

Practice of Epidemiology

Testing Pneumonia Vaccines in the Elderly: Determining a Case Definition for Pneumococcal Pneumonia in the Absence of a Gold Standard

Jukka Jokinen*, Marja Snellman, Arto A. Palmu, Annika Saukkoriipi, Vincent Verlant, Thierry Pascal, Jeanne-Marie Devaster, William P. Hausdorff, and Terhi M. Kilpi

* Correspondence to Dr. Jukka Jokinen, Department of Public Health Solutions, National Institute for Health and Welfare, Mannerheimintie 166, 00300 Helsinki, Finland (e-mail: jukka.jokinen@thl.fi).

Initially submitted June 2, 2017; accepted for publication December 8, 2017.

Clinical assessments of vaccines to prevent pneumococcal community-acquired pneumonia (CAP) require sensitive and specific case definitions, but there is no gold standard diagnostic test. To develop a new case definition suitable for vaccine efficacy studies, we applied latent class analysis (LCA) to the results from 7 diagnostic tests for pneumococcal etiology on clinical specimens from 323 elderly persons with radiologically confirmed pneumonia enrolled in the Finnish Community-Acquired Pneumonia Epidemiology study during 2005–2007. Compared with the conventional use of LCA, which is mainly to determine sensitivities and specificities of different tests, we instead used LCA as an appropriate instrument to predict the probability of pneumococcal etiology for each CAP case based on individual test profiles, and we used the predictions to minimize the sample size that would be needed for a vaccine efficacy trial. When compared with the conventional laboratory criteria of encapsulated pneumococci in culture, in blood culture or high-quality sputum culture, or urine antigen positivity, our optimized case definition for pneumococcal CAP resulted in a trial sample size that was almost 20,000 subjects smaller. We believe that the novel application of LCA detailed here to determine a case definition for pneumococcal CAP could also be similarly applied to other diseases without a gold standard.

case definition; diagnostic tests; latent class analysis; pneumonia; *Streptococcus pneumoniae*; trial sample size

Abbreviations: CAP, community-acquired pneumonia; CbpA, choline-binding protein A; FinCAP Epi, Finnish Community-Acquired Pneumonia Epidemiology; HQSc, high-quality sputum culture; LCA, latent class analysis; *lytA*, autolysin gene; NPS, nasopharyngeal swab; PCR, polymerase chain reaction; *ply*, pneumolysin gene; PsaA, pneumococcal surface adhesin A; VE, vaccine efficacy.

Community-acquired pneumonia (CAP) is a common illness and a significant cause of death, especially in frail populations such as the elderly. One of the most commonly detected pathogens in CAP patients is *Streptococcus pneumoniae* (1–4). The exact proportion of pneumococcal CAP cases in adults and the elderly is, however, unknown. Estimates vary widely, between 12% and 46%, partly depending on the population and case selection, but mainly on how the diagnosis of pneumococcal CAP is made (4, 5). The diagnosis of pneumococcal CAP is difficult and complex for 2 reasons.

First, there is no etiological gold standard (i.e., a diagnostic test that would identify all pneumococcal cases and yet be definitive). Traditionally, positive cultures of blood and high-quality sputum have been considered specific indicators of pneumococcal etiology in radiologically confirmed pneumonia. However,

only a small proportion of pneumonia cases are blood-culture positive, and a high-quality sputum sample is difficult to obtain from all pneumonia patients. The addition of serological, urine antigen, and various polymerase chain reaction (PCR) tests have been suggested to increase the sensitivity, but the specificity of these tests may be suboptimal.

Second, radiological confirmation of CAP itself is also a very difficult and resource-intensive task: Even before starting the review process, several x-ray images are typically required during the disease development for proper case confirmation, all of which need to be collected and typically transformed in digitized format in order to make them easily accessible for the reviewers.

We conducted the Finnish Community-Acquired Pneumonia Epidemiology (FinCAP Epi) study initially in order to construct a case definition for pneumococcal CAP. Our case definition

and the corresponding pneumococcal CAP incidence have been published previously (1). Here we present the methodological approach for constructing the case definition, which was optimized to facilitate conducting a potential vaccine efficacy (VE) trial. Because there is no gold standard for pneumococcal CAP, we applied latent class analysis (LCA), a statistical method for estimating the true disease status based on imperfect tests. We also took advantage of the same methodological framework to investigate whether the criteria aimed at identifying the pathogen could equally be applied to clinically suspected cases, potentially obviating the need for strict radiological confirmation.

METHODS

Pneumonia patients

The FinCAP Epi study is described more comprehensively elsewhere (1). In brief, 490 cases of elderly patients, aged ≥ 65 years and with symptoms suggestive of pneumonia, were enrolled during visits to the 2 hospitals in Tampere, Finland, between May 2005 and May 2007. Radiological review was further conducted for these 490 clinically suspected CAP cases, of which 323 were considered x-ray-confirmed CAP, based on the criteria that at least 2 out of 3 reviewers classified the case as pneumonia. Patients were sampled at the acute visit for blood (for aerobic and anaerobic blood cultures), urine, nasopharyngeal swab (NPS), and sputum, as well as a venous blood sample, with a paired convalescent sample 4–8 weeks later.

