DOI: 10.52950/ES.2024.13.2.002

APPLICATION OF HIERARCHICAL BAYESIAN MODELS FOR MODELING ECONOMIC COSTS IN THE IMPLEMENTATION OF NEW DIAGNOSTIC TESTS

TOMÁŠ KAREL, MIROSLAV PLAŠIL

Abstract:

The COVID-19 pandemic has highlighted the need for reliable and rapid diagnostic tests to control the spread of infection. The introduction of new rapid antigen tests often goes in tandem with the limited data availability, making it challenging to assess their performance at the initial phase of the pandemic. Sensitivity and specificity, the key performance characteristics provided by manufacturers, are typically derived under laboratory conditions and may not accurately reflect the tests' performance in field settings. We use the hierarchical Bayesian model to obtain their realistic estimates in real world conditions and show how it may be used in situations in which new tests with limited history are presented on the market. Proposed methodology allows for the efficient information pooling, thereby improving on the accuracy of parameter estimates for new tests. The results suggest that the application of hierarchical model on the Czech data led to a considerabile reduction in uncertainty associated with the parameter estimates as well as with potential economic cost implied by false positive test results. The model can thus assist in better informed decision-making and financial planning of both the government and corporations.

Keywords:

Bayesian statistics, Hierarchical Bayesian Model, COVID-19, Antigen tests, False Positivity

JEL Classification: C11, C10, C19

Authors:

TOMÁŠ KAREL, Prague University of Economics and Business, Czech Republic, Email: tomas.karel@vse.cz
MIROSLAV PLAŠIL, Prague University of Economics and Business, Czech Republic, Email: miroslav.plasil@vse.cz

Citation:

TOMÁŠ KAREL, MIROSLAV PLAŠIL (2024). Application of Hierarchical Bayesian Models for Modeling Economic Costs in the Implementation of New Diagnostic Tests. International Journal of Economic Sciences, Vol. XIII(2), pp. 20-37., 10.52950/ES.2024.13.2.002

1 Introduction

The COVID-19 pandemic has underscored the critical importance of diagnostic testing in controlling the spread of infection. Fast and affordable mass testing enables the timely and accurate detection of infected individuals, which is crucial for their isolation and treatment and for preventing further transmission of the virus (Lopes et al., 2020). However, the introduction of new rapid antigen tests (RADs) often encounters the challenge of their questionable reliability and unknown performance in the real-world conditions given the limited evidence available outside the laboratory settings. The paucity of data at the initial phase of the pandemic makes it difficult to precisely estimate key characteristics of these tests such as sensitivity and specificity, which are fundamental quantities for assessing their performance.

The sensitivity of a diagnostic test determines the extent to which it is able to identify individuals who are truly infected (i.e., those who will yield a positive result), whereas its specificity determines the extent to which it is able to identify individuals who are uninfected (i.e., those who will yield a negative result). It is of paramount importance to obtain precise estimates of these parameters in order to keep negative effects of pandemic under control and minimise the number of false positive and false negative results. Incorrect isolation of healthy individuals due to false positive results can have significant economic and social implications, whereas the spread of infection among the population due to false negative results represents a significant public health concern.

In this article, we employ a hierarchical Bayesian model (HBM) to estimate these parameters and demonstrate its ability to enhance their accuracy compared to alternative methods. We further illustrate how the improved specificity estimation achieved through HBM can facilitate a more accurate assessment of the economic impact of widespread testing following a pandemic outbreak. Our analysis particularly focuses on scenarios where multiple tests from various competing manufacturers are available on the market, despite limited empirical evidence regarding their performance under real-world conditions.

The hierarchical Bayesian model offers an elegant and natural solution to the underlying estimation problem by leveraging information from multiple diagnostic tests to improve parameter estimates for newly introduced tests. HBM facilitates information sharing among tests, enhancing the accuracy of sensitivity and specificity estimates, when limited (or even no) data is available for a new test. This approach is particularly useful in the context of a pandemic, where there is a need to rapidly deploy new diagnostic tests while ensuring their high reliability.

In our practical application, we present a real-life case involving a deployment of the new RAD, Actim SARS-CoV-2 test in the Czech Republic. In addition to reducing uncertainty related to its performance outside of laboratory conditions, we conduct an economic analysis demonstrating how improved parameter estimates can lead to significant cost savings associated with Covid misdiagnosis. This analysis illustrates that the use of HBM is not only theoretically advantageous but also has practical and economic merits. These findings are relevant to public health and can aid in assessing the potential economic costs associated with mass testing during a pandemic.

2 Literature Review

The application of Bayesian techniques for the evaluation of diagnostic tests has undergone a notable evolution over time. The integration of these approaches into medical practice offers a robust framework for interpreting the results of medical tests. The Bayesian approach has been instrumental in the context of the ongoing pandemic, being employed in the modeling of epidemiological scenarios and in the diagnosis of the disease itself. In a 2020 study, Watson, Richter, and Deeks emphasized the practical application of Bayesian point estimates in evaluating diagnostic tests for the novel coronavirus disease (2019-nCoV), also known as SARS-CoV-2. They underscored the significance of incorporating prior probabilities and updating them in accordance with the evolving situation of the pandemic. Sierra et al. (2023) employed bayesian hierarchical modeling for estimating age-specific COVID-19 infection fatality rates in developing countries using seroprevalence studies. Rehms et al. (2024) illustrated the utility of Bayesian hierarchical models in forecasting the number of infections and the propagation of the pandemic.

