



**Marja Snellman**

# **Case Definition of Pneumococcal Pneumonia — a Latent Class Analysis Approach**

Kansanterveyslaitoksen julkaisuja **B** 3/2008

Kansanterveyslaitos, Rokoteosasto, Kliininen yksikkö  
KTL-National Public Health Institute, Finland  
Department of Vaccines, Clinical unit  
Helsinki 2008

**CASE DEFINITION OF PNEUMOCOCCAL PNEUMONIA  
A LATENT CLASS ANALYSIS APPROACH**

MARJA SNELLMAN

Kansanterveyslaitos  
Rokoteosasto  
Kliininen yksikkö

KTL-National Public Health Institute, Finland  
Department of Vaccines  
Clinical unit

Helsinki 2008

**Kansanterveyslaitoksen julkaisuja B3 / 2008**

**Copyright National Public Health Institute**

**Julkaisija - Utgivare - Publisher**

**Kansanterveyslaitos (KTL)**

Mannerheimintie 166

00300 Helsinki

Puh. vaihde (09) 474 41, telefax (09) 4744 8408

**Folkhälsoinstitutet**

Mannerheimvägen 166

00300 Helsingfors

Tel. växel (09) 474 41, telefax (09) 4744 8408

**National Public Health Institute**

Mannerheimintie 166

FIN-00300 Helsinki, Finland

Telephone +358 9 474 41, telefax +358 9 4744 8408

<http://www.ktl.fi>

ISBN 978-951-740-772-4 (pdf)

ISSN 1458-3576

<http://www.ktl.fi/portal/2920>

**Kannen kuva - cover graphic Adam Ciesielski**

Marja Snellman

Case definition of pneumococcal pneumonia - a latent class analysis approach

Publications of the National Public Health Institute, B3/2008, 50 pages

ISBN 978-951-740-772-4 (pdf-version)

ISSN 0359-3576

<http://www.ktl.fi/portal/2920>

## ABSTRACT

Community-acquired pneumonia (CAP) is a severe disease and a major cause of death worldwide especially among the elderly. The most common causative pathogen is *Streptococcus pneumoniae*, pneumococcus. The diagnosis of pneumococcal pneumonia is difficult because there is no gold standard, a diagnostic test that would identify all cases and yet be definite.

National Public Health Institute has launched a Finnish Community-Acquired Pneumonia study investigating the frequency and causes of CAP among the elderly aged 65 years and above. Sputum, urine, blood and nasopharyngeal swab samples are collected from the subjects enrolled in the study and a large number of microbiological assays are performed on the samples. One of the main objectives is to find a case definition for pneumococcal pneumonia in the elderly. For this purpose, the accuracy of diagnostic tests performed in the study need to be evaluated. In the absence of gold standard, the true disease status of the subjects is latent and the sensitivities and the specificities of the diagnostic tests cannot be estimated using conventional methods.

The aim of this report is to estimate the sensitivities and the specificities of diagnostic tests as well as the prevalence of pneumococcal pneumonia among the elderly population in Finland using latent class analysis. The method is applied to data collected in the Finnish Community-Acquired Pneumonia study. Methodological issues in latent class analysis are discussed. In addition, a function for estimating the model parameters using statistical program R is presented.

Keywords: latent class analysis, sensitivity, specificity

Marja Snellman

Case definition of pneumococcal pneumonia - a latent class analysis approach

Kansanterveyslaitoksen julkaisuja, B3/2008, 50 sivua

ISBN 978-951-740-772-4 (pdf-versio)

ISSN 0359-3576

<http://www.ktl.fi/portal/2920>

## TIIVISTELMÄ

Avohoitokeuhkokuume on vakava sairaus ja yleinen kuolinsyy maailmanlaajuisesti erityisesti ikääntyneiden keskuudessa. Suomessa sen arvioidaan aiheuttavan vuosittain yli 2000 kuolemantapausta yli 65-vuotiaissa. Keuhkokuumeen yleisimpänä taudinaiheuttajana pidetään *Streptococcus pneumoniae* -bakteeria, pneumokokkia. Keuhkokuumeen mikrobiologian selvittäminen on hankalaa, sillä täydellistä kultaista standardia -diagnostista testiä, joka ei anna vääriä positiivisia tai negatiivisia tuloksia ei ole olemassa.

Kansanterveyslaitos on aloittanut keuhkokuumeetutkimuksen, jossa pyritään selvittämään keuhkokuumeen yleisyyttä ja aiheuttajia ikäihmisten parissa. Erityisen kiinnostuksen kohteena on pneumokokkibakteerin osuus taudinaiheuttajana. Tutkimukseen osallistuvilta potilailta kerätään yskös-, virtsa-, nenänielu sekä verinäytteitä, jotka analysoidaan käyttäen erilaisia mikrobiologisia menetelmiä. Yksi tutkimuksen päätavoitteista on muodostaa pneumokokkikeuhkokuumeen tapausmäärittely. Tätä varten tarvitaan tietoa tutkimuksessa suoritettujen diagnostisten testien sensitiivisyyksistä ja spesifisyyksistä. Kultaisen standardin puuttuessa sensitiivisyyksien ja spesifisyyksien estimointi ei onnistu tavanomaisin menetelmin, sillä tutkittavien yksilöiden todellinen taudin tila on tuntematon.

Tässä raportissa käsitellään latentin luokan analyysia diagnostisten testien luotettavuuden ja pneumokokkikeuhkokuumeen prevalenssin arvioimisessa, kun kultainen standardi puuttuu. Menetelmää sovelletaan Kansanterveyslaitoksen keuhkokuumeetutkimuksessa kerättyyn aineistoon. Menetelmään liittyvän teorian lisäksi tutkielmassa esitellään estimointia varten R-ohjelmalle kirjoittamani funktio.

Asiasanat: latentin luokan analyysi, sensitiivisyys, spesifisyys

## CONTENTS

1. Introduction	6
2. FinCAP epidemiological study	8
2.1. Introduction	8
2.2. Collected samples and tests performed	8
2.3. Missing samples	10
2.4. Preliminary inspection of the data	11
3. Latent class analysis	13
3.1. Introduction	13
3.2. The unconstrained latent class model assuming local independence	16
3.3. Local dependence in latent class models	17
4. Parameter estimation and model diagnostics	19
4.1. Estimation algorithms	19
4.2. Identifiability	22
4.3. Missing data	23
4.4. Software	25
4.5. Model diagnostics	26
5. Latent class models for FinCAP Epi data	28
6. Discussion	33
References	35
Appendix A. EM-algorithm	38
Likelihood function for complete data	38
Appendix B. Tables	41
Appendix C. Functions	43

## 1. INTRODUCTION

Community-acquired pneumonia (CAP) is pneumonia that is not acquired during hospitalization or institutionalization. CAP is a common illness and a major cause of death worldwide among all age groups, especially among children and the elderly. In Finland, approximately 60 000 cases of community-acquired pneumonia occur yearly, of which approximately 40 % are hospitalized (Korppi et al., 2003). In the elderly population aged 65 years or more, pneumonia causes over 2000 deaths and over 10000 periods of hospital treatment yearly.

Most community-acquired pneumonias are bacterial and the most common causative pathogen is *Streptococcus pneumoniae*, pneumococcus. The exact proportion of CAP cases in adults caused by pneumococcus is unknown: estimates vary widely between 12 % and 41 %, partly depending on the population but mainly on how the diagnosis of pneumococcal CAP was made (Koivula, 2001, and references therein). The diagnosis of pneumococcal CAP is difficult, not only because the general definition of community-acquired pneumonia is lacking, but also because there is no gold standard, a diagnostic test that would identify all pneumococcal cases and yet be definite. The ideal gold standard is 100 % specific and 100 % sensitive, which means that there are neither false positive nor false negative test results. For example, blood culture is generally regarded as 100 % specific but not very sensitive, and therefore will not identify all patients with pneumococcal CAP.

In the spring of 2005 the National Public Health Institute of Finland (KTL) launched a 2-year study - the Finnish Community-Acquired Pneumonia Epidemiological (Fin-CAP Epi) study - investigating the frequency and causes of community-acquired pneumonia in the elderly. One of the main objectives is to investigate the definition of pneumococcal pneumonia. Blood, urine, sputum and nasopharyngeal samples are collected from patients enrolled in the study, and analysed using a number of different diagnostic tests. The accuracy of the tests is evaluated by estimating the sensitivities and the specificities of the tests. The aim is to find a pattern of diagnostic tests that could be used for the case definition of pneumococcal CAP. This case definition is planned to be later used in a study of efficacy of a pneumococcal vaccine against pneumonia in the elderly people.

The sensitivity of a test is defined as the proportion of true positives of all diseased cases in the population, or as the probability of the test result being positive given that the subject has the disease. Specificity is the proportion of true negatives of all cases without the disease in the population, or the probability of the test result being negative given that the subject does not have the disease. Let  $D$  indicate the

true disease status of the subject and  $X$  indicate the result of test  $X$ .  $D$  and  $X$  can only take one of two values:  $D = 1$  when the subject has the disease,  $D = 0$  when the subject does not have the disease. Similarly  $X = 1$  when the test result is positive,  $X = 0$  when the test result is negative. For test  $X$  the sensitivity is the conditional probability

$$P(X = 1|D = 1) = \frac{P(X = 1, D = 1)}{P(D = 1)} = \frac{P(\text{true positive})}{P(\text{diseased})},$$

and the specificity is the conditional probability

$$P(X = 0|D = 0) = \frac{P(X = 0, D = 0)}{P(D = 0)} = \frac{P(\text{true negative})}{1 - P(\text{diseased})},$$

where  $P(D = 1)$  is the prevalence of the disease: the proportion of diseased subjects at a given time point.

Estimating sensitivity and specificity is not complicated when the true disease status of subjects is known. However, as mentioned above, pneumococcal CAP does not have a gold standard and therefore the true disease status remains latent for each subject. In the absence of a gold standard, specificity and sensitivity of a test cannot be calculated directly and more advanced methods are needed.

Latent class analysis (LCA) is a statistical method for identifying unobserved class memberships using the information from a set of observed discrete variables that imperfectly measure the true class membership. The observed variables are dependent, but the association can be explained by the latent class variable. The concept of LCA was introduced by Paul F. Lazarsfeld in 1950 under the name latent structure model, but the main breakthrough was not seen until 1974 when Leo Goodman developed an algorithm for obtaining maximum likelihood estimates of the model parameters making it possible to use LCA in practice. For a historical review of LCA, see, e.g. Langeheine (1988) and Goodman (2002).

LCA was first developed for social sciences and psychiatric research, but it has gained popularity in medical research when assessing the accuracy of diagnostic tests when no gold standard is available. LCA has been applied to study the performance of diagnostic tests for myocardial infarction (Rindskopf and Rindskopf, 1986), visceral leishmaniasis (Boelaert et al., 1999), human herpesvirus (Engels et al., 2000) and pneumococcal community-acquired pneumonia (Butler et al., 2003), to name a few. In this thesis, LCA is applied to the data collected in the FinCAP Epi study.



## 2. FINCAP EPIDEMIOLOGICAL STUDY

**2.1. Introduction.** One of the primary objectives of the Finnish community-acquired pneumonia epidemiological (FinCAP Epi) study is to find a case definition for pneumococcal community-acquired pneumonia (pnc CAP) in the elderly. The study is carried out in Tampere, where the population size of non-institutionalized elderly residents aged 65 years or older is approximately 28 000. Based on the incidence estimates from a previous study conducted in Finland (Koivula, 2001, and the references therein), the expected number of CAP cases in this population is around 700 during the two years of follow-up. A free of charge study clinic has been established at Hatanpää Health Centre in order to capture as many potential cases of CAP as possible. The study clinic services are available every day for all elderly having symptoms suggestive of pneumonia. Local health care personnel have been encouraged to refer all patients with suspected pneumonia directly to the clinic. Referral is not a requirement and the patients can walk into the clinic even without an appointment. The patients entering the study clinic are examined and interviewed by the study physician. Also diagnosed CAP patients from Tampere University Hospital are recruited if possible. The study is conducted by the National Health Institute (KTL) in co-operation with GlaxoSmithKline (GSK) Biologicals.

In order to be eligible to be enrolled in the study as a CAP case the subject must be non-institutionalized, at least 65 years of age, permanently living in Tampere and not have been treated in hospital within one week, except if the patient is enrolled within 48 hours of hospitalization. In addition, the subject must obviously have symptoms or signs suggesting pneumonia and the radiological findings need to be consistent with pneumonia. The subjects fulfilling all of the enrollment criteria are offered to enter the study. Within 4 to 8 weeks of the first visit, i.e. the acute visit, a follow-up visit is scheduled.

The study was launched in mid-May 2005 and, by the end of January 2007, altogether 411 patients with clinical suspicion of community-acquired pneumonia with an informed consent have been enrolled in the study. Of these, 64 have been rejected on the basis of further radiological evaluations. The remaining 347 subjects are assumed to be CAP cases and 277 of them have attended the follow-up visit within 4 to 8 weeks of the acute visit.

**2.2. Collected samples and tests performed.** Sputum, blood, urine and nasopharyngeal swab (NPS) samples are collected if possible from each enrolled CAP case. A number of microbiological assays are performed on the samples including

culture, polymerase chain reaction (PCR), antigen detection and serology. All laboratory analyses are performed by KTL, except for blood culture and serology, which are performed by Tampere University Hospital and GSK respectively.