Microbiological, serological, and other laboratory methods

The following assays were performed for the detection of pneumococcus: culture of blood, sputum, and NPS; quantitative real-time PCR on sputum and NPS, including pneumolysin (*ply*) and autolysin (*lytA*) target genes; antigen detection in urine by *Streptococcus pneumoniae* BinaxNOW (Alere Scarborough, Inc., Scarborough, Maine) and in sputum by latex agglutination and Quellung reaction; and serology for pneumococcal surface adhesin A (PsaA) and choline-binding protein A (CbpA) antigens with paired sera. In addition, clinical laboratory assays, such as a C-reactive protein test, were performed at the acute visit. For methodological details, see Palmu et al. (1). The recently conducted Community-Acquired Pneumonia Immunization Trial in Adults (CAPITA) investigated VE against pneumococcal CAP in the elderly (6). The case definition of vaccine-type pneumococcal CAP in CAPITA used serotype-specific urine antigen detection assay (7). This urine test covers only 13 of more than 90 pneumococcal serotypes, and it is not commercially available; it was therefore not investigated in our study.

Latent class analysis

LCA is a statistical method for identifying unobserved disease status using the information from a set of observed tests that imperfectly measure the disease (8). Denote X as the observable test variable, with possible realizations $x = 0$ or 1 , and D as the unobservable true disease status. The probability (pr) of

the test being positive or negative can be expressed in terms of unobserved (latent) disease status by using the rule of total probability:

$$\text{pr}(X = x) = \text{pr}(D = 1) \times \text{pr}(X = x | D = 1) + \text{pr}(D = 0) \times \text{pr}(X = x | D = 0), \quad (1)$$

where $\text{pr}(D = 1)$ denotes the disease probability, in our case the probability of pneumococcus as the causative pathogen among pneumonia patients, whereas $\text{pr}(X = 1 | D = 1)$ and $\text{pr}(X = 0 | D = 0)$ denote, respectively, the sensitivity and the specificity of test X . More generally, these unobserved quantities (disease prevalence, specificities, sensitivities) can be estimated from data consisting of k tests ($k > 2$) by maximizing the likelihood of data for a given latent class model (8).

There are several studies in which LCA has been used to estimate the test performances (5, 8–13), including pneumococcal CAP (5, 12, 13). Specific approaches to LCA model building are required to achieve each goal, with a focus on the concurrent parameter(s) of interest (5). In the FinCAP Epi study, our focus was to elaborate a case definition that would be suitable for VE trial settings. Instead of the conventional use of LCA, which is to estimate test specificities and sensitivities, we treated LCA as a mechanistically appropriate model (see equation (1)) to derive the best possible predictions for future cases. The quantity of interest in such situations is the predictive probability of an individual case being caused by pneumococcus, conditional on the observed test results. In the simplest example, this corresponds to positive predictive value: $\text{pr}(D = 1 | X = 1)$. A straightforward generalization of the positive predictive value is the predictive probability of a case being caused by pneumococcus, conditional on the whole test result profile: $\text{pr}(D = 1 | X_1 = x_1, \dots, X_k = x_k)$, where k denotes the number of tests included in LCA model.

Missing data

A ubiquitous problem in analyses of diagnostic tests is that some test results are missing. Missing data are also a special concern when determining a case definition, because they inevitably lead to a less-sensitive case definition compared with a situation in which all test results are available. Consider for example high-quality sputum culture (HQSc) positive for *S. pneumoniae* as a case definition for pneumococcal etiology of CAP: Assume that HQSc specificity is 99% and sensitivity 80%. Further, assume that half of the cases have no high-quality sputum. The specificity of the case definition would be the same as for HQSc test (99%), but the case definition's sensitivity is half the HQSc test sensitivity (40%). In other words, missing values are automatically interpreted as negative results. Therefore, it is crucial to incorporate not only the test performances but also the information that the result is missing into the process. Fortunately, a coherent methodological framework exists for handling missing data when estimating predictive probabilities based on a likelihood-based LCA model: Consider, for example, conditional probability for a profile of 3 tests: $\text{pr}(X_1 = x_1, X_2 = x_2, X_3 = x_3 | D = 1)$. Further, assume that the result of X_2 is missing (i.e., $\text{pr}(X_1 = x_1, X_2 = *, X_3 = x_3 | D = 1)$), where $*$ denotes a missing value. The likelihood can be obtained as a sum of 2 possible realizations of X_2 :

$$\begin{aligned} & \text{pr}(X_1 = x_1, X_2 = 0, X_3 = x_3 | D = 1) + \\ & \text{pr}(X_1 = x_1, X_2 = 1, X_3 = x_3 | D = 1) \quad (2) \\ & = \text{pr}(X_1 = x_1, X_3 = x_3 | D = 1). \end{aligned}$$

By maximizing the likelihood for both complete and partially observed test profiles in equation (2), one is borrowing strength from the complete profiles to the partially observed ones. The practical consequence of this is that if X_2 is a highly specific and sensitive test, but results are missing for a large proportion of cases, one can, for example, investigate the usefulness of a case definition “both X_1 and X_3 positive” in those with missing X_2 not by assuming X_2 to be negative but rather based on how frequently X_1 and X_3 provided support for X_2 positivity/negativity in completely observed profiles.

