA.

[3] An expanded gold standard is positive if the test result is positive by 1 amplification test and at least one serology test (either IgM or a significant rise of IgG antibodies as defined according to the instructions of the manufacturer).

[4] An expanded gold standard is positive if the test result is positive by a significant rise of IgG antibodies in at least two different EIA's.

Now the results of the latent class models are more in agreement with the expanded gold standard using the two nucleic acid amplification tests ($EGS^2$).

### 5.5.2 Evaluation of Tests Used for the Detection of *Chlamydophila Pneumoniae*

In the same way as we have applied the conditional independence and dependence models for tests used for the diagnosis of *M. pneumoniae*, we also applied these models

for tests used to detect *C. pneumoniae*. The results of the conditional independence, conditional dependence and the expanded gold standard for the Medac tests, ImmunoWELL tests, ANILab tests and EUROIMMUN tests used for the diagnosis of *C. pneumoniae* are given in Tables 5.11–5.14, respectively. To compare the nucleic acid sequence-based amplification techniques with the serology tests for the detection of *C. pneumoniae*, the two amplification techniques (NASBA and PCR) were also included in the models.

**Medac Tests**

**Table 5.11:** Estimates and standard errors of prevalence, sensitivities and specificities for the Medac tests used for the diagnosis of *C. pneumoniae*.

| Parameters | Expanded Gold Standard† | | | | Latent Class Model | |
|---|---|---|---|---|---|---|
| | EGS[1] | EGS[2] | EGS[3] | EGS[4] | Conditional independence model | Conditional dependence model‡ |
| Prevalence | 0.268 | 0.037 | 0.044 | 0.261 | 0.046(0.019) | 0.045(0.018) |
| Sens. Me-IgM | 0.277 | 1.000 | 0.833 | 0.285 | 0.926(0.097) | 0.909(0.107) |
| Sens. Me-IgG-titer | 0.722 | 0.400 | 0.500 | 0.714 | 0.426(0.196) | 0.407(0.186) |
| Sens. Me-IgG-seroc-titer | 0.388 | 0.600 | 0.500 | 0.400 | 0.599(0.193) | 0.595(0.193) |
| Sens. NASBA | 0.166 | 1.000 | 1.000 | 0.142 | 0.907(0.124) | 0.916(0.111) |
| Sens. PCR | 0.138 | 1.000 | 0.833 | 0.142 | 0.905(0.125) | 0.915(0.113) |
| Spec. Me-IgM | 0.979 | 0.945 | 0.945 | 0.979 | 0.945(0.021) | 0.942(0.021) |
| Spec. Me-IgG-titer | 0.346 | 0.317 | 0.320 | 0.343 | 0.319(0.041) | 0.318(0.042) |
| Spec. Me-IgG-seroc-titer | 1.000 | 0.915 | 0.914 | 1.000 | 0.914(0.025) | 0.913(0.024) |
| Spec. NASBA | 1.000 | 0.992 | 1.000 | 0.989 | 0.989(0.009) | 0.989(0.009) |
| Spec. PCR | 1.000 | 1.000 | 1.000 | 1.000 | 0.995(0.006) | 0.995(0.006) |
| ME\|D=0 | | | | | | 1.769(0.815) |
| ME\|D=1 | | | | | | 1.099(0.808) |
| $pD$ | | | | | 6.010 | 5.269 |
| $DIC$ | | | | | 57.410 | 51.867 |
| Bayesian $p$-value | | | | | 0.239 | 0.425 |

[1] An expanded gold standard is positive if the test result is positive by PCR and NASBA or positive by one amplification test and at least one serology test (either IgM or a significant rise of IgG antibodies as defined according to the instructions of the manufacturer) or a significant rise of IgG antibodies in at least two different EIA's.

[2] An expanded gold standard is positive if the test result is positive by PCR and NASBA.

[3] An expanded gold standard is positive if the test result is positive by 1 amplification test and at least one serology test (either IgM or a significant rise of IgG antibodies as defined according to the instructions of the manufacturer).

[4] An expanded gold standard is positive if the test result is positive by a significant rise of IgG antibodies in at least two different EIA's.

† These expanded gold standards are not real gold standards but rather optimal subjective standards based on clinical experience and evidence.

‡ The two Medac tests (Me-IgM and Me-IgG-titer) are assumed to be dependent in the disease subjects and in the non-disease subjects.

**Microimmunofluorescence Tests**

**Table 5.12:** Estimates and standard errors for prevalence, sensitivities and specificities for the Microimmunofluorescence tests used for the diagnosis of *C. pneumoniae*.

| Parameters | Expanded Gold Standard† | | | | Latent Class Model | |
|---|---|---|---|---|---|---|
| | EGS[1] | EGS[2] | EGS[3] | EGS[4] | Conditional independence model | Conditional dependence model‡ |
| Prevalence | 0.274 | 0.037 | 0.051 | 0.266 | 0.045(0.018) | 0.045(0.018) |
| Sens. MIF-IgM | 0.324 | 1.000 | 0.714 | 0.333 | 0.920(0.104) | 0.898(0.118) |
| Sens. MIF-IgG | 0.729 | 0.400 | 0.428 | 0.722 | 0.407(0.194) | 0.397(0.184) |
| Sens. MIF-IgG-seroconv | 1.000 | 1.000 | 1.000 | 1.000 | 0.931(0.091) | 0.930(0.093) |
| Sens. NASBA | 0.189 | 1.000 | 1.000 | 0.166 | 0.925(0.099) | 0.923(0.102) |
| Sens. PCR | 0.135 | 1.000 | 0.714 | 0.138 | 0.917(0.106) | 0.911(0.115) |
| Spec. MIF-IgM | 0.816 | 0.807 | 0.804 | 0.818 | 0.807(0.034) | 0.805(0.035) |
| Spec. MIF-IgG | 0.316 | 0.292 | 0.289 | 0.313 | 0.293(0.040) | 0.293(0.040) |
| Spec. MIF-IgG-seroconv | 0.000 | 0.000 | 0.000 | 0.000 | 0.005(0.006) | 0.005(0.005) |
| Spec. NASBA | 1.000 | 0.984 | 1.000 | 0.989 | 0.982(0.012) | 0.982(0.012) |
| Spec. PCR | 1.000 | 1.000 | 1.000 | 1.000 | 0.995(0.006) | 0.995(0.006) |
| MIF\|D=0 | | | | | | 1.074(0.496) |
| MIF\|D=1 | | | | | | 1.127(0.828) |
| $pD$ | | | | | 4.091 | 4.520 |
| $DIC$ | | | | | 40.213 | 38.410 |
| Bayesian $p$-value | | | | | 0.542 | 0.672 |

[1] An expanded gold standard is positive if the test result is positive by PCR and NASBA or positive by one amplification test and at least one serology test (either IgM or a significant rise of IgG antibodies as defined according to the instructions of the manufacturer) or a significant rise of IgG antibodies in at least two different EIA's.

[2] An expanded gold standard is positive if the test result is positive by PCR and NASBA.

[3] An expanded gold standard is positive if the test result is positive by 1 amplification test and at least one serology test (either IgM or a significant rise of IgG antibodies as defined according to the instructions of the manufacturer).

[4] An expanded gold standard is positive if the test result is positive by a significant rise of IgG antibodies in at least two different EIA's.

† These expanded gold standards are not real gold standards but rather optimal subjective standards based on clinical experience and evidence.

‡ The two Microimmunofluorescence tests (MIF-IgM and MIF-IgG) are assumed to be dependent in the disease subjects and in the non-disease subjects.

**ANILab Tests**

**Table 5.13:** Estimates and standard errors for prevalence, sensitivities and specificities for the ANILab tests used for the diagnosis of *C. pneumoniae*.

| Parameters | Expanded Gold Standard† | | | | Latent Class Model | |
|---|---|---|---|---|---|---|
| | EGS[1] | EGS[2] | EGS[3] | EGS[4] | Conditional independence model | Conditional dependence model‡ |
| Prevalence | 0.268 | 0.037 | 0.044 | 0.261 | 0.045(0.018) | 0.045(0.018) |
| Sens. AL-IgM | 0.222 | 1.000 | 0.833 | 0.228 | 0.928(0.094) | 0.909(0.109) |
| Sens. AL-IgG | 0.750 | 0.600 | 0.666 | 0.742 | 0.599(0.193) | 0.571(0.188) |
| Sens. AL-IgA | 0.527 | 0.800 | 0.666 | 0.542 | 0.772(0.163) | 0.745(0.167) |
| Sens. AL-IgG-sero | 0.444 | 0.600 | 0.500 | 0.457 | 0.595(0.193) | 0.593(0.193) |
| Sens. NASBA | 0.166 | 1.000 | 1.000 | 0.142 | 0.910(0.114) | 0.924(0.099) |
| Sens. PCR | 0.138 | 1.000 | 0.833 | 0.142 | 0.911(0.113) | 0.921(0.103) |
| Spec. AL-IgM | 1.000 | 0.976 | 0.976 | 1.000 | 0.975(0.014) | 0.970(0.015) |
| Spec. AL-IgG | 0.326 | 0.302 | 0.304 | 0.323 | 0.303(0.040) | 0.300(0.041) |
| Spec. AL-IgA | 0.418 | 0.441 | 0.437 | 0.424 | 0.442(0.043) | 0.441(0.045) |
| Spec. AL-IgG-sero | 1.000 | 0.899 | 0.898 | 1.000 | 0.898(0.026) | 0.898(0.027) |
| Spec. NASBA | 1.000 | 0.992 | 1.000 | 0.989 | 0.989(0.009) | 0.989(0.009) |
| Spec. PCR | 1.000 | 1.000 | 1.000 | 1.000 | 0.995(0.006) | 0.995(0.006) |
| AL\|D=0 | | | | | | 2.912(0.515) |
| AL\|D=1 | | | | | | 0.885(0.691) |
| $pD$ | | | | | 7.694 | 6.961 |
| $DIC$ | | | | | 125.663 | 77.019 |
| Bayesian $p$-value | | | | | <0.001 | 0.127 |

[1] An expanded gold standard is positive if the test result is positive by PCR and NASBA or positive by one amplification test and at least one serology test (either IgM or a significant rise of IgG antibodies as defined according to the instructions of the manufacturer) or a significant rise of IgG antibodies in at least two different EIA's.

[2] An expanded gold standard is positive if the test result is positive by PCR and NASBA.

[3] An expanded gold standard is positive if the test result is positive by 1 amplification test and at least one serology test (either IgM or a significant rise of IgG antibodies as defined according to the instructions of the manufacturer).

[4] An expanded gold standard is positive if the test result is positive by a significant rise of IgG antibodies in at least two different EIA's.

† These expanded gold standards are not real gold standards but rather optimal subjective standards based on clinical experience and evidence.

‡ The three ANILab tests (AL-IgM, AL-IgG and AL-IgA) are assumed to be dependent in the disease subjects and in the non-disease subjects.

**EUROIMMUN Tests**

**Table 5.14:** Estimates and standard errors for prevalence, sensitivities and specificities for the EUROIMMUN tests used for the diagnosis of *C. pneumoniae*.

| Parameters | Expanded Gold Standard† | | | | Latent Class Model | |
|---|---|---|---|---|---|---|
| | EGS[1] | EGS[2] | EGS[3] | EGS[4] | Conditional independence model | Conditional dependence model‡ |
| Prevalence | 0.268 | 0.037 | 0.044 | 0.261 | 0.044(0.018) | 0.044(0.018) |
| Sens. EI-IgM | 0.166 | 1.000 | 0.833 | 0.171 | 0.928(0.095) | 0.893(0.111) |
| Sens. EI-IgG | 0.527 | 0.600 | 0.500 | 0.542 | 0.592(0.193) | 0.544(0.185) |
| Sens. EI-IgA | 0.222 | 0.600 | 0.500 | 0.228 | 0.590(0.194) | 0.546(0.186) |
| Sens. EI-IgG-sero | 0.500 | 0.600 | 0.500 | 0.514 | 0.589(0.194) | 0.583(0.194) |
| Sens. NASBA | 0.166 | 1.000 | 1.000 | 0.142 | 0.926(0.097) | 0.927(0.096) |
| Sens. PCR | 0.138 | 1.000 | 0.833 | 0.142 | 0.924(0.099) | 0.920(0.103) |
| Spec. EI-IgM | 0.969 | 0.968 | 0.968 | 0.969 | 0.966(0.016) | 0.963(0.017) |
| Spec. EI-IgG | 0.438 | 0.449 | 0.445 | 0.444 | 0.450(0.043) | 0.447(0.044) |
| Spec. EI-IgA | 0.887 | 0.875 | 0.875 | 0.888 | 0.875(0.029) | 0.871(0.030) |
| Spec. EI-IgG-sero | 1.000 | 0.883 | 0.882 | 1.000 | 0.882(0.028) | 0.882(0.028) |
| Spec. NASBA | 1.000 | 0.992 | 1.000 | 0.989 | 0.988(0.009) | 0.989(0.009) |
| Spec. PCR | 1.000 | 1.000 | 1.000 | 1.000 | 0.995(0.006) | 0.995(0.006) |
| EI\|D=0 | | | | | | 1.899(0.472) |
| EI\|D=1 | | | | | | 1.825(0.951) |
| $pD$ | | | | | 7.784 | 6.662 |
| $DIC$ | | | | | 93.501 | 73.974 |
| Bayesian $p$-value | | | | | 0.007 | 0.196 |

[1] An expanded gold standard is positive if the test result is positive by PCR and NASBA or positive by one amplification test and at least one serology test (either IgM or a significant rise of IgG antibodies as defined according to the instructions of the manufacturer) or a significant rise of IgG antibodies in at least two different EIA's.

[2] An expanded gold standard is positive if the test result is positive by PCR and NASBA.

[3] An expanded gold standard is positive if the test result is positive by 1 amplification test and at least one serology test (either IgM or a significant rise of IgG antibodies as defined according to the instructions of the manufacturer).

[4] An expanded gold standard is positive if the test result is positive by a significant rise of IgG antibodies in at least two different EIA's.

† These expanded gold standards are not real gold standards but rather optimal subjective standards based on clinical experience and evidence.

‡ The three EUROIMMUN tests (EI-IgM, EI-IgG and EI-IgA) are assumed to be dependent in the disease subjects and in the non-disease subjects.

## 5.6 Discussion

In this chapter, we applied latent class models to evaluate tests used for the diagnosis of *Mycoplasma pneumoniae* and *Chlamydophila pneumoniae* infections. First we considered a conditional independence model and afterwards conditional dependence models with different dependency structures between the tests are considered. In order to evaluate the dependency of the tests on covariates, we extend the conditional independence and the conditional dependence models by including age and sex. From the results of these models, the performance of the tests used for the diagnosis of *Mycoplasma pneumoniae* and *Chlamydophila pneumoniae* infections do not depend on the age and sex. In the following sections, some discussions are presented.

### 5.6.1 Diagnosis of *Mycoplasma Pneumoniae*

Both serology and NAATS are widely used for diagnosis of *M. pneumoniae* in respiratory tract infections, but studies comparing the different methods are rare (Loens *et al.*, 2003; Loens *et al.*, 2010). In this study, 4 IgG, 4 IgM and 2 IgA commercially available serology assays from 4 manufacturers and 2 in-house nucleic acid amplification tests for the detection of *M. pneumoniae* were compared using paired sera and corresponding respiratory specimens from 201 patients. Important differences between the performances of the different assays were found.

Some commercialized assays lack both sensitivity and specificity, emphasizing the need for more validation and quality control (Beersma *et al.*, 2005; Nir-Paz *et al.*, 2006). The differences could be related to the use of different antigens and/or antigen preparations. A great number of antigen preparations have been proposed: whole organisms, protein fractions, glycoprotein fractions, recombinant antigens. The Medac ELISA is based upon a recombinant antigen which makes cross-reactivities with other bacteria less likely and might explain why only 35.5% of acute sera reacted positively in the IgG assay. This is in contrast to the other assays where higher percentages of positive results were found, 49.8–71.6%. These findings confirm previous studies such as the one by Montagnani *et al.* (2010).

Since IgM antibodies appear earlier than IgG antibodies the detection of IgM in serum is a widely used approach for the early serology diagnosis of a *M. pneumoniae* infection, especially in children. It should be realized that IgM antibodies are often not produced in children under 6 months of age, in a proportion of primary infections and during reinfections. A single IgM measurement may detect an acute infection with higher sensitivity if the test is performed at least 7 days after onset of disease

(Liu *et al.*, 2008). In some patients, IgM antibodies appear even later. Ozaki *et al.* (2007) found that a single assay using the IgM Immunocard (Meridian Biosciences) had a sensitivity of 31.8% for detection of an acute *M. pneumoniae* infection which increased to 88.6% when paired sera were analyzed from seropositive children with pneumonia. Furthermore, an elevated IgM may persist for months after the acute infection. IgM tests are usually less sensitive and specific than 4-fold changes in antibody titres between paired specimens separated by several weeks.

It has been reported that the detection of IgA-specific antibody seems to be a good indicator of a recent *M. pneumoniae* infection in both children and adults. On the other hand, when evaluating the Medac IgM, IgG and IgA assay on 159 serum samples from 113 patients with acute RTIs, Narita (2005) did not find a significant advantage of detecting IgA in children. The results of the measurements of IgA in the sera from the adult patients in this study confirm these findings.

In earlier studies, when applying NAATs and serology, not all patients found to be positive for *M. pneumoniae* by NAATs were confirmed by serology and vice versa as is the case in this study as well. During a community outbreak of *M. pneumoniae*, Nilsson *et al.* (2008) compared seminested and real-time PCR of oropharyngeal swabs with serology for diagnosis of *M. peumoniae* infections at different time points after onset of disease. The authors concluded that PCR was superior to serology for diagnosis of a *M. pneumoniae* infection during the early phase of infection. When examining 73 children with RTIs for *M. pneumoniae* by real-time PCR, and 2 serology assays (a passive agglutination test and the Immunocard assay), Otomo *et al.* (2008) confirmed the results of Nilsson *et al.* (2008). They found a sensitivity of 100% and 33.3% and a specificity of 100% and 82.1% for PCR and the Immunocard assay, respectively. According to the authors, real-time PCR or a related molecular assay is suitable for rapid diagnosis as a first screening test. These data confirm the lack of correlation of serology methods with culture and/or PCR.

In this study, it was found that the seroprevalence of *M. pneumoniae*, as well as the relationship between seropositivity for *M. pneumoniae* and LRTI, is influenced by the assay applied since the serology assays vary greatly in sensitivity and specificity. Thus the choice of the serology test has important implications when performing seroepidemiological studies and for the management of the individual patient. The clinical significance of positive test results, obtained by serology and/or NAATs, should be further defined by studies of patients with a documented infection and for whom detailed information concerning the time lapses between onset of disease and the collection of

the serum specimens are known. A combination of a nucleic acid amplification test and a serology test might be the best choice for an accurate *M. pneumoniae* diagnosis in adult patients presenting with an LRTI.

### 5.6.2 Diagnosis of *Chlamydophila Pneumoniae*

In this study, four commercially available serology assays (2 species specific, 1 genus specific, and 1 genus specific MIF test) for the detection of *C. pneumoniae* IgM and IgG antibodies and 2 commercially available serology assays for the detection of *C. pneumoniae* IgA antibodies were also compared with PCR and NASBA on serum samples and throat swabs, respectively, from 134 adult patients with community-acquired pneumonia. Substantial differences between the performances of the assays were found. Using serum samples of 80 healthy volunteers to evaluate 11 different *C. pneumoniae* IgG tests, 10 and 1 being species and genus specific respectively, Hermann *et al.* (2002) demonstrated that serology assays for the detection of anti-*C. pneumoniae*-specific IgG vary greatly in their sensitivities and specificities. Since it is not clear which criteria were used in the IgG EIAs to define a significant rise in titre, serology results are difficult to interpret. Some might represent false positive IgG results. Here it is demonstrated that *C. pneumoniae* seroprevalence, as well as the relationship between seropositivity for *C. pneumoniae* and community-acquired pneumonia, is influenced by the assay applied. serology assays vary greatly in sensitivity and specificity. We conclude that the choice of serology test has important implications when performing seroepidemiological studies. We can only agree with Persson *et al.* (2000) that the use of a proper gold standard is critical (Persson *et al.*, 2000). serology tests should preferably be confirmed by tests that demonstrate the organism, like cell culture or PCR, although these methods have their limitations too.

# Chapter 6

# Simulation Study

We have conducted simulation studies under different scenarios to investigate the impact of misspecifying the conditional dependency of the tests, and to look into the performance of the conditional dependence model under small, moderate and large dependency. In Section 6.1, we generate data under the conditional independence assumption and we analyze them using conditional independence and conditional dependence models; while in Section 6.2, we generate data under the conditional dependence assumption and we analyze them using conditional independence and conditional dependence models. We consider these scenarios using different true values in Sections 6.3 and 6.4.

## 6.1 Scenario 1: Data are generated using the conditional independence model

First using the multinomial probabilities (6.1), we generate 250 datasets with N=201 under Model 1 (in Chapter 5). Afterwards, the datasets are analyzed using the conditional independence and conditional dependence models. The estimates of the conditional independence model (Model 1) for the Medac tests in Table 5.4 were used as true values. Model 1 states that

$$\begin{aligned} &P(Y_{i1} = y_{i1}, \ldots, Y_{iJ} = y_{iJ}) \\ &\qquad = \pi \Pi_{j=1}^{J} S_j^{y_j} (1 - S_j)^{(1-y_j)} + (1 - \pi) \Pi_{j=1}^{J} C_j^{(1-y_j)} (1 - C_j)^{y_j}, \end{aligned} \tag{6.1}$$

where $\pi$ is the prevalence, $P(D = 1)$, $S_j$ is the sensitivity for the $j$th test, $P(Y_{ij} =$

$1|D_i = 1) = \eta^{-1}(\alpha_{j1})$, and $C_j$ is the specificity for the $j$th test, $P(Y_{ij} = 0|D_i = 0) = 1 - \eta^{-1}(\alpha_{j0})$. The values of the joint probabilities are given in Table B.1 (Appendix B). The simulation averages for the posterior means, simulation standard errors and simulation averages of the estimated standard errors obtained by fitting Models 1 and 2 are given in Tables 6.1 and 6.2.

**Table 6.1:** Parameter estimates: simulation averages of the posterior means, simulation standard errors and simulation averages of the estimated standard errors for the parameters obtained by fitting Models 1 and 2 (data are generated under Model 1).

| Parameters | True Values | Model 1 | Model 2 |
|---|---|---|---|
| $\alpha_{11}$ | -1.541 | -1.8950(0.9546)(0.7782) | -2.1022(1.0192)(0.8717) |
| $\alpha_{21}$ | -0.263 | -0.2833(0.5913)(0.5332) | -0.2972(0.6752)(0.6336) |
| $\alpha_{31}$ | -1.542 | -1.8455(0.8744)(0.7667) | -2.0528(0.9536)(0.8600) |
| $\alpha_{41}$ | 1.110 | 1.2059(0.7508)(0.6548) | 1.2062(0.7504)(0.6549) |
| $\alpha_{51}$ | 4.446 | 4.1894(0.5393)(1.6169) | 4.1921(0.5397)(1.6160) |
| $\alpha_{61}$ | 4.180 | 3.8021(0.7038)(1.6298) | 3.8026(0.7039)(1.6269) |
| $\alpha_{10}$ | -2.989 | -3.0495(0.3787)(0.3615) | -3.1564(0.3871)(0.3816) |
| $\alpha_{20}$ | -0.226 | -0.2227(0.1453)(0.1495) | -0.2346(0.1537)(0.1585) |
| $\alpha_{30}$ | -1.426 | -1.4509(0.1810)(0.1894) | -1.5212(0.1996)(0.2085) |
| $\alpha_{40}$ | -2.992 | -3.1014(0.3675)(0.3708) | -3.1011(0.3672)(0.3709) |
| $\alpha_{50}$ | -3.650 | -3.9495(0.6295)(0.6593) | -3.9472(0.6274)(0.6571) |
| $\alpha_{60}$ | -6.360 | -5.9907(0.5339)(1.3173) | -5.9910(0.5357)(1.3168) |
| $\beta_{10}$ | 0 | | 0.4115(0.1598)(0.2469) |
| $\beta_{11}$ | 0 | | 0.8793(0.3524)(0.5692) |

**Table 6.2:** Parameter estimates: simulation averages for the posterior means, simulation standard errors and simulation averages of the estimated standard errors for the prevalence, sensitivities and specificities obtained by fitting Models 1 and 2 (data are generated under Model 1).

| Parameters | True Values | Model 1 | Model 2 |
|---|---|---|---|
| Prev. | 0.0850 | 0.0912(0.0193)(0.0205) | 0.0912(0.0193)(0.0205) |
| Sens. $T_1$ | 0.1764 | 0.1797(0.0925)(0.0849) | 0.1880(0.0922)(0.0869) |
| Sens. $T_2$ | 0.4346 | 0.4388(0.1244)(0.1136) | 0.4450(0.1211)(0.1127) |
| Sens. $T_3$ | 0.1762 | 0.1821(0.0896)(0.0861) | 0.1903(0.0892)(0.0881) |
| Sens. $T_4$ | 0.7521 | 0.7318(0.1063)(0.1026) | 0.7318(0.1065)(0.1025) |
| Sens. $T_5$ | 0.9884 | 0.9625(0.0189)(0.0436) | 0.9626(0.0191)(0.0435) |
| Sens. $T_6$ | 0.9849 | 0.9439(0.0354)(0.0596) | 0.9441(0.0353)(0.0593) |
| Spec. $T_1$ | 0.9521 | 0.9497(0.0156)(0.0158) | 0.9495(0.0156)(0.0159) |
| Spec. $T_2$ | 0.5563 | 0.5549(0.0355)(0.0365) | 0.5547(0.0354)(0.0366) |
| Spec. $T_3$ | 0.8063 | 0.8069(0.0271)(0.0289) | 0.8065(0.0270)(0.0290) |
| Spec. $T_4$ | 0.9522 | 0.9521(0.0145)(0.0155) | 0.9521(0.0145)(0.0155) |
| Spec. $T_5$ | 0.9747 | 0.9747(0.0116)(0.0118) | 0.9747(0.0116)(0.0118) |
| Spec. $T_6$ | 0.9983 | 0.9949(0.0024)(0.0051) | 0.9949(0.0024)(0.0051) |

The results in Table 6.1 indicate that there are only slight differences between the parameter estimates obtained by fitting Models 1 and 2. The results obtained by fitting Model 1 are close to the true values. Model 2 overestimates the $\alpha$ parameters for the first three tests. The dependency parameters ($\beta_{10}$ and $\beta_{11}$) are also overestimated by Model 2. From the results given in Table 6.1, we also observe a difference between the estimated and empirical standard errors. For the parameters with high $\alpha$ values, the simulation averages of the estimated standard errors are higher than the simulation standard errors.

The estimates for prevalence, sensitivities and specificities (Table 6.2) indicate that the estimates obtained by fitting Model 1 are quite close to the true values. When comparing the results of Model 2 with the results of Model 1, Model 2 slightly overestimates the sensitivities for the first three tests. As it was the case in Table 6.1, there is a difference between the estimated and empirical standard errors for the parameters with a high $\alpha$ values.

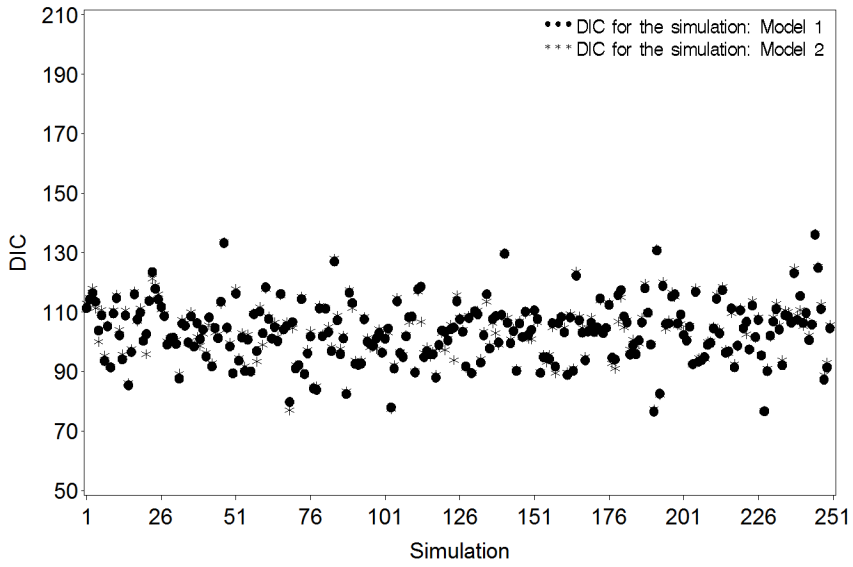



**Figure 6.1:** DIC-values (dots and stars) for the simulation runs (data are generated under Model 1).

Figure 6.1 shows the DIC-values for the simulation runs. As can be seen in Figure 6.1, the DIC-values for the simulation runs are quite close. The percentage of the DIC-values for Model 1 smaller than the DIC-values for Model 2 is 69%.

## 6.2 Data Generated Using the Conditional Dependence Model

In this section, using the multinomial probabilities (6.2), we generate datasets under the conditional dependence model (Model 2). Afterwards, the datasets are analyzed using the conditional independence and conditional dependence models. The estimates of the conditional dependence model for the Medac tests in Table 5.4 were used as true values. Model 2 states that

$$\begin{aligned} &P(Y_{i1} = y_{i1}, \ldots, Y_{iJ} = y_{iJ}) \\ &= P(D_i = 0) \int_{-\infty}^{\infty} P(Y_{i1} = y_{i1}, \ldots, Y_{iJ} = y_{iJ} | D_i = 0, Z_i = z_{1i}) f(z_{1i}) dz_{1i} \\ &\quad + P(D_i = 1) \int_{-\infty}^{\infty} P(Y_{i1} = y_{i1}, \ldots, Y_{iJ} = y_{iJ} | D_i = 1, Z_i = z_{2i}) f(z_{2i}) dz_{2i}. \end{aligned} \tag{6.2}$$

The values of the joint probabilities are included in Table B.1 (Appendix B). To generate data from the conditional dependence model, we consider different degrees of dependency of the tests (i.e. weak dependence, moderate dependence and high dependence), using increasing values for the parameters $\beta_{10}$ and $\beta_{11}$. The dependency only applies on the first three tests, in the diseased and non-diseased subjects. The last three tests remain independent.

### 6.2.1 Scenario 2A: Data are generated using the conditional dependence model with a weak dependency

In this scenario, we generate 250 datasets using the conditional dependence model with N=201 and a weak degree of dependency of the tests with $\beta_{10}$=0.5 and $\beta_{11}$=0.5. The simulation averages for the posterior means, simulation standard errors and simulation averages of the estimated standard errors obtained by fitting Models 1 and 2 (Table 6.3) indicate that there are slight differences between the parameter estimates obtained by fitting Models 1 and 2. As there is a very weak dependency between the first three tests, the results of Model 1 are more closer to the true values than the results of Model 2. Model 2 over estimates the $\alpha$ parameters for the first three tests and the dependence parameters. As we have observed in the simulation study discussed in Section 6.1, there is also a difference between the estimated and empirical standard errors. For the $\alpha_{51}, \alpha_{61}$ and $\alpha_{60}$ parameters, the simulation averages of the estimated standard errors are higher than the simulation standard errors.