Blood and sputum cultures are conventional diagnostic tests performed when establishing the etiology of community-acquired pneumonia. Blood culture is the only test generally regarded as 100 % specific, but typically it has a very low sensitivity. Sputum culture is also lacking sensitivity, especially if the patient has received antimicrobial treatment. The sputum sample is therefore also analysed using PCR, a relatively new technique in which the pneumococcal DNA is amplified. PCR detects even low numbers of the pathogen and, compared to culture, it is believed to be less affected by antimicrobial treatment because it can detect non-viable pathogens. This, however, may cause false positives, which implies low specificity. In this study pneumolysin (ply) PCR is used. Positive PCR ply test results are considered negative in the case when the number of genomes found is not sufficient: the cut-point is set to 5 genomes per microlitre in this study.

Diagnosis based only on sputum is considered controversial due to possible contamination from the upper airways. However, the results from NPS culture and NPS PCR ply indicate that contamination is not a major issue in this study; see section 2.4. *Streptococcus pneumoniae* isolates in blood, NPS and sputum culture are serotyped, that is, classified according to the structure of the polysaccharide capsule enveloping the pneumococcal cell. The findings lacking a capsule could not be serotyped and are regarded as negative. Quantitative analysis was performed also for sputum and NPS culture: the number of pneumococcal colonies on the culture plate was counted. In contrast to sputum PCR ply, no cut-point is set to NPS PCR ply, and all positive results are considered positive, regardless of the number of genomes.

Serology and urine antigen test are not affected by antimicrobial treatment. Antibodies against Pnc surface adhesin A (PsaA) and choline binding protein A (CbpA) are measured from blood serum samples taken on the acute visit and the follow-up visit. An increase in the antibody levels indicates the presence of pneumococcus. In this study, at least a 2-fold increase is regarded as a positive test result. Since antibody levels from two blood samples taken at least a few weeks apart are needed, serology is not a practical method to make a rapid diagnosis of pnc CAP. In a vaccine trial, however, this does not pose a problem.

The following notations and abbreviations are used in the subsequent sections for the eight diagnostic tests described above:

- Sputum culture (Sc)
- Sputum PCR ply (Sp)
- NPS culture (Nc)
- NPS PCR ply (Np)
- Urine antigen (U)
- Blood culture (B)
- PsaA 2-fold increase (PFi2)
- Cbpa 2-fold increase (CFi2)

**2.3. Missing samples.** Out of the 347 CAP cases, only 152 have all samples obtained and analysed. The proportion of missing samples are presented in Table 1. Since the number of missing samples is quite remarkable, it is important to consider the reasons for missingness, for example, whether a sample is more likely to be missing from patients with pneumococcal CAP, or vice versa.

TABLE 1. Missing test results

Sample	Test	n Missing	% Missing
<b>Sputum</b>	culture	110	31.7
	PCR ply	137	39.5
<b>NPS</b>	culture	25	7.2
	PCR ply	25	7.2
<b>Urine</b>	antigen	42	12.1
<b>Blood</b>	culture	18	5.2
	PsaA 2-fold increase	92	26.5
	Cbpa 2-fold increase	75	21.6

Sputum sample is missing from one third of the CAP cases. In addition, sputum sample volume has not been sufficient to perform PCR for 15 CAP cases, and therefore 40 % of the sputum PCR ply results are missing. Sputum sample is hard to collect. Not all cases produce purulent sputum, and as purulent sputum suggests the presence of bacterial infection, these cases are less likely to have CAP caused by pneumococcus. On the other hand, sputum sample is hard to obtain from subjects in poor condition. Since pneumococcus is the most common cause of CAP in patients admitted to hospital, and accounts for more deaths than any

other pathogen, one can assume that a large part of these subjects have pneumonia caused by pneumococcus.

Blood serum has been obtained from all subjects at the acute visit, but since not all subjects have attended the follow-up visit, serology results are missing from one fifth of the subjects. In addition, fold-increase against PsaA is missing from further 17 subjects due to reasons in serological assay techniques. Follow-up visits are missing for several reasons: subjects with a severe condition have either deceased, or are still hospitalized when the follow-up visit is scheduled to take place. Some have decided to withdraw from the study. Unfortunately, the data on reasons for missing follow-up visit are not yet available for all, and therefore the actual reasons for missing serology results remain unknown for the time being.

Urine antigen is missing for approximately 12% of the subjects. Although all patients produce urine, the sample is hard to collect from severe cases unless the patient has been catheterized. A blood sample has been obtained from each subject. The reason for blood culture results missing is unlikely to be related to the severity of the disease or the presence of pneumococcus: 18 blood culture results are missing simply because they have not been received from the Tampere University Hospital register. NPS sample is missing only for 7.2 % of the subjects, and therefore the reasons for missing NPS samples are unlikely to affect the inferences.

**2.4. Preliminary inspection of the data.** The descriptives of the eight tests described in Section 2.2 are presented in Table 2. The test prevalence is the proportion of positive test results out of analysed samples. Since a large proportion of test results is missing, also the proportion of positive test results out of all CAP cases is of interest. These are tabulated in the rightmost column of Table 2.

Pneumolysin PCR for sputum and NPS are the two most prevalent tests with 0.386 and 0.199 test prevalences respectively. If the missing test results are regarded as negative, the prevalence of sputum PCR ply drops remarkably to 0.233. As mentioned above, there are two concerns with the sputum sample. Firstly, it is missing for a third of the subjects and thus is not very useful as a sole diagnostic test for case definition purposes, because the missing samples cannot be assumed to be positive. Secondly, the sample may be contaminated, which would mean that the pneumococcus found in the sample is not from the lower, but from the upper respiratory tract. Since pneumococcus can also be found in the nasopharynx of healthy subjects, diagnosis of pnc CAP based on pneumococcal findings in the upper airways is not considered reliable. However, the data from the FinCAP Epi study suggest that contamination from nasopharynx is minimal: The test prevalence

TABLE 2. Proportion of positive test results out of obtained samples and of all CAP cases

Test	n +	Test prevalence	Proportion positives out of 347 CAP cases
Sputum culture	38	0.160	0.110
Sputum PCR ply	81	0.386	0.233
NPS culture	32	0.099	0.092
NPS PCR ply	64	0.199	0.184
Urine Antigen	24	0.079	0.069
Blood culture	9	0.027	0.026
PsaA 2-fold increase	27	0.106	0.078
CbpA 2-fold increase	29	0.107	0.084

of NPS culture is considerably lower than the prevalence of sputum culture. There are only five CAP cases with a negative sputum culture but a positive NPS culture, see Table 3. In addition, the number of pnc colonies in sputum culture is high for the cases with pneumococcal findings in both NPS and sputum culture, which is unlikely to be the case if the sputum sample was contaminated.

TABLE 3. Sputum culture result compared to NPS culture result

Sputum culture	NPS culture	
	-	+
-	183	5
+	14	22

The yield of blood culture is low as expected. However, the yield of positive urine samples was considerably lower than expected, approximately 8 % tested positive. The urine sample is easy to collect compared to sputum, antibiotic treatment does not diminish the yield of positives and the test is simple to perform. In addition, the result is available within 15 minutes of the sampling. Low test prevalence was thus a disappointment, especially since in previous studies the urine antigen test has shown considerably higher prevalences, see for example Butler et al. (2003).

Both serological tests had also a relatively low prevalence. The cross-tabulation of PsaA 2-fold increase and CbpA 2-fold increase shows that at least 2-fold increase

in antibodies against one protein does not necessarily mean an increase in the antibodies against the other protein as well. Since the prevalences of the two tests are almost identical, and the main interest is in the increase in antibody levels rather than in the rise of the antibody levels against a specific protein, a composite test of the two serological tests was formed, denoted by 2Fi in subsequent sections. When at least 2-fold increase in antibody levels against either PsaA or CbpA is considered positive, the test prevalence of the composite test is 0.143.

TABLE 4. PsaA 2-fold increase compared to CbpA 2-fold increase

CbpA 2-fold increase	PsaA 2-fold increase	
	-	+
-	216	11
+	10	17

The sensitivities and specificities of these tests are unknown. In the absence of a gold standard, the sensitivities and specificities can be estimated using latent class analysis. This method is described in the following section.

### 3. LATENT CLASS ANALYSIS

**3.1. Introduction.** Consider three tests: sputum culture, urine antigen and blood culture, denoted by  $X_1$ ,  $X_2$  and  $X_3$  respectively, and let each test have two possible realisations: 1 indicating positive and 0 negative result. There are altogether 8 ( $2^3$ ) different combinations of the test results, represented in Table 5. Denote the probability of each combination of the test results by  $\pi_s$ ,  $s = 1, \dots, 8$ . For example  $\pi_1 = P(X_1 = 0, X_2 = 0, X_3 = 0)$ .

For simplicity, concentrate on the 179 CAP cases with results available from all three tests. We have observed the frequencies in different combinations of the test results  $\mathbf{n} = [153, 26, 10, 3, 0, 0, 3, 2]$ , but for each CAP case the true disease status  $D$ , where 1 is indicating pneumococcal CAP and 0 is indicating CAP caused by other pathogen, is unknown.

It is assumed that the observed frequencies in the cross-classification of the test results follow a multinomial distribution with 8 probabilities  $\pi_s$ ,  $s = 1, \dots, 8$  where

TABLE 5. All possible test result combinations of three tests

$X_1$	$X_2$	$X_3$	Observed frequency	Probability
0	0	0	153	$\pi_1$
1	0	0	26	$\pi_2$
0	1	0	10	$\pi_3$
1	1	0	3	$\pi_4$
0	0	1	0	$\pi_5$
1	0	1	0	$\pi_6$
0	1	1	3	$\pi_7$
1	1	1	2	$\pi_8$

$\sum_{s=1}^8 \pi_s = 1$ . The multinomial log-likelihood function is

$$(3.1) \quad l(\boldsymbol{\pi}) = \sum_{s=1}^8 n_s \times \log \pi_s$$

where  $n_s$  is the observed frequency of the  $s$ th pattern of test results, see e.g. Agresti (2002, p. 21).

By utilizing the rule of total probability, the probability that a pattern of test results is  $\{X_1 = i, X_2 = j, X_3 = k\}$ , where  $i, j, k = 0, 1$ , can be expressed using the latent class  $D$  as follows:

$$(3.2) \quad P(X_1 = i, X_2 = j, X_3 = k) = \sum_{d=0}^1 P(X_1 = i, X_2 = j, X_3 = k | D = d) \times P(D = d).$$

In the standard latent class model it is assumed that, given the true disease status, the test results are independent of each other:

$$(3.3) \quad \begin{aligned} P(X_1 = i, X_2 = j, X_3 = k | D = d) &= P(X_1 = i | D = d) \times \\ &P(X_2 = j | D = d) \times \\ &P(X_3 = k | D = d). \end{aligned}$$

In other words, conditional on the disease status, knowledge of one of test results gives no information about the other (Pepe and Janes, 2007). This assumption is often referred to as the local independence or the conditional independence assumption. An implication of the local independence assumption is that the only parameters in the model are the prevalence of the disease and the sensitivities and

specificities of the tests. Denote

$$(3.4) \quad P(D = 1) = \mu,$$

$$(3.5) \quad P(X_k = 1|D = 1) = \psi_{X_k},$$

$$(3.6) \quad P(X_k = 0|D = 0) = \phi_{X_k}.$$

In words,  $\mu$  is the prevalence of the disease,  $\psi_{X_k}$  is the sensitivity of the  $k$ th test and  $\phi_{X_k}$  the specificity of the  $k$ th test.

By using the equations (3.2) – (3.6),  $\pi_1$  can expressed as

$$\begin{aligned} \pi_1 &= P(X_1 = 0, X_2 = 0, X_3 = 0|D = 1) \times P(D = 1) + \\ &\quad P(X_1 = 0, X_2 = 0, X_3 = 0|D = 0) \times P(D = 0) \\ &= P(X_1 = 0|D = 1) \times P(X_2 = 0|D = 1) \times P(X_3 = 0|D = 1) \times P(D = 1) + \\ &\quad P(X_1 = 0|D = 0) \times P(X_2 = 0|D = 0) \times P(X_3 = 0|D = 0) \times P(D = 0) \\ &= (1 - \psi_{X_1}) \times (1 - \psi_{X_2}) \times (1 - \psi_{X_3}) \times \mu + \phi_{X_1} \times \phi_{X_2} \times \phi_{X_3} \times (1 - \mu) \end{aligned}$$

and similarly for  $\pi_2, \dots, \pi_8$ .