LCA model building and diagnostics

Full utilization of all available tests in the modeling stage is challenging, because, for example, 10 tests generates $2^{10} = 1,024$ possible outcomes for an individual test profile, the majority of which would not be observed with any feasible sample size. A suitable balance is therefore needed regarding how many tests should be included in the model. One way to reduce the number of tests is to form composite variables from 2 or more tests (e.g., X_1 or X_2 positive). However, caution is needed in forming these types of variables: If their performances or, alternatively, the numbers of missing results differ between the tests, this is likely to introduce complicated dependencies in LCA modeling.

Missing test results were assumed missing at random, under which strength is borrowed from completely observed profiles to the partially observed ones (see equation (2)). Under the missing-at-random assumption, the probability that a missing test is positive can differ for those with partly observed test profiles as long as this probability is equal to the probability for those with complete data, conditional on the other observed test results (5, 14). One practical consequence of this is that if, for example, some samples are difficult to obtain from severely ill patients, and those test results that are available from these patients are more frequently pneumococcal, this type of association between missing and observed test results is inherently taken into account in the model.

All tests analyzed in this study are obviously associated with each other, because they are designed to measure the same pathogen. However, in the standard LCA model, it is assumed that dependence between the tests is fully explained by the true disease status, so that within those (unobserved) subgroups of pneumococcal CAP and nonpneumococcal CAP patients ($D = 1$ or 0) the tests are independent of each other (8). In order to investigate the appropriateness of the model and the variables to be included, this standard conditional independence assumption in LCA (8) was tested by imposing local dependencies (15, 16) and by comparing the fit with the one from a standard model using the likelihood ratio test. Fit of the model to the data was assessed using a conditional likelihood ratio test (17) and scaled fit indicators (18) in order to account for missing data. For full methodological details of the model fitting, diagnostics, and handling of missing data, as well as the software used, see Snellman (19).

Optimizing the case definition

Predictive probabilities for each of the 323 CAP cases were derived from the final LCA model fit parameters using the Bayes formula. The resulting distribution of predictive probabilities was assumed to represent the distribution of pneumococcal CAP probabilities in a future vaccine trial. By using a cutoff for the probability of CAP being caused by pneumococcus, the proportion of true and false positives and of true and false negatives could be derived for each cutoff point. The optimal cutoff was defined as the one producing the smallest sample size for a future VE trial. The central assumptions used for sample size calculation were: 1) radiologically confirmed CAP incidence 5.5/1,000 person-years (1) and 3-year follow-up, 2) VE against pneumococcal CAP of 40%, and 3) α of 5% and power of 80%. Apart from VE, altering these assumptions had no impact on the optimal cutoff. When changing the VE assumption, the optimal cutoff decreased with increasing VE. The details of the sample size calculations are presented in Web Appendix 1 (available at <https://academic.oup.com/aje>).

Forming a case definition based purely on predictive probabilities may be impractical. Therefore, our final case definition was based on available test combinations. The match with the predicted probabilities guided the selection of the most suitable test combination, supplemented by other empirical evidence of the test performances (1). The resulting sample size was calculated for this case definition similarly as for the cutoff points, as well as for the case definition “blood culture positive or HQSc positive or urine antigen positive,” generally regarded as a specific case definition.

Sensitivity analysis of the relative gain of x-ray confirmation

The primary focus in our study was to create a case definition for pneumococcal pneumonia. We approached the case definition sequentially: first through radiological confirmation of suspected CAP cases ($n = 490$) and subsequently through LCA of tests among the x-ray-confirmed CAP cases ($n = 323$). In order to investigate the possibility that the case definition for pneumococcal CAP could be adequately defined even in the absence of radiological confirmation, we used the latent-class modeling framework described above by determining the predictive probability of pneumococcal disease among 490 suspected CAP cases, and further categorized the predictive probabilities according to radiological confirmation status.

RESULTS

Sample availability and positivity

Table 1 reports the number of performed assays and the percentage of positive results for *S. pneumoniae* out of performed assays. More than 50% of cases were missing at least 1 test result ($n = 174$). Blood for serum was obtained from nearly every patient at the acute visit, but due to missed follow-up visits, paired serology results were not available for one-fifth of the CAP episodes. Sputum sample was missing from nearly one-third of the CAP episodes, which is similar to proportions reported in other studies (20, 21).