**Table 6.3:** Parameter estimates: simulation averages for the posterior means, simulation standard errors and simulation averages of the estimated standard errors for the parameters obtained by fitting Models 1 and 2 (data are generated under Model 2 with a weak degree of dependency).

| Parameters | True Values | Model 1 | Model 2 |
|---|---|---|---|
| $\alpha_{11}$ | -2.241 | -2.4652(1.0861)(0.9421) | -2.7216(1.1098)(1.0386) |
| $\alpha_{21}$ | -0.229 | -0.1436(0.5672)(0.5247) | -0.1248(0.6648)(0.6424) |
| $\alpha_{31}$ | -2.269 | -2.5075(1.0753)(0.9594) | -2.7744(1.1178)(1.0538) |
| $\alpha_{41}$ | 1.108 | 1.2232(0.6827)(0.6514) | 1.2255(0.6846)(0.6514) |
| $\alpha_{51}$ | 4.437 | 4.1935(0.5825)(1.6198) | 4.1951(0.5827)(1.6180) |
| $\alpha_{61}$ | 4.180 | 3.7867(0.6544)(1.6311) | 3.7900(0.6489)(1.6319) |
| $\alpha_{10}$ | -4.049 | -4.1699(0.7138)(0.6199) | -4.3763(0.7429)(0.6521) |
| $\alpha_{20}$ | -0.343 | -0.3317(0.1487)(0.1509) | -0.3619(0.1666)(0.1681) |
| $\alpha_{30}$ | -2.077 | -2.0350(0.2370)(0.2329) | -2.1933(0.2825)(0.2809) |
| $\alpha_{40}$ | -2.993 | -3.0962(0.3762)(0.3713) | -3.0958(0.3759)(0.3712) |
| $\alpha_{50}$ | -3.644 | -3.8980(0.6472)(0.6461) | -3.8950(0.6473)(0.6436) |
| $\alpha_{60}$ | -6.355 | -6.0031(0.5120)(1.3282) | -6.0005(0.5138)(1.3241) |
| $\beta_{10}$ | 0.500 | | 0.5689(0.2409)(0.3141) |
| $\beta_{11}$ | 0.500 | | 0.9523(0.3337)(0.6182) |

**Table 6.4:** Parameter estimates: simulation averages for the posterior means, simulation standard errors and simulation averages of the estimated standard errors for the prevalence, sensitivities and specificities obtained by fitting Models 1 and 2 (data are generated under Model 2 with a weak degree of dependency).

| Parameters | True Values | Model 1 | Model 2 |
|---|---|---|---|
| Prev. | 0.0840 | 0.0914(0.0185)(0.0206) | 0.0913(0.0185)(0.0206) |
| Sens. $T_1$ | 0.1126 | 0.1290(0.0805)(0.0722) | 0.1382(0.0823)(0.0754) |
| Sens. $T_2$ | 0.4625 | 0.4694(0.1213)(0.1142) | 0.4767(0.1181)(0.1133) |
| Sens. $T_3$ | 0.1099 | 0.1251(0.0770)(0.0715) | 0.1339(0.0782)(0.0747) |
| Sens. $T_4$ | 0.7518 | 0.7373(0.1013)(0.1017) | 0.7376(0.1013)(0.1016) |
| Sens. $T_5$ | 0.9883 | 0.9618(0.0223)(0.0437) | 0.9619(0.0224)(0.0436) |
| Sens. $T_6$ | 0.9849 | 0.9445(0.0313)(0.0598) | 0.9448(0.0309)(0.0596) |
| Spec. $T_1$ | 0.9810 | 0.9792(0.0101)(0.0100) | 0.9789(0.0102)(0.0102) |
| Spec. $T_2$ | 0.5823 | 0.5813(0.0357)(0.0363) | 0.5808(0.0357)(0.0365) |
| Spec. $T_3$ | 0.8805 | 0.8802(0.0236)(0.0237) | 0.8796(0.0237)(0.0239) |
| Spec. $T_4$ | 0.9523 | 0.9517(0.0152)(0.0156) | 0.9517(0.0152)(0.0156) |
| Spec. $T_5$ | 0.9745 | 0.9735(0.0120)(0.0121) | 0.9735(0.0120)(0.0121) |
| Spec. $T_6$ | 0.9983 | 0.9950(0.0022)(0.0051) | 0.9950(0.0023)(0.0051) |

As it is a very weak degree of dependency, the estimates for prevalence, sensitivities and specificities (Table 6.4) indicate that the estimates obtained by fitting Model 1 are closer to the true values than Model 2. Model 2 slightly overestimates the sensitivities for the first three tests.

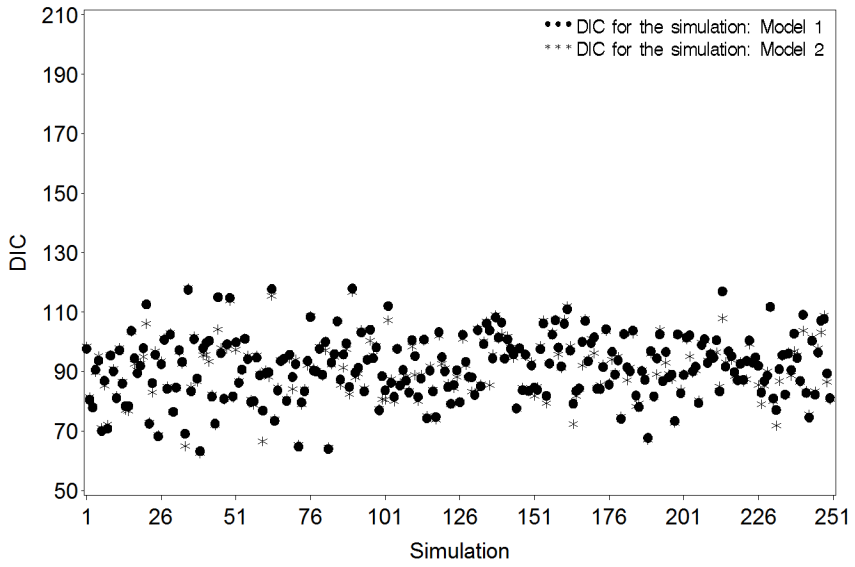



**Figure 6.2:** DIC-values (dots and stars) for the simulation runs (data are generated under Model 2 with a weak degree of dependency).

Figure 6.2 shows the DIC-values for the simulation runs. Alike the simulation study with conditionally independence assumption, the DIC-values for both models are very close. The percentage of the DIC-values for Model 2 smaller than the DIC-values for Model 1 is 52%.

### 6.2.2 Scenario 2B: Data are generated using the conditional dependence model with a moderate dependency

In this scenario, we generate 250 datasets using the conditional dependence model with N=201 and a moderate degree of dependency obtained by fitting Model 2 for the Medac tests $\beta_{10}$=1.685 and $\beta_{11}$=2.098 (Table 5.4). Afterwards, the datasets are analyzed using the conditional independence and conditional dependence models.

The parameter estimates obtained by fitting Models 1 and 2 are given in Table 6.5. From the results (Table 6.5), there are differences between the parameter estimates obtained by fitting Models 1 and 2. The parameter estimates of Model 2 are now much closer to the true values than those of Model 1. The parameter $\alpha_{21}$ was over-estimated by both models. As we have seen in the previous simulation scenarios, the simulation averages of the estimated standard errors are higher than the simulation standard errors for the $\alpha_{51}, \alpha_{61}$ and $\alpha_{60}$ parameters. Model 2 estimates the dependency parameters very well.

**Table 6.5:** Parameter estimates: simulation averages for the posterior means, simulation standard errors and simulation averages of the estimated standard errors for the parameters obtained by fitting Models 1 and 2 (data are generated under Model 2 with a moderate degree of dependency).

| Parameters | True Values | Model 1 | Model 2 |
|---|---|---|---|
| $\alpha_{11}$ | -2.241 | -1.2191(0.7344)(0.6407) | -1.7347(0.8993)(0.9660) |
| $\alpha_{21}$ | -0.229 | 0.1205(0.5896)(0.5392) | 0.3069(0.8730)(0.8613) |
| $\alpha_{31}$ | -2.269 | -1.2421(0.7710)(0.6509) | -1.7891(0.9775)(0.9806) |
| $\alpha_{41}$ | 1.108 | 1.2281(0.7601)(0.6754) | 1.2282(0.7588)(0.6727) |
| $\alpha_{51}$ | 4.437 | 4.1516(0.5699)(1.6153) | 4.1595(0.5668)(1.6149) |
| $\alpha_{61}$ | 4.180 | 3.7383(0.7383)(1.6191) | 3.7528(0.7280)(1.6184) |
| $\alpha_{10}$ | -4.049 | -3.1247(0.4058)(0.3753) | -4.1581(0.5391)(0.5346) |
| $\alpha_{20}$ | -0.343 | -0.2351(0.1473)(0.1495) | -0.3343(0.2190)(0.2233) |
| $\alpha_{30}$ | -2.077 | -1.4523(0.1897)(0.1894) | -2.0665(0.3075)(0.3147) |
| $\alpha_{40}$ | -2.993 | -3.0925(0.3649)(0.3695) | -3.0923(0.3647)(0.3697) |
| $\alpha_{50}$ | -3.644 | -3.9337(0.6636)(0.6458) | -3.9332(0.6684)(0.6442) |
| $\alpha_{60}$ | -6.355 | -5.9993(0.5552)(1.3180) | -6.0013(0.5521)(1.3168) |
| $\beta_{10}$ | 1.685 | | 1.6202(0.2973)(0.3234) |
| $\beta_{11}$ | 2.098 | | 1.9908(0.5777)(0.7631) |

The simulation averages for the posterior means, simulation standard errors and simulation averages of the estimated standard errors for the prevalence, sensitivities and specificities obtained by fitting Models 1 and 2 are given in Table 6.6. Although the estimates for the $\alpha$ parameters differ considerably, the estimates in Table 6.6 indicate that the estimates for sensitivities and specificities do not differ much.

**Table 6.6:** Parameter estimates: simulation averages for the posterior means, simulation standard errors and simulation averages of the estimated standard errors for the prevalence, sensitivities and specificities obtained by fitting Models 1 and 2 (data are generated under Model 2 with a moderate degree of dependency).

| Parameters | True Values | Model 1 | Model 2 |
|---|---|---|---|
| Prev. | 0.0840 | 0.0893(0.0210)(0.0203) | 0.0892(0.0211)(0.0203) |
| Sens. $T_1$ | 0.2495 | 0.2646(0.1071)(0.1010) | 0.2774(0.1037)(0.1009) |
| Sens. $T_2$ | 0.5084 | 0.5269(0.1267)(0.1160) | 0.5372(0.1187)(0.1127) |
| Sens. $T_3$ | 0.2465 | 0.2629(0.1118)(0.1007) | 0.2736(0.1074)(0.1003) |
| Sens. $T_4$ | 0.7518 | 0.7338(0.1118)(0.1035) | 0.7340(0.1113)(0.1033) |
| Sens. $T_5$ | 0.9883 | 0.9607(0.0219)(0.0453) | 0.9610(0.0218)(0.0449) |
| Sens. $T_6$ | 0.9849 | 0.9404(0.0404)(0.0619) | 0.9414(0.0398)(0.0610) |
| Spec. $T_1$ | 0.9543 | 0.9526(0.0158)(0.0154) | 0.9510(0.0159)(0.0159) |
| Spec. $T_2$ | 0.5585 | 0.5579(0.0360)(0.0365) | 0.5559(0.0357)(0.0373) |
| Spec. $T_3$ | 0.8088 | 0.8070(0.0283)(0.0289) | 0.8046(0.0283)(0.0297) |
| Spec. $T_4$ | 0.9523 | 0.9517(0.0151)(0.0156) | 0.9517(0.0151)(0.0156) |
| Spec. $T_5$ | 0.9745 | 0.9743(0.0117)(0.0119) | 0.9742(0.0118)(0.0118) |
| Spec. $T_6$ | 0.9983 | 0.9949(0.0026)(0.0051) | 0.9949(0.0026)(0.0051) |

Both models overestimate the sensitivities for the first three tests, and they underestimate the sensitivities for the last three tests. The estimates for the specificities obtained by fitting Models 1 and 2 are very close to each other, and also to the true values.

Figure 6.3 shows the the DIC-values for the simulation runs. As the data was generated under the conditional dependence model with a moderate degree of dependency, all the DIC-values for Model 2 are better than the DIC-values for Model 1.

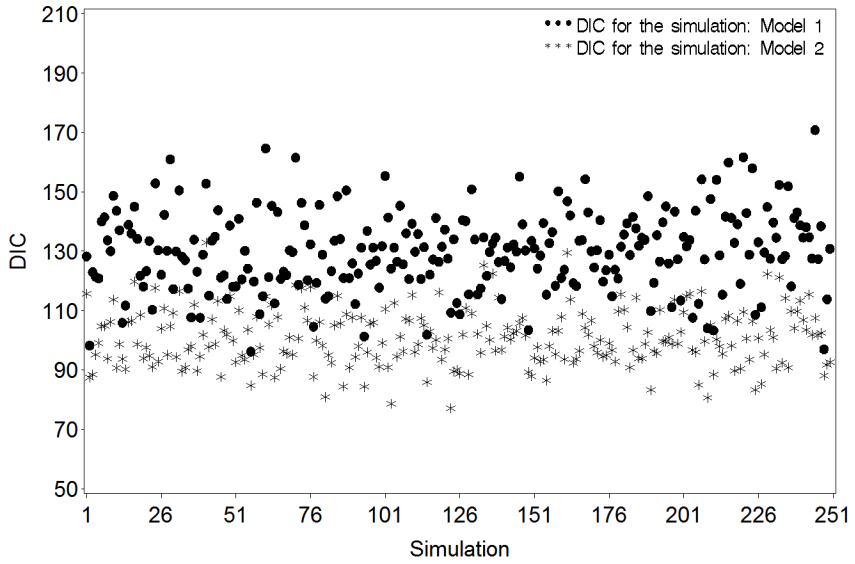



**Figure 6.3:** DIC-values (dots and stars) for the simulation runs (data are generated under Model 2 with a moderate degree of dependency).

Comparing the estimates for the sensitivities and specificities given in Tables 6.4 and 6.6, as the degree of dependency increases the true values and the estimates of sensitivities for the first three tests increase slightly, while their specificities decrease.

### 6.2.3 Scenario 2C: Data are generated using the conditional dependence model with a high dependency

In this scenario, we generate 250 datasets using the conditional dependence model with N=201 and a high degree of dependency of the tests with $\beta_{10}$=6 and $\beta_{11}$=6. From the results obtained by fitting Models 1 and 2 in Table 6.7, the estimates obtained by Model 1 are underestimating the true values, especially $\alpha_{11}$, $\alpha_{31}$, $\alpha_{10}$ and $\alpha_{30}$. The parameter $\alpha_{21}$ is still overestimated by both models. Model 2 severely

underestimates the dependency parameters. Again there is also a difference between the estimated and empirical standard errors.

**Table 6.7:** Parameter estimates: simulation averages for the posterior means, simulation standard errors and simulation averages of the estimated standard errors for the parameters obtained by fitting Models 1 and 2 (data are generated under Model 2 with a high degree of dependency).

| Parameters | True Values | Model 1 | Model 2 |
|---|---|---|---|
| $\alpha_{11}$ | -2.241 | -0.2336(0.8137)(0.5410) | -0.6274(1.0558)(1.1140) |
| $\alpha_{21}$ | -0.229 | 0.5847(1.3703)(0.6735) | 0.6066(1.1193)(1.1157) |
| $\alpha_{31}$ | -2.269 | -0.0574(1.2028)(0.6214) | -0.6696(1.0599)(1.1157) |
| $\alpha_{41}$ | 1.108 | 0.9829(1.1775)(0.7167) | 1.2792(0.7539)(0.6803) |
| $\alpha_{51}$ | 4.437 | 3.6062(1.7700)(1.6473) | 4.1968(0.5362)(1.6218) |
| $\alpha_{61}$ | 4.180 | 3.1534(1.8236)(1.6318) | 3.7666(0.7138)(1.6244) |
| $\alpha_{10}$ | -4.049 | -1.3894(1.1657)(0.3725) | -3.2472(0.5135)(0.5852) |
| $\alpha_{20}$ | -0.343 | -0.2154(0.5381)(0.2030) | -0.1856(0.4918)(0.4784) |
| $\alpha_{30}$ | -2.077 | -0.7921(0.9395)(0.2919) | -1.6596(0.5027)(0.5107) |
| $\alpha_{40}$ | -2.993 | -3.0507(0.3551)(0.3771) | -3.0782(0.3276)(0.3661) |
| $\alpha_{50}$ | -3.644 | -3.8685(0.7722)(0.6670) | -3.9382(0.6932)(0.6538) |
| $\alpha_{60}$ | -6.355 | -5.7285(0.9092)(1.2969) | -5.9975(0.5379)(1.3183) |
| $\beta_{10}$ | 6.000 | | 4.8683(0.4101)(0.6174) |
| $\beta_{11}$ | 6.000 | | 3.0896(0.4825)(0.8276) |

**Table 6.8:** Parameter estimates: simulation averages for the posterior means, simulation standard errors and simulation averages of the estimated standard errors for the prevalence, sensitivities and specificities obtained by fitting Models 1 and 2 (data are generated under Model 2 with a high degree of dependency).

| Parameters | True Values | Model 1 | Model 2 |
|---|---|---|---|
| Prev. | 0.0840 | 0.1222(0.0858)(0.0310) | 0.0899(0.0199)(0.0204) |
| Sens. $T_1$ | 0.4281 | 0.4545(0.1524)(0.1101) | 0.4337(0.1133)(0.1114) |
| Sens. $T_2$ | 0.5377 | 0.5786(0.1776)(0.1111) | 0.5609(0.1196)(0.1109) |
| Sens. $T_3$ | 0.4265 | 0.4697(0.1907)(0.1133) | 0.4303(0.1138)(0.1112) |
| Sens. $T_4$ | 0.7518 | 0.6868(0.2006)(0.1087) | 0.7429(0.1087)(0.1017) |
| Sens. $T_5$ | 0.9883 | 0.8807(0.2287)(0.0678) | 0.9625(0.0192)(0.0438) |
| Sens. $T_6$ | 0.9849 | 0.8581(0.2350)(0.0820) | 0.9423(0.0373)(0.0605) |
| Spec. $T_1$ | 0.7300 | 0.7512(0.0873)(0.0388) | 0.7333(0.0325)(0.0341) |
| Spec. $T_2$ | 0.5261 | 0.5499(0.0983)(0.0444) | 0.5142(0.0380)(0.0384) |
| Spec. $T_3$ | 0.6288 | 0.6566(0.1085)(0.0439) | 0.6250(0.0370)(0.0374) |
| Spec. $T_4$ | 0.9523 | 0.9497(0.0159)(0.0166) | 0.9515(0.0140)(0.0156) |
| Spec. $T_5$ | 0.9745 | 0.9706(0.0177)(0.0134) | 0.9739(0.0125)(0.0119) |
| Spec. $T_6$ | 0.9983 | 0.9910(0.0119)(0.0073) | 0.9949(0.0025)(0.0051) |

From the results in Table 6.8, the parameter estimates obtained by fitting Model 2

are quite close to the true values. Simulation standard errors for Model 2 are much smaller than the simulation standard errors for Model 1. The prevalence is overestimated by Model 1. The sensitivity for the second test is overestimated by both models. The DIC-values for the simulation runs are shown in Figure B.1 (in Appendix B). Alike the simulation study with a moderate degree of dependency of the tests, all the DIC values for Model 2 are smaller than the DIC values for Model 1.

When comparing the sensitivities and specificities for this scenario with the results of simulation studies discussed in Sections 6.2.1 and 6.2.2, the sensitivities for the first three tests slightly increase and their specificities decreased a little bit.

## 6.3 Scenario 3A: Data are generated using the conditional independence model

In this scenario, we have conducted the simulation study discussed in Section 6.1 using other true values, where the true values for the parameters of the dependent tests are more in line in sign and magnitude. Only the first three tests will become dependent in the next sections.

**Table 6.9:** Parameter estimates: simulation averages for the posterior means, simulation standard errors and simulation averages of the estimated standard errors for the parameters obtained by fitting Models 1 and 2 (data are generated under Model 1 using different true values).

| Parameters | True Values | Model 1 | Model 2 |
|---|---|---|---|
| $\alpha_{11}$ | 1.831 | 1.9485(0.5192)(0.5026) | 2.0826(0.5503)(0.5510) |
| $\alpha_{21}$ | 1.812 | 1.9515(0.5249)(0.4868) | 2.0827(0.5651)(0.5359) |
| $\alpha_{31}$ | 1.355 | 1.4464(0.4268)(0.3879) | 1.5493(0.4688)(0.4372) |
| $\alpha_{41}$ | 1.308 | 1.3954(0.4061)(0.3867) | 1.4119(0.4175)(0.3963) |
| $\alpha_{51}$ | 2.196 | 2.4641(0.6985)(0.6556) | 2.5360(0.7227)(0.7071) |
| $\alpha_{61}$ | 1.742 | 1.8899(0.5710)(0.4749) | 1.9078(0.5814)(0.4885) |
| $\alpha_{10}$ | -2.425 | -2.5049(0.3620)(0.3391) | -2.6131(0.3765)(0.3639) |
| $\alpha_{20}$ | -1.690 | -1.7122(0.2392)(0.2392) | -1.8033(0.2668)(0.2657) |
| $\alpha_{30}$ | -1.279 | -1.2944(0.2107)(0.2061) | -1.3662(0.2286)(0.2299) |
| $\alpha_{40}$ | -2.057 | -2.0559(0.2567)(0.2718) | -2.0668(0.2603)(0.2760) |
| $\alpha_{50}$ | -2.178 | -2.2197(0.3052)(0.2972) | -2.2422(0.3144)(0.3080) |
| $\alpha_{60}$ | -1.435 | -1.4406(0.2111)(0.2166) | -1.4465(0.2144)(0.2189) |
| $\beta_{10}$ | 0 | | 0.4896(0.1894)(0.2978) |
| $\beta_{11}$ | 0 | | 0.6924(0.2664)(0.4472) |

From the parameter estimates obtained by fitting Models 1 and 2 given in Ta-

ble 6.9, the parameter estimates obtained by fitting Model 1 are close to the true values, while Model 2 overestimates the parameters especially the dependence parameters. Unlike the simulation studies discussed in the previous sections, the simulation averages of the estimated standard errors are very close to the simulation standard errors.

**Table 6.10:** Parameter estimates: simulation averages for the posterior means, simulation standard errors and simulation averages of the estimated standard errors for the prevalence, sensitivities and specificities obtained by fitting Models 1 and 2 (data are generated under Model 1 using different true values).

| Parameters | True Values | Model 1 | Model 2 |
|---|---|---|---|
| Prev. | 0.2620 | 0.2631(0.0308)(0.0325) | 0.2635(0.0310)(0.0329) |
| Sens. $T_1$ | 0.8619 | 0.8566(0.0499)(0.0517) | 0.8478(0.0507)(0.0544) |
| Sens. $T_2$ | 0.8596 | 0.8571(0.0529)(0.0504) | 0.8477(0.0541)(0.0531) |
| Sens. $T_3$ | 0.7949 | 0.7948(0.0603)(0.0573) | 0.7846(0.0609)(0.0594) |
| Sens. $T_4$ | 0.7872 | 0.7875(0.0597)(0.0587) | 0.7894(0.0603)(0.0592) |
| Sens. $T_5$ | 0.8999 | 0.8988(0.0430)(0.0448) | 0.9025(0.0430)(0.0451) |
| Sens. $T_6$ | 0.8509 | 0.8492(0.0529)(0.0510) | 0.8506(0.0528)(0.0514) |
| Spec. $T_1$ | 0.9187 | 0.9177(0.0238)(0.0234) | 0.9153(0.0241)(0.0240) |
| Spec. $T_2$ | 0.8442 | 0.8421(0.0304)(0.0306) | 0.8394(0.0304)(0.0311) |
| Spec. $T_3$ | 0.7823 | 0.7808(0.0344)(0.0344) | 0.7778(0.0344)(0.0348) |
| Spec. $T_4$ | 0.8867 | 0.8813(0.0253)(0.0272) | 0.8822(0.0254)(0.0273) |
| Spec. $T_5$ | 0.8983 | 0.8958(0.0260)(0.0260) | 0.8975(0.0262)(0.0263) |
| Spec. $T_6$ | 0.8077 | 0.8043(0.0321)(0.0332) | 0.8051(0.0324)(0.0334) |

From the estimates for the prevalence, sensitivities and specificities in Table 6.10, the estimates obtained by fitting Model 1 are quite close to the true values. Model 2 underestimates the sensitivities for the first three tests slightly. The DIC-values for the simulation runs are shown in Figure B.2 (in Appendix B). The percentage of the DIC-values for Model 1 smaller than the DIC-values for Model 2 is 81%.

## 6.4 Data Generated Using the Conditional Dependence Model

In this section, we have conducted the simulation studies discussed in Section 6.2 using different true values, where the true values for the parameters of the dependent tests are more inline in sign and magnitude. In Section 6.4.1, we generate datasets under the conditional dependence model with a moderate degree of dependency and we analyze them using conditional independence and conditional dependence models.

In Section 6.4.2, the degree of dependency of the tests is increased further.

### 6.4.1 Scenario 3B: Data are generated using the conditional dependence model with a moderate dependency

In this scenario, we have conducted the simulation study discussed in Section 6.2.2 using different true values. The simulation averages for the posterior means, simulation standard errors and simulation averages of the estimated standard errors obtained by fitting Models 1 and 2 are given in Tables 6.11 and 6.12.

**Table 6.11:** Parameter estimates: simulation averages for the posterior means, simulation standard errors and simulation averages of the estimated standard errors for the parameters obtained by fitting Models 1 and 2 (data are generated under Model 2 with a moderate degree of dependency and using different true values).

| Parameters | True Values | Model 1 | Model 2 |
|---|---|---|---|
| $\alpha_{11}$ | 1.831 | 1.8225(0.6657)(0.5749) | 2.0077(0.6108)(0.6696) |
| $\alpha_{21}$ | 1.812 | 1.9045(0.6298)(0.5692) | 2.0327(0.6000)(0.6501) |
| $\alpha_{31}$ | 1.355 | 1.4538(0.4569)(0.4243) | 1.4608(0.4827)(0.5740) |
| $\alpha_{41}$ | 1.308 | 0.9405(0.5113)(0.3987) | 1.4305(0.5295)(0.4985) |
| $\alpha_{51}$ | 2.196 | 1.5843(0.7981)(0.6080) | 2.6435(0.8904)(0.9848) |
| $\alpha_{61}$ | 1.742 | 1.3685(0.6119)(0.4817) | 1.9146(0.6266)(0.6148) |
| $\alpha_{10}$ | -2.425 | -2.0400(0.5087)(0.3876) | -2.4429(0.4314)(0.3992) |
| $\alpha_{20}$ | -1.690 | -1.3453(0.3078)(0.2578) | -1.6483(0.3231)(0.3365) |
| $\alpha_{30}$ | -1.279 | -1.0645(0.2809)(0.2260) | -1.2721(0.3245)(0.3129) |
| $\alpha_{40}$ | -2.057 | -1.9942(0.3098)(0.2976) | -2.1157(0.3277)(0.3188) |
| $\alpha_{50}$ | -2.178 | -2.0865(0.3765)(0.3229) | -2.2962(0.4338)(0.3992) |
| $\alpha_{60}$ | -1.435 | -1.4094(0.2452)(0.2327) | -1.4709(0.2564)(0.2388) |
| $\beta_{10}$ | 2.000 | | 1.9548(0.3277)(0.3477) |
| $\beta_{11}$ | 2.000 | | 2.0162(0.5680)(0.6074) |

The results in Table 6.11 indicate that there is a clear difference between the parameter estimates. The results for Model 2 are more close to the true value. Mode 2 estimates the dependence parameters very well, but it slightly overestimates the parameters $\alpha_{11}$, $\alpha_{21}$, $\alpha_{31}$ and $\alpha_{51}$. As there is a moderate degree of dependency between the tests, the estimates for the prevalence, sensitivities and specificities obtained by the conditional dependence model are more close to the true value (Table 6.12). Model 1 overestimates the prevalence, the sensitivities and specificities for the first three tests. It underestimates the sensitivities for the last three tests. The DIC-values for the simulation runs are shown in Figure B.3 (in Appendix B). As expected, all the DIC-values for Model 2 are smaller than the DIC-values for Model 1.

**Table 6.12:** Parameter estimates: simulation averages for the posterior means, simulation standard errors and simulation averages of the estimated standard errors for the prevalence, sensitivities and specificities obtained by fitting Models 1 and 2 (data are generated under Model 2 with a moderate degree of dependency and using different true values).

| Parameters | True Values | Model 1 | Model 2 |
|---|---|---|---|
| Prev. | 0.2620 | 0.2919(0.0471)(0.0417) | 0.2658(0.0373)(0.0382) |
| Sens. $T_1$ | 0.7720 | 0.8341(0.0607)(0.0599) | 0.7639(0.0686)(0.0670) |
| Sens. $T_2$ | 0.7700 | 0.8453(0.0601)(0.0563) | 0.7673(0.0646)(0.0657) |
| Sens. $T_3$ | 0.7178 | 0.7935(0.0651)(0.0601) | 0.7047(0.0666)(0.0690) |
| Sens. $T_4$ | 0.7872 | 0.7018(0.0930)(0.0729) | 0.7850(0.0690)(0.0675) |
| Sens. $T_5$ | 0.8999 | 0.7918(0.0917)(0.0707) | 0.8962(0.0548)(0.0552) |
| Sens. $T_6$ | 0.8509 | 0.7714(0.0847)(0.0684) | 0.8454(0.0570)(0.0596) |
| Spec. $T_1$ | 0.8231 | 0.8708(0.0450)(0.0348) | 0.8194(0.0359)(0.0348) |
| Spec. $T_2$ | 0.7400 | 0.7858(0.0470)(0.0402) | 0.7334(0.0377)(0.0391) |
| Spec. $T_3$ | 0.6867 | 0.7376(0.0513)(0.0415) | 0.6847(0.0408)(0.0405) |
| Spec. $T_4$ | 0.8867 | 0.8731(0.0316)(0.0306) | 0.8849(0.0301)(0.0293) |
| Spec. $T_5$ | 0.8983 | 0.8809(0.0333)(0.0305) | 0.8983(0.0307)(0.0300) |
| Spec. $T_6$ | 0.8077 | 0.7983(0.0379)(0.0360) | 0.8075(0.0377)(0.0354) |

### 6.4.2 Scenario 3C: Data are generated using the conditional dependence model with a high dependency

In this scenario, we have conducted the simulation study discussed in Section 6.2.3 using different true values. The simulation results are given in Tables 6.13 and 6.14.