Let  $\boldsymbol{\theta}$  denote the model parameters:  $\boldsymbol{\theta} = (\mu, \psi_{X_1}, \psi_{X_2}, \psi_{X_3}, \phi_{X_1}, \phi_{X_2}, \phi_{X_3})$ . The log-likelihood function in this example assuming local independence is

$$\begin{aligned} l(\boldsymbol{\theta}) &= \sum_{s=1}^8 n_s \times \log \pi_s \\ &= 153 \times \log \{(1 - \psi_{X_1}) \times (1 - \psi_{X_2}) \times (1 - \psi_{X_3}) \times \mu + \\ &\quad \phi_{X_1} \times \phi_{X_2} \times \phi_{X_3} \times (1 - \mu)\} + \\ &26 \times \log \{\psi_{X_1} \times (1 - \psi_{X_2}) \times (1 - \psi_{X_3}) \times \mu + \\ &\quad (1 - \phi_{X_1}) \times \phi_{X_2} \times \phi_{X_3} \times (1 - \mu)\} + \\ &10 \times \log \{(1 - \psi_{X_1}) \times \psi_{X_2} \times (1 - \psi_{X_3}) \times \mu + \\ &\quad \phi_{X_1} \times (1 - \phi_{X_2}) \times \phi_{X_3} \times (1 - \mu)\} + \\ &3 \times \log \{\psi_{X_1} \times \psi_{X_2} \times (1 - \psi_{X_3}) \times \mu + \\ &\quad (1 - \phi_{X_1}) \times (1 - \phi_{X_2}) \times \phi_{X_3} \times (1 - \mu)\} + \\ &0 \times \log \{(1 - \psi_{X_1}) \times (1 - \psi_{X_2}) \times \psi_{X_3} \times \mu + \\ &\quad \phi_{X_1} \times \phi_{X_2} \times (1 - \phi_{X_3}) \times (1 - \mu)\} + \\ &0 \times \log \{\psi_{X_1} \times (1 - \psi_{X_2}) \times \psi_{X_3} \times \mu + \\ &\quad (1 - \phi_{X_1}) \times \phi_{X_2} \times (1 - \phi_{X_3}) \times (1 - \mu)\} + \\ &3 \times \log \{(1 - \psi_{X_1}) \times \psi_{X_2} \times \psi_{X_3} \times \mu + \\ &\quad (\phi_{X_1} \times (1 - \phi_{X_2}) \times (1 - \phi_{X_3}) \times (1 - \mu)\} + \\ &2 \times \log \{\psi_{X_1} \times \psi_{X_2} \times \psi_{X_3} \times \mu + \\ &\quad (1 - \phi_{X_1}) \times (1 - \phi_{X_2}) \times (1 - \phi_{X_3}) \times (1 - \mu)\}. \end{aligned}$$



There are altogether 7 parameters to estimate on the basis of 8 observed frequencies. Methods for estimating these parameters are discussed in Section 4. In the following section, the latent class model for  $K$  tests is introduced.

### 3.2. The unconstrained latent class model assuming local independence.

Consider  $K$  diagnostic tests, and denote a  $1 \times K$ -vector of random variables as  $\mathbf{X} = (X_1, \dots, X_K)$ , where the realisation of  $X_k$ ,  $k = 1, \dots, K$ , is 0 when the  $k$ th test is negative and 1 when positive. Consequently, there are  $2^K$  different realisations of  $\mathbf{X}$ , which represent the different combinations of the  $K$  test results. These can be conveniently expressed in matrix form. Define the  $2^K \times K$  matrix  $\mathbf{A}$  of binary indicators as

$$(3.7) \quad \mathbf{A} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}$$

where the first column is changing the fastest and the last column the slowest. Each row of matrix  $\mathbf{A}$  corresponds to a different realisation of  $\mathbf{X}$ , denoted by  $\mathbf{a}_s$ ,  $s = 1, \dots, 2^K$ .

In the absence of a gold standard the binary variable  $D$  indicating the presence ( $D = 1$ ) or absence ( $D = 0$ ) of the disease is unobserved. What are observed are the frequencies of tested subjects in different combinations of the  $K$  tests,  $\mathbf{n} = \{n_1, n_2, \dots, n_{2^K}\}$ . When assuming local independence the probabilities of the different test result combinations can be expressed as generalisations of equations (3.2) – (3.3) as follows:

$$\begin{aligned}
P(\mathbf{X} = \mathbf{a}_s.) &= P(X_1 = a_{s1}, X_2 = a_{s2}, \dots, X_K = a_{sK}) \\
&= \sum_{d=0}^1 P(X_1 = a_{s1}, X_2 = a_{s2}, \dots, X_K = a_{sK} | D = d) P(D = d) \\
&= \sum_{d=0}^1 \prod_{k=1}^K P(X_k = a_{sk} | D = d) P(D = d) \\
&= \prod_{k=1}^K P(X_k = a_{sk} | D = 0) P(D = 0) + \\
&\quad \prod_{k=1}^K P(X_k = a_{sk} | D = 1) P(D = 1) \\
&= P(D = 0) \prod_{k=1}^K P(X_k = a_{sk} | D = 0) + \\
&\quad P(D = 1) \prod_{k=1}^K P(X_k = a_{sk} | D = 1) \\
&= P(D = 0) \prod_{k=1}^K P(X_k = 0 | D = 0)^{1-a_{sk}} \{1 - P(X_k = 0 | D = 0)\}^{a_{sk}} + \\
&\quad P(D = 1) \prod_{k=1}^K P(X_k = 1 | D = 1)^{a_{sk}} \{1 - P(X_k = 1 | D = 1)\}^{1-a_{sk}}.
\end{aligned}$$

Recall from Section 3.1 that  $\mu = P(D = 1)$ ,  $\psi_{X_k} = P(X_k = 1 | D = 1)$  and  $\phi_{X_k} = P(X_k = 0 | D = 0)$ , i.e disease prevalence, sensitivity of the  $k$ th test and the specificity of the  $k$ th test respectively. Then the  $1 \times K$ -vectors  $\boldsymbol{\psi} = (\psi_{X_1}, \dots, \psi_{X_K})$  and  $\boldsymbol{\phi} = (\phi_{X_1}, \dots, \phi_{X_K})$  denote the sensitivities and the specificities of the  $K$  tests respectively.

Let  $\boldsymbol{\theta}$  denote the model parameters:  $\boldsymbol{\theta} = (\mu, \boldsymbol{\psi}, \boldsymbol{\phi})$ . Assuming local independence and using a generalisation of Equation 3.1, the log-likelihood function can be expressed as

$$\begin{aligned}
l(\boldsymbol{\theta}) &= \sum_{s=1}^{2^K} n_s \times \log \{P(\mathbf{X} = \mathbf{a}_s.)\} \\
&= \sum_{s=1}^{2^K} n_s \cdot \log \left\{ (1 - \mu) \cdot \prod_{k=1}^K \{1 - \phi_{X_k}\}^{a_{sk}} \cdot \phi_{X_k}^{1-a_{sk}} + \right. \\
&\quad \left. \mu \cdot \prod_{k=1}^K \psi_{X_k}^{a_{sk}} \cdot \{1 - \psi_{X_k}\}^{1-a_{sk}} \right\}.
\end{aligned}$$

**3.3. Local dependence in latent class models.** As mentioned above, local independence is a basic assumption of the standard latent class model. However, this assumption often fails in practice, for example, in a situation where the same sample is analysed using two different tests measuring the same condition. Also, local dependence may occur in situations where there exists a spectrum of severity of the disease: the severe cases are likely to be positive for most tests whereas the less severe are more likely to be negative for more than one test (Hui and Zhou, 1998).

If local independence is falsely assumed, the model is misspecified and the model estimates of sensitivities and specificities may be biased.

Latent class analysis has been criticized for the local independence assumption by Alonzo and Pepe (1999), among others. However, local dependence can be accounted for. If two diagnostic tests are locally dependent, one can delete one of the tests from the model or, alternatively, form a composite test, assuming that the remaining tests are locally independent. This method is not very useful in the case when there are several dependent tests in the model, or in the case when the sensitivities and specificities of both tests are of interest. One could also increase the number of categories in the latent variable to improve the model fit; for example, instead of categories non-diseased/diseased have three categories non-diseased/probably diseased/definitely diseased. However, this would inflict the interpretation of the model parameters: the parameters would no longer correspond to sensitivities and specificities of the tests nor the prevalence of the disease. Alternative and more feasible methods are to increase the number of latent variables, or to model the dependence of one or several pairs of tests explicitly. The latter approach is described in section 3.3.2.

3.3.1. *Detecting local dependence.* Several methods to detect local dependence have been suggested over the years by Hagenars (1988), Garret and Zeger (2000) and Qu et al. (1996) among others. These methods are based on comparison of the observed frequencies and the model-predicted frequencies for a pair of diagnostic tests. For example, a large difference between the observed and the expected log odds ratio for a pair of tests implies local dependence. When calculating the expected frequencies, one must obviously first fit a model where local independence is assumed. Therefore, detecting local dependence is rather a part of model diagnostics than a preliminary step before fitting the model.

3.3.2. *Modelling local dependence.* Assume that the tests  $X_1$  and  $X_2$  are locally dependent. In other words the sensitivities and specificities are not independent:

$$\begin{aligned} P(X_1 = 1, X_2 = 1|D = 1) &\neq P(X_1 = 1|D = 1) \times P(X_2 = 1|D = 1), \\ P(X_1 = 0, X_2 = 0|D = 0) &\neq P(X_1 = 0|D = 0) \times P(X_2 = 0|D = 0). \end{aligned}$$

One way to model local dependence between the two tests is to replace them with a joint test,  $X_{12}$ , with four possible realisations: 00, 10, 01 and 11. Uebersax (2000) refers to this as the joint item method, but it is also known as direct effects method,

first introduced by Harper (1972) and Hagenaaars (1988). Variable  $X_{12}$  has four possible realisations and the following eight conditional probabilities:

$$\begin{aligned}
&P(X_1 = 0, X_2 = 0|D = 1) \\
&P(X_1 = 1, X_2 = 0|D = 1) \\
&P(X_1 = 0, X_2 = 1|D = 1) \\
&P(X_1 = 1, X_2 = 1|D = 1) \\
&P(X_1 = 0, X_2 = 0|D = 0) \\
&P(X_1 = 1, X_2 = 0|D = 0) \\
&P(X_1 = 0, X_2 = 1|D = 0) \\
&P(X_1 = 1, X_2 = 1|D = 0).
\end{aligned}$$

Since  $\sum_{i,j} P(X_1 = i, X_2 = j|D = 1) = 1$ , and  $\sum_{i,j} P(X_1 = i, X_2 = j|D = 0) = 1$ , only six of these eight parameters need to be estimated, which is two more compared to a model with a local independence assumption.

Sensitivities  $\psi_{X_i}$ ,  $i = 1, 2$ , and specificities  $\phi_{X_i}$ ,  $i = 1, 2$ , can be calculated as marginal sums from the conditional probabilities:

$$\begin{aligned}
\psi_{X_1} &= P(X_1 = 1, X_2 = 0|D = 1) + P(X_1 = 1, X_2 = 1|D = 1) \\
\psi_{X_2} &= P(X_1 = 0, X_2 = 1|D = 1) + P(X_1 = 1, X_2 = 1|D = 1) \\
\phi_{X_1} &= P(X_1 = 0, X_2 = 0|D = 0) + P(X_1 = 0, X_2 = 1|D = 0) \\
\phi_{X_2} &= P(X_1 = 0, X_2 = 0|D = 0) + P(X_1 = 1, X_2 = 0|D = 0).
\end{aligned}$$

A latent class model with a local independence assumption requires at least three different tests for all model parameters to be identifiable. If local dependence between two tests is modeled with the direct effects method, the minimum number of tests is four.

Other methods for dealing with local dependence have been proposed. The joint item method is a special case of the random effects model, first introduced by Qu et al. (1996).

## 4. PARAMETER ESTIMATION AND MODEL DIAGNOSTICS

**4.1. Estimation algorithms.** For many statistical models an explicit expression for obtaining maximum likelihood (ML) estimates is lacking. For such models, the likelihood function is typically maximized using iterative methods.

Although Pepe and Janes (2007) have derived analytic expressions for the parameter estimates in a 3-test latent class model, iterative methods are, in general, needed to

find the maximum likelihood estimates of parameters in a latent class model. The main optimization algorithms used are the Expectation-Maximization algorithm (EM) and the Newton-Raphson method.

4.1.1. *Expectation-maximization algorithm.* The EM-algorithm is an iterative method for finding ML estimates of parameters when the data are incomplete. The method had been proposed for many years in special contexts before it was generalized and named by Dempster, Laird and Rubin in 1977, see for example Little and Rubin (1987). In fact, the proportional fitting algorithm for fitting latent class models proposed by Goodman in 1974 is a variant of the EM-algorithm (Formann and Kohlmann, 1996).

The EM-algorithm is actually not a specific algorithm but a general description for an algorithm. The basic idea behind the algorithm is to first fill in the missing data (E-step) using the observed data and the current parameter estimates and then maximize the likelihood for complete data (M-step) in order to get improved estimates of the model parameters. The two steps are repeated until convergence of the likelihood. The formal definition of the algorithm is described in detail by Dempster et al. (1977).