Obtaining precise estimates may limit economic and social risks associated with wrong results. The consequences and economic impacts of erroneous diagnostic test evaluations are described by authors such as Mayers and Baker (2020), who discussed the causes of false positive results, estimated the rate of false positives in the UK, and outlined the consequences of false positive results. According to the authors, some infections considered asymptomatic are caused by false positive results. This increases the risk of overestimating the incidence of COVID-19 and adds strain to contact tracing efforts (Mayers and Baker, 2020). Basile et al. (2020) focused on the consequences of false positivity in Australia. They reported that laboratories implemented measures, such as targeting asymptomatic individuals without identified epidemiological links, which led to a reduction in false positive tests. They emphasized that false positive results can lead to unnecessary quarantine measures, unjustified contact tracing, delays in identifying truly infected individuals, and psychological impacts on patients (Basile et al., 2020).

Alsheikh et al. (2021) described impacts of false positivity, including the postponement of surgical procedures, the transfer of patients to COVID-19 isolation wards, unnecessary quarantines, and medical leave. They emphasized the need for thorough evaluation of suspected cases. Similar conclusions were drawn by Healy et al. (2021). Surkova et al. (2020) described the health impacts, psychological, and financial aspects of false positive results. They stressed the economic losses related to isolation, loss of income, canceled trips, and the misallocation of public resources for testing and tracing (Surkova, Nikolayevskyy, and Drobniewski, 2020).

Czech experience is largly in line with these findings. Demographic analyses of the impacts of COVID-19, such as those by Huliková et al. (2021) or Klimkovský et al. (2021), indicate that the Czech Republic was one of the most affected countries. The economic aspects of the pandemic were analyzed by Němec and Špaček (2020). The economic consequences of erroneous diagnostic test evaluations in the form of false positive results during the COVID-19 pandemic, including the estimation of explicit costs in the Czech economy, were described by Karel, Mazouch, and Fischer (2022).

3 Methodology

3.1 Confusion Matrix

A fundamental tool for assessing the quality of diagnostic tests is the Confusion Matrix. This matrix provides an intuitive representation of the reliability characteristics of a test. Its first column represents the positive condition (in this case, "infected"), and the second column represents the negative condition (in this case, "healthy"). Test results, which can be either positive or negative, are displayed in the rows. The confusion matrix therefore displays the test results as TruePositive (TP), TrueNegative (TN), FalsePositive (FP) and FalseNegative (FN). The values in the matrix can be displayed either as absolute values, where the total is the number of trials conducted, or as relative values, where each element of the matrix represents the probability of each trial result given the current state. The primary objective of any clinical test is to maximize true positive (TP) and true negative (TN) results while minimizing false negative (FN) and false positive (FP) outcomes.

Table 1: Confusion Matrix

	Infected	**Healthy**
Negative test result	FN	TN
Positive test result	TP	FP

3.2 Sensitivity and Specificity

Sensitivity and specificity are key metrics for evaluating the performance of diagnostic tests. Sensitivity is the test's ability to correctly identify positive cases (infected individuals). It is defined as:

$$Sensitivity = \frac{TP}{TP+FN}, \quad (1)$$

where TP is the number of true positive results and FN is the number of false negative results.

Specificity is the test's ability to correctly identify negative cases (uninfected individuals) and is defined as:

$$Specificity = \frac{TN}{TN+FP}, \quad (2)$$

where TN is the number of true negative results and FP is the number of false positive results.

The sensitivity and specificity characteristics of tests are provided by the manufacturer. However, as evidenced by the research of Kliegr et al. (2022, 2023), there can be significant discrepancies

between the performance parameters reported by manufacturers and the actual results obtained under field conditions. In practice, assessment of the test performance is further complicated by the fact that the true status of the patient's condition (healthy vs. infected) is unknown. As a result, performance is typically defined relative to a 'gold standard' (best available benchmark), which, in the case of Covid 19 pandemic, is the PCR test (RT-PCR method). In our analysis, we also adhere to this convention and treat the result of a PCR test as definitive standard.

3.3 Bayesian inference

To estimate key characteristics of the test, we stick to Bayesian techniques. In practical terms, Bayesian inference is built on two fundamental sources of information (see, e.g. Koop, 2003 for a textbook treatment). Similar to classical statistics, it utilizes information from the data to make statistical inference. However, unlike traditional statistics, Bayesian inference also incorporates prior knowledge about the parameters to be estimated in the form of a probabilistic distribution. This prior knowledge may stem from earlier studies, similar data samples, or reflect researchers' beliefs based on the available experience. The characteristics of the prior distribution, such as its mean and variance, offer enough flexibility to align unknown parameters subject to estimation with the prior knowledge and the associated uncertainty surrounding it.