**Table 6.13:** Parameter estimates: simulation averages for the posterior means, simulation standard errors and simulation averages of the estimated standard errors for the parameters obtained by fitting Models 1 and 2 (data are generated under Model 2 with a high degree of dependency and using different true values).

| Parameters | True Values | Model 1 | Model 2 |
|---|---|---|---|
| $\alpha_{11}$ | 1.831 | 2.5573(0.6517)(0.5968) | 1.5909(0.8825)(0.8956) |
| $\alpha_{21}$ | 1.812 | 3.1061(0.7259)(0.7270) | 1.6031(0.8547)(0.8878) |
| $\alpha_{31}$ | 1.355 | 3.1856(0.7457)(0.7235) | 1.2572(0.8143)(0.8627) |
| $\alpha_{41}$ | 1.308 | -0.5204(0.2510)(0.2297) | 1.5939(0.7182)(0.7645) |
| $\alpha_{51}$ | 2.196 | -0.3678(0.2689)(0.2300) | 2.9611(0.9181)(1.3390) |
| $\alpha_{61}$ | 1.742 | -0.2057(0.2423)(0.2255) | 2.1987(0.8224)(0.9268) |
| $\alpha_{10}$ | -2.425 | -3.5158(0.7052)(0.7801) | -1.9292(0.5342)(0.5966) |
| $\alpha_{20}$ | -1.690 | -2.8359(0.5808)(0.5650) | -1.3270(0.5505)(0.5760) |
| $\alpha_{30}$ | -1.279 | -2.4911(0.4698)(0.4487) | -0.9537(0.5525)(0.5652) |
| $\alpha_{40}$ | -2.057 | -1.2690(0.2532)(0.2383) | -2.0399(0.3409)(0.3590) |
| $\alpha_{50}$ | -2.178 | -1.2031(0.2480)(0.2345) | -2.2559(0.5576)(0.5237) |
| $\alpha_{60}$ | -1.435 | -0.8676(0.2120)(0.2153) | -1.4372(0.2713)(0.2703) |
| $\beta_{10}$ | 6.000 | | 4.8900(0.4348)(0.6496) |
| $\beta_{11}$ | 6.000 | | 3.9055(0.5084)(0.7779) |

From the results in Table 6.13, there is now a clear difference between the parameter estimates. The parameter estimates obtained by fitting Model 2 are much closer to the true values than Model 1 does. Model 1 estimates all parameters poorly. Model 2 slightly underestimates the $\alpha$ parameters for the first three tests. Model 2 also underestimates the dependence parameters $\beta_{10}$ and $\beta_{11}$.

**Table 6.14:** Parameter estimates: simulation averages for the posterior means, simulation standard errors and simulation averages of the estimated standard errors for the prevalence, sensitivities and specificities obtained by fitting Models 1 and 2 (data are generated under Model 2 with a high degree of dependency and using different true values).

| Parameters | True Values | Model 1 | Model 2 |
|---|---|---|---|
| Prev. | 0.2620 | 0.4479(0.0365)(0.0373) | 0.2569(0.0450)(0.0453) |
| Sens. $T_1$ | 0.6452 | 0.9100(0.0369)(0.0349) | 0.6431(0.0792)(0.0755) |
| Sens. $T_2$ | 0.6442 | 0.9405(0.0311)(0.0285) | 0.6445(0.0774)(0.0750) |
| Sens. $T_3$ | 0.6202 | 0.9438(0.0319)(0.0276) | 0.6155(0.0756)(0.0756) |
| Sens. $T_4$ | 0.7872 | 0.3758(0.0563)(0.0522) | 0.7914(0.0794)(0.0796) |
| Sens. $T_5$ | 0.8999 | 0.4105(0.0578)(0.0531) | 0.9046(0.0570)(0.0645) |
| Sens. $T_6$ | 0.8509 | 0.4494(0.0554)(0.0537) | 0.8577(0.0682)(0.0669) |
| Spec. $T_1$ | 0.6477 | 0.9585(0.0222)(0.0217) | 0.6440(0.0392)(0.0426) |
| Spec. $T_2$ | 0.6071 | 0.9313(0.0301)(0.0277) | 0.6005(0.0416)(0.0434) |
| Spec. $T_3$ | 0.5831 | 0.9121(0.0318)(0.0303) | 0.5729(0.0424)(0.0436) |
| Spec. $T_4$ | 0.8867 | 0.7749(0.0430)(0.0401) | 0.8762(0.0302)(0.0327) |
| Spec. $T_5$ | 0.8983 | 0.7636(0.0433)(0.0409) | 0.8886(0.0392)(0.0357) |
| Spec. $T_6$ | 0.8077 | 0.7004(0.0432)(0.0440) | 0.8012(0.0384)(0.0388) |

The estimates for prevalence, sensitivities and specificities in Table 6.14 indicate that the estimates obtained by fitting Model 2 are quite close to the true values. Model 1 poorly estimates the prevalence, sensitivities and specificities of the tests. Here, we can clearly see that for the data generated under the conditional dependence model, the conditional dependence model is much better than the conditional independence model.

Figure B.4 (in Appendix B) shows the DIC-values for the simulation runs. As it was a case for the simulation study with a moderate degree of dependency (Section 6.4.1), all the DIC-values for Model 2 are better than the DIC-values for Model 1.

Again we observe that the sensitivities and specificities of the dependent tests seem to tend to 0.5 as the level of dependency increases. This will be investigated further in the next discussion section.

## 6.5 Discussion

We have conducted simulation studies first by using the estimates of the conditionally independent and conditionally dependent models of the Medac tests (Table 5.4) as true values for the parameters. Then, as the true values for the dependent tests are very different in magnitude and some of the sensitivities and specificities are very low, we used other true values which for the dependent tests are more in line and tests with higher sensitivities and specificities are considered. In these simulation settings, we considered different degree of dependency (i.e. weak dependency, moderate dependency and high dependency) without changing the $\alpha$ parameters.

From the results of simulation settings, the conditional independence assumption is better when the data are generated under the conditional independence model. When the data are generated under the conditional dependence model, with moderate to high dependency, the conditional dependence model outperforms the conditional independent model.

From the simulation results given in Tables 6.10, 6.12 and 6.14 we observed that the true values of the sensitivities and specificities tend to the value of 0.5 as the degree of dependency between the tests increases. In order to confirm these, we calculate the true values for the sensitivities and specificities for different values of $\beta_{11}$ and $\beta_{10}$, respectively. The plot of specificities for the first test as a function of $\beta_{10}$, for different values of $\alpha_{10}$, is given in Figure 6.4.

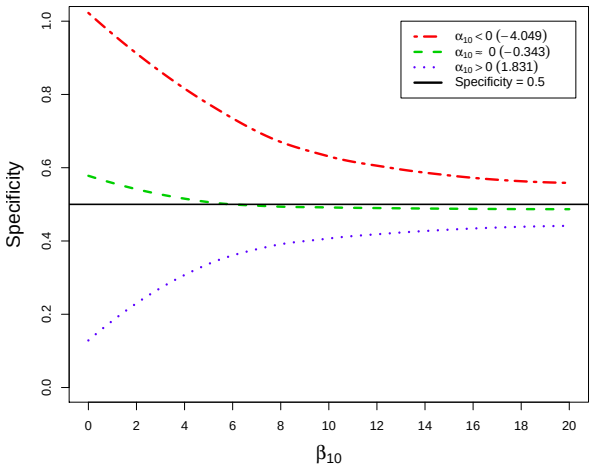



**Figure 6.4:** Smooth plot of specificities for the first test and $\beta_{10}$ for different values of $\alpha_{10}$ parameters.

We can clearly see from Figure 6.4 that, when increasing the dependency parameter; the specificity will increase to 0.5 for a test with a higher $\alpha_{10}$ value, the specificity will be around 0.5 for a test with $\alpha_{10}$ value close to 0, while the specificity will decrease to 0.5 for a test with a very small $\alpha_{10}$ value. Figure 6.5 shows the plot of sensitivities for the first test as a function of $\beta_{11}$ for different values of $\alpha_{11}$.

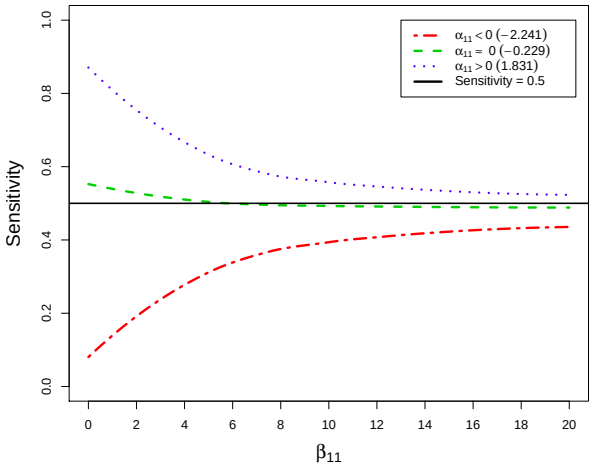



**Figure 6.5:** Smooth plot of sensitivities for the first test and $\beta_{11}$ for different values of $\alpha_{11}$ parameters.

From Figure 6.5, when increasing the dependency parameter in the non-diseased subjects: the sensitivity will increases to 0.5, it will be around 0.5 and decreases to 0.5 respectively for $\alpha_{11} < 0$, $\alpha_{11} \approx 0$ and $\alpha_{11} > 0$. As can be seen from Figures 6.4 and 6.5, the results of sensitivities and specificities depend on the degree of dependency of the tests and the $\alpha$ parameters.

We also have conducted simulation studies discussed in Sections 6.3 and 6.4 with N=500. The results are given in Appendix B. Increasing N from 201 to 500, does not have much influence on the results. The results of these simulation studies are more in line with the simulation studies discusses in Sections 6.3 and 6.4.

## 6.6 Concluding Remarks

In this chapter, we have conducted simulation studies under different scenarios to investigate the impact of misspecifying the conditional dependency of the tests. We generate data under the conditional independence and conditional dependence assumptions with different degree of dependency, and we analyze them using conditional independence and conditional dependence models.

The parameter estimates of the conditional independence model and the conditional dependence model for the case study (in Chapter 5) were used as true values. We consider simulation studies by increasing and by decreasing the dependency of the tests. We also consider these scenarios using other true values, where the true values for the parameters of the dependent tests are more inline in sign and magnitude.

From the simulation results, when the data are generated under the conditional independence model and analyzed using the conditional independence and conditional dependence models, there are slight differences between the results of these models. But when the data are generated under the conditional dependence model and analyzed using the conditional independence and conditional dependence models, there is a clear difference between these models for moderate to high dependency.

# Chapter 7

# Diagnosis of Coronary Heart Disease

As individual patient data (IPD) meta-analyses of diagnostic tests are still very rare (Broeze *et al.*, 2009; Mant *et al.*, 2009; Khan *et al.*, 2003), data analyses and model building is still confronted with several challenging statistical issues. Studies of diagnostic accuracy require more sophisticated methods for their meta-analysis than studies of therapeutic interventions (Harbord *et al.*, 2007). Methods must reflect the negative correlation between sensitivity and specificity, and, in contrast to meta-analysis of data from randomized controlled trials, substantial between-study heterogeneity is to be expected and should be incorporated into the models. Analysis is also complicated by the presence of missing covariate data on several levels: covariate values might be missing at the study level as not all variables are available for all studies; they might be missing not at random, at the general practitioner level, and at the individual patient level.

The main objectives of this study are:

(a) To perform a systematic review assessing accuracy in diagnosing coronary heart disease in patients presenting with chest pain in primary care of clinical features.

(b) To perform an individual patient data analysis to establish whether clinical prediction rules based on signs and symptoms usefully predict the presence of coronary heart disease.

The diagnostic accuracy of each sign and symptom is examined separately in a series of bivariate analyses in Haasenritter *et al.* (2012c). To explore how accuracy of each sign and symptom is affected by age and sex, these basic patient characteristics are included as covariates in meta-regression analysis in Haasenritter *et al.* (2012c). In this chapter of the dissertation, the diagnostic characteristics of the optimal combination of signs and symptoms are jointly studied. A stage-wise modelling procedure is applied. First, study-specific models are built. Afterwards, an IPD meta-analysis is used to combine the data of all five studies (Higgins and Green, 2011).

## 7.1 Study-specific Analyses

As the identified studies differ in the number of predictors (Table A.1), it is very difficult to combine the five studies and conduct variable selection in order to fit a good parsimonious prediction model. To select the most important variables in each study, we first conducted a study-specific analysis. Using the selected sets of predictors, we then conducted an IPD meta-analysis. In Section 7.1.1, we introduced the imputation mechanism that we used to impute the missing values. The variable selection method is described in Section 7.1.2. Section 7.1.3 introduced the logistic regression model that we fitted using all the selected variables and their two way interaction terms.

### 7.1.1 Imputation

The well-established method of multiple imputations is applied to resolve the missing data issue (Rubin *et al.*, 1978; Rubin *et al.*, 1987; Buuren *et al.*, 2011). For the imputation of the missing values, we used Multivariate Imputation by Chained Equations (MICE) which is available from `CRAN` as an `R` package **`mice`**(Buuren *et al.*, 2011). MICE is a flexible and general methodology for generating multiple imputations in multivariate data. Figure 7.1 provides a graphic representation of the different steps for the multiple imputation analyses: generation of imputations, repeated analyses on the imputed data, and combination of the results.

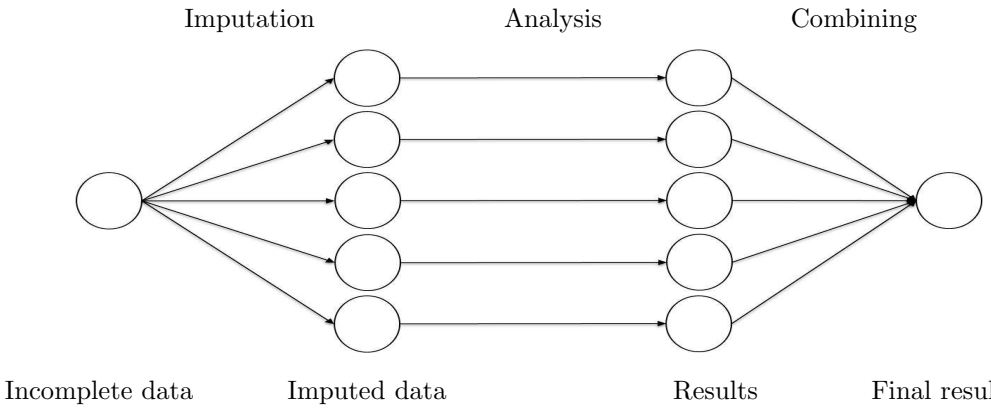



**Figure 7.1:** Graphical representation of the different steps for the multiple imputation analyses: generation of imputations, repeated analyses on the imputed data, and combination of the results.

The results of five imputed datasets are combined according to the theory of multiple imputations.

### 7.1.2 Variable Selection

Random forests are applied to select subset of predictor variables. The method of random forests is an ensemble method that combines several individual classification trees in the following way: from the original sample several bootstrap samples of the data are drawn, and an unpruned classification tree is fit to each bootstrap sample; the variable selection for each split in the classification tree is conducted only from a small random subset of predictor variables; from the complete forest the status of the response variable is predicted as an average or majority vote of the predictions for all trees (Breiman *et al.*, 2001).

Random forests can highly increase the prediction accuracy as compared to individual classification trees, because the ensemble adjusts for the instability of the individual trees induced by small changes in the learning sample, that impairs the prediction accuracy in test samples.

As part of the algorithm, a random forest analysis returns measures of variable importance. The measures of variable importance can be used to perform variable selection. Variable importance measures for random forests have been receiving increased attention as a means of variable selection in many classification tasks in bioinformatics and related scientific fields, for instance to select a subset of genetic markers relevant for the prediction of a certain disease (Strobl *et al.*, 2007).

### 7.1.3 Logistic Regression Model

In the first stage of the multivariate analysis, for each imputed dataset, the variable importance list produced by the random forest algorithm is used in order to select a list of variables representing a list of candidate models. The random forest algorithm is a powerful data-driven approach commonly used for identifying important variables. Next, a logistic regression model with all selected variables and their two way interaction terms is fitted (Agresti, 2002; Hosmer and Lemeshow, 2000). For model selection, an automatic model selection procedure (backward selection procedure) is used. All variables which are significant in at least one imputed dataset are retained. Using these variables, a final logistic regression model is fitted to all imputed datasets separately for each study, and the results over the imputed datasets are combined.

Here, in our setting the logistic regression model is given by:

$$\text{logit}\Big(P(Y_{ij}=1)\Big) = \beta_0 + \textstyle\sum_k^P \beta_k X_{kij}, \quad i = 1,\ldots,5; j = 1,\ldots,n_i, \tag{7.1}$$

where $Y_{ij}$ is the outcome variable for a patient $j$ in study $i$ (0 indicate that CHD is absent, or 1 indicate that CHD is present in the individual), $X_{kij}$ represents the $k$th covariate, $\beta_0$ is the intercept, $\beta_k$ is the coefficient for the covariate/predictor $X_{kij}$, $P$ is the number of covariates, and $n_i$ is the number of patients in study $i$.

### 7.1.4 Receiver Operating Characteristic Curve

Receiver operating characteristic (ROC) curve is an effective and widely used method for evaluating the discriminating power of a diagnostic test or statistical model. ROC curve is a plot of Sensitivity, i.e. true positive rate (TPR) vs. 1-Specificity, i.e. the false positive rate (FPR) at different thresholds.

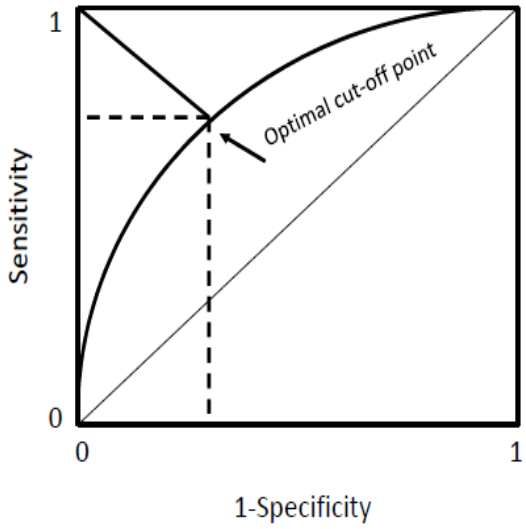



**Figure 7.2:** The ROC curve obtained by plotting Sensitivity vs. 1-Specificity at different cut-offs values.

The closer the curve comes to the left-hand border and then the top border of the graph, the more accurate is the model: i.e. it has high sensitivity and specificity. The closer the curve comes to the diagonal, the less accurate is the model (this suggests no discrimination). The best cut-off that maximizes sensitivity and specificity is a cut-off which is close towards the upper left hand corner of the curve.

The ROC curve can be summarized by the area under the curve (AUC). The area under the curve measures the discriminating power of the model (Hosmer and Lemeshow, 2000).

## 7.2 Individual Patient Data Meta-analysis

In the second stage, the five imputed datasets for each study are combined to five meta-datasets, and an IPD meta-analysis is conducted using the variables which are significant in at least one study-specific analysis. Meta-analysis is a statistical literature review of magnitude of an effect. Meta-analysis combines the effects from all studies to give an overall mean effect and other important statistics (Higgins and Green, 2011).

The potential advantage of meta-analysis includes an increase in power, an improvement in precision, the ability to answer question not posted by individual studies,

and the opportunity to settle controversies arising from conflicting claims. However, they also have the potential to mislead seriously, particularly if specific study designs, within-study biases, variation across studies, and reporting biases are not carefully considered (Hunter *et al.*, 1982).

Random-effect (mixed-model) meta-analysis assume there are real differences between all studies in the magnitude of the effect. The random effect is the standard deviation representing the variation in the true magnitude from study to study. For the random effects model, we need more studies than for traditional (fixed-effects) meta-analysis. As we have only 5 studies, fixed-effects models are used rather than random-effects models. To allow heterogeneity across studies, we have included study-specific intercepts.

$$\text{logit}\Big(P(Y_{ij}=1)\Big) = \beta_{0i} + \textstyle\sum_k^P \beta_k X_{kij} * I_{ki}, \quad i = 1,\ldots,5; j = 1,\ldots,n_i, \tag{7.2}$$

where $\beta_{0i}$ $(i = 1,\ldots,5)$ is the study-specific intercept for the $i$th study, $I_{ki}$ is the study indicator for the $k$th covariate,

$$I_{ki} = \begin{cases} 1 & \text{if the covariate } X_k \text{ is available in individual study } i \\ 0 & \text{if the covariate } X_k \text{ is not available in individual study } i. \end{cases} \tag{7.3}$$

## 7.3 Combining Predictors for Classification

Based on the estimated regression coefficients of an IPD meta-analysis, all predictors (signs & symptoms, age, gender) are combined to one linear score, $L(X)$, to classify the patients. If the score is greater than a threshold value $c \in (-\infty, \infty)$, then the patient is CHD positive; if the score is less than a threshold value, then the patient is CHD negative.

We consider linear score of the form

$$L_\beta(X) = X_1 + \beta_2 X_2 + ... + \beta_P X_P. \tag{7.4}$$

The linear score does not include an intercept and that the coefficient associated with $X_1$ is 1. This is not a restriction since with $\alpha_1 > 0$ (and we can redefine $X_1$ as $-X_1$ to ensure $\alpha_1 > 0$), rules based on the linear predictor $L_\alpha(X) = \alpha_0 + \alpha_1 L_\beta(X)$ exceeding a threshold are equivalent to rules based on $L_\beta(X)$ exceeding a threshold. The ROC curves of $L_\beta(X)$ and $L_\alpha(X)$ are the same, so it is enough to consider $L_\beta(X)$ (Pepe *et al.*, 2006).

Under what circumstances is $L_\beta(X)$ the "right" combination score for classification to CHD=1 or 0 based on X? If the score is some monotone increasing function of $L_\beta(X)$,

$$P(D=1|X) = f(X_1 + \beta_2 X_2 + \cdots + \beta_P X_P) = f\big(L_\beta(X)\big). \tag{7.5}$$

It follows from the Neyman-Pearson lemma (Neyman and Pearson, 1933) that rules based on $L_\beta(X) > c$ are optimal. They are optimal in the sense that no other classification rule based on X can have even a single accuracy point (FPR, TPR) that lies above the ROC curve for $L_\beta(X)$. Thus for a fixed FPR, the TPR of the rule $L_\beta(X) > c$ is higher than the TPR of any other rule with the same FPR. Similarly for fixed TPR, the rule $L_\beta(X) > c$ has lowest FPR among all rules based on X with the same TPR (Pepe *et al.*, 2006).

The Logistic regression is a special case of the general linear model (7.5) with $f(x) = logit^{-1}(\alpha_0 + \alpha_1 x)$. If we assume that the logistic regression model holds and calculate the maximum likelihood estimates $(\hat{\alpha}_1^L, \ldots, \hat{\alpha}_P^L)$, this yields maximum likelihood estimates of $(\beta_2, \ldots, \beta_P)$, namely $\beta_P^L = \hat{\alpha}_P^L / \hat{\alpha}_1^L$. In summary, the logistic likelihood can be used as an objective function to derive linear predictor

$$L_{\hat{\beta}^L}(X) = X_1 + \hat{\beta}_2^L X_2 + \cdots + \hat{\beta}_P^L X_P. \tag{7.6}$$

Another approach is, optimality of $L_\beta(X)$ implies that the ROC curve for any other function of X cannot be higher at any point than the ROC curve for $L_\beta(X)$. Since $L_\beta(X)$ has the best ROC curve among all functions of X, it certainly has the best ROC curve among all linear predictors of the form $L_b(X) = X_1 + b_2 X_2 + \cdots + b_P X_P$. The idea is to select choices of coefficients $(b_2, \ldots, b_P)$ that yield the best empirical ROC curve for $\{L_b(X_{Di}), i = 1, \ldots, n_D; L_b(X_{\bar{D}j}), j = 1, \ldots, n_{\bar{D}}\}$. These are then interpreted as estimates of $(\beta_2, \ldots, \beta_P)$.

The area under the ROC curve is the most popular ROC summary index. It can be interpreted as the probability that, for a random case-control pair, the score for the case exceeds that of the control, $P\Big(L_b(X_{Di}) > L_b(X_{\bar{D}j})\Big)$. The optimal ROC curve has maximum AUC, so we can use it as the basis for an objective function of the data to estimate $\beta$. It is easy to show that the AUC of the empirical ROC curve is the Mann-Whitney U statistic

$$\widehat{AUC}(b) = \frac{\sum_{i=1}^{n_D} \sum_{j=1}^{n_{\bar{D}}} I\big[L_b(X_{Di}) > L_b(X_{\bar{D}j})\big]}{n_D n_{\bar{D}}}. \tag{7.7}$$

We write the corresponding AUC based estimator of $\beta$ as

$$\widehat{\beta}^{AUC} = \text{argmax}\big(\widehat{AUC}(b)\big). \tag{7.8}$$

As the maximization of AUC is not available in standard software, we developed the code in R. The main part of the program is included in Appendix D.

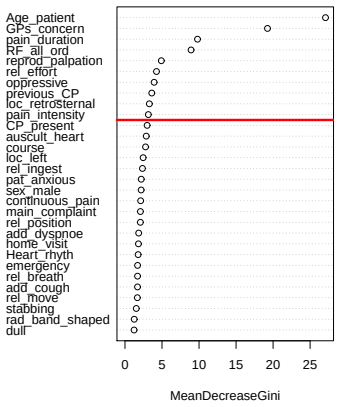
## 7.4 Results

### 7.4.1 Study-specific Analysis

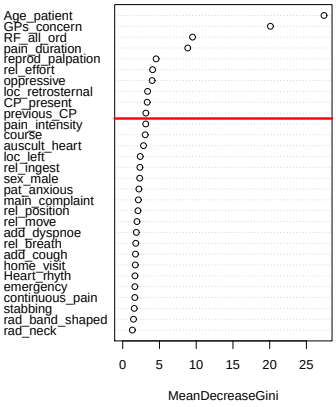
For each study separately, the variable selection method has been applied for the imputed datasets in order to get the most important set of predictors. The variable importance plot from random forest and the results of the logistic regression model for the finally selected set of variables are given in Figures 7.3–7.7 and Tables 7.1–7.5, respectively.

## Swiss Study

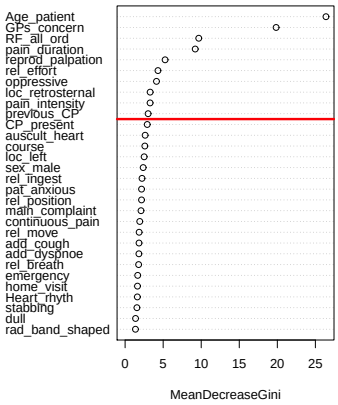
The variable importance list from the random forest for the Swiss study is given in Figure 7.3, separately for each imputed dataset.



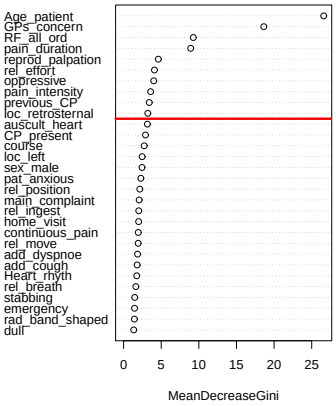

(a). Imputed data 1



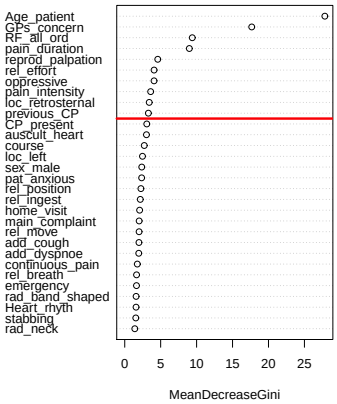

(b). Imputed data 2




(c). Imputed data 3




(d). Imputed data 4




(e). Imputed data 5

**Figure 7.3:** Variable importance plot for the Swiss study.

From the importance list (Figure 7.3), we chose the top ten predictors. We could have included more predictors, but the fit of the logistic regression model is questionable as we also include the interaction term between the predictors. Automated variable selection method has been applied to identify predictors for developing parsimonious regression model.

Using the finally selected set of predictors from the automated variable selection method, we fit a logistic regression model. The results of the logistic regression for each imputed data are given in Table 7.1. In Table 7.1, we also include the combined results of the logistic regression model for the set of variables which are significant in at least one of the imputed dataset.

**Table 7.1:** Parameter estimates and standard errors for the parameters obtained from the study-specific analysis of the Swiss study.

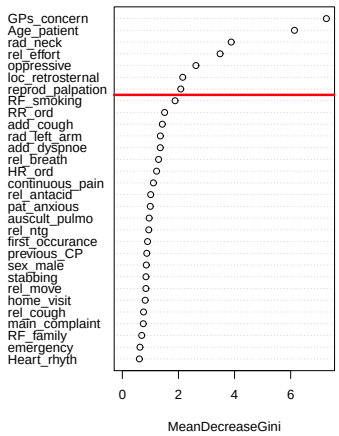
| Parameter | | Analysis by imputation | | | | | Combined results |
|---|---|---|---|---|---|---|---|
| | | Imputation 1 | Imputation 2 | Imputation 3 | Imputation 4 | Imputation 5 | |
| Intercept | | -6.672(0.621)** | -5.999(0.546)** | -6.028(0.549)** | -6.657(0.625)** | -5.840(0.536)** | -6.998(0.652)** |
| Age-patient | 1 | 2.699(0.458)** | 2.857(0.454)** | 2.858(0.453)** | 2.739(0.458)** | 2.905(0.456)** | 2.768(0.462)** |
| Previous-CP | 1 | 1.135(0.344)** | - | - | 1.113(0.339)** | - | 1.122(0.352)** |
| GPs-concern | 1 | 2.640(0.347)** | 2.687(0.342)** | 2.647(0.341)** | 2.607(0.347)** | 2.547(0.340)** | 2.493(0.359)** |
| Loc-retrosternal | 1 | 1.204(0.348)** | 1.104(0.340)** | 1.159(0.338)** | 1.179(0.348)** | - | 1.038(0.357)** |
| Oppressive | 1 | - | - | - | - | 1.094(0.305)** | 0.943(0.329)** |

** The variable is significant at 5% level of significance ($p$-value $< 0.05$).

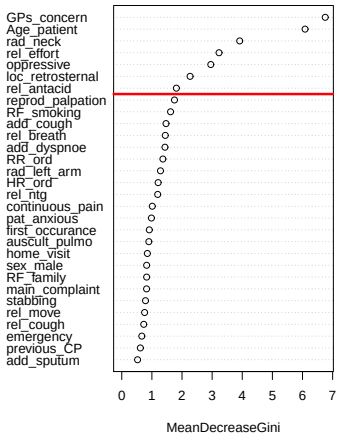
From the results of the logistic regression model (Table 7.1), the signs and symptoms related to previous-CP (had the patient experienced chest pain before?), GPs-concern (was the GPs concern that the underlying cause of the chest pain is something serious?), loc-retrosternal (was the pain located in the area behind the sternum?), and oppressive (was the pain characterized as "oppressive"?) have significant effect on CHD, at least in one of the imputed data. Age of patient (Group 1: male $<$ 55 year, female $<$ 65 year; Group 2: male $\geq$ 55 year, female $\geq$ 65 year) have also a significant effect on CHD, at least in one of the imputed data.

## Belgian Study

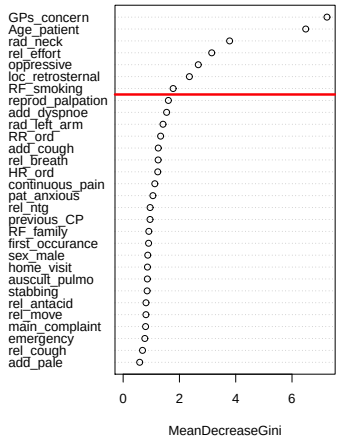
The variable importance list from the random forest for the Belgian study is given in Figure 7.4, separately for each imputed dataset.