The EM-algorithm is well suited for latent class models since the models are naturally formulated in terms of missing, that is latent, data. Consider the latent class model described in Section 3.2. The missing data are the numbers of diseased subjects in the different test result combinations. In the E-step the unobserved frequencies are replaced by working frequencies which are derived from the observed data and the current estimates of the prevalence of the disease and the sensitivities and specificities of the tests. In the M-step the likelihood function for complete data, that is, the likelihood function we would have if the true disease status of each subject was known is maximized over the model parameters using as data the working frequencies calculated in the E-step. The log-likelihood function for complete data and the maximum likelihood estimators of the parameters are derived in the Appendix. More specifically, the algorithm is the following:

- (1) Define starting values for  $\boldsymbol{\theta}^{(0)} = (\mu^{(0)}, \psi_{X_1}^{(0)}, \dots, \psi_{X_K}^{(0)}, \phi_{X_1}^{(0)}, \dots, \phi_{X_K}^{(0)})$ .
- (2) The E step: in the  $(r+1)$ st iteration calculate  $P(\mathbf{X} = \mathbf{a}_s | D = d)^{(r+1)}$ ,  $s = 1, \dots, 2^K$ , and  $d = 0, 1$  using,  $\boldsymbol{\theta}^{(r)} = (\mu^{(r)}, \psi_{X_1}^{(r)}, \dots, \psi_{X_K}^{(r)}, \phi_{X_1}^{(r)}, \dots, \phi_{X_K}^{(r)})$

$$(4.1) \quad P(\mathbf{X} = \mathbf{a}_s | D = 0)^{(r+1)} = \prod_{k=1}^K \phi_{X_k}^{(r) 1 - a_{sk}} [1 - \phi_{X_k}^{(r)}]^{a_{sk}},$$

$$(4.2) \quad P(\mathbf{X} = \mathbf{a}_s | D = 1)^{(r+1)} = \prod_{k=1}^K \psi_{X_k}^{(r) a_{sk}} [1 - \psi_{X_k}^{(r)}]^{1-a_{sk}},$$

and

$$(4.3) \quad P(\mathbf{X} = \mathbf{a}_s)^{(r+1)} = \mu^{(r)} \cdot P(\mathbf{X} = \mathbf{a}_s | D = 1)^{(r+1)} + (1 - \mu^{(r)}) \cdot P(\mathbf{X} = \mathbf{a}_s | D = 0)^{(r+1)}.$$

Then calculate the working frequencies, denoted by

$\mathbf{f}^{(r+1)} = [f_1^{(r+1)}, f_2^{(r+1)}, \dots, f_{2^K}^{(r+1)}]$  using the Equations (4.1-4.3) and the observed frequencies  $\mathbf{n} = [n_1, n_2, \dots, n_{2^K}]$ :

$$f_s^{(r+1)} = n_s \cdot \frac{P(\mathbf{X} = \mathbf{a}_s | D = 1)^{(r+1)} \cdot \mu^{(r)}}{P(\mathbf{X} = \mathbf{a}_s)^{(r+1)}}, s = 1, \dots, 2^K.$$

- (3) The M step: Using the working frequencies  $\mathbf{f}^{(r+1)}$ , calculate new values  $\mu^{(r+1)}$ ,  $\boldsymbol{\psi}^{(r+1)} = (\psi_{X_1}^{(r+1)}, \dots, \psi_{X_K}^{(r+1)})$  and  $\boldsymbol{\phi}^{(r+1)} = (\phi_{X_1}^{(r+1)}, \dots, \phi_{X_K}^{(r+1)})$ :

$$\mu^{(r+1)} = \sum_{s=1}^{2^K} \frac{f_s^{(r+1)}}{N},$$

and

$$\boldsymbol{\psi}^{(r+1)} = \frac{1}{\mu^{(r+1)}} \text{diag}\left(\frac{f_1^{(r+1)}}{N}, \dots, \frac{f_{2^K}^{(r+1)}}{N}\right) \mathbf{A},$$

and

$$\boldsymbol{\phi}^{(r+1)} = \frac{1}{(1 - \mu^{(r+1)})} \text{diag}\left(\frac{(n_1 - f_1)^{(r+1)}}{N}, \dots, \frac{(n_{2^K} - f_{2^K})^{(r+1)}}{N}\right) (\mathbf{B} - \mathbf{A}),$$

where  $N = \sum_{s=1}^{2^K} n_s$ ,  $\mathbf{B}$  is a  $2^K \times K$  matrix of ones and  $\mathbf{A}$  is the matrix 3.7 in Section 3.2.

- (4) Repeat steps 2 and 3 until convergence.

4.1.2. *Newton-Raphson method.* The Newton method, or the Newton-Raphson (NR) method as it is often called, is also an iterative method, and it begins with a set of initial parameter values  $\boldsymbol{\theta}^{(0)}$ . On each iteration, a gradient vector of the log-likelihood function, denoted by  $\mathbf{g}(\boldsymbol{\theta})$ , and a matrix of second order derivatives, called the Hessian matrix and denoted by  $\mathbf{H}$ , are calculated. The parameter values on the  $(r + 1)$ th iteration are calculated using the gradient vector and the inverse of the

Hessian matrix from the previous iteration  $r$  as follows:

$$\boldsymbol{\theta}^{(r+1)} = \boldsymbol{\theta}^{(r)} - \mathbf{H}^{(r)-1} \cdot \mathbf{g}^{(r)}$$

where  $-\mathbf{H}^{(r)-1} \cdot \mathbf{g}^{(r)}$  is called the Newton step, see e.g. McCutcheon (2002).

Seber and Wild (2003) do not recommend this unmodified Newton-Raphson method for maximizing the likelihood function for two reasons. If the starting values  $\boldsymbol{\theta}^{(0)}$  are not close enough to the final solution, the Newton step may be too long and even decrease the likelihood. Secondly, the Newton-Raphson method does not require the Hessian to be negative definite at each iteration and therefore does not ensure an increase in the likelihood at each Newton step. This may lead to that the method converges to a local minimum. Instead of using the Newton method described above one should use a modified version of it, which is called the Quasi-Newton method.

EM-algorithm is less sensitive to the choice of starting values compared to the Newton-Raphson algorithm. However, the Newton-Raphson algorithm is faster when close to the maximum. NR also produces standard errors for parameter estimates as a by-product by inverting the Hessian matrix. Thus, some software begin the estimation process with the EM-algorithm and switch to Newton-Raphson when approaching the maximum likelihood estimates. This way the starting values are not too far from the final solution and the Newton-Raphson method most likely will converge.

The log-likelihood function may have multiple local maxima and both Newton-Raphson and EM-algorithm may converge to a local maximum solution instead of the global maximum. The algorithm stops when a maximum is reached, but it cannot distinguish the global maximum from a local maximum. One should repeat the estimation procedure using different starting values to ensure that the same parameter estimates are reached with each of the start values (McCutcheon, 2002).

**4.2. Identifiability.** A model is identifiable when it has a unique maximum likelihood solution, that is, no other parameter estimates yield the same log-likelihood value. Identifiability may be a problem when estimating latent class models, see, e.g. Formann (2003a).

The necessary and rather obvious condition for identifiability is that the number of different test combinations must be larger than the number of parameters to be estimated. Therefore, when local independence is assumed, the minimum number of tests is three. This condition depends on the model specification. However, some models which fulfill this criterion are yet not identified. Whether the model is

identifiable or not, may also depend on the data. If the observed information matrix is not of full rank, and therefore not invertible, one can be sure that the model is unidentified. However, also large standard errors of the parameters imply that there are problems with identifiability. Vermunt and Magidson (2005) refer to this as weak identification, meaning that although the parameters are uniquely determined, the data are not informative enough to obtain stable estimates. The formal necessary and sufficient condition for local identifiability is provided by Goodman (1974) and later formulated differently by Formann (1985); namely that if the eigenvalues of the information matrix are smaller than 0, the model is not identified.

If the software used for estimation of a latent class model does not provide the information matrix, Uebersax (2000) and McCutcheon (2002) suggest to run the estimation algorithm using different starting values. If a model is not identifiable, different starting values yield the same log-likelihood value but different parameter estimates. This method is not very practical. In fact, the information matrix can be calculated from the maximum likelihood estimates, see Section 4.4.

If the model is unidentifiable one can try to fix the sensitivity or the specificity of a test to a certain value to reduce the number of parameters. For example, it may be reasonable to fix blood culture specificity as 100%. Also fixing the parameters that lie on the boundary of the parameter space to the closest boundary may lead to an identified model.

**4.3. Missing data.** In most types of data analysis incomplete data create difficulties. Latent class analysis is no exception. Little and Rubin (1987) classified missing data as missing at random (MAR), missing completely at random (MCAR), and or neither MAR nor MCAR, which is often referred to as missing not at random (MNAR). When the data are missing at random or completely at random they are said to be ignorable.

If the probability of a test result being missing depends neither on the missing test result nor on the results of the other tests, the test result is missing completely at random. Assuming MCAR means that the population missing the result of a certain test is equal to the population not missing the test result in terms of the results of the other tests in the model (Kolb and Dayton, 1996). In other words, an equal proportion of positive test results would occur in the missing results as in the actually observed results. Assuming that the test prevalences are equal implies that the disease prevalence is equal in these two subpopulations, those missing and those not missing the result of a certain test. An example of a MCAR situation is



when the test result is missing for a patient because the sample tube was dropped in the laboratory.

The main reason for a test result being missing is, however, that the sample to be analysed was not obtained. Assume that the probability of a test result being missing is higher for patients in poor condition than for patients who are in good condition. Furthermore assume that the diagnostic tests are more likely to be positive for patients in poor condition. In a situation like this, the MCAR assumption does not hold. The unobserved proportion of positive results in the missing samples would probably be higher than the proportion positives in the obtained samples.

In a circumstances like this it is more realistic to assume that the test result is missing at random. Assuming MAR means that the probability of a test result being missing may depend on the observed data but not on the missing data. The test prevalences and therefore the disease prevalences are now allowed to be different in the two subpopulations as long as the test prevalences are the same conditional on the results of the other tests.

One approach to handle the missing data in the model estimation is to exclude the subjects with missing test results from the analysis. This, the so-called complete-case analysis, is problematic for two reasons. Firstly, dropping cases leads to decrease in statistical power and secondly, the estimates are unbiased only if the assumption of MCAR is valid, see e.g. Schafer and Graham (2002).

However, when using all available information from each subject, likelihood based inference is valid provided that the missing data mechanism is ignorable, that is, MAR or MCAR (Little and Rubin, 1987). For example, consider the three tests  $X_1$ ,  $X_2$  and  $X_3$  introduced in Section 3.1 and assume that for a group of patients the result of test  $X_1$  is missing. The number of possible combinations of test results for complete data is 8. The possible combinations of the test results for the incomplete data are  $\{\star, 0, 0\}$ ,  $\{\star, 1, 0\}$ ,  $\{\star, 0, 1\}$  and  $\{\star, 1, 1\}$  where  $\star$  denotes the missing result of the test  $X_1$ . A patient with the test result pattern  $\{\star, 0, 0\}$  contribute to the likelihood with

$$\begin{aligned}
 P(X_1 = \star, X_2 = 0, X_3 = 0) &= P(X_1 = 0, X_2 = 0, X_3 = 0) + \\
 &P(X_1 = 1, X_2 = 0, X_3 = 0) \\
 &= P(X_2 = 0, X_3 = 0) \\
 &= (1 - \psi_{X_2}) \times (1 - \psi_{X_3}) \times \mu + \\
 &\phi_{X_2} \times \phi_{X_3} \times (1 - \mu)
 \end{aligned}
 \tag{4.4}$$

As can be seen, the contribution to the likelihood does not depend on the missing data  $X_1$ , but on  $X_2$  and  $X_3$ .

As the latent class itself can be regarded missing information, there is a double missing data problem when the data set is incomplete. When estimating the model using the EM-algorithm, the M-step remains the same as for complete data as described in Section 4.1.1, but the E-step has to be modified; for more detail, see Vermunt 1997, section 3.3.2. Missing data do not affect the identifiability, since models, which are identified for complete data, are also identified when applied to incomplete data, assuming that the missing mechanism is ignorable (Hagenaars, 1990, p. 259). Thus missing data do not complicate the estimating procedure notably. The problems arise when assessing the model fit, see section 4.5

**4.4. Software.** There are several programs available for latent class analysis. We used  $\ell$ EM, which is a general program for the analysis of categorical data developed by Jeroen K. Vermunt. The program is available free of charge from the following site: <http://spitswww.uvt.nl/~vermunt/>.

The name  $\ell$ EM stands for log-linear and event history analysis with missing data using the EM algorithm. With  $\ell$ EM, it is possible to add direct effects between tests and fix any parameter, except for the prevalence, to a certain value. However, if a parameter value is fixed the program neither calculates standard errors for parameter estimates nor the eigenvalues of the information matrix, and thus does not check whether the model is identifiable. Also, the program produces estimates even for models that are clearly not identifiable (negative  $df$ ).

For these reasons, I wrote a function where the likelihood function is maximized using both EM-algorithm and a quasi-Newton algorithm using statistical software R (R Development Core Team, 2006), which is freely available at <http://www.R-project.org>. For the quasi-Newton algorithm, function *nlm* in R base package was used. The EM-algorithm used is based on function *lca* in package *e1071*. Unlike the *lca* function, this function can handle incomplete data sets. All parameter values including prevalence can be fixed at a certain value and direct effects are possible as well. Since the function uses a Newton type algorithm, the identifiability is checked by inverting the Hessian matrix. Standard errors are produced for all non-fixed parameters. The downside with the function is that it is very sensitive to starting values and parameters close to boundaries must often be fixed for convergence. As  $\ell$ EM turned out to be more robust in terms of starting values, the function was used for validation of the model and for calculating standard errors when direct effects are added in to the model.

In addition, this function, along with the function *dfHess* in package *nlme*, was used for calculating the information matrix from the fits obtained by  $\ell$ EM. The function is described in more detail in the Appendix. For a review of other software available for latent class analysis, see Uebersax (2006).

#### 4.5. Model diagnostics.

4.5.1. *Goodness-of-fit.* Model fit of a latent class model is usually evaluated by comparing the observed frequencies to the model-predicted frequencies. If the expected frequencies differ a lot from the observed ones, the model does not fit well. This can be tested using a Pearson chi-square statistic:

$$(4.5) \quad X^2 = \sum_{s=1}^{2^K} \frac{(n_s - \hat{m}_s)^2}{\hat{m}_s},$$

or the likelihood ratio statistic:

$$(4.6) \quad G^2 = 2 \sum_{s=1}^{2^K} n_s \cdot \ln \frac{n_s}{\hat{m}_s},$$

where  $n_s$  is the observed frequency in the  $s$ th test result combination, and  $\hat{m}_s$  the corresponding expected frequency. The  $X^2$  statistic and  $G^2$  statistic are asymptotically  $\chi^2$ -distributed, with  $2^K - p - 1$  degrees of freedom, where  $p$  is the number of parameters to be estimated and  $K$  is the number of tests.