Table 1. Availability and Pneumococcal Positivity of Assays for 323 Confirmed Community-Acquired Pneumonia Cases in the Finnish Community-Acquired Pneumonia Epidemiology Study, 2005–2007

Assay	No. of Samples Analyzed (<i>n</i> = 323 Cases)	No. of Samples Positive for <i>S. pneumoniae</i>	% Positive
Encapsulated pneumococci from blood culture	319	9	3
Urine BinaxNOW antigen test	281	27	10
Encapsulated pneumococci from sputum culture	226	40	18
<i>ply</i> PCR from sputum	223	103	46
<i>lytA</i> PCR from sputum	224	57	25
Encapsulated pneumococci from NPS culture	306	32	10
<i>ply</i> PCR from NPS	306	67	22
<i>lytA</i> PCR from NPS	306	44	14
Two-fold increase in PsaA antibody	248	29	12
Two-fold increase in CbpA antibody	263	34	13
Two-fold increase in either CbpA and/or PsaA	264	43	16
CRP over 150 mg/L	323	102	32

Abbreviations: CbpA, choline-binding protein A; CRP, C-reactive protein; *lytA*, autolysin gene; NPS, nasopharyngeal swab; PCR, polymerase chain reaction; *ply*, pneumolysin gene; PsaA, pneumococcal surface adhesin A.

Blood culture, generally regarded as 100% specific (12), was positive for only 9 cases. Conversely, pneumococcal culture from high-quality sputum, which is also considered highly specific, was positive in 21% of the samples obtained (1). However, because high-quality sputum could be obtained only from approximately half of the cases ($n = 170$), the percentage of positive results from high-quality sputum out of all 323 CAP cases was only 11%.

Positive cultures of *S. pneumoniae* from high-quality and low-quality sputa showed similar concordance with other diagnostic tests if the other tests were positive. If the other tests were negative for *S. pneumoniae*, encapsulated pneumococci were isolated from low-quality sputum approximately half as often as from high-quality sputum (22). This suggests that culture of low-quality sputum for demonstration of pneumococcal etiology of CAP may have low sensitivity rather than low specificity.

Latent class analysis

Because several tests (e.g., sputum culture and *lytA* and *ply* PCR) were performed on the same sample, it was important to investigate the assumption of conditional independence (6). Indeed, complicated dependencies were found between the tests, mainly in those using samples taken from the respiratory tract. In addition, blood culture was found to be locally dependent with several tests. However, when blood culture was left out of the model, 7 of 9 cases with positive blood culture had predictive probabilities of >0.99 . Hence, blood culture did not add relevant information to the predictions, but it introduced complicated conditional dependencies and was therefore left out of the final model. For the other tests to be included in the model, selection was guided by including a wide spectrum of test performances, as well as by covering all 323 cases with a reasonable amount of observed test results. Sputum culture was included regardless of

quality; if only HQSc results had been used in the model, information on a large portion of the patients with sputum sample would have been omitted.

Two-fold increases in PsaA and CbpA antibodies showed similar performances, being 12% and 13% positive, respectively. Furthermore, the available results were practically from the same cases. Therefore, a composite variable “at least 2-fold increase in serum antibodies against PsaA and/or CbpA” was formed.

The following variables were included in the final model: culture of encapsulated pneumococci from any quality sputum; *ply* PCR from sputum and NPS; *lytA* PCR from sputum; BinaxNOW pneumococcal urine antigen; 2-fold increase in PsaA and/or CbpA antibodies; and C-reactive protein above 150 mg/L.

Significant dependencies were found in sensitivities between culture of encapsulated pneumococci from any quality sputum and *lytA* PCR from sputum and in specificities between *ply* PCR from sputum and *ply* PCR from NPS. Both were plausible findings, considering the loci from where the specimens were sampled and based on the assays performed. In order to account for the dependencies, these parameters were estimated jointly (15). The goodness of fit indicated that the final model fitted the data well ($\chi^2 = 91.0$ with 106 degrees of freedom; $P = 0.85$).

The estimates of the LCA model are reported in Table 2. True pneumococcal CAP prevalence was estimated at 24%. Each of the sputum tests performed quite differently, with culture being highly specific and, conversely, *ply* PCR highly sensitive. Urine antigen sensitivity was very low, which was anticipated, considering the relatively low yield of positives (10%, Table 1). Two-fold increases in PsaA and/or CbpA antibodies, NPS *ply* PCR, and especially C-reactive protein >150 mg/L were neither highly sensitive nor specific, but they provided valuable information for prediction purposes, because samples were available for 82%, 95%, and 100% of the subjects, respectively.

Table 2. Latent Class Analysis Model Estimates for the 7 Selected Tests, Using Data From the Finnish Community-Acquired Pneumonia Epidemiology Study, 2005–2007

Test	Pneumococcal CAP Prevalence 0.24 (SE, 0.07)	
	Specificity (SE)	Sensitivity (SE)
Sputum culture (encapsulated pneumococci)	0.99 (0.009)	0.64 (0.067)
Sputum <i>lytA</i> PCR	0.98 (0.015)	0.90 (0.055)
Sputum <i>ply</i> PCR	0.72 (0.036)	0.98 (0.022)
NPS <i>ply</i> PCR	0.92 (0.020)	0.66 (0.066)
Urine antigen	0.97 (0.013)	0.32 (0.017)
Two-fold increase in PsaA and/or CbpA	0.94 (0.019)	0.46 (0.024)
CRP over 150 mg/L	0.76 (0.028)	0.55 (0.064)

Abbreviations: CAP, community-acquired pneumonia; CbpA, choline-binding protein A; CRP, C-reactive protein; *lytA*, autolysin gene; NPS, nasopharyngeal swab; PCR, polymerase chain reaction; *ply*, pneumolysin gene; PsaA, pneumococcal surface adhesin A; SE, standard error.