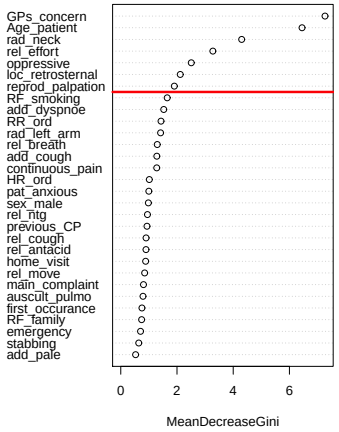

(a). Imputed data 1



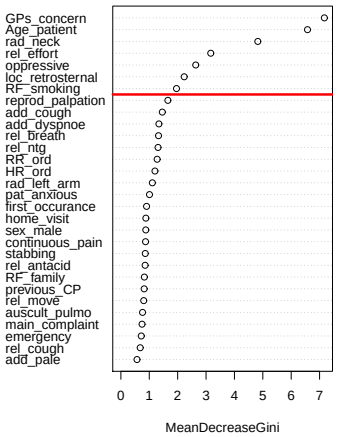

(b). Imputed data 2




(c). Imputed data 3




(d). Imputed data 4




(e). Imputed data 5

**Figure 7.4:** Variable importance plot for the Belgian study.

From the importance list (Figure 7.4), we chose the top seven predictors. We could have included more predictors, but the fit of the logistic regression model is questionable as we also include the interaction term between the predictors. Automated variable selection method has been applied to identify predictors for developing parsimonious regression model.

Using the finally selected set of predictors from the automated variable selection method, we fit a logistic regression model. The results of the logistic regression for each imputed data are given in Table 7.2. In Table 7.2, we also include the combined results of the logistic regression model for the set of variables which are significant in at least one of the imputed dataset.

**Table 7.2:** Parameter estimates and standard errors for the parameters obtained from the study-specific analysis of the Belgian study.

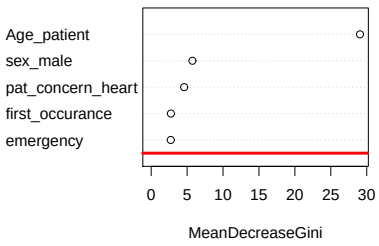
| Parameter | | Analysis by imputation | | | | | Combined results |
|---|---|---|---|---|---|---|---|
| | | Imputation 1 | Imputation 2 | Imputation 3 | Imputation 4 | Imputation 5 | |
| Intercept | | -6.757(1.072)** | -7.377(1.041)** | -8.031(1.134)** | -6.821(1.078)** | -8.014(1.133)** | -7.374(1.168)** |
| GPs-concern | 1 | 3.214(0.727)** | 3.249(0.703)** | 3.025(0.685)** | 3.196(0.728)** | 2.987(0.689)** | 3.052(0.717)** |
| Loc-retrosternal | 1 | 1.635(0.697)** | 1.959(0.698)** | 1.784(0.710)** | 1.605(0.703)** | 1.914(0.748)** | 1.598(0.747)** |
| Rad-neck | 1 | 1.567(0.693)** | 1.568(0.633)** | 1.714(0.676)** | 1.715(0.678)** | 1.986(0.698)** | 1.795(0.714)** |
| Oppressive | 1 | 1.556(0.656)** | 1.614(0.649)** | 2.075(0.692)** | 1.554(0.666)** | 1.900(0.703)** | 1.799(0.718)** |
| Rel-effort | 1 | 2.344(0.666)** | 1.956(0.618)** | 2.116(0.649)** | 2.416(0.675)** | 1.994(0.654)** | 2.267(0.698)** |
| Reprod-palpation | 1 | -2.109(0.881)** | - | - | -2.010(0.887)** | - | -1.670(0.916)* |
| RF-smoking | 1 | - | - | 1.424(0.610)** | - | 1.493(0.629)** | 1.319(0.660)** |

** The variable is significant at 5% level of significance ($p$-value $< 0.05$),
* The variable is significant at 10% level of significance ($p$-value $< 0.10$).

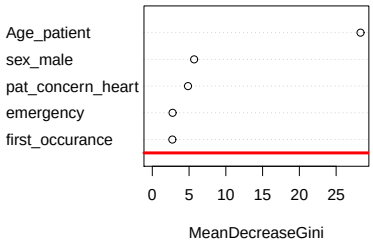
From the results of the logistic regression model (Table 7.2), the signs and symptoms related to GPs-concern (was the GPs concern that the underlying cause of the pain was something serious?), loc-retrosternal (was the pain located in the area behind the sternum?), rad-neck (did the pain radiate to the neck/jaw/bottom side of the face?), oppressive (was the pain characterized as oppressive?), rel-effort (was the pain related to effort or exercise?), reprod-palpation (was the pain reproducible by palpitation?), and RF-smoking (did the patient smoke?) have significant effect on CHD, at least in one of the imputed data.

## Swedish Study

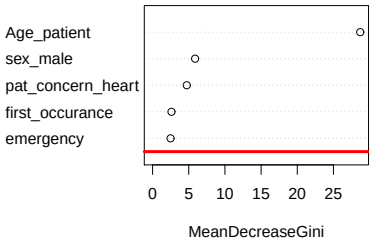
The variable importance list from the random forest for the Swedish study is given in Figure 7.5, separately for each imputed dataset.



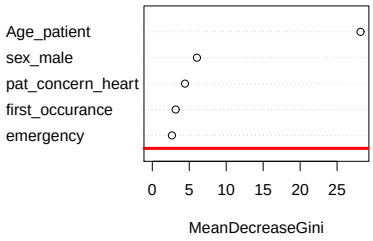

(a). Imputed data 1



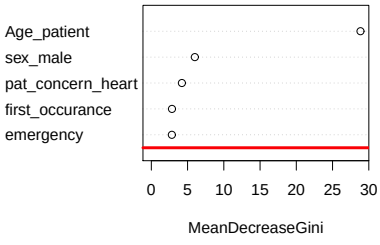

(b). Imputed data 2




(c). Imputed data 3




(d). Imputed data 4




(e). Imputed data 5

**Figure 7.5:** Variable importance plot for the Swedish study.

As there are very few variables available in Swedish study, we include all the five predictors and the two way interaction between these predictors. Then, automated variable selection method has been applied to identify predictors for developing parsimonious regression model.

Using the finally selected set of predictors from the automated variable selection method, we fit a logistic regression model. The results of the logistic regression for each imputed data are given in Table 7.3. In Table 7.3, we also include the combined results of the logistic regression model for the set of variables which are significant in at least one of the imputed dataset.

**Table 7.3:** Parameter estimates and standard errors for the parameters obtained from the study-specific analysis of the Swedish study.

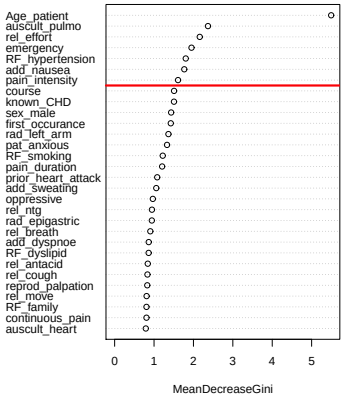
| Parameter | | Analysis by imputation | | | | | Combined results |
|---|---|---|---|---|---|---|---|
| | | Imputation 1 | Imputation 2 | Imputation 3 | Imputation 4 | Imputation 5 | |
| Intercept | | -4.674(0.510)** | -4.816(0.533)** | -4.816(0.533)** | -4.672(0.511)** | -4.705(0.514)** | -4.736(0.527)** |
| Age-patient | 1 | 1.727(0.327)** | 1.717(0.328)** | 1.717(0.328)** | 1.717(0.327)** | 1.727(0.327)** | 1.721(0.327)** |
| Sex-male | 1 | 0.959(0.322)** | 0.960(0.322)** | 0.960(0.322)** | 0.971(0.322)** | 0.983(0.322)** | 0.966(0.322)** |
| Pat-concern-heart | 1 | 1.440(0.426)** | 1.604(0.453)** | 1.604(0.453)** | 1.435(0.426)** | 1.459(0.427)** | 1.508(0.448)** |

** The variable is significant at 5% level of significance ($p$-value $< 0.05$).

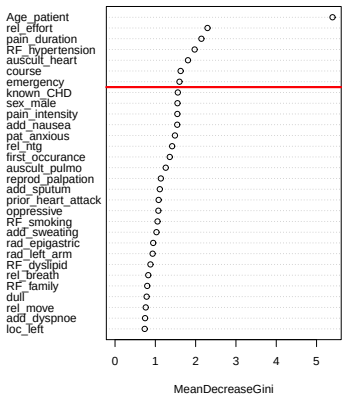
From the results of the logistic regression model (Table 7.3), the signs and symptoms related to pat-concern-heart (did the patient think the pain is related to the heart?) has significant effect on CHD, at least in one of the imputed data. Sex-male (sex of the patient: male?) and age of patient (Group 1: male $< 55$ year, female $< 65$ year; Group 2: male $\geq 55$ year, female $\geq 65$ year) have also a significant effect on CHD, at least in one of the imputed data.

## US Study

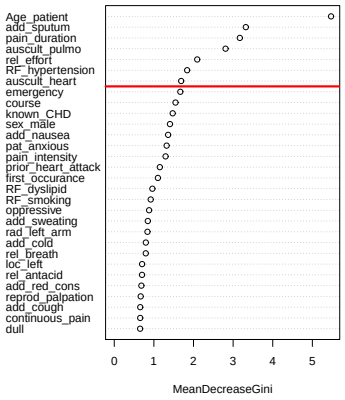
The variable importance list from the random forest for the US study is given in Figure 7.6, separately for each imputed dataset.



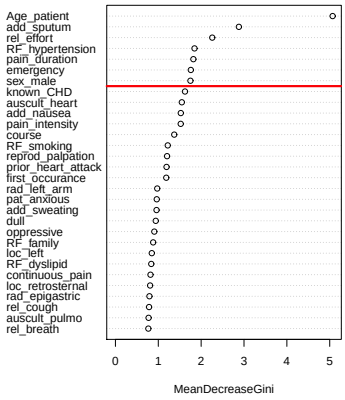

(a). Imputed data 1



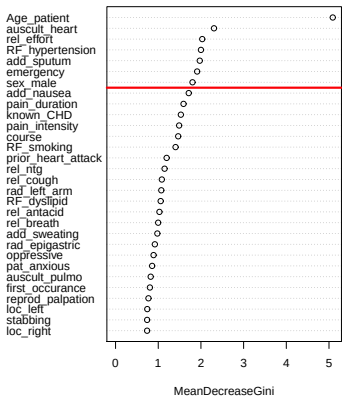

(b). Imputed data 2




(c). Imputed data 3




(d). Imputed data 4




(e). Imputed data 5

**Figure 7.6:** Variable importance plot for the US study.

From the importance list (Figure 7.6), we chose the top seven predictors. We could have included more predictors, but the fit of the logistic regression model is questionable as we also include the interaction term between the predictors. Automated variable selection method has been applied to identify predictors for developing parsimonious regression model.

Using the finally selected set of predictors from the automated variable selection method, we fit a logistic regression model. The results of the logistic regression for each imputed data are given in Table 7.4. In Table 7.4, we also include the combined results of the logistic regression model for the set of variables which are significant in at least one of the imputed dataset.

**Table 7.4:** Parameter estimates and standard errors for the parameters obtained from the study-specific analysis of the US study.

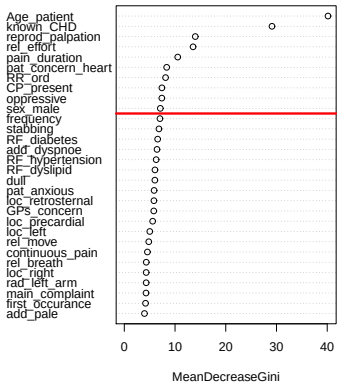
| Parameter | | Analysis by imputation | | | | | Combined results |
|---|---|---|---|---|---|---|---|
| | | Imputation 1 | Imputation 2 | Imputation 3 | Imputation 4 | Imputation 5 | |
| Intercept | | -5.763(0.721)** | -3.478(0.427)** | -9.861(1.389)** | -3.662(0.548)** | -9.197(1.236)** | -7.873(4.554) |
| Age-patient | 1 | - | 0.679(0.512)** | 1.896(0.707)** | | 0.577(0.616)** | 0.068(0.684) |
| Sex-male | 1 | - | - | | 1.292(0.509)** | 2.322(0.615)** | 1.648(1.052) |
| Emergency | 1 | 1.318(0.470)** | - | - | - | - | 1.165(1.014) |
| Rel-effort | 1 | 1.683(0.470)** | 1.788(0.463)** | 2.190(0.562)** | 1.327(0.457)** | 2.197(0.536)** | 2.025(0.876)** |
| Add-nausea | 1 | 0.988(0.532)** | - | - | - | - | 0.789(0.654) |
| Add-sputum | 1 | - | - | 2.679(0.733)** | -2.735(0.756)** | 2.308(0.594)** | 0.473(3.163) |
| Auscult-pulmo | 1 | 2.235(0.567)** | - | 3.072(0.699)** | - | - | 1.468(1.978) |
| Auscult-heart | 1 | - | -2.912(0.790)** | 1.880(0.679)** | - | 2.834(0.656)** | 0.468(3.082) |
| RF-hypertension | 1 | 1.000(0.481)** | 1.932(0.489)** | 2.292(0.596)** | 1.504(0.462)** | 1.751(0.537)** | 1.581(0.868)* |

** The variable is significant at 5% level of significance ($p$-value $< 0.05$),
* The variable is significant at 10% level of significance ($p$-value $< 0.10$).

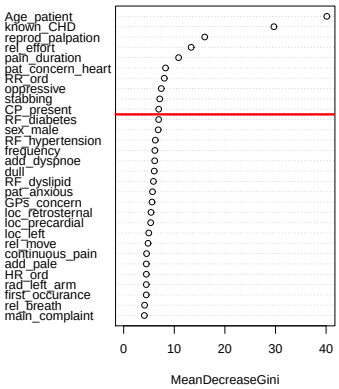
From the results of the logistic regression model (Table 7.4), the signs and symptoms related to emergency (was it an emergency visit? did the patient claim the consultation to be urgent?), rel-effort (was the pain related to effort or exercise?), add-nausea (additional symptoms: nausea), add-sputum (additional symptoms: sputum), auscult-pulmo (pulmonary auscultation), auscult-heart (cardial auscultation), RF-hypertension (had the patient a history of hypertation?) have significant effect on CHD, at least in one of the imputed data. Sex-male (sex of the patient: male?) and age of patient (Group 1: male $<$ 55 year, female $<$ 65 year; Group 2: male $\geq$ 55 year, female $\geq$ 65 year) have also significant effect on CHD, at least in one of the imputed data.

## German Study

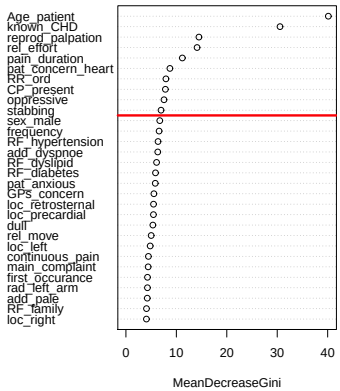
The variable importance list from the random forest for the German study is given in Figure 7.7, separately for each imputed dataset.



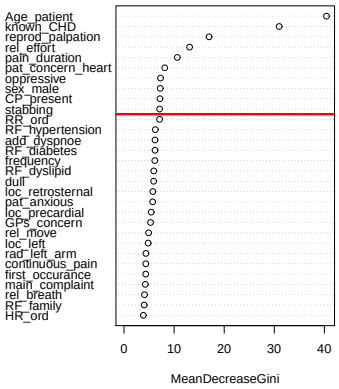

(a). Imputed data 1



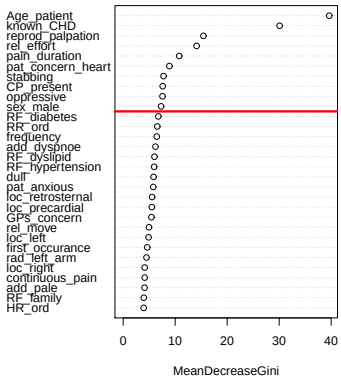

(b). Imputed data 2




(c). Imputed data 3




(d). Imputed data 4




(e). Imputed data 5

**Figure 7.7:** Variable importance plot for the German study.

From the importance list (Figure 7.7), we chose the top ten predictors. We could have included more predictors, but the fit of the logistic regression model is questionable as we also include the interaction term between the predictors. Automated variable selection method has been applied to identify predictors for developing parsimonious regression model.

Using the finally selected set of predictors from the automated variable selection method, we fit a logistic regression model. The results of the logistic regression for each imputed data are given in Table 7.5. In Table 7.5, we also include the combined results of the logistic regression model for the set of variables which are significant in at least one of the imputed dataset.

**Table 7.5:** Parameter estimates and standard errors for the parameters obtained from the study-specific analysis of the German study.

| Parameter | | Analysis by imputation | | | | | Combined results |
|---|---|---|---|---|---|---|---|
| | | Imputation 1 | Imputation 2 | Imputation 3 | Imputation 4 | Imputation 5 | |
| Intercept | | -4.124(0.325)** | -3.352(0.297)** | -4.156(0.325)** | -3.700(0.347)** | -3.922(0.349)** | -3.792(0.368)** |
| Age-patient | 1 | 1.201(0.229)** | 1.133(0.231)** | 1.229(0.231)** | 1.135(0.234)** | 1.199(0.234)** | 1.168(0.234)** |
| Pat-concern-heart | 1 | 1.042(0.239)** | 1.027(0.237)** | 1.035(0.239)** | 0.956(0.246)** | 1.201(0.246)** | 1.045(0.264)** |
| Stabbing | 1 | - | -0.740(0.228)** | - | -0.544(0.246)** | -0.598(0.247)** | -0.555(0.248)** |
| Rel-effort | 1 | 1.167(0.209)** | 1.161(0.209)** | 1.236(0.210)** | 1.187(0.212)** | 1.145(0.214)** | 1.163(0.217)** |
| Reprod-palpation | 1 | -1.680(0.257)** | -1.889(0.278)** | -1.823(0.258)** | -1.960(0.286)** | -1.774(0.261)** | -1.794(0.309)** |
| Known-CHD | 1 | 1.739(0.216)** | 1.837(0.218)** | 1.803(0.219)** | 1.866(0.225)** | 1.933(0.229)** | 1.837(0.234)** |
| Oppressive | 1 | 0.751(0.221)** | - | 0.780(0.221)** | 0.511(0.239)** | 0.545(0.239)** | 0.530(0.244)** |

** The variable is significant at 5% level of significance ($p$-value $< 0.05$).

From the results of the logistic regression model (Table 7.5), the signs and symptoms related to pat-concern-heart (did the patient think the pain is related to the heart?), stabbing (was the pain characterized as "stabbing"), rel-effort (was the pain related to effort or exercise?), reprod-palpation (was the pain reproducible by palpitation?), known-CHD (history of coronary heart disease?), and oppressive (was the pain characterized as "oppressive") have significant effect on CHD, at least in one of the imputed data. Age of patient (Group 1: male $<$ 55 year, female $<$ 65 year; Group 2: male $\geq$ 55 year, female $\geq$ 65 year) have also a significant effect on CHD, at least in one of the imputed data.

### 7.4.2 Individual Patient Data Meta-analysis

We combined the imputed datasets together in order to conduct an IPD meta-analysis. The variables which were significant in at least one of the study-specific analysis were kept. Using this list of variables, a meta-model is fitted for the combined dataset. The country-specific analysis is re-fitted by including the predictors which are significant in another studies. The estimates and standard errors for the parameters obtained from the study-specific analyses and from the IPD meta-analysis are shown in Table 7.6.

The results in Table 7.6 indicate that the symptoms and signs for diagnosis of coronary heart disease related to medical history of the chest pain, pain radiation, pain characteristics, smoking and previous medical history of coronary heart disease have a significant effect on the presence or absence of CHD. The results also suggest that older people, male patients and patients with a known previous CHD medical history have a higher chance of having CHD.

**Table 7.6:** Parameter estimates and standard errors for the parameters obtained from the study-specific analyses and the IPD meta-analysis.

| | | Study-specific analyses | | | | | |
|---|---|---|---|---|---|---|---|
| Parameter | | Switzerland | Belgium | Sweden | USA | Germany | Meta-Model |
| | $S_1$ | | | | | | -3.992(0.330)** |
| | $S_2$ | | | | | | -4.027(0.678)** |
| Intercept | $S_3$ | -6.870(0.813)** | -6.138(1.756)** | -4.584(0.572)** | -8.946(5.760) | -3.710(0.413)** | -3.643(0.306)** |
| | $S_4$ | | | | | | -4.753(1.023)** |
| | $S_5$ | | | | | | -5.094(0.355)** |
| Age-patient | 1 | 2.700(0.488)** | 0.429(0.823) | 1.691(0.331)** | -0.288(1.059) | 1.099(0.247)** | 1.433(0.157)** |
| Sex-male | 1 | 0.750(0.367)** | 0.082(0.781) | 0.986(0.324)** | 1.905(1.468) | -0.089(0.213) | 0.283(0.141)** |
| Emergency | 1 | -0.697(0.419)* | -0.898(0.894) | -0.214(0.317) | 0.736(1.216) | - | -0.176(0.194) |
| Previous-CP | 1 | 1.063(0.398)** | -0.042(0.802) | - | 0.268(0.940) | - | 0.425(0.250)* |
| GPs-concern | 1 | 2.487(0.404)** | 3.647(0.949)** | - | - | 0.513(0.268)* | 1.302(0.187)** |
| Pat-concern-heart | 1 | - | - | 1.512(0.448)** | - | 1.033(0.271)** | 1.127(0.230)** |
| Loc-retrosternal | 1 | 0.885(0.388)** | 1.360(0.800)* | - | 0.662(0.845) | -0.218(0.233) | 0.251(0.171) |
| Rad-neck | 1 | 1.185(1.133) | 1.533(0.827)* | - | 0.054(1.510) | 0.125(0.378) | 0.606(0.287)** |
| Stabbing | 1 | 0.644(0.533) | 0.198(1.240) | - | -2.814(1.609)* | -0.563(0.253)** | -0.426(0.208)** |
| Oppressive | 1 | 0.855(0.363)** | 2.022(1.135)* | - | 0.407(0.784) | 0.532(0.251)** | 0.635(0.175)** |
| Rel-effort | 1 | 0.462(0.383) | 2.310(0.769)** | - | 2.454(1.044)** | 1.162(0.219)** | 1.194(0.166)** |
| Add-nausea | 1 | - | -0.095(1.209) | - | 1.095(0.810) | -0.377(0.429) | -0.087(0.315) |
| Add-sputum | 1 | -1.342(0.940) | -1.391(1.198) | - | 0.675(3.948) | - | -0.746(1.128) |
| Auscult-pulmo | 1 | -1.307(0.790)* | -1.748(1.013)* | - | 0.894(2.245) | - | -0.447(0.624) |
| Auscult-heart | 1 | 1.669(0.775)** | - | - | -0.057(3.334) | - | 0.748(1.093) |
| Reprod-palpation | 1 | -0.861(0.424)** | -1.896(1.010)* | - | -2.222(1.144)* | -1.819(0.310)** | -1.542(0.239)** |
| RF-hypertension | 1 | - | - | - | 2.092(1.021)* | 0.113(0.216) | 0.349(0.210)* |
| RF-smoking | 1 | - | 1.676(0.853)* | - | 0.921(0.841) | 0.057(0.367) | 0.630(0.261)** |
| Known-CHD | 1 | - | - | - | 1.195(0.878) | 1.884(0.245)** | 1.732(0.217)** |

CP: chest pain, CHD: coronary heart disease, GP: general practitioner, RF: risk factor, $S_1$: Switzerland, $S_2$: Belgium, $S_3$: Sweden, $S_4$: USA, $S_5$: Germany,

** The variable is significant at 5% level of significance ($p$-value $< 0.05$),

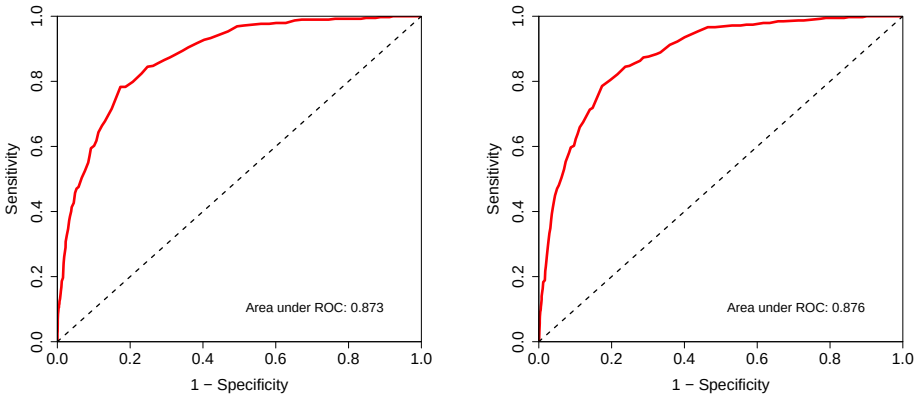
* The variable is significant at 10% level of significance ($p$-value $< 0.10$).

The areas under the ROC curves are given in Table 7.7. The area under the ROC curve provides a measure of the models ability to discriminate between those patients who experience the outcome of interest versus who do not.

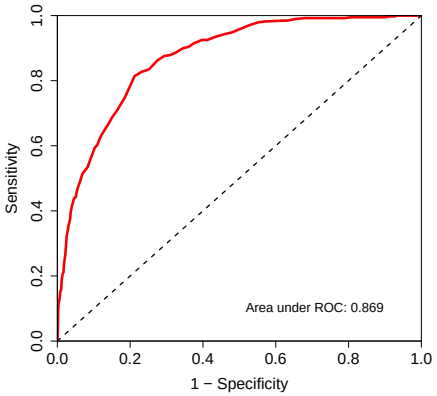
**Table 7.7:** Area under the ROC curves obtained from study-specific analyses and IPD meta-analysis.

| | Study-specific analysis | | | | | |
|---|---|---|---|---|---|---|
| Imputation | Switzerland | Belgium | Sweden | USA | Germany | Meta-Model |
| 1 | 0.938 | 0.948 | 0.683 | 0.902 | 0.864 | 0.873 |
| 2 | 0.938 | 0.947 | 0.671 | 0.932 | 0.875 | 0.876 |
| 3 | 0.938 | 0.945 | 0.670 | 0.950 | 0.877 | 0.869 |
| 4 | 0.938 | 0.941 | 0.667 | 0.913 | 0.878 | 0.881 |
| 5 | 0.936 | 0.946 | 0.668 | 0.924 | 0.882 | 0.875 |

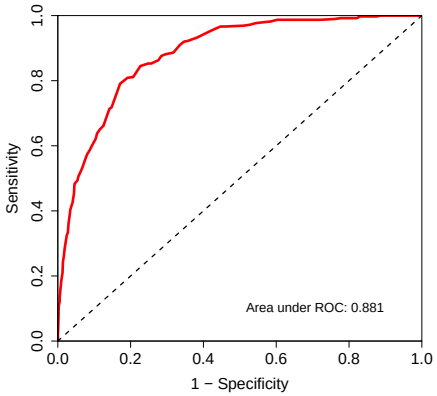
The results given in Table 7.7 suggest that the fitted model has good discrimination ability. The ROC curves for the IPD meta-mode are given in Figure 7.8.



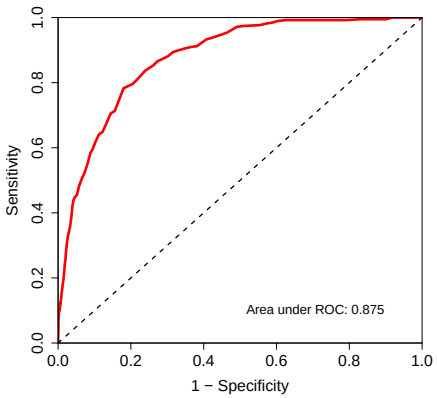

(a). ROC curve for the imputed data 1 (b). ROC curve for the imputed data 2




(c). ROC curve for the imputed data 3




(d). ROC curve for the imputed data 4




(e). ROC curve for the imputed data 5

**Figure 7.8:** ROC curves obtained from fitting the IPD meta-model.

Since the variables which have been included in the IPD meta-model in Table 7.6 are complicated for clinical use, we reduced the IPD meta-model step by step to 6 predictors by removing the non-significant and the least significant variables. The estimates for the parameters in the final model are given in the Table 7.8.

**Table 7.8:** Parameter estimates and standard errors for the parameters obtained from the reduced IPD meta-model.

| Parameter | | Estimates (Standard errors) | $p$-value |
|---|---|---|---|
| Intercept | $S_1$ | -3.736(0.233) | <0.0001 |
| | $S_2$ | -3.935(0.288) | <0.0001 |
| | $S_3$ | -2.806(0.170) | <0.0001 |
| | $S_4$ | -3.933(0.265) | <0.0001 |
| | $S_5$ | -3.976(0.209) | <0.0001 |
| Age-patient | 1 | 1.547(0.148) | <0.0001 |
| GPs-concern | 1 | 1.351(0.178) | <0.0001 |
| Oppressive | 1 | 0.839(0.159) | <0.0001 |
| Rel-effort | 1 | 1.252(0.155) | <0.0001 |
| Reprod-palpation | 1 | -1.697(0.214) | <0.0001 |
| Known-CHD | 1 | 1.711(0.201) | <0.0001 |

CHD: coronary heart disease, GP: general practitioner, $S_1$: Switzerland, $S_2$: Belgium, $S_3$: Sweden, $S_4$: USA, $S_5$: Germany.

The variables related to the age of patients, GPs first concern that the underlying causes of the chest pain was something serious, characterization of the pain as "oppressive", pain related to the effort or exercise, pain reproducible by palpation and history of coronary heart disease have significant effect on CHD. The areas under the ROC curves for the reduced IPD meta-mode are given in Table 7.9.

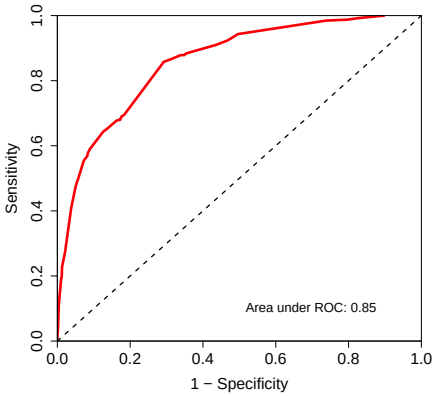
**Table 7.9:** Area under the ROC curves obtained from the reduced IPD meta-model.

| Imputation | Area under the ROC curve | Area under the ROC curve$^\dagger$ |
|---|---|---|
| 1 | 0.850 | 0.840 |
| 2 | 0.853 | 0.842 |
| 3 | 0.853 | 0.842 |
| 4 | 0.856 | 0.844 |
| 5 | 0.853 | 0.842 |

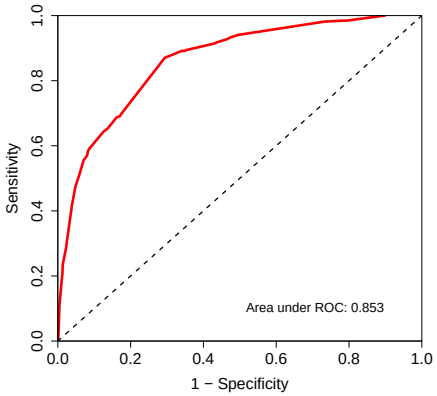
$^\dagger$ Area under the ROC curves for the simplified calculation of the score.

When reducing the number of predictors to 6, the AUC slightly reduced. Similar, simplifying the calculation of the score by replacing the parameter estimates in Table 7.8 by 1 (if the parameter estimates are greater than zero) and -1 (if the parameter estimate are less than zero) had also only a minor effect on the discriminative power of the diagnostic model.