There are two limitations for which neither the Pearson chi-square statistic nor the likelihood ratio statistic as such are useful when assessing the goodness-of-fit of the models in Section 5. Firstly, for sparse data the true distribution of  $G^2$  and  $X^2$  is unknown, and possibly deviates a lot from the chi-square distribution, see e.g. Collins et al. (1993). Sparse data often occur when the data set is small or when the number of possible test result combinations is large. Some solutions for dealing with the sparseness of the data have been suggested, see e.g. Formann (2003b) and Reiser and Lin (1999). These methods are not, however, discussed further here since they do not solve the second problem which occurs when the data set is incomplete.

Though it is possible to extend equations 4.1 and 4.2 to handle incomplete data, the tests as such are not of great interest as they test not only the fit of the model but also the assumption of MCAR (Hagenaars, 1988, pp. 275-258). Fortunately, if

the assumption of MAR is valid, the goodness-of-fit of the model of interest can be assessed using the conditional likelihood ratio test introduced by Fuchs (1982):

$$(4.7) \quad G_r^2 - G_s^2,$$

where  $G_r^2$  is the likelihood ratio statistic of the restricted model, that is, the model of interest, and  $G_s^2$  is the likelihood ratio statistic obtained from fitting the saturated model as explained in Section 5. Just as for the likelihood statistic for complete data, the conditional likelihood ratio test statistic is asymptotically  $\chi^2$ -distributed with  $2^K - 1 - p$  degrees of freedom provided that the data are not too sparse.

In addition to performing overall goodness-of-fit tests, it may be useful to examine the Pearson residuals for the different test result combinations to detect misfit. However, Pearson residuals are calculated assuming MCAR and therefore cannot be utilized for diagnosing fit problems if the data are missing at random. A possible solution, yet not widely discussed in the current literature, is to calculate the difference between the expected frequencies under the saturated model and the expected frequencies under the model of interest. These, so-called fit indicators, can be scaled in order to obtain scaled fit indicators which are analogous to Pearson residuals. (Collins et al. 2002, and the references therein)

4.5.2. *Detecting local dependence.* The usual methods to detect local dependence between the pairs of tests are based on comparison of expected and observed frequencies between pairs of tests and are thus not applicable when data are missing at random. For a review of these methods, see Zhang (2004). To detect local dependence one can compare two nested models. The general idea is that, if adding a direct effect between two tests improves the log-likelihood significantly the two tests are probably locally dependent. To test this, one can use the likelihood ratio test:

$$-2 \cdot \log \Lambda = -2(L_{M_1} - L_{M_2}),$$

where  $L_{M_1}$  and  $L_{M_2}$  denote the maximized log-likelihoods of models  $M_1$  and  $M_2$  and  $M_1$  is the local independence model and  $M_2$  the model with direct effects. If the data are not too sparse, the test statistic has a chi-squared distribution with degrees of freedom equal to the number of additional parameters in the more complex model. Again, sparseness of the data causes complications. However, it has been suggested that the magnitude of the likelihood ratio test is often large enough to show that the two models are substantially different (Uebersax, 2000, and the references therein).

## 5. LATENT CLASS MODELS FOR FINCAP EPI DATA

The selection of tests to be included in the model is twofold. Too many tests increase the risk of unidentifiability and local dependence. On the other hand we need to have enough tests from different locations, e.g. respiratory tract, blood and urine, in order to catch the whole spectrum of severity of the disease.

Of 128 possible result combinations of the seven tests (Sc,Sp,Nc,Np,U,B and the composite test Fi2) introduced in Section 2.2, altogether 103 have a zero observed frequency and only 4 have a frequency above five. Despite the sparseness, the latent class model fitted for the seven tests was identified. However, several parameter estimates were on the boundary of the parameter space. The specificity of blood culture and the sensitivity of sputum PCR ply were estimated to be 100 %. The specificity estimate for blood culture is reasonable and expected, but the perfect sensitivity of sputum PCR ply is a concern. The model was fitted assuming local independence between tests. As the model includes several tests from the same location, it is unlikely that this assumption holds. Because of the severe sparseness and the missing data, limited methods exist to detect which tests are locally dependent. However, according to initial exploration of the data, NPS culture seemed to overlap with several other tests. Also likelihood ratio tests performed for models with direct effects between pairs of tests implied that NPS culture was locally dependent with NPS PCR ply and sputum culture, which, on the other hand, were locally independent. As this kind of complex dependency is hard to take into account, NPS culture was dropped out from the model.

The second model for the six remaining tests, denoted by  $\mathcal{M}_6$ , was also fitted assuming local independence. The maximum likelihood estimates and standard errors are presented in Table 6. The prevalence of pneumococcal pneumonia was estimated to be 22 %, which is somewhat lower than expected. The tests with the lowest prevalence, urine, blood culture and serology, showed poor sensitivity but high specificity. Sputum PCR ply on the other hand had a high sensitivity of 98 %, but a specificity below 80 %.

In order to perform a goodness-of-fit test assuming MAR, a saturated model, denoted by  $\mathcal{S}_6$ , has to be estimated. The saturated model is obtained by estimating the  $2^K$  probabilities of the multinomial distribution so that the partially missing test result combinations are summed over the possible realisations of the missing test results as explained in Equation 4.4. As many of the observed frequencies are zero, a small constant, 0.01 was added to each cell to ensure convergence.

TABLE 6. Estimates and asymptotic standard errors for model  $\mathcal{M}_6$

		Estimate (S.E)	
<b>Prevalence</b>		0.220	(0.033)
<b>Sensitivity</b>	Sc	0.609	(0.076)
	Sp	0.980	(0.042)
	Np	0.731	(0.076)
	U	0.260	(0.063)
	B	0.123	(0.041)
	2Fi	0.450	(0.076)
<b>Specificity</b>	Sc	0.978	(0.016)
	Sp	0.786	(0.037)
	Np	0.948	(0.021)
	U	0.969	(0.013)
	B	1.000	(0.000)
	2Fi	0.944	(0.018)

According to the conditional likelihood ratio test, the model fit was adequate:  $G_r^2 - G_s^2 = 192.06 - 145.84 = 46.22$  for  $df = 63 - 13 = 50$ ,  $p = 0.63$ . However, since 43 out of 64 possible test result combinations have a zero frequency, the p-value should be interpreted with caution. In addition, from the scaled fit indicators it appears that especially test result patterns  $\{0, 1, 1, 1, 1, 1\}$  and  $\{1, 1, 1, 1, 1, 1\}$  are underestimated, see Table 8. Though the sparseness surely affects the validity of the likelihood ratio tests presented in Table 9, the magnitude of differences in the maximum likelihoods implies that blood culture, urine antigen and serology are dependent.

The fact that blood culture is locally dependent with several tests is expected, since the test is usually positive only for severe cases of CAP who are likely to have other tests positive aswell. As mentioned in Section 3.3, local dependence often

TABLE 7. Likelihood ratio, log-likelihood and degrees of freedom for LC models

	Model	$G^2$	Log-likelihood	$p$
1	$\mathcal{S}_6$	145.84	-527.13	63
2	$\mathcal{M}_{6I}$	192.06	-550.24	13
3	$\mathcal{S}_5$	104.80	-511.32	31
4	$\mathcal{M}_{5I}$	124.43	-521.13	11
5	$\mathcal{M}_{5D}$	114.36	-516.09	13

$p$ : number of parameters

TABLE 8. Scaled fit indicators greater than 1.5 in absolute value for model  $\mathcal{M}_6$

Sc	Sp	Test				2Fi	$\mathcal{S}_6$	$\mathcal{M}_6$	Scaled
		Np	U	B	Expected		Expected	fit indicator	
1	1	0	0	0	0	9.77	5.47	-1.84	
1	0	1	0	0	0	1.60	0.46	-1.66	
1	1	1	0	0	0	19.71	11.97	-2.24	
0	1	1	1	0	0	0.04	2.76	1.64	
0	1	0	1	0	1	2.66	0.90	-1.86	
0	0	0	0	1	1	0.20	0.01	-2.34	
1	0	0	0	1	1	0.18	0.01	-1.66	
0	1	1	1	1	1	3.40	0.31	-5.59	
1	1	1	1	1	1	4.55	0.48	-5.88	

TABLE 9. Likelihood ratio tests for models with direct effect between a pair of tests compared to local independence model

Joint tests	Statistic	p-value
Sc & Sp	2.98	0.22
Sc & Np	2.23	0.33
Sp & Np	0.37	0.83
Sc & U	3.79	0.15
Sp & U	0.59	0.75
Np & U	0.43	0.80
Sc & B	0.10	0.95
Sp & B	0.26	0.88
Np & B	1.29	0.52
U & B	12.64	0.00
Sc & 2Fi	3.96	0.14
Sp & 2Fi	0.87	0.65
Np & 2Fi	0.58	0.75
U & 2Fi	9.15	0.01
B & 2Fi	13.48	0.00

occurs in situations of this kind. The simple solution is, yet again, to drop one of the dependent tests instead of incorporating a complicated dependence structure in the model. As blood culture seems to be the major source of dependency, it was excluded from the final models.

The estimates for the model assuming local independence and the model with direct effect between urine antigen and serology, denoted by  $\mathcal{M}_{5I}$  and  $\mathcal{M}_{5D}$  respectively,

TABLE 10. Estimates and 95 % bootstrap confidence limits for 5 test models

		$\mathcal{M}_{5I}$			$\mathcal{M}_{5D}$		
		Estimate	LCL	UCL	Estimate	LCL	UCL
<b>Prevalence</b>		0.235	0.172	0.307	0.235	0.176	0.306
<b>Sensitivity</b>	Sc	0.590	0.443	0.734	0.590	0.446	0.735
	Sp	0.966	0.859	1.000	0.964	0.863	1.000
	Np	0.715	0.564	0.887	0.725	0.577	0.882
	U	0.245	0.135	0.380	0.234	0.124	0.355
	2Fi	0.422	0.285	0.571	0.408	0.278	0.548
<b>Specificity</b>	Sc	0.983	0.952	1.000	0.985	0.957	1.000
	Sp	0.797	0.723	0.869	0.799	0.729	0.871
	Np	0.956	0.919	0.991	0.959	0.924	0.993
	U	0.969	0.940	0.999	0.965	0.937	0.990
	2Fi	0.943	0.905	0.980	0.939	0.900	0.976

are reported in Table 10. The estimates of the two models are almost identical. The prevalence is somewhat higher and the sensitivities lower than compared to model  $\mathcal{M}_6$ , but the differences are not large. Neither model had problems with identifiability. Both models were fitted using several starting values and as same estimates were obtained each time, the solutions were assumed to be the global maximum. There was one estimate on the boundary of the parameter space for the model  $\mathcal{M}_{5D}$ : the joint probability  $P(U = 1, 2Fi = 1 | D = 0)$  was estimated to be zero. According to the model  $\mathcal{M}_{5D}$ , non-pneumococcal CAP cases have never both urine antigen and serology positive, in other words, all patients with both tests positive are true cases of pneumococcal CAP. Though estimates on the boundaries of the parameter space are a concern, this estimate does not seem implausible.

The  $G^2$  values for the two models and for the saturated model,  $\mathcal{S}_5$ , are presented in Table 7. Comparison to the saturated model is again problematic because of a high number of zero frequencies, see Table 12 in the Appendix B. However, when comparing the two nested models  $\mathcal{M}_{5I}$  and  $\mathcal{M}_{5D}$ , the chi-squared distribution of the differences between the  $G^2$  values can be expected to be satisfactory even for sparse data, see Hagenaars (1990, p. 59). According to the difference  $G^2$  test, the model which takes into account the local dependence of urine antigen and serology fits better than the local independence model:  $G_{\mathcal{M}_{5I}} - G_{\mathcal{M}_{5D}} = 10.07$  with  $2 df$  and  $p = 0.0065$ . Also the scaled fit indicators imply that  $\mathcal{M}_{5D}$  fits well as for none of the 32 test result patterns the scaled fit indicators exceeds 2 in absolute value.



The 95 % confidence limits in Table 10 are computed using nonparametric bootstrap percentile method. The bootstrap is a resampling technique introduced by Efron in 1979. It is often used to obtain estimates for confidence intervals in situations, where the assumption of asymptotic normality is questionable. 2000 data sets of 347 cases were sampled with replacement from the original data. For each bootstrap sample, the latent class model was fitted. The lower and upper confidence limits for the parameter estimates were determined as the 2.5th and 97.5th percentile of the obtained 2000 estimates. For local independence model there were no problems with identifiability. For model  $\mathcal{M}_{5D}$ , 196 of 2000 bootstrap models had either positive eigenvalues or non-invertible Hessian matrix, both of which imply that these models were not identified.

The prevalence of pneumococcal CAP in the elderly population was estimated to be 23.5 %. Many of the results were expected. Sputum culture is the most specific test in the model, but it is lacking sensitivity. Analysing the same sample using PCR increased the sensitivity remarkably but at the expense of specificity. Urine antigen and serology are also specific tests but not as sensitive as the tests from the respiratory tract. Though urine antigen and serology seem to lose out in both sensitivity and specificity to sputum culture, the tests are still valuable as the tests are not affected by antimicrobial treatment and the samples are easier to obtain and process than sputum.

The only result that was surprising, and even regarded as implausible, was the high specificity of NPS PCR ply. Diagnostic tests based on NPS samples are not generally considered reliable for reasons described in Section 2.4. Thus, the specificity of 95.9 % seemed too high. However, out of 152 CAP cases with all results available, only three are NPS PCR ply positive but negative by other tests. This implies that if the other tests are highly specific, NPS PCR ply must also be specific as it finds the same cases.