Effect of antimicrobials

Antimicrobial exposure was reported in 39% of patients before sputum and NPS collection (1). However, the prevalence of culture-based and *lytA* PCR tests was notably lower only in those 45 cases (14%) who had antimicrobial treatment started within

2 weeks prior to the hospital visit and still ongoing (23). A robustness analysis was conducted to investigate the effect of antimicrobial treatment by coding the test results that are presumably affected by antimicrobials as missing (5). No material difference in the estimates with or without excluding results affected by antimicrobials was found.

Case definition based on predictive probabilities

Predictive probabilities of the 323 confirmed CAP cases, estimated based on the LCA model in Table 2, are depicted in Figure 1. The majority of CAP cases could clearly be classified as pneumococcal CAP or nonpneumococcal CAP according to the pneumococcal CAP probabilities: More than 15% of the CAP cases had pneumococcal CAP probability above 99% and almost 70% below 10% (Web Table 1).

The optimal cutoff for the predictive probability, minimizing the sample size for purposes of a vaccine trial, was estimated to be between 44%–52% (no cases within that interval). When assuming 90% VE, the optimal cutoff was 35%–38%, and with 99% VE, it was 24%–30%. Using these predictions, laboratory criteria were derived for pneumococcal CAP:

- Encapsulated pneumococci cultured from blood

OR

- Encapsulated pneumococci cultured from high-quality sputum

OR

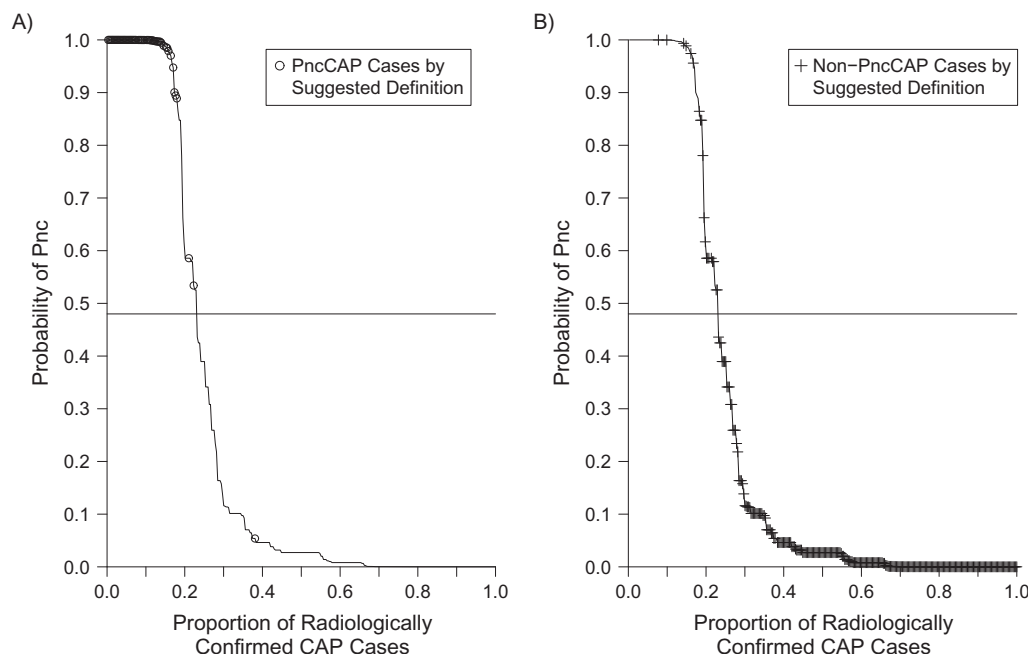


Figure 1. Predictive probabilities of pneumococcal (Pnc) community-acquired pneumonia (CAP) in 323 confirmed CAP cases based on their observed test profiles, Finnish Community-Acquired Pneumonia Epidemiology study, 2005–2007. Pneumococcal CAP cases, denoted by a \circ (A), and nonpneumococcal CAP cases, denoted by a $+$ (B), according to suggested case definition are superimposed on the curves. The horizontal line corresponds to the optimal cutoff minimizing the sample size of a vaccine efficacy trial.

- At least 2 of the following:
 - Urine pneumococcal antigen positive
 - At least 2-fold increase in anti-PsaA and/or anti-CbpA
 - Detection of pneumococci from any quality sputum or NPS, by culture (encapsulated) or *lytA* PCR

The fit of the suggested laboratory criteria with the predicted probabilities is depicted in Figure 1 with the optimal cutoff minimizing the sample size for a VE trial. The 1 case with low predictive probability (Figure 1A) had results available for all tests: The case was positive in sputum *lytA* PCR and had more than 2-fold increase in anti-PsaA and/or anti-CbpA but was negative in every other test result. The cases with high predictive probabilities but not fulfilling the case definition (Figure 1B) were typically missing results for several tests and were often positive in sputum tests (both culture and PCR).