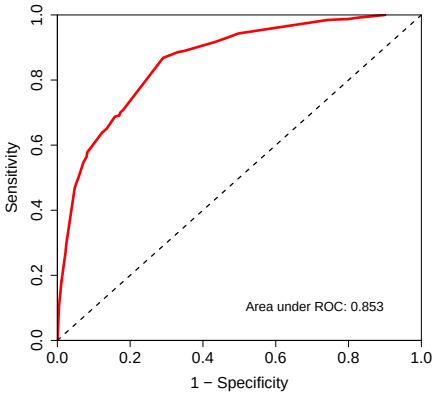
Figure 7.9 show the ROC curves obtained from fitting the study-specific analysis and the IPD meta-analysis.



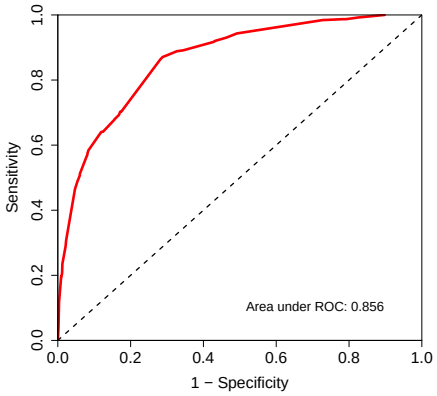

(a). ROC curve for the imputed data 1



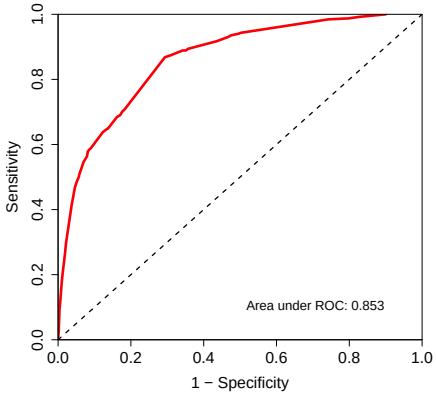

(b). ROC curve for the imputed data 2




(c). ROC curve for the imputed data 3




(d). ROC curve for the imputed data 4




(e). ROC curve for the imputed data 5

**Figure 7.9:** ROC curves obtained from fitting the reduced IPD meta-model.

### 7.4.3 Interchest Survey Results

To get more insight on the variables which have been selected by the statistical variable selection methods, the INTERCHEST collaborators have conducted a survey on 24 physicians (17 General Practitioners and 7 Cardiologists) from Sweden, Switzerland, Belgium and Germany. The physicians rated each sign and symptom using this scale:

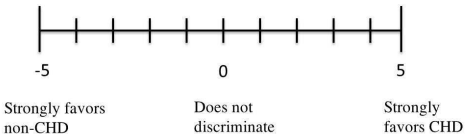



**Figure 7.10:** The scale used to rate each sign and symptom used for the diagnosis of coronary heart disease.

Clinical findings rated to be helpful by the majority of the respondents if the median is above 2 and below -2. The results of the interchest survey are included in Table A.2. From the survey results, the variables which have been selected by the statistical variable selection methods are rated to be helpful by the respondents of the survey.

### 7.4.4 New Clinical Decision Rule

Based on the predictors which are in the reduced IPD meta-model, we have constructed a new clinical prediction rule (CPR) for the diagnosis of CHD in primary care. A linear score of the form

$$L_{\hat{\beta}^L}(X) = X_1 + \hat{\beta}_2^L X_2 + \cdots + \hat{\beta}_P^L X_P \tag{7.9}$$

is used to define the new prediction rule. Where expression 7.9 is as expressed in Section 7.3.

$$CPR = \begin{cases} 1 & L_{\hat{\beta}^L}(X) \geq c \\ 0 & L_{\hat{\beta}^L}(X) < c \end{cases} \tag{7.10}$$

The best cut-off that maximizes sensitivity and specificity is 1.6, i.e. a cut-off value which is close towards the upper left hand corner of the curve. As measures of calibration (classification) we calculated the sensitivity, specificity and likelihood ratios for different thresholds for the reduced model and for the most simplified diagnostic model. The results of the new clinical decision rule for different threshold values ($c$) are given in Table 7.10.

**Table 7.10:** Combined parameter estimates and 95% confidence intervals for the measures of diagnostic accuracy for comparing classification based on the score with the reference diagnosis for different threshold values.

| Threshold (c) | Measures of diagnostic accuracy | Estimates and 95% confidence intervals | Estimates and 95% confidence intervals† |
|---|---|---|---|
| c = 1 | Sensitivity | 0.8879 (0.8499, 0.9172) | 0.9421 (0.9136, 0.9615) |
| | Specificity | 0.6516 (0.6328, 0.6700) | 0.5044 (0.4850, 0.5239) |
| | Positive likelihood | 2.5493 (2.3852, 2.7241) | 1.9013 (1.8152, 1.9915) |
| | Negative likelihood | 0.1719 (0.1274, 0.2319) | 0.1147 (0.0763, 0.1724) |
| c = 1.6 | Sensitivity | 0.7018 (0.6514, 0.7477) | 0.6889 (0.6380, 0.7356) |
| | Specificity | 0.8219 (0.8050, 0.8377) | 0.8284 (0.8135, 0.8424) |
| | Positive likelihood | 3.9429 (3.5403, 4.3907) | 4.0162 (3.5909, 4.4914) |
| | Negative likelihood | 0.3627 (0.3087, 0.4260) | 0.3755 (0.3218, 0.4380) |
| c = 2 | Sensitivity | 0.6821 (0.6326, 0.7278) | 0.6889 (0.6395, 0.7343) |
| | Specificity | 0.8391 (0.8219, 0.8550) | 0.8284 (0.8135, 0.8424) |
| | Positive likelihood | 4.2423 (3.7609, 4.7854) | 4.0162 (3.5909, 4.4914) |
| | Negative likelihood | 0.3787 (0.3260, 0.4398) | 0.3755 (0.3218, 0.4380) |

† Measures of diagnostic accuracy for the simplified calculation of the score,

CHD negative if score $<$ c; CHD positive if score $\geq$ c.

### 7.4.5 Cross-validation of Decision Rule

Since an independent dataset is not available, we could not conduct an external validation of the score. In order to have an internal validation, we used a 3-fold cross-validation approach. The whole sample is randomly partitioned in three sets (1, 2 and 3). Then we iterate three times the following procedure:

a) Take one of the sets as test sample, the other two as learning sample.

b) Using the learning sample, refit the full model as shown in Table 7.6, and simplify it gradually to a simplified model with the 6 most important predictors, and associated further simplified clinical tool (with all coefficients rounded to 1 and -1).

c) For each of the simplified models (with original and rounded coefficients), measure sensitivity, specificity etc using the test sample. So the model built with the learning sample is then tested with an independent test sample.

The results of the internal cross-validation approach for the 3-folds are included in Appendix C.

## 7.5 Discussion

This study was motivated by the need to assess the available evidence regarding the diagnostic accuracy of the medical history and physical examination for CHD in patients presenting with chest pain in primary care. Systematic reviews on the accuracy of diagnostic tests with subsequent meta-analysis of the measures of diagnostic accuracy can play an important role in decision making. They allow more precise estimates of diagnostic efficacy. However, the interpretation of the results is not straightforward. A high degree of heterogeneity, or between-study variance, that is not due to chance is a frequent finding in diagnostic accuracy reviews. Therefore, investigating the different sources is of particular importance (Buntinx *et al.*, 2009b; Rutjes *et al.*, 2006). Heterogeneity can be caused by study-level characteristics like methodological shortcomings in design or conduct of the study (bias), varied definitions of test positives, or other design-related characteristics (Buntinx *et al.*, 2009b; Rutjes *et al.*, 2006). Furthermore, patient-level characteristics can act as modifiers of diagnostic accuracy (Riley *et al.*, 2008). For example, data from secondary care suggests that the accuracy of the history and physical findings may vary according to patient characteristics like age (Gorelik *et al.*, 2007) or sex (McSweeney *et al.*, 2003; Goldberg *et al.*, 1998). Individual studies often lack statistical power to reliably estimate such effect modifications. Aggregate data meta-analysis can neither disentangle these levels nor investigate patient-level modifiers of diagnostic efficacy (Riley *et al.*, 2008; Riley *et al.*, 2010).

The problem of the insufficient discriminative power of individual signs and symptoms may be overcome by combining several criteria, for example, by developing a clinical decision rule. We performed an IPD meta-analysis to explore the combined diagnostic value of several signs and symptoms. As result we provided three diagnostic models based only on information gathered during the initial clinical examination. The most comprehensive model based on 11 predictors: age; sex; history of chest pain; GP assumed something serious; patient assumed that pain is related to heart; radiation to neck/jaw; stabbing pain, oppressive pain; pain related to effort; pain reproducible by palpitation; history of hypertension; history of smoking, history of CHD. This model showed a good discriminative power. The final and most simplified model based only on 6 predictors showed a similar discriminative power than the more complex models.

Since the maximum likelihood (ML) estimates of a logistic regression model are not necessarily maximizing the area under the receiver operating characteristic curve (AUC), the alternative approach of Pepe *et al.* (2006) was also applied and compared

with the ML approach. As there is no much different between the two approaches, results of the ML approach are used for the determination of the final model and corresponding final diagnostic tool.

We conceded that our analysis had several limitations. The studies included in the analysis had been conducted over a span of almost thirty years during which the diagnostic routines for investigating chest pain have changed. Matching of the predictors used in the different studies might be affected by semantic and cultural differences. The small number of studies limited the possibilities of the statistical analysis. Especially, using a random effects model to estimate the effect sizes for the predictors was not feasible. The respective studies investigated different sets of the predictors. The prevalence of CHD varied between 7.4 and 12.5% across studies. This might reflect differences in the study populations e.g. because of varying inclusion criteria like age as well as differences between primary care in different health care systems. A strength of our study might be that we used a statistical approach that accounted for these differences. However, regarding the novelty of this approach our findings should be considered as explanatory.

Two of the studies included in this analysis had already been used to derive a clinical prediction rule (CPR) (Gencer *et al.*, 2010; Bösner *et al.*, 2010b). Not surprisingly, the items of both CPRs showed substantial but not complete overlapping with the items of the diagnostic models developed in our study. We suggest that the score derived from the analysis presented here is more robust than the above mentioned CPRs and should have broader applicability across countries and health care systems. However, a check of this assumption by external validation using an independent data set remains desirable.

For the final and most simplified model we provided sensitivity, specificity and likelihood ratios for three different thresholds. However, we did not recommend which of these thresholds should be used in clinical practice. Choosing such an optimal threshold is often based on determining the tradeoff between sensitivity and specificity which maximize both likewise.

In this paper, we found evidence that 11 out of 61 items of the clinical examination were independent predictors for CHD in primary care. Moreover, in our analysis using only a subset of six predictors showed a similar discriminative power than using all 11 predictors. However, clinicians should consider that none of the three diagnostic models we provided in this chapter were externally validated in an independent sample. Instead we have conducted an internal validation approach.

# Chapter 8

# Discussions and Concluding Remarks

In this thesis, we introduced and explored different statistical modelling strategies for public health research. In the following sections, some discussions and concluding remarks are presented.

## 8.1 Outpatient Antibiotics Use in Europe

Quality assessments and improvement in healthcare are a major issue in many countries. The available ESAC data on outpatient antibiotics use in Europe enable countries to audit their antibiotics use by creating and maintaining a comprehensible, comparable and reliable reference database. The ESAC data have been shown to be a valuable data source for the evaluation of guidelines and policies and for the assessment of the outcomes of national interventions.

In Chapter 3, mixed-effects models were used to assess the total outpatient antibiotics use in Europe from 1997 to 2009, to analyse the trend of total antibiotics use and to analyse the seasonal variation. The applications of the models yields new important insights in the evolution of outpatient antibiotics use in Europe.

The observed differences between European countries in the levels of tetracycline use suggest that this subgroup of antibiotics is prescribed inappropriately in many countries. Seasonality of outpatient tetracycline use was also observed in all countries and was significantly related with the total outpatient tetracycline consumption.

Some of the countries implemented a national programme to control antimicrobial resistance and to improve the national use of antibiotics. This includes media coverage, such as national public campaigns in Belgium and France, repeated media reports in Slovenia and Sweden. In Chapter 4, we proposed a change-point mixed model to assess the changes in the trend of outpatient antibiotics use. The location of the change-points may be related to points in time where public-health strategies aiming at increasing the awareness of the public to a more rational use of antibiotics or targeting to reduce overconsumption of antibiotics were initiated.

For future research, more detailed data on antibiotics use linked to the patient's age and gender, the indication and prescriber characteristics could substantially broaden interpretation of the striking variations between and within European countries. Although ESAC focused on national outpatient antibiotics use, regional data can display different and more meaningful results.

## 8.2 Diagnosis of Acute Infections

Paired sera from 201 adult patients with community-acquired pneumonia were tested for the presence of antibodies against *M. pneumoniae* by several commercial test kits: *M. pneumoniae*-ELISA medac (IgM, IgG and IgA); ANILabsystems *M. pneumoniae* (IgM and IgG); EUROIMMUN Anti-*M. pneumoniae* ELISA (IgM, IgG, and IgA); and ImmunoWELL *M. pneumoniae* IgM and IgG EIA. Enzyme immunoassays (EIAs) for the detection of *Mycoplasma pneumoniae* antibodies were compared to nucleic acid sequence-based amplification (NASBA) and PCR.

Four commercially available serology assays (2 species specific, 1 genus specific, and 1 genus specific MIF test) for the detection of *C. pneumoniae* IgM and IgG antibodies and 2 commercially available serology assays for the detection of *C. pneumoniae* IgA antibodies were also compared with PCR and NASBA from 134 adult patients.

In Chapter 5, latent class models were used to evaluate tests used for the diagnosis of *Mycoplasma pneumoniae* and *Chlamydophila pneumoniae* in adult patients with lower respiratory tract infections in order to identify the most appropriate test. In Chapter 5, we considered the conditional independence model which assumes all the diagnostic tests are independent conditional on the true disease status, and the conditional dependence model which assumes some or all the diagnostic tests are dependent conditional on the true disease status. In order to evaluate the dependency of the tests on covariates, these models were extended by including age and sex. In

Chapter 6, simulation studies were set up to compare the conditional independence and conditional dependence models. For highly correlated tests, it was shown that the conditional dependence model performs better than the conditional independence model.

Differences in IgG seroprevalence were noticed when applying the 4 different IgG-assays.When comparing the serology IgG assays, the best results in terms of sensitivities and specificities were obtained by using the Medac kits. The choice of the serology assay has important implications for the detection of *M. pneumoniae* infection. A combination of a nucleic acid amplification test and a serology test might be the best choice for an accurate *M. pneumoniae* diagnosis in adult patients presenting with an LRTI.

Substantial differences between the performances of the assays were found for the detection of *C. pneumoniae* infection. We advise that the use of a proper gold standard is critical for the detection of *C. pneumoniae*.

## 8.3 Diagnosis of Coronary Heart Disease

In order to give GPs evidence-based recommendations, INTERCHEST collaborators have conducted a systematic review of studies evaluating the diagnostic accuracy of signs and symptoms for diagnosing coronary heart disease in primary care. In this study, we aim to estimates the diagnostic accuracy for clusters of signs and symptoms for the diagnosis of CHD in unselected patients presenting with chest pain in primary care. Several methodological challenges make it difficult to anticipate the results of the study. The individual studies were conducted over a span of almost thirty years accompanied by several changes, e.g., the definition of myocardial infarction.

Most individual signs and symptoms are not sufficient to reliably diagnose CHD. This problem may be overcome by combining several findings into a clinical prediction rule. In Chapter 7, IPD meta-analyses were proposed to explore the combined diagnostic value of all signs and symptoms. Based on the data of all studies, we have constructed a new clinical prediction rule for the diagnosis of CHD in primary care. An internal cross-validation approach was used to validate the new clinical prediction rule.

Based on our findings, we provided recommendations regarding the need of further research including the investigation of diagnostic algorithms based on combinations of findings of the history, physical examination, and technological devices (ECG, point of care blood tests for troponins or other biomarkers). Based on the findings, we also

provide recommendations regarding the design of future diagnostic studies in primary care and the conduct of diagnostic accuracy reviews based on individual patient data.

## 8.4 General Discussion

In this research study, we proposed and studied appropriate statistical modelling strategies for public health research with applications in the surveillance of antimicrobial consumption, the diagnosis of acute infections and the diagnosis of coronary heart disease. To analyze antimicrobial consumption datasets and to predict the true disease status of patients in the absence of a true gold standard test, random-effects models were proposed.

The random-effects models were applied to the total outpatient antibiotics use datasets to analyse the country-specific antibiotics use, to analyse the seasonal variation and to assess the change in the trend of outpatient antibiotics use. Random effects latent class models were also used to assess the diagnostic accuracy of tests used for the diagnosis of acute infections in order to identify the most appropriate test.

For the third application, as we have only 5 studies fixed-effects models are used rather than random-effects models to explore the combined diagnostic value of all signs and symptoms for diagnosing coronary heart disease in primary care. Study-specific intercepts were included in order to allow for heterogeneity across the studies.

The applications of the proposed methods will give important insights in (a) the evolution of total outpatient antibiotics use in Europe, (b) the evaluation of tests used for diagnosis of *Chlamydophila pneumoniae* and *Mycoplasma pneumoniae* infections, and (c) the diagnosis of coronary heart disease for clinicians, policymakers and others concerned with public health.

# Bibliography

Adriaenssens, N., Coenen, S., Versporten, A., Muller, A., Minalu, G., Faes, C., Vankerckhoven, V., Aerts, M., Hens, N., Molenberghs, G., Goossens, H. (2011a). European Surveillance on Antimicrobial Consumption (ESAC): outpatient antibiotic use in Europe (1997–2009). *Journal of Antimicrobial Chemotherapy*, **66** Suppl 6:vi3–vi12.

Adriaenssens, N., Coenen, S., Versporten, A., Muller, A., Minalu, G., Faes, C., Vankerckhoven, V., Aerts, M., Hens, N., Molenberghs, G., Goossens, H. (2011b). European Surveillance on Antimicrobial Consumption (ESAC): outpatient macrolide, lincosamide and streptogramin use in Europe (1997–2009). *Journal of Antimicrobial Chemotherapy*, **66** Suppl 6:vi37–vi45.

Adriaenssens, N., Coenen, S., Versporten, A., Muller, A., Minalu, G., Faes, C., Vankerckhoven, V., Aerts, M., Hens, N., Molenberghs, G., Goossens, H. (2011c). European Surveillance on Antimicrobial Consumption (ESAC): outpatient quinolone use in Europe (1997–2009). *Journal of Antimicrobial Chemotherapy*, **66** Suppl 6:vi47–vi56.

Agresti, A. (2002). *Categorical Data Analysis, second edition.* Hoboken: John Wiley & Sons.

Beersma, M.F., Dirven, K., van Dam, A.P., Templeton, K.E., Claas, E.C., Goossens, H.(2005). Evaluation of 12 commercial tests and the complement fixation test for Mycoplasma pneumoniae-specific immunoglobulin G (IgG) and IgM antibodies, with PCR used as the "gold standard". *Journal of Clinical Microbiology*, **43**: 2277–2285.

Bösner, S., Becker, A., Haasenritter, J., Abu Hani, M., Keller, H., Sönnichsen, A.C., Karatolios, K., Schaefer, J.R., Seitz, G., Baum, E. *et al.* (2009). Chest pain in

primary care: epidemiology and pre-work-up probabilities. *European Journal of General Practice*, **15(3)**:141–146.

Bösner, S., Becker, A., Abu Hani, M., Keller, H., Sönnichsen, A.C., Haasenritter, J., Karatolios, K., Schaefer, J.R., Baum, E., Donner-Banzhoff, N. (2010). Accuracy of symptoms and signs for coronary heart disease assessed in primary care. *British Journal of General Practice*, **60(575)**:246–257.

Bösner, S., Haasenritter, J., Becker, A., Karatolios, K., Vaucher, P., Gencer, B., Herzig, L., Heinzel-Gutenbrunner, M., Schaefer, J.R., Abu Hani, M., et al. (2010). Ruling out coronary artery disease in primary care:development and validation of a simple prediction rule. *Canadian Medical Association Journal*, **182(12)**, 1295–300.

Breiman L. (2001). Random Forests. *MachineLearning*, **45 (1)**:5–32.

Broeze, K.A., Opmeer, B.C., Bachmann, L.M., Broekmans, F.J., Bossuyt, P.M., Coppus, S.F., Johnson, N.P., Khan, K.S. *et al.* (2009). Individual patient data meta-analysis of diagnostic and prognostic studies in obstetrics, gynaecology and reproductive medicine. *BMC medical research methodology*, **9**:22.

Bruyninckx, R., Aertgeerts, B., Bruyninckx, P., Buntinx, F. (2008). Signs and symptoms in diagnosing acute myocardial infarction and acute coronary syndrome: a diagnostic meta-analysis. *British Journal of General Practice*, **58(547)**:105–111.

Buntinx, F., Truyen, J., Embrechts, P., Moreel, G., Peeters, R. (1992). Evaluating patients with chest pain using classification and regression trees. *Family Practice*, **9(2)**:149–153.

Buntinx, F., Aertgeerts, B., Aerts, M., Bruyninckx, R., Knottnerus, J.A., Van den Bruel, A., Van den Ende, J. (2009). *Multivariable analysis in diagnostic accuracy studies: what are the possobilities? In The evidence base of clinical diagnosis.* 2nd edition. Edited by Knottnerus J.A., Buntinx F. Oxford, Hoboken NJ: Wiley-Blackwell Pub./BMJ Books:146–166.

Buntinx, F., Aertgeerts, B., Macaskill, P. (2009). *Guidelines for conducting systematic reviews of studies evaluating the accuracy of diagnostic tests.* In The evidence base of clinical diagnosis. 2nd edition. Edited by Knottnerus J.A., Buntinx F. Oxford, Hoboken NJ: Wiley-Blackwell Pub./BMJ Books:180–212

Van Buuren, S., Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, **45(3)**:1–67.

Carlin, B.P., Gelfand, A.E., Smith, A.F.M., (1992). Hierarchical Bayesian anaysis of change points problems. *Applied Statistics*, **41**:389–405.

Clogg, C.C. (1995). *Handbook of Statistical Modeling for the Social and Behavioral Sciences*. New York: Plenum.

Coenen, S., Adriaenssens, N., Versporten, A., Muller, A., Minalu, G., Faes, C., Vankerckhoven, V., Aerts, M., Hens, N., Molenberghs, G., Goossens, H. (2011). European Surveillance of Antimicrobial Consumption (ESAC): outpatient use of tetracyclines, sulphonamides and trimethoprim, and other antibacterials use in Europe (1997–2009). *Journal of Antimicrobial Chemotherapy*, **66** Suppl 6:vi57–vi70.

Collins, L.M., and Lanza, S.T. (2010). *Latent Class and Latent Transition Analysis for the Social, Behavioral, and Health Sciences*. New York: Wiley.

Costelloe C, Metcalfe C, Lovering A. *et al*. (2010) Effect of antibiotic prescribing in primary care on antimicrobial resistance in individual patients: systematic review and meta-analysis. *BMJ*, **340**: c2096.

Chun, A.A., McGee, S.R. (2004). Bedside diagnosis of coronary artery disease: a systematic review. *The American Journal of Medicine*, **117(5)**:334–343.

Davey, P., Ferech, M., Ansari, F., Muller, A., and Goossens, H. (2008). Outpatient antibiotic use in the four administrations of the UK: cross-sectional and longitudinal analysis. *Journal of Antimicrobial Chemotherapy*, **62**:1441–1447.

Daxboeck, F., Krause, R., Wenisch, C. (2003). Laboratory diagnosis of *Mycoplasma pneumoniae* infection. *Clinical Microbiology and Infection*, **9**:263–273.

Dendukuri, N. and Joseph, L. (2001). Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics*, **57**:158–167.

Dominicus, A., Ripatti, S., Pedersen, N.L., and Palmgren, J. (2008). A random change-point model for assessing variability in repeated measures of cognitive function. *Statistics in Medicine*, **27**:5786–5798.

Doust, J. (2009). Diagnosis in General Practice. Using probabilistic reasoning. *BMJ*, **339**:b3823.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (2004). *Bayesian Data Analysis* (2nd edn). Florida: Chapman & Hall/CRC.

Gencer, B., Vaucher, P., Herzig, L., Verdon, F., Ruffieux, C., Bösner, S., Burnand, B., Bischoff, T., Donner-Banzhoff, N., Favrat, B. (2010). Ruling out coronary heart disease in primary care patients with chest pain: a clinical prediction score. *BMC Medicine*, **8**:9.

Ghosh, P. and Vaida, F. (2007). Random change-point modelling of HIV immunological responses. *Statistics in Medicine*, **26**:2074–2087.

Goldberg, R.J., O'Donnell, C., Yarzebski, J., Bigelow, C., Savageau, J., Gore, J.M. (1998). Sex differences in symptom presentation associated with acute myocardial infarction: a population-based perspective. *American Heart Journal*, **136**:189–195.

Goossens, H., Ferech, M., Vander Stichele, R. *et al*. (2005). Outpatient antibiotic use in Europe and association with resistance: a cross-national database study. *Lancet*, **365**:579–87.

Gorelik, O., Almoznino-Sarafian, D., Yarovoi, I., Alon, I., Shteinshnaider, M., Tzur, I., Modai, D., Cohen, N. (2007). Patient-dependent variables affecting treatment and prediction of acute coronary syndrome are age-related. A study performed in Israel. *International Journal of Cardiology*, **121**:163–170.

Grijseels, E.W., Deckers, J.W., Hoes, A.W., Hartman, J.A., Van der Does, E., Van Loenen, E., Simoons, M.L. (1995). Pre-hospital triage of patients with suspected myocardial infarction. Evaluation of previously developed algorithms and new proposals. *European Heart Journal*, **16**:325–332.

Haasenritter, J., Aerts, M., Bösner, S., Buntinx, F., Burnand, B., Herzig, L., Knottnerus, J.A., Minalu, G., Nilsson, S., Renier, W., Sox, C., Sox, H., Donner-Banzhoff, N. (2012a). Coronary heart disease in primary care: accuracy of medical history and physical findings in patients with chest pains - a study protocol for a systematic review with individual patient data. *BMC Family Practice*, **13**:81.

Haasenritter, J., Bösner, S., Vaucher, P., Herzig, L., Heinzel-Gutenbrunner, M., Baum, E., Donner-Banzhoff, N. (2012b). Ruling out coronary artery disease in primary care: external validation of the Marburg Heart Score. *British Journal of General Practice*, **62(599)**,415–421.

Haasenritter, J., Nilsson, S., Aerts, M., Bösner, S., Buntinx, F., Burnard, B., Herzig, L., Knottnerus, A., Minalu, G., Renier, W., Sox, C., Sox, H., Donner-Banzhoff, N. (2012c). Diagnostic accuracy of signs and symptoms for diagnosing CHD in primary care. *Working paper*.

Hadgu, A. and Qu, Y. (1998). A biomedical application of latent class models with random effects. *Journal of Royal Statistical Society: Series C (Applied Statistics)*, **47(4)**:603–616.

Hagenaars, J.A., and McCutcheon, A.L. (2002). *Applied Latent Class Analysis*. Cambridge: Cambridge University Press.

Hall, C.B., Lipton, R.B., Sliwinski, M., Stewart, W.F. (2000). A change-point model for eastimating the onset of cognitive decline in preclinical Alzheimer's disease. *Statistics in Medicine*, **19**:1511–1566.

Hall, C.B., Ying, J., Kuo, L., Lipton R.B. (2003). Bayesian and profile likelihood change-point methods for modeling cognitive function over time. *Computational Statistics & Data Analysis*, **42**:91–109.

Harbord, R.M., Deeks, J.J., Egger, M., Whiting, P., Sterne, J.A. (2007). A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics*, **8**:239–251.

Hermann, C., Graf, K., Groh, A., Straube, E., Hartung, T. (2002). Comparison of eleven commercial tests for *Chlamydia pneumoniae* specific immunoglobulin G in asymptomatic healthy individuals. *Journal of Clinical Microbiology*, **40**:1603–1609.

Hens, N., Minalu, G., Goeyvaerts, N., Aerts, M., Mossong, J., Edmunds, J.W., Beutels, P. (2009). Estimating the Impact of School Closure on Social Mixing Behaviour and the Transmission of Close Contact Infections in Eight European Countries. *BMC Infectious Diseases*, **9**:187

Hens, N., Kvitkovicova, A., Aerts, M., Hlubinka, D., Beutels, P. (2010). Modelling distortions in seroprevalence data using change-point fractional polynomials. *Statistical Modelling*, **10(2)**:159–175.

Higgins, J.P.T., and Green, S. (2011). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0*. Chichester: John Wiley & Sons Ltd.

Hinkley, D.V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika*, **57**:1–17.

Hosmer, D.W., and Lemeshow, S. (2000). *Applied Logistic Regression*. New York : Wiley.

Hoymans, V.Y., Bosmans, J.M., Van Renterghem, L., Mak, R., Ursi, D., Wuyts, F., Vrints, C.J., Ieven, M. (2003). Importance of methodology in determination

of Chlamydia pneumoniae seropositivity in healthy subjects and in patients with coronary atherosclerosis. *Journal of Clinical Microbiology*, **41**:4049–4053.

Hunter, J.E., Schmidt, F.L., and Jackson, G.B. (1982). *Meta-analysis: cumulating research findings across studies*. Beverly Hills: Sage Publications.

Huttner, B., Goossens, H., Verheij, T., Harbarth, S. (2010). Characteristics and outcomes of public campaigns aimed at improving the use of antibiotics in outpatients in high-income countries. *The Lancet Infectious Diseases*, **10(1)**:17–31.

Khan, K.S., Bachmann, L.M., ter Riet, G. (2003). Systematic reviews with individual patient data meta-analysis to evaluate diagnostic tests. *European Journal of Obstetrics, Gynecology, and Reproductive Biology*, **108**:121–125.

Kiuchi, A.S., Hartigan, J.A., Holford, T.R., Rubinstein, P., Stevens, C.E. (1995). Change-points in the series of T4 counts prior to AIDS. *Biometrics*, **51**:236-248.

Klinkman, M.S., Stevens, D., Gorenflo, D.W. (1994). Episodes of care for chest pain. a preliminary report from MIRNET. Michigan Research Network. *Journal Family Practice*, **38(4)**:345–352.

Knottnerus, J.A., Muris, J.W. Assessment of the accuracy of diagnostic tests: the cross-sectional study. (2009). In The evidence base of clinical diagnosis. 2nd ed. edition. Edited by Knottnerus, J.A., Buntinx, F. Oxford: Hoboken NJ: Wiley-Blackwell Pub./BMJ Books.

Lange, N., Carlin, B.P., and Gelfand, A.E. (1992). Hierarchical bayes models for the progression of HIV infection using longitudinal CD4 T-cell numbers (with discussion). *Journal of American Statistical Association*, **87**:615–632.

Little, R. and Rubin, D. (1987). *Statistical Analysis with Missing Data.* New York: Wiley.

Liu H., Li G., Cumberland G.W. and Wu T. (2005).Testing Statistical Significance of the Area under a Receiving Operating Characteristics Curve for Repeated Measures Design with Bootstrapping. *Journal of Data Science*, **3**:257–278.

Liu, F.C., Chen, P.Y., Huang, F.L., Tsai, C.R., Lee, C.Y., Lin, C.F. (2008). Do serological tests provide adequate rapid diagnosis of *Mycoplasma pneumoniae* infection? *Japan Journal of Infectious Diseases*, **61**:397–399.

Loader, C. (1999). *Local regression and Likelihood. Statistics and Computing.* Springer-Verlag: New York.