To back up this rationalization, the model estimates were used to predict the proportion of pneumococcal CAP cases for each test result pattern using Bayes' theorem:

$$P(D = 1|\mathbf{X} = \mathbf{a}_s) = \frac{P(\mathbf{X} = \mathbf{a}_s|D = 1) \times \mu}{P(\mathbf{X} = \mathbf{a}_s|D = 1) \times \mu + P(\mathbf{X} = \mathbf{a}_s|D = 0) \times (1 - \mu)}.$$

These proportions using estimates from both five test models are reported in Table 13 in the Appendix B. According to model  $\mathcal{M}_{5D}$ , only 6 % of the subjects with solely NPS PCR ply positive have pneumococcal CAP, but if any other test from the respiratory tract is positive as well, the proportion of pnc CAP cases goes up

TABLE 11. Expected frequencies and scaled fit indicators for 5 test models

Test	$\mathcal{S}_5$	$\mathcal{M}_{5I}$		$\mathcal{M}_{5D}$	
ScSpNpU2Fi	Expected frequency	Expected frequency	Scaled fit indicator	Expected frequency	Scaled fit indicator
0 0 0 0 0	180.358	181.898	0.114	181.378	0.076
1 0 0 0 0	3.084	3.368	0.155	3.097	0.007
0 1 0 0 0	51.321	50.420	-0.127	50.327	-0.140
1 1 0 0 0	9.702	6.577	-1.218	7.705	-0.719
0 0 1 0 0	8.037	8.725	0.233	8.269	0.081
1 0 1 0 0	1.581	0.662	-1.129	0.817	-0.846
0 1 1 0 0	10.972	12.189	0.348	14.768	0.988
1 1 1 0 0	19.355	14.498	-1.276	18.478	-0.204
0 0 0 1 0	6.595	5.784	-0.337	7.010	0.157
1 0 0 1 0	0.001	0.167	0.405	0.130	0.356
0 1 0 1 0	2.951	2.767	-0.111	2.132	-0.561
1 1 0 1 0	0.002	1.899	1.377	0.569	0.752
0 0 1 1 0	0.003	0.381	0.612	0.338	0.576
1 0 1 1 0	0.002	0.172	0.411	0.059	0.235
0 1 1 1 0	0.003	3.331	1.824	1.068	1.031
1 1 1 1 0	3.207	4.696	0.687	1.431	-1.485
0 0 0 0 1	13.345	11.070	-0.684	12.340	-0.286
1 0 0 0 1	0.002	0.341	0.581	0.295	0.540
0 1 0 0 1	3.269	5.728	1.028	4.987	0.770
1 1 0 0 1	2.508	4.261	0.849	2.796	0.172
0 0 1 0 1	1.821	0.767	-1.203	0.717	-1.304
1 0 1 0 1	0.001	0.386	0.619	0.281	0.528
0 1 1 0 1	4.547	7.469	1.069	5.167	0.273
1 1 1 0 1	8.471	10.560	0.643	7.250	-0.454
0 0 0 1 1	0.001	0.380	0.614	0.064	0.247
1 0 0 1 1	0.044	0.055	0.045	0.092	0.157
0 1 0 1 1	2.680	1.039	-1.610	1.693	-0.759
1 1 0 1 1	1.130	1.369	0.205	2.437	0.837
0 0 1 1 1	0.003	0.101	0.308	0.168	0.402
1 0 1 1 1	0.001	0.123	0.347	0.242	0.490
0 1 1 1 1	8.135	2.387	-3.720	4.466	-1.736
1 1 1 1 1	3.867	3.428	-0.237	6.429	1.010

notably. In fact, this applies to all the tests: if no other test support the positive result, the result is unlikely to be true positive.

## 6. DISCUSSION

Latent class analysis has become a popular method in the evaluation of diagnostic tests in the absence of gold standard. Though criticized by Alonzo and Pepe (1999) and Pepe and Janes (2007) for being a 'black-box' for clinicians, the simplicity of a standard latent class model is attractive: the observed probabilities of

the combinations of test results can be expressed in terms of the disease prevalence and the sensitivities and the specificities of the diagnostic tests using basic rules of probability.

Behind this simplicity lies, however, the rather strong assumption of local independence. Though methods that relax this often unrealistic assumption are available, it has been demonstrated that misspecified dependence structures may also lead to biased estimates (Albert and Dodd, 2004). Thus, it may be more reasonable to avoid too complex dependence structures and strive for a model in which the local independence assumption holds by forming composite tests and excluding the tests that are locally dependent with several other tests. If all dependency cannot be eliminated, the joint item method is a simple way to take into account local dependence between pairs of tests.

Local dependence is only one of the many issues in latent class analysis: unidentifiability, local maximum solution and the effect of sparse data on model diagnostics are all well recognized potential problems in LCA and need to be taken into account in the modelling process. In the FinCAP Epi study, however, many complications arose from missing data, a topic which is overlooked in the current literature concerning latent class analysis. Especially model diagnostics for incomplete data has received remarkably little attention, considering that missing data occur in almost every study. This is certainly a field in latent class analysis that needs to be explored in the future.

There has been a rather lively discussion between Garret and Zeger (2000, 2003) and Formann (2003a) concerning frequentist and Bayesian approaches to latent class analysis. In this thesis, a frequentist point of view was adopted. Equally well Bayesian framework could have been chosen, but it is unlikely that this would have changed the conclusions.

Estimating sensitivities and specificities using latent class analysis is only one part of the case definition project. Using these results and other data collected in the study along with substantive information, a group of experts will form the final case definition of pneumococcal CAP.

## REFERENCES

- Agresti, A. (2002). *Categorical Data Analysis*. Wiley, New York.
- Albert, P. S. and Dodd, L. E. (2004). A cautionary note on the robustness of latent class models for estimating diagnostic error without gold standard. *Biometrics*, 60:427–435.
- Alonzo, T. and Pepe, M. (1999). Using a combination of reference tests to assess the accuracy of a new diagnostic test. *Statistics in Medicine*, 18:2987–3003.
- Boelaert, M., Safi, S., Goetghebeur, E., Gomes-Pereira, S., Ray, D., and Stuyft, P. (1999). Latent class analysis permits unbiased estimates of the validity of data for the diagnosis of visceral leishmaniasis. *Tropical Medicine International Health*, 4(5):395–401.
- Butler, J. C., Bosshardt, S. C., Phelan, M., Moroney, S. M., Tondella, L. M., Farley, M. M., Schuchat, A., and Fields, B. S. (2003). Classical and latent class analysis evaluation of sputum polymerase chain reaction and urine antigen testing for diagnosis of pneumococcal pneumonia in adults. *The Journal of Infectious Diseases*, 187:1416–1423.
- Collins, L. M., Fidler, P. L., Wugalter, S. E., and Long, J. D. (1993). Goodness-of-fit testing for latent class models. *Multivariate Behavioral Research*, 28(3):375–389.
- Collins, L. M., Lanza, S. T., Schafer, J. L., and Flaherty, B. P. (2002). *WinLTA User's Guide, Version 3.0*. The Methodology Center, The Pennsylvania State University.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the *em* algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38.
- Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 7:1–26.
- Engels, E. A., Sinclair, M. D., Biggar, R. J., Whitby, D., Ebbesen, P., Goedert, J. L., and Gastwirth, J. L. (2000). Latent class analysis of human herpesvirus 8 assay performance and infection prevalence in sub-saharan africa and malta. *International Journal of Cancer*, 88(6):1003–1008.
- Formann, A. and Kohlmann, T. (1996). Latent class analysis in medical research. *Statistical Methods in Medical Research*, 5:179–211.
- Formann, A. K. (1985). Constrained latent class models: Theory and applications. *British Journal of Mathematical and Statistical Psychology*, 38:87–111.
- Formann, A. K. (2003a). Latent class model diagnosis from a frequentist point of view - a review and some proposals. *Biometrics*, 59:189–196.
- Formann, A. K. (2003b). Latent class model diagnostics - a review and some proposals. *Computational statistics & Data Analysis*, 41:549–559.

- Fuchs, C. (1982). Maximum likelihood estimation and model selection in contingency tables with missing data. *Journal of the American Statistical Association*, 77(378):270–278.
- Garret, E. S. and Zeger, S. L. (2000). Latent class model diagnosis. *Biometrics*, 56:1056–1067.
- Garret, E. S. and Zeger, S. L. (2003). Rejoinder to "latent class model diagnosis from a frequentist point of view. *Biometrics*, 59:197–198.
- Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61(2):215–231.
- Goodman, L. A. (2002). Latent class analysis: The empirical study of latent types, latent variables, and latent structures. In Hagenaaars, J. A. and McCutcheon, A. L., editors, *Applied Latent Class Analysis*, chapter 1, pages 27–31. Cambridge University Press, Cambridge.
- Hagenaaars, J. A. (1988). Latent structure models with direct effects between indicators. *Sociological Methods & Research*, 16(3):379–405.
- Hagenaaars, J. A. (1990). *Categorical Longitudinal Data*. Sage Publications, London.
- Harper, D. (1972). Local dependence latent structure models. *Psychometrika*, 37(1):53–59.
- Hui, S. L. and Zhou, X. H. (1998). Evaluation of diagnostic tests without gold standards. *Statistical Methods in Medical Research*, 7:354–370.
- Koivula, I. H. (2001). Epidemiology of community-acquired pneumonia. In Marrie, T. J., editor, *Community-Acquired Pneumonia*. Kluwer Academic Publishers, New York.
- Kolb, R. R. and Dayton, C. M. (1996). Correcting for nonresponse in latent class analysis. *Multivariate behavioral research*, 31:7–32.
- Korppi, M., Koivula, I., and Jokinen, C. (2003). Alahengitystieinfektioit. In Huovinen, P., Meri, S., Peltola, H., Vaara, M., Vaheri, A., and Valtonen, V., editors, *Mikrobiologia ja infektiosairaudet*. Duodecim, Helsinki.
- Langeheine, R. (1988). New developments in latent class theory. In Langeheine, R. and Rost, J., editors, *Latent Trait and Latent Class models*, chapter 4. Plenum Press, New York.
- Little, R. J. A. and Rubin, D. B. (1987). *Statistical analysis with missing data*. Wiley, New York.
- McCutcheon, A. L. (2002). Basic concepts and procedures in single- and multiple-group latent class analysis. In Hagenaaars, J. A. and McCutcheon, A. L., editors, *Applied Latent Class Analysis*, chapter 2, pages 64–66. Cambridge University Press, Cambridge.
- Pepe, M. and Janes, H. (2007). Insights into latent class analysis of diagnostic test performance. *Biostatistics*, 8:474–484.

- Qu, Y., Tan, M., and Kutner, M. (1996). Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics*, 52:948–958.
- R Development Core Team (2006). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available from: <http://www.R-project.org>.
- Reiser, M. and Lin, Y. (1999). A goodness-of-fit test for the latent class model when expected frequencies are small. *Sociological Methodology*, 29:81–111.
- Rindskopf, D. and Rindskopf, W. (1986). The value of latent class analysis in medical diagnosis. *Statistics in Medicine*, 5:21–27.
- Schafer, J. L. and Graham, J. W. (2002). Missing data: Our view of the state of art. *Psychological Methods*, 7(2):47–177.
- Seber, G. A. F. and Wild, C. J. (2003). *Nonlinear Regression*. Wiley, New York.
- Uebersax, J. A practical guide to local dependence in latent class models [online]. (2000). Available from: <http://ourworld.compuserve.com/homepages/jsuebersax/condep.htm>.
- Uebersax, J. Lca software [online]. (2006). Available from: <http://ourworld.compuserve.com/homepages/jsuebersax/soft.htm>.
- Vermunt, J. K. (1997). *Log-Linear Models for Event Histories*. Sage Publications, Thousand Oaks.
- Vermunt, J. K. and Magidson, J. (2005). Technical guide for latent gold 4.0. basic and advanced. Belmont Massachusetts: Statistical Innovations Inc.
- Zhang, N. L. (2004). Hierarchical latent class models for cluster analysis. *Journal of Machine Learning Research*, 5(697–723).

APPENDIX A. EM-ALGORITHM

**Likelihood function for complete data.** Suppose, that the true disease status of each subject is known. For each of the  $2^K$  different test result combinations, observed is the number of diseased subjects  $f_s$  and the number of non-diseased subjects  $n_s - f_s$ , where  $n_s$  is the observed frequency of the diseased and the non-diseased subjects and  $s = 1, \dots, 2^K$ .