Applying the LCA analysis framework to a population of suspected CAP cases

We also fitted the model to all 490 suspected CAP cases regardless of x-ray confirmation, and we investigated the distribution of predictive probabilities among confirmed (Figure 2A) and nonconfirmed CAP cases (Figure 2B). In the group of nonconfirmed CAP ($n = 167$), all but 5 cases had a probability below 0.5 (Figure 2B). These 5 cases were also the only ones who fulfilled our laboratory criteria in the group of nonconfirmed cases.

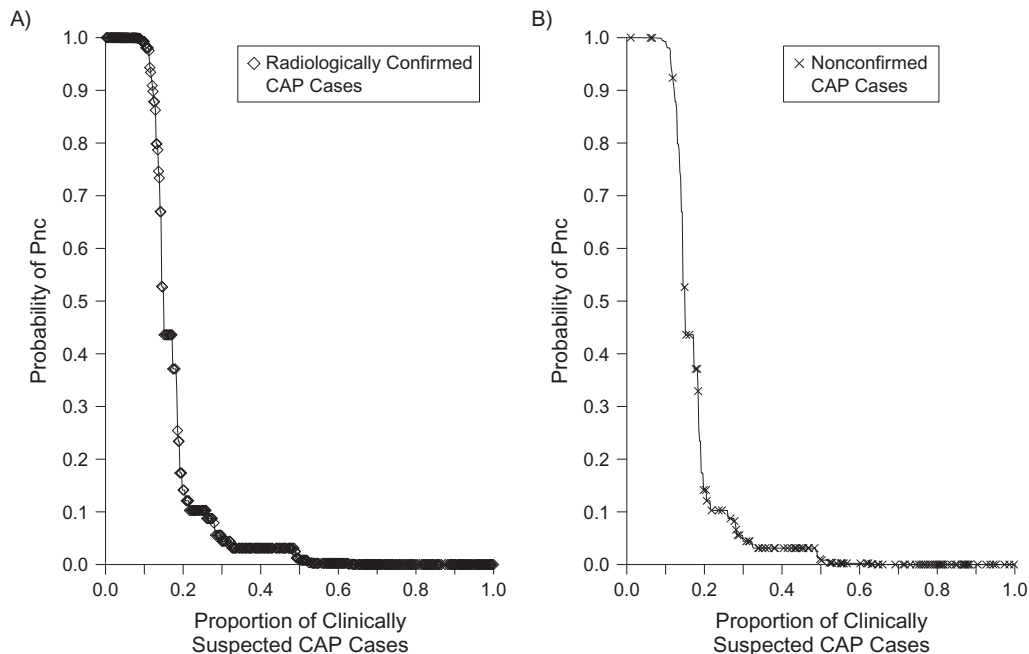


Figure 2. Predictive probabilities of pneumococcal (Pnc) community-acquired pneumonia (CAP) in 490 clinically suspected CAP cases based on their observed test profiles, Finnish Community-Acquired Pneumonia Epidemiology study, 2005–2007. Radiologically confirmed CAP cases, denoted by a \diamond (A), and nonconfirmed CAP cases, denoted by a \times (B), according to predictive pneumococcal probabilities are superimposed on the curves.

Trial sample size based on different case definition scenarios

Relevant quantities for a vaccine trial are presented in Table 3 and compared with different case definitions. Because none of the case definitions is perfect, the observed VE does not match the true VE (40%) due to misclassification. The optimal cutoff case definition was defined so that cases with predictive probability $>52\%$ were defined as pneumococcal CAP. Out of the 3 compared choices, this case definition has the highest prevalence and, by definition, also the lowest sample size. Our suggested case definition and the “blood culture positive or HQSc positive or urine antigen positive” case definition both have similar prevalence, but the observed VE in our case definition is higher. In addition, the sample size is 17,000 lower with our suggested case definition. When the case definition was applied to the whole set of suspected CAP cases ($n = 490$), the sample size was estimated at 74,000.

DISCUSSION

The primary aim of the FinCAP Epi study was to develop a case definition for pneumococcal pneumonia suitable for a vaccine trial. Based on our estimated predictive probabilities, the causal pathogen in the majority of CAP cases could be clearly classified as pneumococcal or nonpneumococcal. For example, 13% could be classified as being caused by pneumococcus based on positive results from a culture of blood or high-quality sputum (1). However, blood culture lacks sensitivity, and half the cases were missing high-quality sputum. Therefore, we added further

Table 3. Relevant Quantities for a Vaccine Efficacy Trial Using 3 Different Case Definitions, Using Data From the Finnish Community-Acquired Pneumonia Epidemiology Study, 2005–2007

Case Definition Estimate	Optimal Cutoff Case Definition, %	Suggested Study Case Definition, %	Positive in HQ Sputum, Blood Culture, or Urine, %
Prevalence	22.9	17.0	17.6
True positives	21.0	16.3	15.1
False positives	2.0	0.7	2.5
Observed VE	36.6	38.3	34.3
Specificity	97.4	99.0	96.7
Sensitivity	85.9	67.0	62.0
Total sample size ^a	62,000	76,000	93,000

Abbreviations: HQ, high-quality; VE, vaccine efficacy.