Loens, K., Ursi, D., Ieven, M., van Aarle, P., Sillekens, P., Oudshoorn, P., Goossens, H.(2002). Detection of *Mycoplasma pneumoniae* in spiked clinical samples by nucleic acid sequence-based amplification. *Journal of Clinical Microbiology*, **40**:1339–1345.

Loens, K., Ursi, D., Goossens, H., Ieven, M.(2003). Molecular diagnosis of *Mycoplasma pneumoniae* respiratory tract infections. *Journal of Clinical Microbiology*, **41**:4915–4923.

Loens, K., Goossens, H., Ieven, M. (2010). Acute respiratory infection due to *Mycoplasma pneumoniae* current status of diagnostic methods. *European Journal of Clinical Microbiology & Infectious Diseases*, **29**:1055–1069.

Loens, K., Minalu, G., Hens, N., Ursi, D., Goossens, H., Aerts, M., and Ieven, M. (2012a). Importance of methodology in detection of *Mycoplasma pneumoniae* infections in patients with lower respiratory tract infections. *Working paper*.

Loens, K., Minalu, G., Hens, N., Ursi, D., Goossens, H., Aerts, M., and Ieven, M. (2012b). Evaluation of 2 amplification methods and 4 serological tests for the detection of *Chlamydophila pneumoniae* in patients with community-acquired pneumonia. *Working paper*.

Malhotra-Kumar, S., Lammens, C., Coenen, S. *et al*. (2007) Impact of azithromycin and clarithromycin therapy on pharyngeal carriage of macrolide-resistant streptococci among healthy volunteers: a randomised, double-blind, placebo-controlled study. *Lancet*, **369**:482–90.

Mant, J., McManus, R.J., Oakes, R.A., Delaney, B.C., Barton, P.M., Deeks, J.J., Hammersley, L., Davies, R.C., Davies, M.K., Hobbs, F.D. (2004). Systematic review and modelling of the investigation of acute and chronic chest pain presenting in primary care. *Health Technol Assess*, **8(2)**, iii: 1–158.

Mant, J., Doust, J., Roalfe, A., Barton, P., Cowie, M., Glasziou, P., Mant, D. *et al*. (2009). Systematic review and individual patient data meta-analysis of diagnosis of heart failure, with modelling of implications of different diagnostic strategies in primary care. *Health technology assessment (Winchester, England)*, **13**:1–232.

McCutcheon, A.C. (1987). *Latent class analysis*. Beverly Hills: Sage Publications.

McSweeney, J.C., Cody, M., O'Sullivan, P., Elberson, K., Moser, D.K., Garvin, B.J. (2003). Women's early warning symptoms of acute myocardial infarction. *Circulation*, **108**:2619–2623.

Mebis, J., Jansens, H., Minalu, G., Molenberghs, G., Schroyens, W., Gadisseur, A., Velde, A., Vrelust, I., Goossens, H., Berneman, Z. (2010). Long-term epidemiology of bacterial susceptibility profiles in adults suffering from febrile neutropenia with hematologic malignancy after antibiotic change. *Infection and Drug Resistance*, **3**: 53–61.

Menten, J., Boelaert, M., and Lesaffre, E.(2008). Bayesian latent class models with conditionally dependent diagnostic tests: A case study.*Statistics in Medicine*, **27**:4469–4488.

Minalu, G., Aerts, M., Coenen, S., Versporten, A., Muller, A., Adriaenssens, N., Beutels, P., Molenberghs, G., Goossens, H., Hens, N. (2011). Application of mixed-effects models to study the country-specific outpatient antibiotic use in Europe: a tutorial on longitudinal data analysis. *Journal of Antimicrobial Chemotherapy*, **66** Suppl 6:vi79–vi87.

Minalu, G., Aerts, M., Coenen, S., Versporten, A., Muller, A., Adriaenssens, N., Beutels, P., Molenberghs, G., Goossens, H., Hens, N. (2012a). Adaptive Change-point Mixed Models Applied to Data on Outpatient Antibiotic Use in Europe. Hasselt: University of Hasselt, Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BIOSTAT). *In revision*.

Minalu, G., Aerts, M., Bösner, S., Buntinx, F., Burnand, B., Haasenritter, J., Herzig, L., Knottnerus, J.A., Nilsson, S., Renier, W., Sox, C., Sox, H., Donner-Banzhoff, N. (2012b). Medical history and physical findings for coronary heart disease in primary care - a meta-analysis with individual patient data. *In revision*.

Molenberghs, G. and Verbeke, G. (2005).*Models for Discrete Longitudinal Data*. New York: Springer.

Montagnani, F., De Paolis, F., Beghetto, E., Gargano, N. (2010). Use of recombinant chimeric antigens for the serodiagnosis of *Mycoplasma pneumoniae* infection. *European Journal of Clinical Microbiology & Infectious Diseases*, **29**:1377–1386.

Morrell, H. (1998). Likelihood Ratio Testing of Variance Components in the Linear Mixed-Effects Model Using Restricted Maximum Likelihood.*Biometrics*, **54**:1560–8.

Muggeo, V.M.R. (2003). Estimating regression models with unknown break-points. *Statistics in Medicine*, **22**:3055–3071.

Narita, M. (2005). Evaluation of ELISA kits for detection of Mycoplasma pneumoniae specific IgG, IgA and IgM antibodies on the diagnosis of *M. pneumoniae* in children. *The Journal of Japanese Association for Infectious diseases*, **79**:457–463.

Neyman, J. and Pearson, E.S. (1933). On the problem of the most efficient tests of statistical haypothesis. *Philosophical Transactions of the Royal Society of London, Series A*, **231**:289–337.

Nilsson, S., Scheike, M., Engblom, D., Karlsson, L.G., Molstad, S., Akerlind, I., Ortoft, K., Nylander, E. (2003). Chest pain and ischaemic heart disease in primary care. *British Journal of General Practice*, **53(490)**:378–382.

Nilsson, A.C., Bjorkman, P., Persson, K. (2008). Polymerase chain reaction is superior to serology for the diagnosis of acute *Mycoplasma pneumoniae* infection and reveals a high rate of persistent infection. *BMC Microbiology*, **8**:93.

Nir-Paz, R., Michael-Gayego, A., Ron, M., Block, C. (2006). Evaluation of eight commercial tests for *Mycoplasma pneumoniae* antibodies in the absence of acute infection. *Clinical Microbiology and Infection*, **12**:685–688.

Otomo, S., Yamamura, J., Hayashi, E., Nakamura, T., Kakinuma, H., Nakamoto, Y., Takahashi, H., Karasawa, T. (2008). Analysis of children with *Chlamydophila (Chlamydia) pneumoniae* and *Mycoplasma pneumoniae* respiratory infections by real-time PCR assay and serological tests. *APMIS*, **116**:477–483.

Ozaki, T., Nishimura, N., Ahn, J., Watanabe, N., Muto, T., Saito, A., Koyama, N., Nakane, K., Funahashi, K. (2007). Utility of a rapid diagnosis kit for *Mycoplasma pneumoniae* in children, and the antimicrobial susceptibility of the isolates. *Journal of Infection and Chemotherapy*, **13**:204–207.

Pepe M.S., Cai T., and Longton G. (2006). Combining predictors for classification using the area under the receiver operating characteristic curve. *Biometrics*, **62**:221–229.

Persson, K., Bomans, J. (2000). Comparison of five serologic tests for diagnosis of acute infections by Chlamydia pneumoniae. *Clinical and Diagnostic Laboratory Immunology*, **7**:739–744.

Qu, Y., Tan, M., Kutner, M.H. (1996). Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics*, **52(3)**:797–810.

Riley, R.D., Dodd, S.R., Craig, J.V., Thompson. J.R., Williamson, P.R. (2008). Meta-analysis of diagnostic test studies using individual patient data and aggregate data. *Statistics in Medicine*, **27**:6111–6136.

Riley, R.D., Lambert, P.C., Abo-Zaid, G. (2010). Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ*, **340**:521–525.

Rubin, D.B. (1978). Multiple imputations in sample surveys - a phenomenological Bayesian approach to nonresponse. In Imputation and Editing of Faulty or Missing Survey Data Edited by U.S. Department of Commerce. Washington, DC: U.S. Department of Commerce, 1–23

Rubin, D.B. (1987). Multiple Imputation for Nonresponse in Surveys. New York: John Wiley & Sons.

Rutjes, A.W., Reitsma, J.B., Di Nisio, M., Smidt, N., van Rijn, J.C., Bossuyt, P.M. (2006). Evidence of bias and variation in diagnostic accuracy studies. *Canadian Medical Association Journal*, **174**:469–476.

Slate, E.H., Turnbull, B.W. (2007). Statistical models for longitudinal biomarkers of disease onset. *Statistics in Medicine*, **19**:617–637.

Smith, A.F.M. (1975). A Bayesian approach to inference about a change-point in a sequence of random variables. *Biometrika*, **62**:407–416.

Sox, H.C., Hickam, D.H., Marton, K.I., Moses, L., Skeff, K.M., Sox, C.H., Neal, E.A. (1990). Using the patient's history to estimate the probability of coronary artery disease: a comparison of primary care and referral practices. *American Journal of Medicine*, **89(1)**:7–14.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Linde A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*, **64(4)**:583–639.

Strobl, C., Boulesteix, A.L., Zeileis, A., Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, **8**:25.

Sturtz, S., Ligges, U., and Gelman, A. (2005). R2WinBUGS: A Package for Running WinBUGS from R. *Journal of Statistical Software*, **12(3)**, 1–16.

Svavarsdottir, A.E., Jonasson, M.R., Gudmundsson, G.H., Fjeldsted, K. (1996). Chest pain in family practice. Diagnosis and long-term outcome in a community setting. *Canadian Family Physician*, **42**: 1122–1128.

Van der Does, E., Lubsen, J., Pool, J. (1980). Acute myocardial infarction: an easy diagnosis in general practice? *Journal of the Royal College General Practitioners*, **30(216)**:405–409.

Verbeke, G., and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer.

Versporten, A., Coenen, S., Adriaenssens, N., Muller, A., Minalu, G., Faes, C., Vankerckhoven, V., Aerts, M., Hens, N., Molenberghs, G., Goossens, H. (2011a). European Surveillance on Antimicrobial Consumption (ESAC): outpatient penicillin use in Europe (1997–2009). *Journal of Antimicrobial Chemotherapy*, **66** Suppl 6:vi13–vi23.

Versporten, A., Coenen, S., Adriaenssens, N., Muller, A., Minalu, G., Faes, C., Vankerckhoven, V., Aerts, M., Hens, N., Molenberghs, G., Goossens, H. (2011b). European Surveillance on Antimicrobial Consumption (ESAC): outpatient cephalosporin use in Europe (1997–2009). *Journal of Antimicrobial Chemotherapy*, **66** Suppl 6:vi25–vi35.

Verdon, F., Herzig, L., Burnand, B., Bischoff, T., Pecoud, A., Junod, M., Muhlemann, N., Favrat, B. (2008). Chest pain in daily practice: occurrence, causes and management. *Swiss Medical Weekly*, **138(23–24)**: 340–347.

Whiting, P., Rutjes, A.W., Reitsma, J.B., Bossuyt, P.M., Kleijnen, J. (2003). The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Medical Research Methodology*, **3**:25.

Whiting, P.F., Weswood, M.E., Rutjes, A.W., Reitsma, J.B., Bossuyt, P.N., Kleijnen, J. (2006). Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. *BMC Medical Research Methodology*, **6**:9.

Woodhead, M., Blasi, F., Ewig, S., Garau, J., Huchon, G., Ieven, M., Ortqvist, A., Schaberg, T., Torres, A., van der Heijden, G., Read, R., Verheij, T.J. (2011). Guidelines for the management of adult lower respiratory tract infections-full version. *Clinical Microbiology & Infection*, **17** Suppl 6:E1–59.

# Appendices

# Appendix A

# Supplementary Materials

## I. Variables which have been included in the merged data set

**Table A.1:** List of variables which have been included in the merged data set based on studies that the INTERCHEST collaborators identified to date in five different countries.

| Predictor | Country | | | | |
|---|---|---|---|---|---|
| | Switzerland | Belgium | Sweden | USA | Germany |
| GP ID | X | - | - | X | X |
| Age patient | X | X | X | X | X |
| Sex male | X | X | X | X | X |
| Home visit | X | X | - | - | X |
| Pat known | X | - | - | X | X |
| Emergency | X | X | X | X | - |
| Main complaint | X | X | - | - | X |
| CP present | X | - | - | - | X |
| Previous CP | X | X | - | X | - |
| GPs concern | X | X | - | - | X |
| Prob CHD | - | - | X | - | X |
| Pat anxious | X | X | - | X | X |
| Pat concern heart | - | - | X | - | X |
| Loc retrosternal | X | X | - | X | X |
| Loc precardial | X | X | - | X | X |
| Loc left | X | - | - | X | X |
| Loc right | X | - | - | X | X |
| Rad band shaped | X | X | - | X | X |
| Rad left arm | X | X | - | X | X |
| Rad right arm | X | - | - | X | X |

Continued on Next Page...

Table A.1 – Continued

| | | | | | |
|---|---|---|---|---|---|
| Rad neck | X | X | - | X | X |
| Rad epigastric | X | X | - | X | X |
| Pain intensity | X | - | - | X | - |
| Stabbing | X | X | - | X | X |
| Oppressive | X | X | - | X | X |
| Burning | X | X | - | X | X |
| Dull | X | - | - | X | X |
| First occurance | - | X | X | X | X |
| Continuous pain | X | X | - | X | X |
| Pain duration | X | - | - | X | X |
| Frequency | - | - | - | X | X |
| Course | X | - | - | X | - |
| Typical angina | - | X | - | X | - |
| Rel breath | X | X | - | X | X |
| Rel move | X | X | - | X | X |
| Rel swallow | X | X | - | X | - |
| Rel ingest | X | - | - | - | X |
| Rel effort | X | X | - | X | X |
| Rel cough | X | X | - | X | - |
| Rel position | X | - | - | X | - |
| Rel ntg | - | X | - | X | - |
| Rel antacid | - | X | - | X | - |
| Add fever | X | X | - | X | - |
| Add cough | X | X | - | X | X |
| Add dyspnoe | X | X | - | X | X |
| Add sweating | X | X | - | X | X |
| Add pale | X | X | - | - | X |
| Add nausea | - | X | - | X | X |
| Add cold | - | - | - | X | X |
| Add sputum | X | X | - | X | - |
| Add red cons | X | - | - | X | - |
| HR con | - | X | - | - | X |
| HR ord | X | X | - | - | X |
| RR sys con | - | X | - | - | X |
| RR dia con | - | X | - | - | X |
| RR ord | - | X | - | - | X |
| Heart rhyth | X | X | - | - | - |
| Auscult pulmo | X | X | - | X | - |
| Auscult heart | X | - | - | X | - |
| Reprod palpation | X | X | - | X | X |
| RF dyslipid | - | - | - | X | X |
| RF diabetes | - | - | - | X | X |
| RF family | - | X | - | X | X |

Continued on Next Page...

Table A.1 – Continued

| | | | | | |
|---|---|---|---|---|---|
| RF hypertension | - | - | - | X | X |
| RF smoking | - | X | - | X | X |
| Prior heart attack | - | - | - | X | - |
| Known CHD | - | - | - | X | X |
| RF all con | - | - | - | X | X |
| RF all ord | X | - | - | X | X |
| RD any CAD† | X | X | X | X | X |
| Number of patients | 644 | 299 | 523 | 395 | 1238 |

X: The regarding variable/ question was asked in the individual study, CP: chest pain, CHD: coronary heart disease, GP: general practitioner, RF: risk factor,

† Reference diagnosis: Coronary heart disease (cases with stable CHD and cases with an acute coronary syndrome).

## II. Interchest survey results

**Table A.2:** Interchest survey results.

| Clinical finding | Median |
|---|---|
| typical angina | 4 |
| related to effort | 4 |
| history of heart attack | 4 |
| history of angina | 4 |
| history of revascularization. | 3.5 |
| GPs concern: something serious | 3 |
| rad:left arm | 3 |
| oppressive | 3 |
| pain relief: NTG | 3 |
| history of stroke | 3 |
| history of pad | 2.75 |
| higher age | 2.5 |
| loc: retrosternal | 2.5 |
| risk factor diabetes m | 2.5 |
| emergency | 2 |
| CP was maincomplaint | 2 |
| rad: band-shaped | 2 |
| rad: neck | 2 |
| add. symp: sweating | 2 |
| add. symp: pale | 2 |
| arrhythmia | 2 |
| risk factor dyslipidaemia | 2 |
| risk factor family history of MI | 2 |
| risk factor hypertension | 2 |
| risk factor smoking | 2 |
| pain intensity | 1.5 |
| patient anxious | 1.25 |
| sex male | 1 |
| CP present during consultation | 1 |
| CP in the past | 1 |
| patient concern: related to heart | 1 |
| loc: precardial | 1 |
| duration of episode less than 10m | 1 |
| duration of episode 10 to 30m | 1 |
| hypotension | 1 |
| auscultation heart: abnormal findings | 1 |
| atypical angina | 0.75 |
| loc: left side | 0.5 |
| add. symp: dyspnoe | 0.5 |

Continued on Next Page...

Table A.2 – Continued

| | |
|---|---|
| tachycardia | 0.5 |
| bradycardia | 0.5 |
| rad: epigastric | 0.25 |
| duration of episode 30 to 60m | 0.25 |
| home visit required | 0 |
| rad: right arm | 0 |
| burning | 0 |
| dull | 0 |
| first occurrence less than 24h | 0 |
| first occurrence more than 24h | 0 |
| continuous pain | 0 |
| freqency: less than 2/d | 0 |
| frequency: more than 2/d | 0 |
| pain increases with time | 0 |
| add. symp: nausea | 0 |
| add. symp: red.consciousness | 0 |
| hypertension | 0 |
| auscultation lung: abnormal findings | 0 |
| duration of episode hours | -0.5 |
| loc: right side | -1 |
| unspecific CP | -1 |
| add. symp: cough | -1.25 |
| stabbing | -2 |
| add. symp: fever | -2 |
| add. symp: cold | -2 |
| related to breathing | -3 |
| related to movement | -3 |
| related to swallowing | -3 |
| related to ingestion | -3 |
| related to cough | -3 |
| pain relief: antacid | -3 |
| add. symp: sputum | -3 |
| pain reprod. by palpation | -3 |
| related to body position | -4 |

# Appendix B

# Simulation Results

In this appendix, we present additional tables and figures for the simulation studies discussed in Chapter 6.

## B.1 Simulation Results where 250 Datasets are Generated Under Models 1 and 2 with N=201

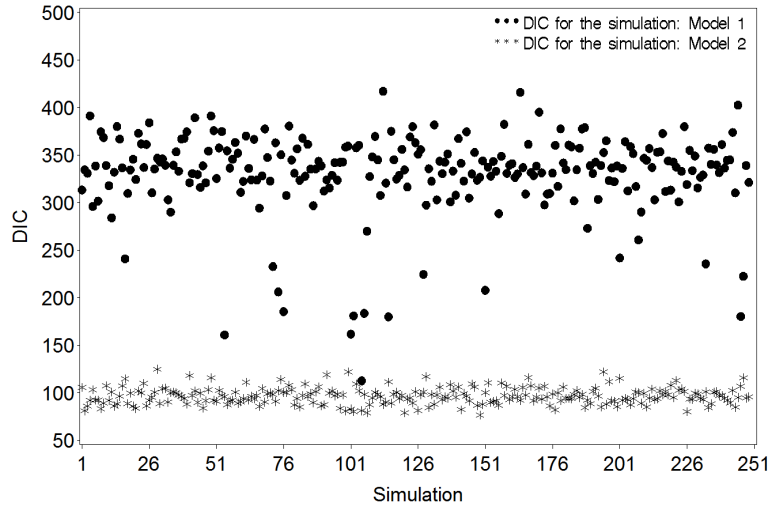



**Figure B.1:** DIC-values (dots and stars) for the simulation runs (data are generated under Model 2 with a high degree of dependency).

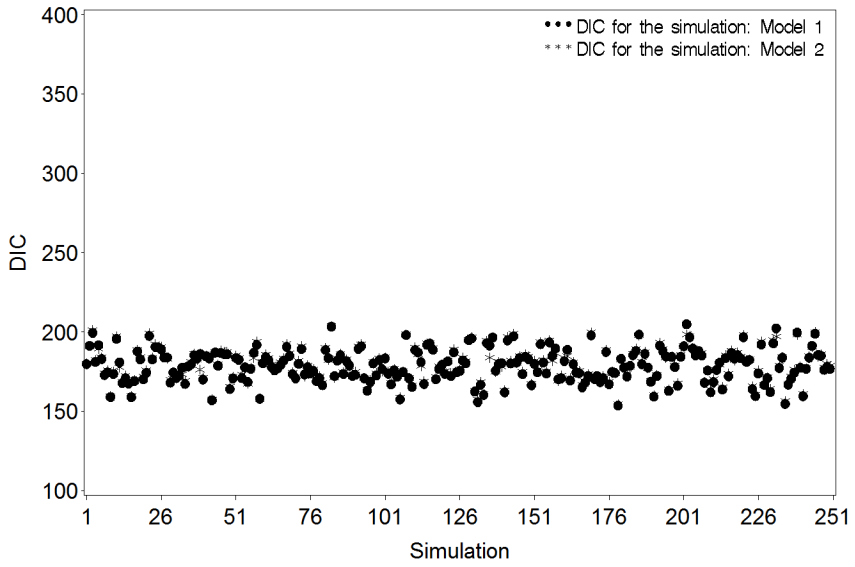



**Figure B.2:** DIC-values (dots and stars) for the simulation runs (data are generated under Model 1 using different true values). The percentage of the DIC-values for Model 1 smaller less than the DIC-values for Model 2 is 81%.

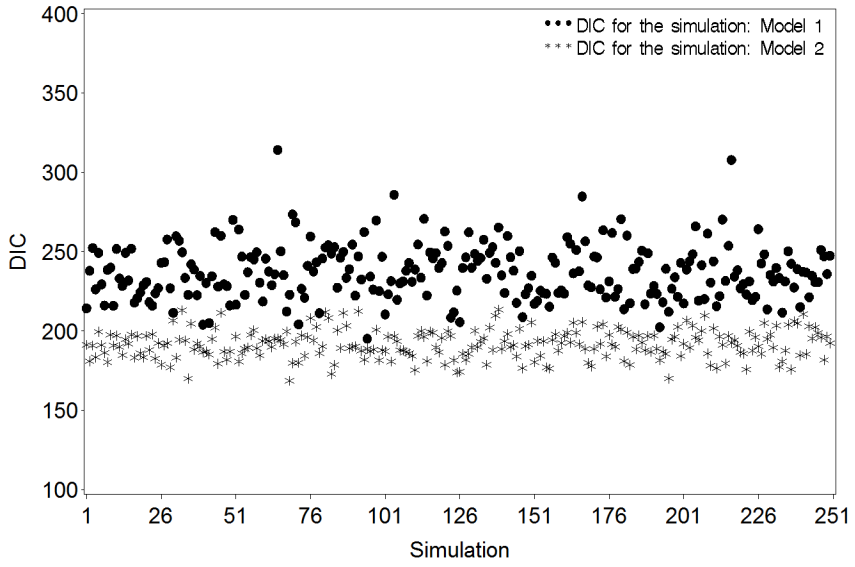



**Figure B.3:** DIC-values (dots and stars) for the simulation runs (data are generated under Model 2 with a moderate degree of dependency and using different true values).

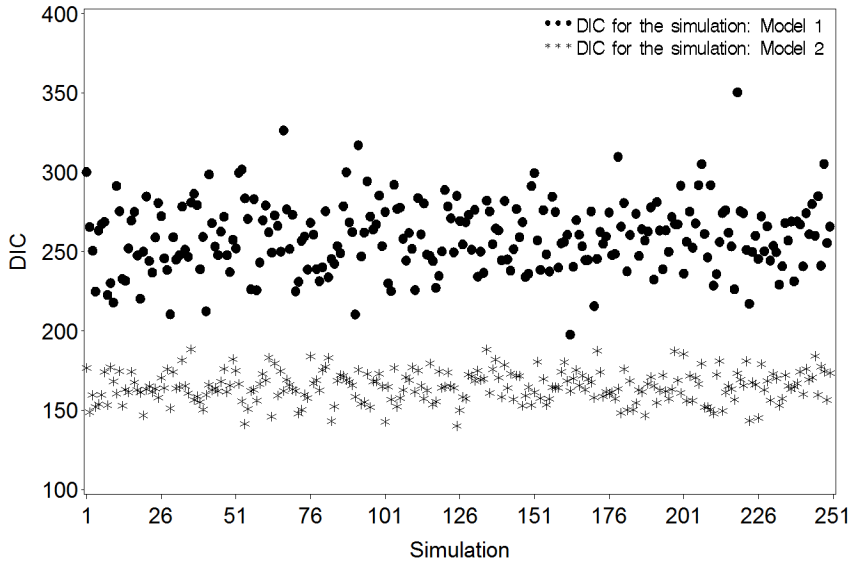



**Figure B.4:** DIC-values (dots and stars) for the simulation runs (data are generated under Model 2 with a high degree of dependency and using different true values).

The values of the joint probabilities for the simulation studies discussed in Sections 6.1 and 6.2 are given in Table B.1; while the values of the joint probabilities for the simulation studies discussed in Sections 6.3 and 6.4 are given in Table B.2.

**Table B.1:** The values of the joint probabilities for each outcome patterns for Scenario 1, Scenario 2A, Scenario 2B and Scenario 2C.

| | Outcome patterns | | | | | | Joint probabilities | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $T_1$ | $T_2$ | $T_3$ | $T_4$ | $T_5$ | $T_6$ | 1 | 2A | 2B | 2C |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.00084 | 0.00053 | 0.00786 | 0.02234 |
| 2 | 1 | 1 | 1 | 1 | 1 | 0 | 0.00002 | 0.00001 | 0.00014 | 0.00062 |
| 3 | 1 | 1 | 1 | 1 | 0 | 1 | 0.00001 | 0.00001 | 0.00009 | 0.00028 |
| 4 | 1 | 1 | 1 | 1 | 0 | 0 | 0.00018 | 0.00006 | 0.00082 | 0.01069 |
| 5 | 1 | 1 | 1 | 0 | 1 | 1 | 0.00028 | 0.00018 | 0.00259 | 0.00739 |
| 6 | 1 | 1 | 1 | 0 | 1 | 0 | 0.00010 | 0.00003 | 0.00047 | 0.00569 |
| 7 | 1 | 1 | 1 | 0 | 0 | 1 | 0.00001 | 0.00001 | 0.00006 | 0.00046 |
| 8 | 1 | 1 | 1 | 0 | 0 | 0 | 0.00349 | 0.00119 | 0.01634 | 0.21317 |
| 9 | 1 | 1 | 0 | 1 | 1 | 1 | 0.00393 | 0.00307 | 0.00503 | 0.00293 |
| 10 | 1 | 1 | 0 | 1 | 1 | 0 | 0.00008 | 0.00006 | 0.00010 | 0.00006 |
| 11 | 1 | 1 | 0 | 1 | 0 | 1 | 0.00005 | 0.00004 | 0.00006 | 0.00004 |
| 12 | 1 | 1 | 0 | 1 | 0 | 0 | 0.00073 | 0.00032 | 0.00070 | 0.00058 |
| 13 | 1 | 1 | 0 | 0 | 1 | 1 | 0.00130 | 0.00101 | 0.00166 | 0.00097 |
| 14 | 1 | 1 | 0 | 0 | 1 | 0 | 0.00040 | 0.00018 | 0.00039 | 0.00032 |
| 15 | 1 | 1 | 0 | 0 | 0 | 1 | 0.00004 | 0.00002 | 0.00004 | 0.00003 |
| 16 | 1 | 1 | 0 | 0 | 0 | 0 | 0.01454 | 0.00640 | 0.01397 | 0.01156 |
| 17 | 1 | 0 | 1 | 1 | 1 | 1 | 0.00109 | 0.00040 | 0.00065 | 0.00038 |
| 18 | 1 | 0 | 1 | 1 | 1 | 0 | 0.00002 | 0.00001 | 0.00001 | 0.00001 |
| 19 | 1 | 0 | 1 | 1 | 0 | 1 | 0.00001 | 0.00001 | 0.00001 | 0.00001 |
| 20 | 1 | 0 | 1 | 1 | 0 | 0 | 0.00022 | 0.00006 | 0.00012 | 0.00010 |
| 21 | 1 | 0 | 1 | 0 | 1 | 1 | 0.00036 | 0.00013 | 0.00022 | 0.00013 |
| 22 | 1 | 0 | 1 | 0 | 1 | 0 | 0.00012 | 0.00003 | 0.00007 | 0.00006 |
| 23 | 1 | 0 | 1 | 0 | 0 | 1 | 0.00001 | 0.00001 | 0.00001 | 0.00001 |
| 24 | 1 | 0 | 1 | 0 | 0 | 0 | 0.00438 | 0.00113 | 0.00247 | 0.00204 |
| 25 | 1 | 0 | 0 | 1 | 1 | 1 | 0.00511 | 0.00292 | 0.00180 | 0.00066 |
| 26 | 1 | 0 | 0 | 1 | 1 | 0 | 0.00010 | 0.00005 | 0.00004 | 0.00001 |
| 27 | 1 | 0 | 0 | 1 | 0 | 1 | 0.00006 | 0.00004 | 0.00002 | 0.00001 |
| 28 | 1 | 0 | 0 | 1 | 0 | 0 | 0.00092 | 0.00037 | 0.00030 | 0.00012 |
| 29 | 1 | 0 | 0 | 0 | 1 | 1 | 0.00169 | 0.00096 | 0.00060 | 0.00022 |
| 30 | 1 | 0 | 0 | 0 | 1 | 0 | 0.00050 | 0.00021 | 0.00017 | 0.00006 |
| 31 | 1 | 0 | 0 | 0 | 0 | 1 | 0.00005 | 0.00002 | 0.00002 | 0.00001 |
| 32 | 1 | 0 | 0 | 0 | 0 | 0 | 0.01822 | 0.00740 | 0.00600 | 0.00232 |
| 33 | 0 | 1 | 1 | 1 | 1 | 1 | 0.00393 | 0.00298 | 0.00489 | 0.00285 |
| 34 | 0 | 1 | 1 | 1 | 1 | 0 | 0.00015 | 0.00011 | 0.00021 | 0.00015 |
| 35 | 0 | 1 | 1 | 1 | 0 | 1 | 0.00005 | 0.00004 | 0.00007 | 0.00004 |
| 36 | 0 | 1 | 1 | 1 | 0 | 0 | 0.00348 | 0.00230 | 0.00503 | 0.00417 |
| 37 | 0 | 1 | 1 | 0 | 1 | 1 | 0.00130 | 0.00099 | 0.00162 | 0.00095 |
| 38 | 0 | 1 | 1 | 0 | 1 | 0 | 0.00182 | 0.00122 | 0.00265 | 0.00219 |
| 39 | 0 | 1 | 1 | 0 | 0 | 1 | 0.00014 | 0.00009 | 0.00019 | 0.00016 |

Continued on Next Page...