Recall from Section 3.2

$$\begin{aligned} P(D = 1) &= \mu \\ P(\mathbf{X} = \mathbf{a}_s | D = 0) &= \prod_{k=1}^K \phi_{X_k}^{1-a_{sk}} \{1 - \phi_{X_k}\}^{a_{sk}} \\ P(\mathbf{X} = \mathbf{a}_s | D = 1) &= \prod_{k=1}^K (1 - \psi_{X_k})^{1-a_{sk}} \{\psi_{X_k}\}^{a_{sk}} \end{aligned}$$

Let  $\boldsymbol{\theta}$  denote the model parameters:  $\boldsymbol{\theta} = (\mu, \boldsymbol{\psi}, \boldsymbol{\phi})$ . Assuming local independence, the log-likelihood function for complete data can now be expressed as

$$\begin{aligned} l_c(\boldsymbol{\theta}) &= \sum_{s=1}^{2^K} n_s \times \log P(\mathbf{X} = \mathbf{a}_s) \\ &= \sum_{s=1}^{2^K} (n_s - f_s + f_s) \times \log P(\mathbf{X} = \mathbf{a}_s) \\ &= \sum_{s=1}^{2^K} (n_s - f_s) \times \log P(\mathbf{X} = \mathbf{a}_s) + \sum_{s=1}^{2^K} f_s \times \log P(\mathbf{X} = \mathbf{a}_s) \\ &= \sum_{s=1}^{2^K} (n_s - f_s) \times \log P(\mathbf{X} = \mathbf{a}_s, D = 0) + \sum_{s=1}^{2^K} f_s \times \log P(\mathbf{X} = \mathbf{a}_s, D = 1) \\ &= \sum_{s=1}^{2^K} (n_s - f_s) \times \log \{P(\mathbf{X} = \mathbf{a}_s | D = 0) \times P(D = 0)\} + \\ &\quad \sum_{s=1}^{2^K} f_s \times \log \{P(\mathbf{X} = \mathbf{a}_s | D = 1) \times P(D = 1)\} \end{aligned}$$

$$\begin{aligned}
&= \sum_{s=1}^{2^K} (n_s - f_s) \times \log \left\{ \prod_{k=1}^K (1 - \phi_{X_k})^{a_{sk}} \cdot \phi_{X_k}^{1-a_{sk}} \times (1 - \mu) \right\} + \\
&\quad \sum_{s=1}^{2^K} f_s \times \log \left\{ \prod_{k=1}^K \psi_{X_k}^{a_{sk}} \cdot (1 - \psi_{X_k})^{1-a_{sk}} \times \mu \right\} \\
&= \sum_{s=1}^{2^K} (n_s - f_s) \times \left\{ \sum_{k=1}^K \log \left\{ (1 - \phi_{X_k})^{a_{sk}} \cdot \phi_{X_k}^{1-a_{sk}} \right\} + \log(1 - \mu) \right\} + \\
&\quad \sum_{s=1}^{2^K} f_s \times \left\{ \sum_{k=1}^K \log \left\{ \psi_{X_k}^{a_{sk}} \cdot (1 - \psi_{X_k})^{1-a_{sk}} \right\} + \log \mu \right\}
\end{aligned}$$

The maximum likelihood estimates, that is the parameter values that maximize the log-likelihood function, are found by differentiating the log-likelihood function with respect to the unknown parameters  $\mu, \phi_{X_k}$  and  $\psi_{X_k}$ , where  $k = 1, \dots, K$  and setting the derivatives equal to zero.

The derivatives are

$$(A.1) \quad \frac{\partial l_c(\boldsymbol{\theta})}{\partial \mu} = -\frac{\sum_{s=1}^{2^K} (n_s - f_s)}{1 - \mu} + \frac{\sum_{s=1}^{2^K} f_s}{\mu}$$

$$(A.2) \quad \frac{\partial l_c(\boldsymbol{\theta})}{\partial \phi_{X_k}} = \sum_{s=1}^{2^K} (n_s - f_s) \times \left\{ \frac{1 - a_{sk}}{\phi_k} - \frac{a_{sk}}{1 - \phi_k} \right\}$$

$$(A.3) \quad \frac{\partial l_c(\boldsymbol{\theta})}{\partial \psi_{X_k}} = \sum_{s=1}^{2^K} f_s \times \left\{ \frac{a_{sk}}{\psi_k} - \frac{1 - a_{sk}}{1 - \psi_k} \right\}$$

Setting the equations A.1 - A.3 equal to zero yield the maximum likelihood estimates:

$$\begin{aligned}
\hat{\mu} &= \frac{\sum_{s=1}^{2^K} f_s}{\sum_{s=1}^{2^K} f_s + \sum_{s=1}^{2^K} (n_s - f_s)} \\
&= \frac{\sum_{s=1}^{2^K} f_s}{\sum_{s=1}^{2^K} n_s}
\end{aligned}$$



$$\begin{aligned}
\hat{\phi}_{X_k} &= \frac{\sum_{s=1}^{2^K} (n_s - f_s) \times (1 - a_{sk})}{\sum_{s=1}^{2^K} (n_s - f_s)} \\
&= \frac{\sum_{s=1}^{2^K} (n_s - f_s) \times (1 - a_{sk})}{(1 - \hat{\mu}) \sum_{s=1}^{2^K} n_s}.
\end{aligned}$$

$$\begin{aligned}
\hat{\psi}_{X_k} &= \frac{\sum_{s=1}^{2^K} f_s \times a_{sk}}{\sum_{s=1}^{2^K} f_s} \\
&= \frac{\sum_{s=1}^{2^K} f_s \times a_{sk}}{\hat{\mu} \times \sum_{s=1}^{2^K} n_s}
\end{aligned}$$

APPENDIX B. TABLES

TABLE 12. Observed frequencies of different test result combinations

	Sc	Sp	Np	U	2Fi	Observed frequency
1	0	0	0	0	0	79
2	1	0	0	0	0	2
3	0	1	0	0	0	25
4	1	1	0	0	0	5
5	0	0	1	0	0	3
6	1	0	1	0	0	1
7	0	1	1	0	0	6
8	1	1	1	0	0	9
9	0	0	0	1	0	1
10	1	0	0	1	0	0
11	0	1	0	1	0	1
12	1	1	0	1	0	0
13	0	0	1	1	0	0
14	1	0	1	1	0	0
15	0	1	1	1	0	0
16	1	1	1	1	0	1
17	0	0	0	0	1	5
18	1	0	0	0	1	0
19	0	1	0	0	1	1
20	1	1	0	0	1	1
21	0	0	1	0	1	1
22	1	0	1	0	1	0
23	0	1	1	0	1	2
24	1	1	1	0	1	4
25	0	0	0	1	1	0
26	1	0	0	1	1	0
27	0	1	0	1	1	2
28	1	1	0	1	1	0
29	0	0	1	1	1	0
30	1	0	1	1	1	0
31	0	1	1	1	1	1
32	1	1	1	1	1	2

TABLE 13. Estimated proportion of pneumococcal and non-pneumococcal CAP cases for each test result combination

Test	$\mathcal{M}_{5I}$		$\mathcal{M}_{5D}$	
	Pnc CAP	not Pnc CAP	Pnc CAP	not Pnc CAP
0 0 0 0 0	0.001	0.999	0.001	0.999
1 0 0 0 0	0.061	0.939	0.085	0.915
0 1 0 0 0	0.080	0.920	0.097	0.903
1 1 0 0 0	0.877	0.123	0.908	0.092
0 0 1 0 0	0.041	0.959	0.058	0.942
1 0 1 0 0	0.780	0.220	0.851	0.149
0 1 1 0 0	0.825	0.175	0.868	0.132
1 1 1 0 0	0.997	0.003	0.998	0.002
0 0 0 1 0	0.008	0.992	0.002	0.998
1 0 0 1 0	0.401	0.599	0.157	0.843
0 1 0 1 0	0.471	0.529	0.177	0.823
1 1 0 1 0	0.987	0.013	0.952	0.048
0 0 1 1 0	0.306	0.694	0.111	0.889
1 0 1 1 0	0.973	0.027	0.920	0.080
0 1 1 1 0	0.980	0.020	0.929	0.071
1 1 1 1 0	1.000	0.000	0.999	0.001
0 0 0 0 1	0.009	0.991	0.006	0.994
1 0 0 0 1	0.441	0.559	0.350	0.650
0 1 0 0 1	0.511	0.489	0.383	0.617
1 1 0 0 1	0.989	0.011	0.983	0.017
0 0 1 0 1	0.342	0.658	0.265	0.735
1 0 1 0 1	0.977	0.023	0.971	0.029
0 1 1 0 1	0.983	0.017	0.974	0.026
1 1 1 0 1	1.000	0.000	1.000	0.000
0 0 0 1 1	0.089	0.911	1.000	0.000
1 0 0 1 1	0.890	0.110	1.000	0.000
0 1 0 1 1	0.915	0.085	1.000	0.000
1 1 0 1 1	0.999	0.001	1.000	0.000
0 0 1 1 1	0.842	0.158	1.000	0.000
1 0 1 1 1	0.998	0.002	1.000	0.000
0 1 1 1 1	0.998	0.002	1.000	0.000
1 1 1 1 1	1.000	0.000	1.000	0.000

## APPENDIX C. FUNCTIONS

```

lca.function <- function(data,theta,fixed=c(rep(NA,(ncol(data)*2+1),1)),
                        const=0.0001,joint,em.niter=50,print.level=0,init.joint=F,...){
#####
## Estimates specificities, sensitivities and disease prevalence
## Arguments:
## data: data in the form of individual records;
##       0 for negative, 1 for positive and NA for missing
## theta: a vector consisting of starting values in following order:
##         specificities, sensitivities and disease prevalence
## fixed: a vector of length theta consisting of fixed parameter values,
##        NA for not fixed
## const: constant added to all cells to ensure convergence
## joint: a list consisting of (pairs of) tests to be replaced with joint test
## em.niter: number of iterations for EM-algorithm
## print: argument for function nlm: '0' means that no printing occurs,
##        a value of '1' means the initial and final details are printed and
##        a value of 2 mean that full tracing information is printed
## init.joint: logical: if false, no starting values given for joint tests
#####
require("Hmisc")
## Test result combinations
testcombsNA=function(samples){
  ntest=ncol(samples)
  al <- expand.grid(rep(list(c(0:2,NA)),ntest))
  names(al) <- colnames(samples)
  all <- apply(al,1,paste, collapse="")
  y <- table(apply(samples,1,paste,collapse=""))
  y <- y[match(all,names(y))]
  y <- ifelse(is.na(y),0,y)
  z=cbind(as.matrix(al),n=y)
  z= z[which(z[,ntest+1]>0),]
  row.names(z)=c(1:nrow(z))
  return(z)
}
## EM-algorithm for complete data with local independence assumption
## Based on function lca in package e1071
emlca <- function(theta,freq,ntest,vec1,niter,full){
  for(i in 1:niter){
    ## Specificity
    spec <- theta[1:ntest]
    ## Sensitivity
    sens <- theta[(ntest+1):(ntest+ntest)]
    ## Prevalence
    pr <- theta[(ntest+ntest+1)]
    se <- expand.grid(lapply(1:ntest,function(i)c(1-sens[i],sens[i])))
    sp <- expand.grid(lapply(1:ntest,function(i)c(spec[i],1-spec[i])))
  }
}

```

```

pas2 <- rbind(exp(vec1%*%t(log(se)))*pr , exp(vec1%*%t(log(sp)))*(1-pr))
## Probabilities of the different test result combinations
testprop <- drop(rep(1, 2) %*% pas2)
## Class membership probabilities given the observed test results
pas2 <- t(t(pas2)/testprop)
## Working frequencies
##(number of diseased and non-diseases in each test result combination)
workf <- t(t(pas2) * freq)
## Class size (number of diseased and non-diseased subjects)
classsize <- drop(workf %*% rep(1, length(freq)))
## Class probabilities (prevalence, 1-prevalence)
classprob <- classsize/sum(freq)
## Sensitivities and 1-specificities
p <- pas2 %*% as.matrix((freq * full))/classsize
theta <- c((1-p[2,]),(p[1,]), classprob[1])
}
return(theta)
}
## number of tests
ntest <- ncol(data)
## All complete and incomplete test result combinations
mat <- testcombsNA(data)[,1:ntest]
## Frequencies for all complete and incomplete test result combinations
freq <- testcombsNA(data)[,ntest+1]
## Complete cases
obs <- mat[apply(mat,1,function(i) {!any(is.na(i))}),]
## Partially missing cases
miss <- matrix(mat[apply(mat,1,function(i) {any(is.na(i))}),],ncol=ntest)
## frequencies, observed and missing
names(freq) <- apply(mat,1,paste,collapse="")
freq <- freq[match(apply(rbind(obs,miss),1,paste,collapse=""),names(freq))]
## All test result combinations
full <- expand.grid(rep(list(0:1),ntest))
b <- apply(full,1,paste,collapse="")
## Indexes for complete cases
indexobs <- (as.numeric(row.names(full))[match(apply(obs,1,paste,collapse=""),
                                                apply(full,1,paste,collapse=""))])

## Incomplete data
if(length(miss)!=0){
  ## Indexes for partially missing test combinations
  miss2 <- lapply(1:nrow(miss),function(a) {
    nas <- which(is.na(miss[a,]))
    a <- matrix(miss[a,], nrow=2^length(nas), ncol=length(miss[a,]), byrow=T)
    b <- as.matrix(expand.grid(rep(list(0:1),length(nas))))
    a[,nas] <- b
    return(a)
  })
  indexmiss <- lapply(miss2, function(i) {
    a <- apply(i,1,paste,collapse="")

```

```

as.numeric(row.names(full))[match(a,apply(full,1,paste,collapse=""))]}
## All indexes
index <- list(indexobs,indexmiss)
## Frequencies for all testcombinations, const added to all cells
freqall <- c(rep(0, 2^(ntest)),freq[(length(indexobs)+1):length(freq)])
freqall[indexobs] <- freq[1:length(indexobs)]
freqall <- freqall+ const
}
## Complete data
else{
  freqall <- c(rep(0, 2^(ntest)))
  freqall[indexobs] <- freq[1:length(indexobs)]
  freqall <- freqall+ const
  index <- indexobs
}
vec1 <- rep(1,ntest)
## All possible combinations of ntest test
full <- expand.grid(rep(list(0:1),ntest))
## Starting values using EM-algorithm
if(em.niter>0){
  emtheta <- round(emlca(theta=theta,freq=(freqall[1:2^ntest]-const),ntest=ntest,
                        vec1=vec1,niter=em.niter,full=full),4)
  theta <- emtheta
}
## Fixed values
theta[!is.na(fixed)] <- fixed[!is.na(fixed)]
## Local independence model
if (missing(joint)){
  ## Log-likelihood for complete cases
  loglikind.comp <- function(theta,index,freq,ntest,vec1,fixed){
    theta[!is.na(fixed)] <- fixed[!is.na(fixed)]
    sens <- theta[(ntest+1):(ntest+ntest)]
    pr <- theta[(ntest+ntest+1)]
    spec <- theta[1:ntest]
    se <- expand.grid(lapply(1:ntest,function(i)c(1-sens[i],sens[i])))
    sp <- expand.grid(lapply(1:ntest,function(i)c(spec[i],1-spec[i])))
    pi <- exp(vec1%*%t(log(se))*pr + exp(vec1%*%t(log(sp)))*(1-pr)
    ## Log-likelihood
    out <- -sum(freq[1:(2^ntest)]*log((pi)))
    if(any(is.na(c(out,range(se,sp)))) || out==Inf || max(unlist(se))>1 ||
        min(unlist(se))<0 || max(unlist(sp))>1 || min(unlist(sp))<0 || pr<0 || pr>1){
      .Machine$double.xmax}
    else {out}
  }
}
## Log-likelihood for incomplete data with local independence assumption
loglikind <- function(theta,index,freq,ntest,vec1,fixed){

  theta[!is.na(fixed)] <- fixed[!is.na(fixed)]
  sens <- theta[(ntest+1):(ntest+ntest)]