^a Values expressed as number of trial participants.

criteria to our case definition, in which 2 test results from different loci were required. Based on our model predictions, this case definition resulted in a sample size requirement for a vaccine trial that was almost 20,000 subjects lower than the more conventional “blood culture positive or HQSc positive or urine antigen positive” definition.

We used LCA to derive our case definition. Alternatives to LCA modeling for evaluation of test performances and for constructing the case definition have been suggested, mainly by means of descriptive analysis, using cross-tabulations of individual tests. For example, composite reference standard is a special case of such a method, based on user-defined gold standard (24). The drawback of these alternative methods is that there is no objective tool to derive the gold standard. For example, in a case where a highly sensitive new assay is compared with a poorly sensitive gold standard, some true positive cases in the new assay would be considered as false positives; thus, the new assay would be interpreted as unspecific rather than highly sensitive.

Furthermore, comprehensive ways of handling missing data are not generally incorporated into these alternative evaluations. As demonstrated in this study, missing data cause complications in exploratory analysis, model estimation, and diagnostics, and especially in construction of a case definition. In our exploratory analysis, by means of cross-tabulations, we have, for example, investigated the impact of sputum quality on the specificity and sensitivity of the test result (22). To complement these analyses, we applied a comprehensive methodological framework through LCA. In addition, instead of focusing on individual test performances, we used LCA as a convenient tool to utilize the entire test result profile and to borrow strength from completely observed profiles to those that were not fully observed.

Our LCA model consisted of 4 tests from respiratory tract samples, 3 of which were from sputum. This implies dominance of sputum samples in the estimation of gold standard, which may call into question the validity of the LCA model. However, as Table 2 showed, the performance of the tests from the respiratory tract varied considerably. In addition, local dependence structures were imposed between these tests, which reduces the individual contribution of each test to the model. Finally, in construction of our laboratory criteria we required that a sputum test (other than

HQSc) needed to be supported by positive result from another locus. Ideally, having more test results from different loci would have increased the power and validity of our model. Based on our modeling experiences, we identified 2 central guidelines for a design of any study trying to develop a case definition based on imperfect tests: 1) the more loci from which the sample could be obtained, the better; and 2) include tests with wide range of performance, both for sensitivity and specificity, in the LCA model. In order to increase the former, we further obtained oropharyngeal samples from the second-year CAP cases, but the yield from oropharyngeal culture proved to be too low to make any significant contribution to the case definition analysis (25). In addition, several modifications to the BinaxNOW antigen test for urine were performed in order to increase the relatively low yield of BinaxNOW (26). Concentration of urine with reading of results at 60 minutes showed some promise (26), but this was not used in the model for prediction purposes because results were available only for a subset of samples.

Interestingly, when applying our method to suspected cases regardless of x-ray confirmation, our laboratory criteria worked as a useful tool not only to distinguish pneumococcal from non-pneumococcal cases but also to distinguish x-ray confirmed CAP from nonconfirmed CAP among patients with clinically suspected CAP. At least in this study population, only a few suspected, but not radiologically confirmed, CAP cases would have been classified as pneumococcal CAP, if the case definition had consisted solely of clinical suspicion of CAP and the laboratory criteria proposed here, without also requiring radiological confirmation. And the predictive probabilities suggested that, although these cases could not be conclusively classified as CAP cases, it is quite plausible that pneumococcal vaccine would be effective in preventing such cases.

In summary, our focus here was to present a general methodological framework for constructing the case definition for a vaccine trial, minimizing the sample size needed. We demonstrated the feasibility and the benefits of the approach, through a practical example of how to utilize a combination of routinely available assays in the most optimal way. We believe this approach could be used for any set of tests, and could be applied to other diseases as well.

ACKNOWLEDGMENTS

Author affiliations: Department of Public Health Solutions, National Institute for Health and Welfare, Helsinki, Finland (Jukka Jokinen, Marja Snellman, Terhi M. Kilpi); Department of Public Health Solutions, National Institute for Health and Welfare, Tampere, Finland (Arto A. Palmu); Department of Public Health Solutions, National Institute for Health and Welfare, Oulu, Finland (Annika Saukkoriipi); and GSK, Wavre, Belgium (Vincent Verlant, Thierry Pascal, Jeanne-Marie Devaster, William P. Hausdorff).

The Finnish Community-Acquired Pneumonia Epidemiology study was funded mainly by GlaxoSmithKline Biologicals SA.

W.P.H. was and V.V., T.P., and J.M.D. are employees of the GSK group of companies. V.V., T.P., and J.M.D. own shares of the GSK group of companies. M.S. was and J.J., A.A.P., A.S., and T.M.K. are employees of the National Institute for Health and Welfare, which has received research funding from the GSK group of companies.