Table B.1 – Continued

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 40 | 0 | 1 | 1 | 0 | 0 | 0 | 0.06938 | 0.04595 | 0.10036 | 0.08308 |
| 41 | 0 | 1 | 0 | 1 | 1 | 1 | 0.01835 | 0.02184 | 0.01348 | 0.00493 |
| 42 | 0 | 1 | 0 | 1 | 1 | 0 | 0.00066 | 0.00073 | 0.00053 | 0.00020 |
| 43 | 0 | 1 | 0 | 1 | 0 | 1 | 0.00024 | 0.00028 | 0.00018 | 0.00007 |
| 44 | 0 | 1 | 0 | 1 | 0 | 0 | 0.01450 | 0.01509 | 0.01224 | 0.00473 |
| 45 | 0 | 1 | 0 | 0 | 1 | 1 | 0.00606 | 0.00723 | 0.00446 | 0.00163 |
| 46 | 0 | 1 | 0 | 0 | 1 | 0 | 0.00760 | 0.00798 | 0.00645 | 0.00249 |
| 47 | 0 | 1 | 0 | 0 | 0 | 1 | 0.00057 | 0.00061 | 0.00048 | 0.00018 |
| 48 | 0 | 1 | 0 | 0 | 0 | 0 | 0.28877 | 0.30094 | 0.24399 | 0.09435 |
| 49 | 0 | 0 | 1 | 1 | 1 | 1 | 0.00511 | 0.00284 | 0.00175 | 0.00064 |
| 50 | 0 | 0 | 1 | 1 | 1 | 0 | 0.00019 | 0.00011 | 0.00008 | 0.00003 |
| 51 | 0 | 0 | 1 | 1 | 0 | 1 | 0.00007 | 0.00004 | 0.00002 | 0.00001 |
| 52 | 0 | 0 | 1 | 1 | 0 | 0 | 0.00437 | 0.00266 | 0.00216 | 0.00084 |
| 53 | 0 | 0 | 1 | 0 | 1 | 1 | 0.00169 | 0.00094 | 0.00058 | 0.00021 |
| 54 | 0 | 0 | 1 | 0 | 1 | 0 | 0.00229 | 0.00140 | 0.00114 | 0.00044 |
| 55 | 0 | 0 | 1 | 0 | 0 | 1 | 0.00017 | 0.00010 | 0.00008 | 0.00003 |
| 56 | 0 | 0 | 1 | 0 | 0 | 0 | 0.08697 | 0.05314 | 0.04308 | 0.01666 |
| 57 | 0 | 0 | 0 | 1 | 1 | 1 | 0.02387 | 0.02688 | 0.02601 | 0.02674 |
| 58 | 0 | 0 | 0 | 1 | 1 | 0 | 0.00084 | 0.00098 | 0.00095 | 0.00097 |
| 59 | 0 | 0 | 0 | 1 | 0 | 1 | 0.00031 | 0.00036 | 0.00034 | 0.00035 |
| 60 | 0 | 0 | 0 | 1 | 0 | 0 | 0.01817 | 0.02169 | 0.02118 | 0.02133 |
| 61 | 0 | 0 | 0 | 0 | 1 | 1 | 0.00788 | 0.00890 | 0.00861 | 0.00885 |
| 62 | 0 | 0 | 0 | 0 | 1 | 0 | 0.00953 | 0.01144 | 0.01118 | 0.01126 |
| 63 | 0 | 0 | 0 | 0 | 0 | 1 | 0.00072 | 0.00086 | 0.00084 | 0.00084 |
| 64 | 0 | 0 | 0 | 0 | 0 | 0 | 0.36199 | 0.43244 | 0.42236 | 0.42539 |

**Table B.2:** The values of the joint probabilities for each outcome patterns for Scenario 3A, Scenario 3B and Scenario 3C.

| | Outcome patterns | | | | | | Joint probabilities | | |
|---|---|---|---|---|---|---|---|---|---|
| | T1 | T2 | T3 | T4 | T5 | T6 | 3A | 3B | 3C |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.09302 | 0.09015 | 0.08845 |
| 2 | 1 | 1 | 1 | 1 | 1 | 0 | 0.01631 | 0.01629 | 0.01747 |
| 3 | 1 | 1 | 1 | 1 | 0 | 1 | 0.01039 | 0.01110 | 0.01413 |
| 4 | 1 | 1 | 1 | 1 | 0 | 0 | 0.00198 | 0.00630 | 0.01994 |
| 5 | 1 | 1 | 1 | 0 | 1 | 1 | 0.02518 | 0.02530 | 0.02763 |
| 6 | 1 | 1 | 1 | 0 | 1 | 0 | 0.00455 | 0.00829 | 0.02032 |
| 7 | 1 | 1 | 1 | 0 | 0 | 1 | 0.00311 | 0.01117 | 0.03660 |
| 8 | 1 | 1 | 1 | 0 | 0 | 0 | 0.00180 | 0.03602 | 0.14306 |
| 9 | 1 | 1 | 0 | 1 | 1 | 1 | 0.02401 | 0.01471 | 0.00555 |
| 10 | 1 | 1 | 0 | 1 | 1 | 0 | 0.00427 | 0.00275 | 0.00107 |
| 11 | 1 | 1 | 0 | 1 | 0 | 1 | 0.00281 | 0.00200 | 0.00082 |
| 12 | 1 | 1 | 0 | 1 | 0 | 0 | 0.00107 | 0.00184 | 0.00098 |
| 13 | 1 | 1 | 0 | 0 | 1 | 1 | 0.00661 | 0.00429 | 0.00168 |
| 14 | 1 | 1 | 0 | 0 | 1 | 0 | 0.00167 | 0.00207 | 0.00104 |
| 15 | 1 | 1 | 0 | 0 | 0 | 1 | 0.00184 | 0.00333 | 0.00180 |
| 16 | 1 | 1 | 0 | 0 | 0 | 0 | 0.00483 | 0.01220 | 0.00689 |
| 17 | 1 | 0 | 1 | 1 | 1 | 1 | 0.01522 | 0.00935 | 0.00353 |
| 18 | 1 | 0 | 1 | 1 | 1 | 0 | 0.00276 | 0.00189 | 0.00076 |
| 19 | 1 | 0 | 1 | 1 | 0 | 1 | 0.00191 | 0.00159 | 0.00070 |
| 20 | 1 | 0 | 1 | 1 | 0 | 0 | 0.00120 | 0.00252 | 0.00139 |
| 21 | 1 | 0 | 1 | 0 | 1 | 1 | 0.00430 | 0.00300 | 0.00122 |
| 22 | 1 | 0 | 1 | 0 | 1 | 0 | 0.00152 | 0.00251 | 0.00134 |
| 23 | 1 | 0 | 1 | 0 | 0 | 1 | 0.00215 | 0.00463 | 0.00257 |
| 24 | 1 | 0 | 1 | 0 | 0 | 0 | 0.00717 | 0.01833 | 0.01036 |
| 25 | 1 | 0 | 0 | 1 | 1 | 1 | 0.00401 | 0.00799 | 0.00494 |
| 26 | 1 | 0 | 0 | 1 | 1 | 0 | 0.00106 | 0.00165 | 0.00097 |
| 27 | 1 | 0 | 0 | 1 | 0 | 1 | 0.00121 | 0.00143 | 0.00077 |
| 28 | 1 | 0 | 0 | 1 | 0 | 0 | 0.00333 | 0.00246 | 0.00105 |
| 29 | 1 | 0 | 0 | 0 | 1 | 1 | 0.00175 | 0.00263 | 0.00153 |
| 30 | 1 | 0 | 0 | 0 | 1 | 0 | 0.00307 | 0.00242 | 0.00107 |
| 31 | 1 | 0 | 0 | 0 | 0 | 1 | 0.00619 | 0.00454 | 0.00192 |
| 32 | 1 | 0 | 0 | 0 | 0 | 0 | 0.02551 | 0.01810 | 0.00746 |
| 33 | 0 | 1 | 1 | 1 | 1 | 1 | 0.01496 | 0.00925 | 0.00351 |
| 34 | 0 | 1 | 1 | 1 | 1 | 0 | 0.00283 | 0.00215 | 0.00091 |
| 35 | 0 | 1 | 1 | 1 | 0 | 1 | 0.00211 | 0.00217 | 0.00104 |
| 36 | 0 | 1 | 1 | 1 | 0 | 0 | 0.00218 | 0.00505 | 0.00282 |
| 37 | 0 | 1 | 1 | 0 | 1 | 1 | 0.00443 | 0.00349 | 0.00151 |
| 38 | 0 | 1 | 1 | 0 | 1 | 0 | 0.00238 | 0.00475 | 0.00261 |
| 39 | 0 | 1 | 1 | 0 | 0 | 1 | 0.00397 | 0.00935 | 0.00524 |

Continued on Next Page...

Table B.2 – Continued

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 40 | 0 | 1 | 1 | 0 | 0 | 0 | 0.01487 | 0.03817 | 0.02159 |
| 41 | 0 | 1 | 0 | 1 | 1 | 1 | 0.00403 | 0.00791 | 0.00488 |
| 42 | 0 | 1 | 0 | 1 | 1 | 0 | 0.00144 | 0.00191 | 0.00107 |
| 43 | 0 | 1 | 0 | 1 | 0 | 1 | 0.00205 | 0.00201 | 0.00101 |
| 44 | 0 | 1 | 0 | 1 | 0 | 0 | 0.00687 | 0.00497 | 0.00208 |
| 45 | 0 | 1 | 0 | 0 | 1 | 1 | 0.00247 | 0.00312 | 0.00172 |
| 46 | 0 | 1 | 0 | 0 | 1 | 0 | 0.00620 | 0.00463 | 0.00198 |
| 47 | 0 | 1 | 0 | 0 | 0 | 1 | 0.01277 | 0.00920 | 0.00384 |
| 48 | 0 | 1 | 0 | 0 | 0 | 0 | 0.05317 | 0.03770 | 0.01553 |
| 49 | 0 | 0 | 1 | 1 | 1 | 1 | 0.00271 | 0.00512 | 0.00314 |
| 50 | 0 | 0 | 1 | 1 | 1 | 0 | 0.00159 | 0.00169 | 0.00087 |
| 51 | 0 | 0 | 1 | 1 | 0 | 1 | 0.00271 | 0.00228 | 0.00105 |
| 52 | 0 | 0 | 1 | 1 | 0 | 0 | 0.01030 | 0.00736 | 0.00305 |
| 53 | 0 | 0 | 1 | 0 | 1 | 1 | 0.00282 | 0.00286 | 0.00146 |
| 54 | 0 | 0 | 1 | 0 | 1 | 0 | 0.00920 | 0.00667 | 0.00279 |
| 55 | 0 | 0 | 1 | 0 | 0 | 1 | 0.01916 | 0.01367 | 0.00566 |
| 56 | 0 | 0 | 1 | 0 | 0 | 0 | 0.08018 | 0.05682 | 0.02340 |
| 57 | 0 | 0 | 0 | 1 | 1 | 1 | 0.00162 | 0.01507 | 0.04557 |
| 58 | 0 | 0 | 0 | 1 | 1 | 0 | 0.00428 | 0.00622 | 0.01141 |
| 59 | 0 | 0 | 0 | 1 | 0 | 1 | 0.00884 | 0.00944 | 0.01249 |
| 60 | 0 | 0 | 0 | 1 | 0 | 0 | 0.03683 | 0.03328 | 0.03245 |
| 61 | 0 | 0 | 0 | 0 | 1 | 1 | 0.00794 | 0.01080 | 0.01875 |
| 62 | 0 | 0 | 0 | 0 | 1 | 0 | 0.03266 | 0.02991 | 0.03010 |
| 63 | 0 | 0 | 0 | 0 | 0 | 1 | 0.06861 | 0.06190 | 0.06017 |
| 64 | 0 | 0 | 0 | 0 | 0 | 0 | 0.28804 | 0.25824 | 0.24728 |

## B.2 Simulation Results where 250 Datasets are Generated Under Models 1 and 2 with N=500

### I. Scenario 3A: Data are generated using the conditional independence model

In this scenario, we have conducted the simulation study discussed in Section 6.3 with N=500. The simulation averages for the posterior means, simulation standard errors and simulation averages of the estimated standard errors obtained by fitting Models 1 and 2 are given in Tables B.3 and B.4.

**Table B.3:** Parameter estimates: simulation averages for the posterior means, simulation standard errors and simulation averages of the estimated standard errors for the parameters obtained by fitting Models 1 and 2 (data are generated under Model 1 with N=500 and using different true values).

| Parameters | True Values | Model 1 | Model 2 |
|---|---|---|---|
| $\alpha_{11}$ | 1.831 | 1.9113(0.2919)(0.2972) | 2.0018(0.3135)(0.3218) |
| $\alpha_{21}$ | 1.812 | 1.8413(0.2873)(0.2810) | 1.9284(0.3078)(0.3061) |
| $\alpha_{31}$ | 1.355 | 1.3959(0.2359)(0.2349) | 1.4645(0.2571)(0.2585) |
| $\alpha_{41}$ | 1.308 | 1.3461(0.2435)(0.2339) | 1.3541(0.2458)(0.2367) |
| $\alpha_{51}$ | 2.196 | 2.2886(0.3830)(0.3565) | 2.3208(0.3984)(0.3704) |
| $\alpha_{61}$ | 1.742 | 1.7890(0.2891)(0.2741) | 1.8000(0.2934)(0.2772) |
| $\alpha_{10}$ | -2.425 | -2.4512(0.1921)(0.2054) | -2.5189(0.2071)(0.2185) |
| $\alpha_{20}$ | -1.690 | -1.6962(0.1545)(0.1492) | -1.7524(0.1617)(0.1627) |
| $\alpha_{30}$ | -1.279 | -1.2898(0.1294)(0.1294) | -1.3345(0.1351)(0.1413) |
| $\alpha_{40}$ | -2.057 | -2.0708(0.1791)(0.1719) | -2.0777(0.1797)(0.1736) |
| $\alpha_{50}$ | -2.178 | -2.1956(0.1846)(0.1838) | -2.2083(0.1868)(0.1874) |
| $\alpha_{60}$ | -1.435 | -1.4330(0.1317)(0.1361) | -1.4372(0.1330)(0.1368) |
| $\beta_{10}$ | | | 0.3731(0.1357)(0.2205) |
| $\beta_{11}$ | | | 0.5101(0.2015)(0.3154) |

The results in Table B.3 indicate that there are differences between the parameter estimates obtained by fitting Models 1 and 2. The parameter estimates obtained by fitting Model 1 are close to the true values. The simulation averages of the estimated standard errors are close to the simulation standard errors.

**Table B.4:** Parameter estimates: simulation averages for the posterior means, simulation standard errors and simulation averages of the estimated standard errors for the prevalence, sensitivities and specificities obtained by fitting Models 1 and 2 (data are generated under Model 1 with N=500 and using different true values).

| Parameters | True Values | Model 1 | Model 2 |
|---|---|---|---|
| Prev. | 0.2620 | 0.2619(0.0212)(0.0206) | 0.2622(0.0211)(0.0208) |
| Sens. $T_1$ | 0.8619 | 0.8642(0.0310)(0.0327) | 0.8599(0.0315)(0.0338) |
| Sens. $T_2$ | 0.8596 | 0.8564(0.0334)(0.0327) | 0.8518(0.0338)(0.0337) |
| Sens. $T_3$ | 0.7949 | 0.7964(0.0364)(0.0368) | 0.7911(0.0359)(0.0377) |
| Sens. $T_4$ | 0.7872 | 0.7882(0.0380)(0.0377) | 0.7894(0.0382)(0.0379) |
| Sens. $T_5$ | 0.8999 | 0.8993(0.0297)(0.0290) | 0.9015(0.0299)(0.0291) |
| Sens. $T_6$ | 0.8509 | 0.8501(0.0339)(0.0331) | 0.8513(0.0340)(0.0331) |
| Spec. $T_1$ | 0.9187 | 0.9183(0.0140)(0.0150) | 0.9171(0.0140)(0.0152) |
| Spec. $T_2$ | 0.8442 | 0.8430(0.0202)(0.0194) | 0.8416(0.0203)(0.0197) |
| Spec. $T_3$ | 0.7823 | 0.7825(0.0217)(0.0218) | 0.7810(0.0218)(0.0221) |
| Spec. $T_4$ | 0.8867 | 0.8857(0.0176)(0.0170) | 0.8864(0.0176)(0.0171) |
| Spec. $T_5$ | 0.8983 | 0.8975(0.0164)(0.0165) | 0.8985(0.0163)(0.0166) |
| Spec. $T_6$ | 0.8077 | 0.8057(0.0201)(0.0211) | 0.8063(0.0202)(0.0211) |

The estimates for prevalence, sensitivities and specificities given in Table B.4 indicate that there is no much difference between the estimates obtained by fitting Models 1 and 2. The DIC-values for the simulation runs are shown in Figure B.5.

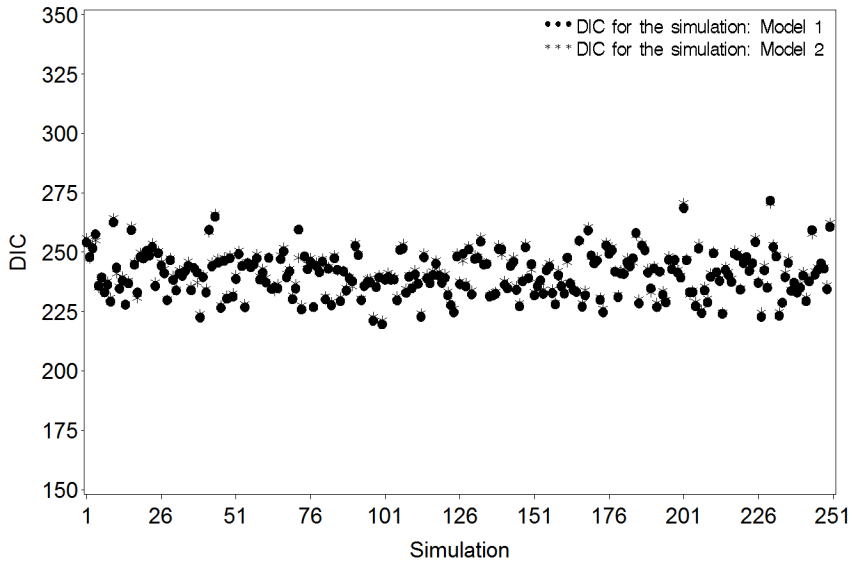



**Figure B.5:** DIC-values (dots and stars) for the simulation runs (data are generated under Model 1 with N=500 and using different true values). The percentage of the DIC-values for Model 1 smaller than the DIC-values for Model 2 is 86%.

## II. Scenario 3B: Data are generated using the conditional dependence model with a moderate dependency

In this scenario, we have conducted the simulation study discussed in Section 6.4.1 with N=500. The simulation results obtained by fitting Models 1 and 2 are given in Tables B.5 indicate that there is a difference between the parameter estimates obtained by fitting Models 1 and 2. There is no much difference between the estimated and empirical standard errors. The results obtained by Model 1 are very close to the true values. Model 2 estimates the dependency parameters very well.

**Table B.5:** Parameter estimates: simulation averages for the posterior means, simulation standard errors and simulation averages of the estimated standard errors for the parameters obtained by fitting Models 1 and 2 (data are generated under Model 2 with a moderate degree of dependency, N=500 and using different true values).

| Parameters | True Values | Model 1 | Model 2 |
|---|---|---|---|
| $\alpha_{11}$ | 1.831 | 1.7143(0.3763)(0.3144) | 2.0421(0.4320)(0.4226) |
| $\alpha_{21}$ | 1.812 | 1.7423(0.3366)(0.2954) | 1.9936(0.3994)(0.4067) |
| $\alpha_{31}$ | 1.355 | 1.4355(0.2863)(0.2500) | 1.5520(0.3629)(0.3667) |
| $\alpha_{41}$ | 1.308 | 0.8548(0.3240)(0.2422) | 1.3507(0.2699)(0.2774) |
| $\alpha_{51}$ | 2.196 | 1.3408(0.4493)(0.3156) | 2.4088(0.5840)(0.5808) |
| $\alpha_{61}$ | 1.742 | 1.2413(0.3945)(0.2770) | 1.8150(0.3351)(0.3335) |
| $\alpha_{10}$ | -2.425 | -1.9276(0.2812)(0.2182) | -2.3557(0.2310)(0.2480) |
| $\alpha_{20}$ | -1.690 | -1.3068(0.2033)(0.1599) | -1.6062(0.2003)(0.2116) |
| $\alpha_{30}$ | -1.279 | -0.9916(0.1704)(0.1406) | -1.1797(0.1832)(0.1953) |
| $\alpha_{40}$ | -2.057 | -1.9864(0.1915)(0.1843) | -2.0717(0.1995)(0.1900) |
| $\alpha_{50}$ | -2.178 | -2.0455(0.2052)(0.1944) | -2.2072(0.2283)(0.2189) |
| $\alpha_{60}$ | -1.435 | -1.3973(0.1416)(0.1457) | -1.4398(0.1404)(0.1468) |
| $\beta_{10}$ | 2.000 | | 1.9653(0.2161)(0.2198) |
| $\beta_{11}$ | 2.000 | | 1.9693(0.4091)(0.4054) |

From the estimates for prevalence, sensitivities and specificities (Table B.6), the estimates obtained by fitting Model 2 are quite close to the true values.

**Table B.6:** Parameter estimates: simulation averages for the posterior means, simulation standard errors and simulation averages of the estimated standard errors for the prevalence, sensitivities and specificities obtained by fitting Models 1 and 2 (data are generated under Model 2 with a moderate degree of dependency, N=500 and using different true values).

| Parameters | True Values | Model 1 | Model 2 |
|---|---|---|---|
| Prev. | 0.2620 | 0.2951(0.0334)(0.0273) | 0.2618(0.0247)(0.0241) |
| Sens. $T_1$ | 0.7765 | 0.8373(0.0447)(0.0383) | 0.7765(0.0443)(0.0431) |
| Sens. $T_2$ | 0.7744 | 0.8425(0.0401)(0.0361) | 0.7718(0.0411)(0.0425) |
| Sens. $T_3$ | 0.7199 | 0.8011(0.0424)(0.0378) | 0.7198(0.0412)(0.0443) |
| Sens. $T_4$ | 0.7872 | 0.6948(0.0660)(0.0486) | 0.7873(0.0420)(0.0436) |
| Sens. $T_5$ | 0.8999 | 0.7789(0.0696)(0.0481) | 0.8997(0.0394)(0.0369) |
| Sens. $T_6$ | 0.8509 | 0.7651(0.0664)(0.0458) | 0.8508(0.0385)(0.0382) |
| Spec. $T_1$ | 0.8173 | 0.8679(0.0294)(0.0233) | 0.8137(0.0214)(0.0230) |
| Spec. $T_2$ | 0.7331 | 0.7838(0.0331)(0.0263) | 0.7293(0.0236)(0.0257) |
| Spec. $T_3$ | 0.6789 | 0.7272(0.0332)(0.0274) | 0.6735(0.0246)(0.0269) |
| Spec. $T_4$ | 0.8867 | 0.8766(0.0195)(0.0194) | 0.8852(0.0194)(0.0187) |
| Spec. $T_5$ | 0.8983 | 0.8825(0.0200)(0.0195) | 0.8975(0.0191)(0.0191) |
| Spec. $T_6$ | 0.8077 | 0.7998(0.0223)(0.0230) | 0.8065(0.0213)(0.0226) |

Figure B.6 shows the DIC-values for the simulation runs. All the DIC-values for Model 2 are better than the DIC-values for Model 1.

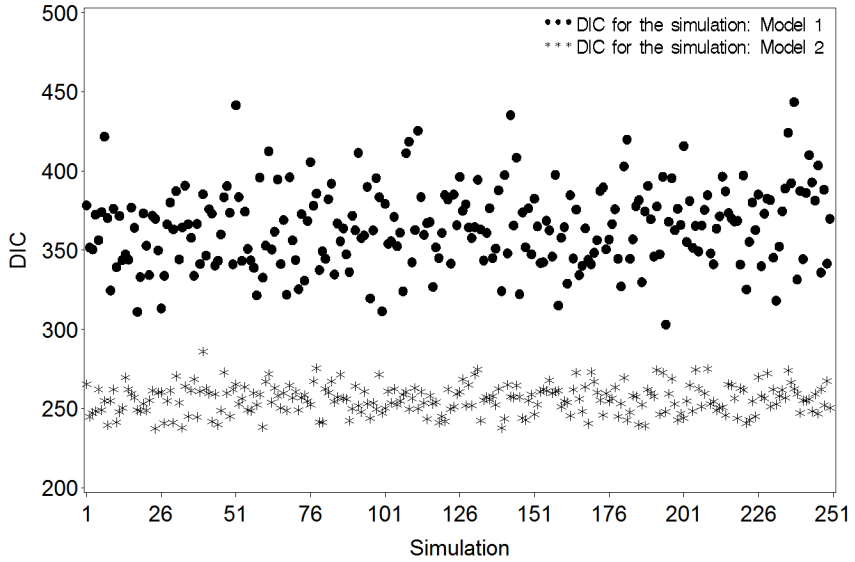



**Figure B.6:** DIC-values (dots and stars) for the simulation runs (data are generated under Model 2 with a moderate degree of dependency, N=500 and using different true values).

## III. Scenario 3C: Data are generated using the conditional dependence model with a high dependency

In this scenario, we have conducted the simulation study discussed in Section 6.4.2 with N=500. The simulation averages for the posterior means, simulation standard errors and simulation averages of the estimated standard errors obtained by fitting Models 1 and 2 are given in Tables B.7 and B.8.

**Table B.7:** Parameter estimates: simulation averages for the posterior means, simulation standard errors and simulation averages of the estimated standard errors for the parameters obtained by fitting Models 1 and 2 (data are generated under Model 2 with a high degree of dependency, N=500 and using different true values).

| Parameters | True Values | Model 1 | Model 2 |
|---|---|---|---|
| $\alpha_{11}$ | 1.831 | 2.1827(0.2644)(0.2573) | 1.8257(0.6388)(0.6269) |
| $\alpha_{21}$ | 1.812 | 2.7498(0.3703)(0.3477) | 1.8446(0.5862)(0.6208) |
| $\alpha_{31}$ | 1.355 | 2.7616(0.3799)(0.3408) | 1.4448(0.6185)(0.6017) |
| $\alpha_{41}$ | 1.308 | -0.5646(0.1447)(0.1385) | 1.4008(0.3542)(0.3625) |
| $\alpha_{51}$ | 2.196 | -0.4163(0.1431)(0.1360) | 2.7622(0.7360)(0.9955) |
| $\alpha_{61}$ | 1.742 | -0.2593(0.1309)(0.1343) | 1.9051(0.5636)(0.4664) |
| $\alpha_{10}$ | -2.425 | -3.3569(0.5072)(0.4788) | -1.8821(0.3827)(0.4124) |
| $\alpha_{20}$ | -1.690 | -2.7873(0.3377)(0.3343) | -1.2359(0.3552)(0.3966) |
| $\alpha_{30}$ | -1.279 | -2.4558(0.2695)(0.2678) | -0.8581(0.3951)(0.3888) |
| $\alpha_{40}$ | -2.057 | -1.2628(0.1458)(0.1527) | -2.0334(0.2021)(0.2066) |
| $\alpha_{50}$ | -2.178 | -1.2277(0.1591)(0.1514) | -2.2218(0.3029)(0.2741) |
| $\alpha_{60}$ | -1.435 | -0.8638(0.1387)(0.1380) | -1.4180(0.1431)(0.1568) |
| $\beta_{10}$ | 6.000 | | 5.3760(0.4290)(0.5262) |
| $\beta_{11}$ | 6.000 | | 4.3746(0.4676)(0.6507) |

The results in Table B.7 indicate that there are differences between the parameter estimates obtained by fitting Models 1 and 2. The estimates obtained by fitting Model 2 are quite close to the true values. There is no much difference between the estimated and empirical standard errors.

From the results given in Table B.8, the estimates obtained by fitting Model 2 are quite close to the true values. Model 2 underestimates the dependency parameters.

**Table B.8:** Parameter estimates: simulation averages for the posterior means, simulation standard errors and simulation averages of the estimated standard errors for the prevalence, sensitivities and specificities obtained by fitting Models 1 and 2 (data are generated under Model 2 with a high degree of dependency, N=500 and using different true values).

| Parameters | True Values | Model 1 | Model 2 |
|---|---|---|---|
| Prev. | 0.2620 | 0.4718(0.0239)(0.0238) | 0.2552(0.0277)(0.0282) |
| Sens. $T_1$ | 0.6520 | 0.8939(0.0240)(0.0229) | 0.6494(0.0505)(0.0485) |
| Sens. $T_2$ | 0.6509 | 0.9340(0.0195)(0.0189) | 0.6513(0.0466)(0.0482) |
| Sens. $T_3$ | 0.6240 | 0.9347(0.0195)(0.0187) | 0.6194(0.0491)(0.0486) |
| Sens. $T_4$ | 0.7872 | 0.3638(0.0330)(0.0317) | 0.7907(0.0519)(0.0504) |
| Sens. $T_5$ | 0.8999 | 0.3984(0.0340)(0.0323) | 0.9117(0.0406)(0.0451) |
| Sens. $T_6$ | 0.8509 | 0.4361(0.0320)(0.0327) | 0.8522(0.0493)(0.0437) |
| Spec. $T_1$ | 0.6346 | 0.9600(0.0147)(0.0143) | 0.6301(0.0259)(0.0290) |
| Spec. $T_2$ | 0.5902 | 0.9370(0.0167)(0.0176) | 0.5864(0.0249)(0.0295) |
| Spec. $T_3$ | 0.5652 | 0.9167(0.0189)(0.0192) | 0.5603(0.0279)(0.0296) |
| Spec. $T_4$ | 0.8867 | 0.7774(0.0243)(0.0261) | 0.8811(0.0201)(0.0207) |
| Spec. $T_5$ | 0.8983 | 0.7711(0.0273)(0.0263) | 0.8967(0.0240)(0.0221) |
| Spec. $T_6$ | 0.8077 | 0.7018(0.0288)(0.0286) | 0.8029(0.0222)(0.0243) |

Figure B.7 shows the the DIC-values for the simulation runs. As it was the case for the simulation study with a moderate degree of dependency, all the DIC-values for Model 2 are less than the DIC-values for Model 1.

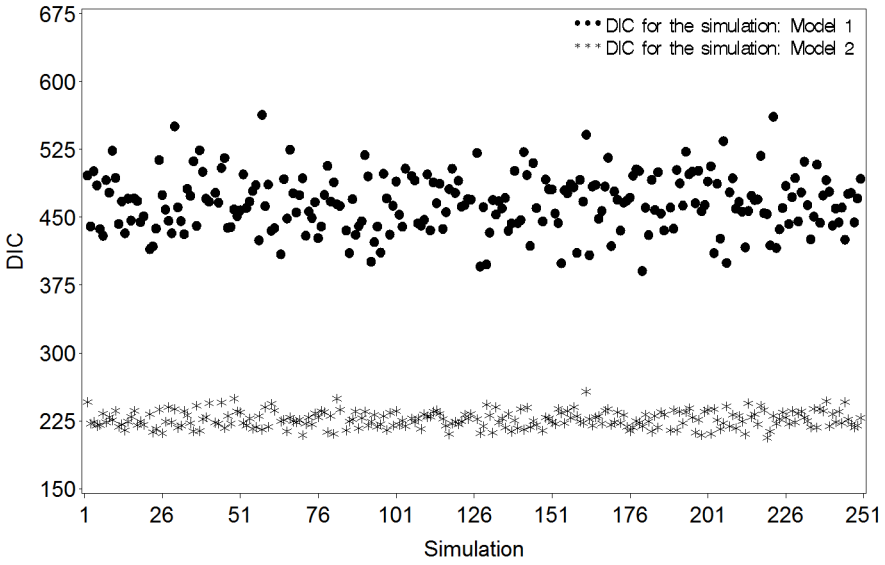



**Figure B.7:** DIC-values (dots and stars) for the simulation runs (data are generated under Model 2 with a high degree of dependency, N=500 and using different true values).