```

```

pr <- theta[(ntest+ntest+1)]
spec <- theta[1:ntest]
se <- expand.grid(lapply(1:ntest,function(i)c(1-sens[i],sens[i])))
sp <- expand.grid(lapply(1:ntest,function(i)c(spec[i],1-spec[i])))
pi <- exp(vec1%*%t(log(se)))*pr + exp(vec1%*%t(log(sp)))*(1-pr)
## Log-likelihood for complete cases
out1 <- sum(freq[1:(2^ntest)]*log(pi))
## Log-likelihood for partially missing cases
out2 <- sum(freq[(2^ntest+1):length(freq)]*
            sapply(index[[2]],function(i)log(sum(pi[i]))))
out <- (-out1-out2)
if(any(is.na(c(out,range(se,sp)))) || out==Inf || max(unlist(se))>1 ||
    min(unlist(se))<0 || max(unlist(sp))>1 || min(unlist(sp))<0 || pr<0 || pr>1){
  .Machine$double.xmax}
else {out}
}
## Complete data
if(length(miss)==0){
  out <- nlm(loglikind.comp,p=theta,index=index,freq=freqall,ntest=ncol(mat),vec1=vec1,
            fixed=fixed,hessian=T,iterlim=500,print.level=print.level,...)
}
## Incomplete data
else {
  out <- nlm(loglikind,p=theta,index=index,freq=freqall,ntest=ncol(mat),vec1=vec1,
            fixed=fixed,hessian=T,iterlim=500,print.level=print.level,...)
}
## Standard errors
std <- vector(length=length(theta))
std[!is.na(fixed)] <- 0 # if fixed, then 0
std[is.na(fixed)] <- sqrt(diag(solve(out$hessian[is.na(fixed),is.na(fixed)])))
out$std2 <- std
names(out$std) <- c(paste("SPEC",colnames(mat),sep=""),
                  paste("SENS",colnames(mat),sep=""),"PREV")
names(out$estimate) <- c(paste("SPEC",colnames(mat),sep=""),
                       paste("SENS",colnames(mat),sep=""),"PREV")
out$est <- out$estimate
out$std <- c(std[(2*ntest+1)],std[(ntest+1):(2*ntest)],std[1:ntest])
out$est <- c(out$est[(2*ntest+1)],out$est[(ntest+1):(2*ntest)],out$est[1:ntest])
out$call <- match.call()
}
## Local dependence, direct effects between tests
else{
fixed <- fixed[1:(2*ntest+1)]
## Loglikelihood function for complete data
loglik.comp <- function(theta,index,freq,ntest,vec1,fixed,joint,jointp){
  theta[!is.na(fixed)] <- fixed[!is.na(fixed)]
  spec <- theta[1:ntest]
  sens <- theta[(ntest+1):(ntest+ntest)]
  pr <- theta[(ntest+ntest+1)]

```

```

joint0 <- matrix(theta[(2*ntest+2):(2*ntest+2+3*length(joint)-1)],
                nrow=length(joint),byrow=T)
joint1 <- matrix(theta[(2*ntest+2+3*length(joint)):length(theta)],
                nrow=length(joint),byrow=T)
joinm0 <- cbind(joint0,apply(joint0,1,function(i) 1-sum(i)))
joinm1 <- cbind(apply(joint1,1,function(i) 1-sum(i)),joint1)
s0 <- lapply(1:length(joint),function(i) cbind(rep(1,2^ntest),
                joinm0[i,][jointp[[i]]]))
s1 <- lapply(1:length(joint),function(i) cbind(rep(1,2^ntest),
                joinm1[i,][jointp[[i]]]))
se <- expand.grid(lapply(1:ntest,function(i)c(1-sens[i],sens[i])))
sp <- expand.grid(lapply(1:ntest,function(i)c(spec[i],1-spec[i])))
se[,unlist(joint)] <- do.call("cbind",s1)
sp[,unlist(joint)] <- do.call("cbind",s0)
pi <- exp(vec1%*%t(log(se))*pr + exp(vec1%*%t(log(sp)))*(1-pr)
## Log-likelihood for complete cases
out <- -sum(freq[1:(2^ntest)]*log((pi)))
if(any(is.na(c(out,range(se,sp)))) || out==Inf || max(unlist(se))>1 ||
    min(unlist(se))<0 || max(unlist(sp))>1 || min(unlist(sp))<0 || pr<0 || pr>1){
  .Machine$double.xmax}
else {out}
}
## Loglikelihood function for incomplete data
loglik <- function(theta,index,freq,ntest,vec1,fixed,joint,jointp){
theta[!is.na(fixed)] <- fixed[!is.na(fixed)]
spec <- theta[1:ntest]
sens <- theta[(ntest+1):(ntest+ntest)]
pr <- theta[(ntest+ntest+1)]
joint0 <- matrix(theta[(2*ntest+2):(2*ntest+2+3*length(joint)-1)],
                nrow=length(joint),byrow=T)
joint1 <- matrix(theta[(2*ntest+2+3*length(joint)):length(theta)],
                nrow=length(joint),byrow=T)
joinm0 <- cbind(joint0,apply(joint0,1,function(i) 1-sum(i)))
joinm1 <- cbind(apply(joint1,1,function(i) 1-sum(i)),joint1)
s0 <- lapply(1:length(joint),function(i) cbind(rep(1,2^ntest),
                joinm0[i,][jointp[[i]]]))
s1 <- lapply(1:length(joint),function(i) cbind(rep(1,2^ntest),
                joinm1[i,][jointp[[i]]]))
se <- expand.grid(lapply(1:ntest,function(i)c(1-sens[i],sens[i])))
sp <- expand.grid(lapply(1:ntest,function(i)c(spec[i],1-spec[i])))
se[,unlist(joint)] <- do.call("cbind",s1)
sp[,unlist(joint)] <- do.call("cbind",s0)
pi <- exp(vec1%*%t(log(se))*pr + exp(vec1%*%t(log(sp)))*(1-pr)
## Log-likelihood for complete cases

out1 <- sum(freq[1:(2^ntest)]*log((pi)))
## Log-likelihood for partially missing cases
out2 <- sum(freq[(2^ntest+1):length(freq)]*sapply(index[[2]],function(i)
                log(sum(pi[i]))))

```



```

out <- (-out1-out2)

if(any(is.na(c(out,range(se,sp)))) || out==Inf || max(unlist(se))>1
  || min(unlist(se))<0
  || max(unlist(sp))>1 || min(unlist(sp))<0 || pr<0 || pr>1){
  .Machine$double.xmax}
  else {out}
}
## Joint item
spec <- theta[1:ncol(mat)]
sens <- theta[ncol(mat)+1):(2*ncol(mat))]
sensj <- sapply(joint,function(i) sens[i])
a <- lapply(1:length(joint),function(j)
  expand.grid(lapply(1:2,function(i)
    c(1-sensj[i,j],sensj[i,j]))))
## Starting values, assuming local independence
sej <- lapply(1:length(joint),function(i)
  {out <- apply(a[[i]],1,prod)
  names(out) <- c("00","10","01","11")
  return(out[2:4])})
specj <- sapply(joint,function(i) spec[i])
b <- lapply(1:length(joint),function(j)
  expand.grid(lapply(1:2,function(i)
    c(specj[i,j],(1-specj[i,j]))))
spj <- lapply(1:length(joint),function(i)
  { out <- apply(b[[i]],1,prod)
  names(out) <- c("00","10","01","11")
  return(out[1:3])})
jointp <- lapply(1:length(joint),function(i)
  match(apply(full[,joint[[i]]],1,paste,collapse=""),
    c("00","10","01","11")))

if(!init.joint)
  theta <- c(theta,unlist(spj),unlist(sej))

## Fixed values
fixed <- c(fixed,rep(NA,2*3*length(joint)))
vec1 <- rep(1,ncol(mat))
theta[!is.na(fixed)] <- fixed[!is.na(fixed)]
## Compete data
if(length(miss)==0){
  out <- nlm(loglik.comp,p=theta,index=index,freq=freqall,ntest=ncol(mat),vec1=vec1,
    fixed=fixed,joint=joint,jointp=jointp,hessian=T,
    iterlim=500,print.level=print.level,...)
}
## Incomplete data
else{
out <- nlm(loglik,p=theta,index=index,freq=freqall,ntest=ncol(mat),vec1=vec1,

```

```

        fixed=fixed, joint=joint, jointp=jointp, hessian=T,
        iterlim=500, print.level=print.level, ...)
}
vecorg <- diag(rep(1, (2*ntest+1)))
vecorg[, c(unlist(joint), (ntest+ unlist(joint)))] <- rep(0, (2*ntest+1))
vecj <- matrix(rep(0, (2*ntest+1)*(6*length(joint))), ncol=(2*ntest+1))
b <- matrix(rep(0, 24*length(joint)^2), nrow=6*length(joint))
apu=0
for(i in 1:length(joint)){
  b[(apu+i):(apu+i+2), (apu/2+i):(apu/2+i+1)] <- cbind(c(1,0,1), c(1,1,0))
  apu=apu+2
}
## Sens
apu2=3*length(joint)
for(i in 1:length(joint)){
  b[(apu2+i):(apu2+i+2), (apu2/2+i+1):(apu2/2+i+2)] <- cbind(c(1,0,1), c(0,1,1))
  apu2=apu2+2
}
vecj[, c(unlist(joint), (ntest+unlist(joint)))] <- b
est <- out$estimate%*%rbind(vecorg, vecj)
out$est <- as.vector(est)
names(out$est) <- c(paste("SPEC", colnames(mat), sep=""),
  paste("SENS", colnames(mat), sep=""), "Prevalence")

## Standard errors
jp <- c(unlist(joint), (ncol(mat)+unlist(joint)))
fixed2 <- fixed
fixed2[jp] <- 1

s <- matrix(rep(0, 36*length(joint)^2), nrow=6*length(joint))
apu=0
for(i in 1:length(joint)){
  s[(apu+i):(apu+i+2), (apu+i):(apu+i+2)] <- cbind(c(1,0,1), c(1,1,0), c(0,0,0))
  apu=apu+2
}
## Sens
apu2=3*length(joint)
for(i in 1:length(joint)){
  s[(apu2+i):(apu2+i+2), (apu2+i):(apu2+i+2)] <- cbind(c(1,0,1), c(0,1,1), c(0,0,0))
  apu2=apu2+2
}

stdvec <- diag(rep(1, (2*ntest+1-length(fixed[!is.na(fixed)])+2*length(joint))))
stdvec[(2*ntest+1-length(fixed[!is.na(fixed)])-4*length(joint)+1):nrow(stdvec),
  (2*ntest+1-length(fixed[!is.na(fixed)])-4*length(joint)+1):nrow(stdvec)] <- s

covar <- solve(out$hessian[is.na(fixed2), is.na(fixed2)])
stdall <- sqrt(diag(covar))
std <- sqrt(diag(t(stdvec)%*%covar%*%stdvec))

```

```

std <- std[std>0]
std2 <- rep(0,2*ntest+1+6*length(joint))
std2[is.na(fixed2)] <- stdall
names(std2) <- c(paste("SPEC",colnames(mat),sep=""),
                paste("SENS",colnames(mat),sep=""),
                "Prevalence",rep(paste(names(spj[[1]]),"|0",sep=""),length(joint)),
                rep(paste(names(sej[[1]]),"|1",sep=""),length(joint)))

std2[c(unlist(joint),unlist(joint)+ntest)] <- std[(2*ntest+1-length(fixed[!is.na(fixed)])
                                                -4*length(joint)+1):length(std)]

out$std <- c(std2[(2*ntest+1)],std2[(ntest+1):(2*ntest)],std2[1:ntest])
out$est <- c(out$est[(2*ntest+1)],out$est[(ntest+1):(2*ntest)],out$est[1:ntest])
out$std2 <- std2
out$call <- match.call()
}
class(out) <- "LCA"
return(out)
print.LCA <- function (x, ...)
{
  cat("Call:\n")
  print(x$call)
  cat("\nLog-likelihood\n")
  cat(-x$minimum,digits=6)
  cat("\nEstimates and standard errors\n")
  ests <- cbind(names(x$est),round(x$est,3),round(x$std,3))
  colnames(ests) <- c("Parameter","Estimate","Standard Error")
  print.char.matrix(ests,col.names=T,row.names=F)
  if (x$code ==2 & sum(abs(x$gradient))>1){
    cat(paste("WARNING: Convergence code",
              x$code, ";See ?nlm for details\n In addition sum of gradient > 1 \n"))
  }
  else if (x$code > 1)
    cat(paste("WARNING: Convergence code", x$code,
              "; See ?nlm for details\n"))
}
}

```