REFERENCES

- Palmu AA, Saukkoriipi A, Snellman M, et al. Incidence and etiology of community-acquired pneumonia in the elderly in a prospective population-based study. *Scand J Infect Dis*. 2014;46(4):250–259.
- File TM. Community-acquired pneumonia. *Lancet*. 2003;362(9400):1991–2001.
- Said MA, Johnson HL, Nonyane BA, et al. Estimating the burden of pneumococcal pneumonia among adults: a systematic review and meta-analysis of diagnostic techniques. *PLoS One*. 2013;8(4):e60273.
- Koivula IH. Epidemiology of community-acquired pneumonia. In: Marrie TJ, ed. *Community-Acquired Pneumonia*. New York, NY: Kluwer Academic Publishers; 2001:13–27.
- Jokinen J, Scott JA. Estimating the proportion of pneumonia attributable to pneumococcus in Kenyan adults: latent class analysis. *Epidemiology*. 2010;21(5):719–725.
- Bonten MJ, Huijts SM, Bolkenbaas M, et al. Polysaccharide conjugate vaccine against pneumococcal pneumonia in adults. *N Engl J Med*. 2015;372(12):1114–1125.
- Pride MW, Huijts SM, Wu K, et al. Validation of an immunodiagnostic assay for detection of 13 *Streptococcus pneumoniae* serotype-specific polysaccharides in human urine. *Clin Vaccine Immunol*. 2012;19(8):1131–1141.
- Formann AK, Kohlmann T. Latent class analysis in medical research. *Stat Methods Med Res*. 1996;5(2):179–211.
- Boelaert M, el Safi S, Goetghebeur E, et al. Latent class analysis permits unbiased estimates of the validity of DAT for the diagnosis of visceral leishmaniasis. *Trop Med Int Health*. 1999;4(5):395–401.
- Albert PS, McShane LM, Shih JH. Latent class modeling approaches for assessing diagnostic error without a gold standard: with applications to p53 immunohistochemical assays in bladder tumors. *Biometrics*. 2001;57(2):610–619.
- Baughman AL, Bisgard KM, Cortese MM, et al. Utility of composite reference standards and latent class analysis in evaluating the clinical accuracy of diagnostic tests for pertussis. *Clin Vaccine Immunol*. 2008;15(1):106–114.
- Butler JC, Bosshardt SC, Phelan M, et al. Classical and latent class analysis evaluation of sputum polymerase chain reaction and urine antigen testing for diagnosis of pneumococcal pneumonia in adults. *J Infect Dis*. 2003;187(9):1416–1423.
- Blake A, Njanpop-Lafourcade BM, Telles JN, et al. Evaluation of chest radiography, *lytA* real-time PCR, and other routine tests for diagnosis of community-acquired pneumonia and estimation of possible attributable fraction of pneumococcus in northern Togo. *Epidemiol Infect*. 2017;145(3):583–594.
- Little RJA, Rubin DB. *Statistical analysis with missing data*. 2nd ed. New York, NY: John Wiley & Sons, Publishers; 2002.
- Hagenaars JA. Latent structure models with direct effects between indicators. *Sociol Methods Res*. 1988;16:379–405.
- Harper D. Local dependence latent structure models. *Psychometrika*. 1972;37(1):53–59.
- Fuchs C. Maximum likelihood estimation and model selection in contingency tables with missing data. *J Am Stat Assoc*. 1982;77(378):270–278.
- Collins LM, Fidler PL, Wugalter SE, et al. Goodness-of-fit testing for latent class models. *Multivariate Behav Res*. 1993;28(3):375–389.
- Snellman M. *Case Definition of Pneumococcal Pneumonia—a Latent Class Analysis Approach*. Helsinki, Finland: National Public Health Institute; 2008. <http://urn.fi/URN:ISBN:978-951-740-772-4>.
- Rosón B, Carratalà J, Verdaguer R, et al. Prospective study of the usefulness of sputum Gram stain in the initial approach to community-acquired pneumonia requiring hospitalization. *Clin Infect Dis*. 2000;31(4):869–874.
- Ruiz M, Ewig S, Marcos MA, et al. Etiology of community-acquired pneumonia: impact of age, comorbidity, and severity. *Am J Respir Crit Care Med*. 1999;160(2):397–405.
- Kilpi TMK, Erkkilä L, Palmu AA, et al. Diagnostic value of culture of high-quality vs low-quality sputum for diagnosis of pneumococcal community-acquired pneumonia in elderly [abstract]. Presented at the 47th Interscience Conference on Antimicrobial Agents and Chemotherapy, Chicago, USA, September 17–20, 2007.
- Saukkoriipi A, Palmu AA, Jokinen J, et al. Effect of antimicrobial use on pneumococcal diagnostic tests in elderly patients with community-acquired pneumonia. *Eur J Clin Microbiol Infect Dis*. 2015;34(4):697–704.
- Alonzo TA, Pepe MS. Using a combination of reference tests to assess the accuracy of a new diagnostic test. *Stat Med*. 1999;18(22):2987–3003.
- Syrjänen RK, Palmu AA, Kaijalainen T, et al. Comparison of nasopharyngeal and oropharyngeal swabs in detecting carriage of *Streptococcus pneumoniae* in elderly [abstract]. Presented at the 6th International Symposium on Pneumococci & Pneumococcal Diseases, Reykjavik, Iceland, June 8–12, 2008.
- Saukkoriipi A, Pascal T, Palmu AA. Evaluation of the BinaxNOW *Streptococcus pneumoniae* antigen test on fresh, frozen and concentrated urine samples in elderly patients with and without community-acquired pneumonia. *J Microbiol Methods*. 2016;121:24–26.