# Appendix C

# Results of External Validation of the Clinical Prediction Rule

In this appendix, we present results of external validation of the clinical prediction rule discussed in Section 7.4.5. A 3-fold cross-validation approach were used to conduct an internal validation of the new clinical prediction rule. The whole sample is randomly partitioned in three sets (1, 2, and 3).

**Cross-validation 1**: Partition 1 and 2 are used as a learning sample, and partition 3 as test sample.

**Cross-validation 2**: Partition 1 and 3 are used as a learning sample, and partition 2 as test sample.

**Cross-validation 3**: Partition 2 and 3 are used as a learning sample, and partition 1 as test sample.

For the three cross-validation, we iterate three times the following procedure:

a) Take one of the sets as test sample, the other two as learning sample.

b) Using the learning sample, refit the full model as shown in Table 7.6, and simplify it gradually to a simplified model with the 6 most important predictors, and associated further simplified clinical tool (with all coefficients rounded to 1 and -1).

c) For each of the simplified models (with original and rounded coefficients), measure sensitivity, specificity etc using the test sample. So the model built with the learning sample is then tested with an independent test sample.

The results of the cross-validation approach for the 3-folds are given below.

## I. Parameter estimates and standard errors for the parameters obtained from the IPD meta-model for the three cross-validations

**Table C.1:** Parameter estimates and standard errors for the parameters obtained from the IPD meta-analysis for the three cross-validations.

| Parameter | | Cross-validation 1 | Cross-validation 2 | Cross-validation 3 |
|---|---|---|---|---|
| | $S_1$ | -4.857 (1.057)** | -4.882 (1.381)** | -4.723 (1.034)** |
| | $S_2$ | -3.972 (0.727)** | -3.960 (0.808)** | -4.319 (0.782)** |
| Intercept | $S_3$ | -3.900 (0.415)** | -3.561 (0.377)** | -3.547 (0.361)** |
| | $S_4$ | -4.049 (0.411)** | -3.918 (0.392)** | -4.169 (0.402)** |
| | $S_5$ | -5.269 (0.466)** | -4.991 (0.416)** | -5.227 (0.442)** |
| Age-patient | 1 | 1.401 (0.195)** | 1.507 (0.193)** | 1.431 (0.190)** |
| Sex-male | 1 | 0.226 (0.177) | 0.179 (0.172) | 0.460 (0.173)** |
| Emergency | 1 | -0.024 (0.234) | -0.268 (0.247) | -0.276 (0.239) |
| Previous-CP | 1 | 0.503 (0.306) | 0.663 (0.303)** | 0.146 (0.316) |
| GPs-concern | 1 | 1.300 (0.232)** | 1.141 (0.226)** | 1.520 (0.232)** |
| Pat-concern-heart | 1 | 1.407 (0.324)** | 1.002 (0.280)** | 1.045 (0.279)** |
| Loc-retrosternal | 1 | 0.480 (0.215)** | 0.034 (0.207) | 0.237 (0.212) |
| Rad-neck | 1 | 0.733 (0.338)** | 0.410 (0.366) | 0.710 (0.363)* |
| Stabbing | 1 | -0.577 (0.259)** | -0.266 (0.252) | -0.445 (0.261)* |
| Oppressive | 1 | 0.289 (0.216) | 0.878 (0.211)** | 0.747 (0.221)** |
| Rel-effort | 1 | 1.390 (0.205)** | 1.174 (0.204)** | 1.081 (0.206)** |
| Add-nausea | 1 | -0.579 (0.412) | 0.125 (0.396) | 0.162 (0.383) |
| Add-sputum | 1 | -0.696 (1.136) | -1.096 (1.366) | -0.524 (1.125) |
| Auscult-pulmo | 1 | -0.668 (0.661) | -0.513 (0.806) | -0.178 (0.612) |
| Auscult-heart | 1 | 0.897 (0.938) | 0.763 (1.492) | 0.639 (1.181) |
| Reprod-palpation | 1 | -1.330 (0.284)** | -1.827 (0.303)** | -1.474 (0.299)** |
| RF-hypertension | 1 | 0.401 (0.280) | 0.335 (0.248) | 0.333 (0.255) |
| RF-smoking | 1 | 0.765 (0.306)** | 0.800 (0.330)** | 0.316 (0.313) |
| Known-CHD | 1 | 1.578 (0.274)** | 1.822 (0.261)** | 1.863 (0.270)** |

CP: chest pain, CHD: coronary heart disease, GP: general practitioner, RF: risk factor, $S_1$: Switzerland, $S_2$: Belgium, $S_3$: Sweden, $S_4$: USA, $S_5$: Germany,

** The variable is significant at 5% level of significance ($p$-value $< 0.05$),

* The variable is significant at 10% level of significance ($p$-value $< 0.10$).

## II. Parameter estimates and standard errors for the parameters obtained from the reduced IPD meta-model for the three cross-validations

**Table C.2:** Parameter estimates and standard errors for the parameters obtained from the reduced IPD meta-model for the three cross validations.

| Parameter | | Estimates (Standard errors) | | |
|---|---|---|---|---|
| | | Cross-validation 1 | Cross-validation 2 | Cross-validation 3 |
| Intercept | $S_1$ | -3.869 (0.324)** | -3.971 (0.329)** | -3.991 (0.326)** |
| | $S_2$ | -3.906 (0.351)** | -3.881 (0.345)** | -4.090 (0.370)** |
| | $S_3$ | -2.729 (0.205)** | -2.955 (0.218)** | -2.755 (0.204)** |
| | $S_4$ | -3.645 (0.285)** | -3.694 (0.278)** | -3.929 (0.296)** |
| | $S_5$ | -3.858 (0.251)** | -4.042 (0.257)** | -4.080 (0.268)** |
| Age-patient | 1 | 1.519 (0.181)** | 1.591 (0.183)** | 1.550 (0.181)** |
| GPs-concern | 1 | 1.371 (0.217)** | 1.155 (0.216)** | 1.547 (0.221)** |
| Oppressive | 1 | 0.635 (0.191)** | 1.003 (0.194)** | 0.892 (0.201)** |
| Rel-effort | 1 | 1.399 (0.188)** | 1.253 (0.197)** | 1.102 (0.193)** |
| Reprod-palpation | 1 | -1.560 (0.258)** | -1.904 (0.276)** | -1.629 (0.274)** |
| Known-CHD | 1 | 1.466 (0.254)** | 1.799 (0.245)** | 1.882 (0.249)** |

CHD: coronary heart disease, GP: general practitioner, $S_1$: Switzerland, $S_2$: Belgium, $S_3$: Sweden, $S_4$: USA, $S_5$: Germany,

** The variable is significant at 5% level of significance ($p$-value $< 0.05$),

## III. Parameter estimates and 95% confidence intervals for the measures of diagnostic accuracy for comparing classification based on the score with the reference diagnosis for different threshold Values

### Results of cross-validation 1

**Table C.3:** Cross-validation 1: Combined parameter estimates and 95% confidence intervals for the measures of diagnostic accuracy for comparing classification based on the score with the reference diagnosis for different threshold values.

| Threshold (c) | Measures of diagnostic accuracy | Estimates and 95% confidence intervals | Estimates and 95% confidence intervals[†] |
|---|---|---|---|
| c = 1 | Sensitivity | 0.8963 (0.8293, 0.9389) | 0.9398 (0.8843, 0.9696) |
| | Specificity | 0.6630 (0.6316, 0.6931) | 0.5104 (0.4771, 0.5435) |
| | Positive likelihood | 2.6606 (2.3863, 2.9655) | 1.9196 (1.7730, 2.0785) |
| | Negative likelihood | 0.1562 (0.0929, 0.2626) | 0.1178 (0.0600, 0.2314) |
| c = 1.6 | Sensitivity | 0.7384 (0.6547, 0.8077) | 0.7368 (0.6545, 0.8053) |
| | Specificity | 0.8377 (0.8094, 0.8624) | 0.8393 (0.8139, 0.8618) |
| | Positive likelihood | 4.5499 (3.7987, 5.4493) | 4.5864 (3.8282, 5.4943) |
| | Negative likelihood | 0.3122 (0.2330, 0.4183) | 0.3134 (0.2348, 0.4184) |
| c = 2 | Sensitivity | 0.7191 (0.6243, 0.7977) | 0.7368 (0.6545, 0.8053) |
| | Specificity | 0.8611 (0.8096, 0.9004) | 0.8393 (0.8139, 0.8618) |
| | Positive likelihood | 5.1800 (3.8094, 7.0490) | 4.5864 (3.8282, 5.4943) |
| | Negative likelihood | 0.3261 (0.2425, 0.4388) | 0.3134 (0.2348, 0.4184) |

[†] Measures of diagnostic accuracy for the simplified calculation of the score, CHD negative if score $<$ c; CHD positive if score $\geq$ c.

## Results of cross-validation 2

**Table C.4:** Cross-validation 2: Combined parameter estimates and 95% confidence intervals for the measures of diagnostic accuracy for comparing classification based on the score with the reference diagnosis for different threshold values.

| Threshold (c) | Measures of diagnostic accuracy | Estimates and 95% confidence intervals | Estimates and 95% confidence intervals[†] |
|---|---|---|---|
| c = 1 | Sensitivity | 0.8791 (0.7921, 0.9327) | 0.9317 (0.8691, 0.9655) |
| | Specificity | 0.6357 (0.5374, 0.7239) | 0.5068 (0.4733, 0.5402) |
| | Positive likelihood | 2.4136 (1.9408, 3.0131) | 1.8892 (1.7389, 2.0521) |
| | Negative likelihood | 0.1901 (0.1141, 0.3178) | 0.1346 (0.0687, 0.2638) |
| c = 1.6 | Sensitivity | 0.6801 (0.4306, 0.8567) | 0.6250 (0.5352, 0.7069) |
| | Specificity | 0.8017 (0.6823, 0.8838) | 0.8225 (0.7947, 0.8472) |
| | Positive likelihood | 3.4301 (2.7373, 4.2101) | 3.5214 (2.8714, 4.3189) |
| | Negative likelihood | 0.3989 (0.2107, 0.7398) | 0.4559 (0.3611, 0.5756) |
| c = 2 | Sensitivity | 0.5950 (0.5035, 0.6803) | 0.6250 (0.5352, 0.7069) |
| | Specificity | 0.8382 (0.8113, 0.8618) | 0.8225 (0.7947, 0.8472) |
| | Positive likelihood | 3.6775 (2.9713, 4.5508) | 3.5214 (2.8714, 4.3189) |
| | Negative likelihood | 0.4831 (0.3870, 0.6030) | 0.4559 (0.3611, 0.5756) |

[†] Measures of diagnostic accuracy for the simplified calculation of the score, CHD negative if score $<$ c; CHD positive if score $\geq$ c.

## Results of cross-validation 3

**Table C.5:** Cross-validation 3: Combined parameter estimates and 95% confidence intervals for the measures of diagnostic accuracy for comparing classification based on the score with the reference diagnosis for different threshold values.

| Threshold (c) | Measures of diagnostic accuracy | Estimates and 95% confidence intervals | Estimates and 95% confidence intervals† |
|---|---|---|---|
| c = 1 | Sensitivity | 0.9114 (0.8313, 0.9555) | 0.9538 (0.9007, 0.9791) |
| | Specificity | 0.6344 (0.5766, 0.6886) | 0.4958 (0.4619, 0.5297) |
| | Positive likelihood | 2.4937 (2.1413, 2.9015) | 1.8917 (1.7508, 2.0436) |
| | Negative likelihood | 0.1395 (0.0728, 0.2670) | 0.0931 (0.0424, 0.2045) |
| c = 1.6 | Sensitivity | 0.7150 (0.6280, 0.7885) | 0.6986 (0.6116, 0.7732) |
| | Specificity | 0.8118 (0.7842, 0.8366) | 0.8227 (0.7955, 0.8469) |
| | Positive likelihood | 3.7996 (3.1819, 4.5347) | 3.9404 (3.2801, 4.7314) |
| | Negative likelihood | 0.3510 (0.2641, 0.4661) | 0.3663 (0.2794, 0.4800) |
| c = 2 | Sensitivity | 0.6779 (0.5767, 0.7648) | 0.6986 (0.6116, 0.7732) |
| | Specificity | 0.8373 (0.7971, 0.8708) | 0.8227 (0.7955, 0.8469) |
| | Positive likelihood | 4.1678 (3.3514, 5.1769) | 3.9404 (3.2801, 4.7314) |
| | Negative likelihood | 0.3845 (0.2899, 0.5094) | 0.3663 (0.2794, 0.4800) |

† Measures of diagnostic accuracy for the simplified calculation of the score, CHD negative if score $<$ c; CHD positive if score $\geq$ c.

# Appendix D

# R and WinBUGS Codes

## I. R code used to fit the adaptive change-point model

The following WinBUGS code were used in R using the R-package R2WinBUGS to fit the change-point model with one unknown common change-point, one country-specific random change-point and a country-specific latent indicator for the change-point.

```
# Model
model{
# Basic model
for (i in 1:N){
Y[i]~dnorm(mu[i],tau)
mu[i]<-(B0+b1[ID[i]])+(B1+b2[ID[i]])*T[i]+(B2+b3[ID[i]])*(T[i]-
       (c1[ID[i]]))*step(T[i]-(c1[ID[i]]))*change[ID[i]]+
       (alpha+b4[ID[i]]+alphaTime*T[i])*sin(omega*T[i]+delta)
       }
# Priors for random effects
for (j in 1:M){
b1[j]~dnorm(0,b0.tau)
b2[j]~dnorm(0,b1.tau)
b3[j]~dnorm(0,b2.tau)
b4[j]~dnorm(0,b3.tau)
c1[j]~dnorm(C1,c1.tau)
change[j]~dbern(changemean)
        }
```

```
# Priors for fixed effects
B0   ~dnorm(0,0.0001)
B1   ~dnorm(0,0.0001)
B2   ~dnorm(0,0.0001)
alpha~dnorm(0,0.0001)
alphaTime~dnorm(0,0.0001)
delta~dnorm(0,0.0001)
C1~dunif(1,52)
changemean ~ dbeta(1,1)

#Hyper priors
tau  ~ dgamma(0.001, 0.001)
b0.tau~dgamma(0.001, 0.001)
b1.tau~dgamma(0.001, 0.001)
b2.tau~dgamma(0.001, 0.001)
b3.tau~dgamma(0.001, 0.001)
c1.tau~dgamma(0.001, 0.001)

sigma   <-1/tau
sigma_b0<-1/b0.tau
sigma_b1<-1/b1.tau
sigma_b2<-1/b2.tau
sigma_b3<-1/b3.tau
sigma_c1<-1/c1.tau
}
```

## II. R code used to fit the conditional independence latent class model

```
# Model
model{
# Basic model
       for (i in 1:nPats){
    status[i]~dbern(prev)
    for(k in 1:nTests){
        Y[i,k]~dbern(P[i,k])
          }
    logit(P[i,1])<-status[i]*alpha[1]+(1-status[i])*beta[1]
    logit(P[i,2])<-status[i]*alpha[2]+(1-status[i])*beta[2]
    logit(P[i,3])<-status[i]*alpha[3]+(1-status[i])*beta[3]
    logit(P[i,4])<-status[i]*alpha[4]+(1-status[i])*beta[4]
    logit(P[i,5])<-status[i]*alpha[5]+(1-status[i])*beta[5]
    logit(P[i,6])<-status[i]*alpha[6]+(1-status[i])*beta[6]
     }

# Priors for fixed effects
  prev ~ dbeta(1,1)
  for(k in 1:nTests){
      logit(sens[k])<-alpha[k]
      logit(spec[k])<--beta[k]
      }

  for(k in 1:nTests){
      alpha[k]~dnorm(0.0,.1)
      beta[k]~dnorm(0.0,.1)
      }
  }
```

## III. R code used to fit the conditional dependence latent class model

```
# Model
  model{
# Basic model
   for (i in 1:nPats){
      status[i]~dbern(prev)
  for(k in 1:nTests){
   Y[i,k]~dbern(P[i,k])
              }
   logit(P[i,1])<-status[i]*alpha[1]+status[i]*Me1*RE1[i]+
                 (1-status[i])*beta[1]+(1-status[i])*Me0*RE2[i]
   logit(P[i,2])<-status[i]*alpha[2]+status[i]*Me1*RE1[i]+
                 (1-status[i])*beta[2]+(1-status[i])*Me0*RE2[i]
   logit(P[i,3])<-status[i]*alpha[3]+status[i]*Me1*RE1[i]+
                 (1-status[i])*beta[3]+(1-status[i])*Me0*RE2[i]
   logit(P[i,4])<-status[i]*alpha[4]+(1-status[i])*beta[4]
   logit(P[i,5])<-status[i]*alpha[5]+(1-status[i])*beta[5]
   logit(P[i,6])<-status[i]*alpha[6]+(1-status[i])*beta[6]

     RE1[i] ~ dnorm(0,1)
     RE2[i] ~ dnorm(0,1)
     }

# Priors for fixed effects
  prev ~ dbeta(1,1)
  Me0 ~ dnorm(0,0.5)I(0,)
  Me1 ~ dnorm(0,0.5)I(0,)
  for(k in 1:nTests){
      logit(sens[k])<-alpha[k]
      logit(spec[k])<--beta[k]
      }

  for(k in 1:nTests){
      alpha[k]~dnorm(0.0,.1)
      beta[k]~dnorm(0.0,.1)
      }
   }
```

## IV. R code used to maximize the area under the receiver operating characters curve

```
#Maximizing the AUC
m=5
Estimate_1<- as.list(1:m)
for (k1 in (1:m))
{
Meta_analysis_1<-Meta_analysis[Meta_analysis$Imputation==k1,]
nD<-nrow(Meta_analysis_1[Meta_analysis_1$RD_any_CAD==1,])
nD_bar<-nrow(Meta_analysis_1[Meta_analysis_1$RD_any_CAD==0,])
y<-Meta_analysis_1$RD_any_CAD

meta.AUC<-function(theta,y,X){
yX1<-cbind(y,X)
yX2<-yX1[order(yX1[,1]),]

XD1<-yX2[yX2[,1]==2,]
XD2<-XD1[,-c(1,2,3,4,5,6,7)]
nD<-length(XD1[,1])

XD_bar1<-yX2[yX2[,1]==1,]
XD_bar2<-XD_bar1[,-c(1,2,3,4,5,6,7)]
nD_bar<-length(XD_bar1[,1])

XBD<-NULL
XBD_bar<-NULL
I2<-matrix(nrow=nD,ncol=nD_bar)
I3<-matrix(nrow=nD,ncol=1)
k<-ncol(XD2)
beta<-theta[1:k]

for (j in (1:nD))
{
XBD[j]<-XD2[j,7] + XD2[j,]%*%beta
}

for (k in (1:nD_bar))
{
XBD_bar[k]<-XD_bar2[j,7] + XD_bar2[k,]%*%beta
```

```
}

for (j in (1:nD)){
for (k in (1:nD_bar)){
if(XBD[j]>XBD_bar[k]) {I1=1}
else if(XBD[j]==XBD_bar[k]) {I1=0.5}
else {I1=0}
I2[j,k]<-I1
}
I3[j]<-sum(I2[j,])
}
I4<-sum(I3)
AUC<-I4/(nD*nD_bar)
return(-AUC)
}
p.AUC<-optim(c(1.433,0.283,-0.176,0.425,1.302,1.127,0.251,
0.606,-0.426,0.635,1.194,-0.087,-0.746,-0.447,0.748,-1.542,
0.349,0.630,1.732),meta.AUC,method="Nelder-Mead",hessian=T,
y=y,X=X,control=list(maxit=20000)
)
Estimate_1[[k1]]<-p.AUC$par
}
Estimate_1
```

# Samenvatting

In dit proefschrift hebben we verschillende methodes voor statistisch modelleren in het kader van volksgezondheidsonderzoek voorgesteld en onderzocht. Hoofdstuk 2 geeft een beknopte beschrijving van de datasets die gebruikt zijn in deze dissertatie. In hoofdstukken 3 en 4 hebben we de datasets geanalyseerd met betrekking tot het totale ambulante antibioticagebruik. Hoofdstuk 3 beschrijft de toepassing van de *gemengde modellen* op het jaarlijkse en driemaandelijkse ambulante antibiotica. In hoofdstuk 4 stellen we een *change-point gemengd model* voor om trendveranderingen in ambulant antibioticagebruik vast te stellen.

In hoofdstuk 5 werden conditioneel onafhankelijke en conditioneel afhankelijke latente klassenmodellen gebruikt om testen te evalueren. Het doel was de meest geschikte test te identificeren voor de diagnose van *Mycoplasma pneumoniae* en *Chlamydophila pneumoniae* bij volwassen patiënten met infecties van de onderste luchtwegen.

In hoofdstuk 7 werden meta-analyses met gegevens van individuele patiënten gebruikt om de gecombineerde diagnostische waarde te bestuderen van alle ziekteverschijnselen en symptomen voor de diagnose van coronaire hartziektes in eerstelijnsgezondheidszorg. In de onderstaande paragrafen worden een aantal conclusies en concluderende opmerkingen gepresenteerd.

## 1. Ambulant Antibioticagebruik in Europa

Kwaliteitscontrole en verbetering van de gezondheidszorg zijn een belangrijk thema in een groot aantal landen. De beschikbare ESAC gegevens over ambulant antibioticagebruik in Europa stellen landen in staat om hun antibioticagebruik te controleren door een duidelijke, vergelijkbare en betrouwbare referentiedatabase op te stellen en te onderhouden. De ESAC gegevens hebben bewezen een waardevolle databron te zijn voor de evaluatie van richt- en beleidslijnen alsook voor de beoordeling van de

uitkomsten van nationale interventies.

In hoofdstuk 3 werden *gemengde modellen* aangewend om een vaststelling te maken van het totale ambulante antibioticagebruik in Europa tussen 1997 en 2009, om de trend van het totale antibioticagebruik te analyseren en om seizoensvariatie te onderzoeken. Het twee-fase model en het *lineaire gemengde model* werden toegepast op de jaarlijkse data betreffende ambulant antibioticagebruik. Wat de driemaandelijkse data inzake ambulant antibioticagebruik betreft, werd een *niet-lineair gemengd model* gebruikt om landspecifieke trends in Europa vast te stellen en werd het seizoenseffect mee in beschouwing genomen. De toepassingen van deze modellen brengen nieuwe belangrijke inzichten aan het licht inzake de evolutie van ambulant antibioticagebruik in Europa.

De waargenomen verschillen tussen Europese landen betreffende het gebruik van tetracycline suggereert dat deze subgroup van antibiotica ongepast wordt voorgeschreven in vele landen. Het gebruik van tetracycline is in de meeste Europese landen gedaald. Seizoensgebonden ambulant gebruik van tetracycline werd waargenomen in alle landen en was significant gerelateerd met de totale ambulante consumptie van tetracycline. De hoogste seizoensgebonden variatie werd vastgesteld in de Europese landen die het meest consumeren, wat suggereert dat tetracycline onnodig gebruikt wordt voor virale infecties. De seizoensgebonden variatie van ambulant tetracycline gebruik is mettertijd gedaald, waardoor we kunnen aannemen dat het op een betere wijze wordt voorgeschreven.

Een aantal landen hebben een nationaal programma geïmplementeerd om antimicrobiële resistentie te controleren en het nationale antibioticagebruik te verbeteren. Dit omvat publiciteit in de media, zoals nationale publieke campagnes in België en Frankrijk en herhaalde berichtgeving in de media in Slovenië en Zweden. In hoofdstuk 4 werd het *niet-lineaire gemengde model* uitgebreid door bekende en onbekende gezamenlijke change-points en landenspecifieke willekeurige change-points mee op te nemen om de veranderingen in trends inzake antibioticagebruik met de tijd vast te stellen. De locatie van de change-points mag gerelateerd zijn aan momenten in tijd waarop volksgezondheidsstrategieën werden geïnitieerd met als doel het bewustzijn bij het publiek te verhogen naar een meer rationeel gebruik van antibiotica of overconsumptie van antibiotica te verminderen.

Voor toekomstig onderzoek zou de interpretatie van de opvallende variaties tussen en in Europese landen substantieel kunnen verbreed worden indien meer gedetailleerde data over antibioticagebruik, gelinkt aan de leeftijd en het geslacht van de patiënt,

indicatie en kenmerken m.b.t. de voorschrijver gebruikt worden. Hoewel ESAC focuste op het totale nationale antibioticagebruik, kunnen regionale gegevens andere en meer betekenisvolle resultaten tonen.

## 2. Diagnose van Acute Infecties

Gepaarde sera van 201 volwassen patiënten met een community-acquired pneumonie werden getest op de aanwezigheid van antistoffen tegen *M. pneumoniae* met verschillende commerciële testkits: *M. pneumoniae*-ELISA medac (IgM, IgG and IgA); ANILabsystems *M. pneumoniae* (IgM and IgG); EUROIMMUN Anti-*M. pneumoniae* ELISA (IgM, IgG, and IgA); en ImmunoWELL *M. pneumoniae* IgM en IgG EIA.

Vier commercieel beschikbare serologische analyses (2 species specifiek, 1 gender specifiek en 1 gender specifiek MIF test) voor de detectie van *C. pneumoniaie* IgM en IgG antistoffen en 2 commercieel beschikbare serologische analyses voor de detectie van *C. pneumoniaie* IgA antistoffen werden ook vergeleken met PCR en NASBA van 134 volwassen patiënten.

In hoofdstuk 5 werden latente klassenmodellen gebruikt om testen te evalueren die aangewend werden voor de diagnose van *Mycoplasma pneumoniae* en *Chlamydophila pneumoniae* bij volwassen patiënten met infecties van de onderste luchtwegen om de meest geschikte test te identificeren. In dit hoofdstuk bekijken we eerst het conditioneel onafhankelijkheidsmodel dat veronderstelt dat alle diagnostische testen onafhankelijk conditioneel zijn van de werkelijke ziektestatus. Vervolgens nemen we het conditioneel afhankelijkheidsmodel in aanmerking dat veronderstelt dat sommige of alle diagnostische testen afhankelijk conditioneel zijn van de werkelijke ziektestatus. We hebben de resultaten van de latente klassenmoddelen vergeleken met de uitgebreide gouden standaard. Om de afhankelijkheid van de testen te evalueren op co-varianten, werden deze modellen uitgebreid door leeftijd en geslacht mee op te nemen. Uit de resultaten van deze modellen is gebleken dat de uitslag van de testen die gebruikt werden voor de diagnose van *Mycoplasma pneumoniae* en *Chlamydophila pneumoniae* infecties niet afhangt van leeftijd en geslacht.

In hoofdstuk 6 werden simulatiestudies opgezet om de conditioneel onafhankelijke en de conditioneel afhankelijke modellen te vergelijken. Voor sterk gecorreleerde testen bleek dat het conditioneel afhankelijke model beter presteert dan het conditioneel onafhankelijk model.

Verschillen in IgG seroprevalentie werden vastgesteld wanneer de 4 verschillende IgG

analyses werden toegepast. Bij het vergelijken van de serologische IgG analyses werden de beste resultaten in termen van gevoeligheden en specificiteit behaald wanneer de Medac kits gebruikt werden. De keuze van de serologische analyse heeft belangrijke implicaties voor de waarneming van *M. pneumoniae* infectie. Een combinatie van een nucleïnezuur amplificatie test en een serologische test is mogelijk de beste keuze voor een nauwkeurige *M. pneumoniae* diagnose bij volwassen patinten die een infectie van de onderste luchtwegen vertonen.

Er werden substantiële verschillen tussen de prestaties van de analyses gevonden voor de waarneming van *C. pneumoniae* infectie. We adviseren dat het gebruik van een geschikte gouden standaard cruciaal is voor de waarneming van *C. pneumoniae*.

## 3. Diagnose van Coronaire Hartziekte

Om huisartsen bewijs-gebaseerde aanbevelingen te geven, hebben INTERCHEST medewerkers een systematische beoordeling uitgevoerd op studies die de diagnostische nauwkeurigheid evalueren van ziekteverschijnselen en symptomen om coronaire hartziekten (CHD) te diagnosticeren in eerstelijnsgezondheidszorg. In deze studie proberen we de diagnostische nauwkeurigheid in te schatten voor clusters van ziektverschijnselen en symptomen voor de diagnose van CHD in eerstelijnsgezondheidszorg bij niet-geselecteerde patiënten met pijn op de borstkas. Verschillende methodologische uitdagingen maken het moeilijk om te anticiperen op de resultaten van de studie. De individuele studies werden uitgevoerd over een periode van bijna dertig jaar waarbij verscheidene veranderingen zijn opgetreden, o.a. de definitie van een myocardinfarct.

De meeste individuele ziekteverschijnselen en symptomen zijn niet voldoende om betrouwbaar een diagnose van CHD te stellen. Dit probleem kan voorkomen worden door verschillende bevindingen te combineren tot een klinische voorspellingsregel. In hoofdstuk 7 werden IPD meta-analyses voorgesteld om de gecombineerde diagnostische waarde van ziekteverschijnselen en symptomen te onderzoeken. Gebaseerd op de data van alle studies hebben we een nieuwe klinische voorspellingsregel voor de diagnose van CHD in eerstelijnsgezondheidszorg ontwikkeld. Een interne kruisvalidatie benadering werd gebruikt om de nieuwe klinische voorspellingsregel te valideren.

Gebaseerd op onze bevindingen, hebben we aanbevelingen gegeven betreffende de noodzaak tot verder wetenschappelijk onderzoek, inclusief onderzoek van diagnostische algoritmen gebaseerd op combinaties van bevindingen van de anamnese,

lichamelijk onderzoek, en technologische apparaten (ECG, point of care bloedtesten voor troponine of andere biomarkers). Gebaseerd op de bevindingen, doen we eveneens aanbevelingen betreffende het opzet van toekomstige diagnostische studies in eerstelijnsgezondheidszorg en het uitvoeren van diagnostische nauwkeursigheidscontroles gebaseerd op data van individuele patiënten.

## 4. Algemene Conclusie

In dit onderzoek hebben we geschikte strategieën voor het statistisch modelleren voor volksgezondheidsonderzoek bepaald, met toepassingen in het toezicht van antimicrobiële consumptie, de diagnose van acute infecties en de diagnose van coronaire hartziekten. *Gemengde modellen* werden voorgesteld om de datasets met betrekking tot antimicrobiële consumptie te analyseren en de werkelijke ziektestatus van de patiënten te voorspellen bij afwezigheid van een echte gouden standaard test.

De *gemengde modellen* werden toegepast op de datasets met het totale ambulante antibioticagebruik om het landenspecifieke antibioticagebruik te analyseren, de seizoensvariatie te analyseren en de trendverandering in ambulant antibioticagebruik te bepalen. *Gemengde latente klassen modellen* werden ook gebruikt om de diagnostische nauwkeurigheid van de testen, die gebruikt werden voor de diagnose van acute infecties, te bepalen en om de meest geschikte test te identificeren.

Aangezien we slechts 5 studies hebben, worden voor de derde toepassing eerder *fixed-effects modellen* gebruikt in plaats van *gemengde modellen* om de gecombineerde diagnostische waarde van alle ziekteverschijnselen en symptomen voor de diagnose van coronaire hartziekten in eerstelijnsgezondheidszorg te onderzoeken. Studie-specifieke intercepten werden mee opgenomen om heterogeniteit tussen de studies te kwantificeren.

De toepassingen van de voorgestelde methodes zullen belangrijke inzichten geven in (a) de evolutie van het totale antibioticagebruik in Europa, (b) de evaluatie van de testen die gebruikt worden voor de diagnose van *Chlamydophila pneumoniae* en *Mycoplasma pneumoniae* infecties, en (c) de diagnose van coronaire hartziektes voor clinici, beleidsmakers en anderen die betrokken zijn bij volksgezondheid.