By combining the information from the data with the prior, one obtains the posterior distribution, which is the central quantity of interest in Bayesian statistics. The posterior distribution encapsulates everything known about the parameters after observing the data, thus forming the foundation for statistical inference. Put it differently, the posterior distribution updates a researcher's opinions about an unknown parameter based on the observed data. This updating process is governed by rigorous probabilistic rules implied by Bayes' theorem.

Let us denote a vector of uknown parameters of interest as θ. Using Bayes' theorem the posterior distribution of θ given the observed data D, $P(\theta|D)$, then emerges as:

$$p(\theta|D) = \frac{p(\theta)p(D|\theta)}{p(D)} \tag{3}$$

where $p(D|\theta)$ is the traditional likelihood function, $p(\theta)$ is the prior distribution of the parameter θ, and $p(D)$ is normalization constant that ensures that the posterior is a valid probabilistic distribution. In many practical applications, this normalization constant can be safely ignored, leading to an alternative version of Bayes' theorem:

$$POSTERIOR \propto LIKELIHOOD \times PRIOR \tag{4}$$

This expression highlights that the posterior distribution is proportional to the product of the prior distribution and the likelihood function, clearly illustrating how the two building blocks of Bayesian statistics are combined to provide a ground for statistical inference.

3.4 Beta-Binomial Bayesian model

As it is clear from (1) and (2), sensitivity and specificity both represent a proportion of true positive and true negative cases, respectively on the specified population. Assuming that the results of individual tests carried out in the field are independent of each other, the number of true positive (true negative) cases, y, in a total of n tests follows a binomial distribution $Binomial(n, \theta)$ with probability mass function

$$p(y) = \binom{n}{y} \theta^y (1-\theta)^{n-y}. \quad (5)$$

Here unknown parameter θ represents true on-field performance of the test, specifically its sensitivity or specificity. By collecting the data, i.e. by collecting the results of the diagnostic tests and comparing them against the benchmark (PCR test), one can simply obtain the estimates of sensitivity (specificity) using (1) and (2) as these also represent a traditional maximum likelihood estimate of parameter θ in a binomial distribution (5). However, these estimates can be very imprecise and unreliable when they are derived from a data sample of limited size. This is where Bayesian statistics may jump in, offering a more robust approach to inference by incorporating prior knowledge along with the observed data.

The starting point for incorporating the prior knowledge into Bayesian analysis is the choice of a prior distribution for the unknown parameter, θ. This choice should be made independently of the data sample and consider all available information. Since the parameters of sensitivity and specificity take values within the interval from 0 to 1 (or equivalently between 0 an 100 %) a suitable and flexible candidate distribution for formulating prior beliefs about them is the Beta distribution, as demonstrated *inter alia* in Albert (2009), Bolstad (2007), and Wagenmakers and Matzke (2023).

The overall shape of the Beta distribution is fully controlled by two parameters, α and β.

$$Beta(\theta;\ \alpha, \beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)},\ \theta \in \langle 0; 1 \rangle \quad (6)$$

where $B(\alpha, \beta)$ is the beta function defined as

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1} dt. \quad (7)$$

Depending on the choice of the parameters α and β the Beta distribution can take different shapes encompassing a wide range of prior beliefs about the values of sensitivity or specificity, respectively. If both parameters α and β are equal to one, the distribution corresponds to a uniform distribution which is frequently used in Bayesian statistics as a non-informative prior distribution. If the parameters are greater than one, the distribution is unimodal.

In addition to its flexibility, another advantage of choosing the Beta distribution is computational. In particular the Beta distribution is the conjugate prior distribution for Bernoulli and binomial

probability distributions, which means that posterior distribution can be derived analytically in a closed form, thereby eliminating a need for costly posterior simulations. It can be shown that when the prior distribution is Beta and the data distribution is binomial, the resulting posterior distribution is also Beta with known parameters (see e.g. Robert, 2007 and Gelman et al., 2013). Combining information from the likelihood function where data y_i represent true positive (true negative) cases in the sample with the prior distribution (6), the posterior distribution takes the form:

$$p(\theta|y) = Beta(\alpha + \sum y_i\,, \beta + n - \sum y_i). \tag{8}$$

3.5 Hierarchical Beta-Binomial Model

The key assumption of the simple Beta-Binomial model outlined in (6)–(8) is that all diagnostic tests are mutually independent and exhibit identical performance. However, this assumption is often unrealistic, as tests from different manufacturers may vary due to differences in virus detection thresholds. Ignoring these differences means overlooking important assumptions of the model. To address this issue, one option would be to estimate test characteristics separately for each manufacturer.

However, since all tests are based on the same underlying biochemical principles, they are not entirely unrelated either. Therefore, estimating test characteristics separately for each manufacturer would result in a loss of valuable information. Hierarchical Bayesian Models are well-suited for this context, as they allow for different test characteristics across manufacturers while also accounting for their similarities. HBMs enable the pooling of information across diagnostic tests, providing a more nuanced and accurate estimation by leveraging both the shared and unique aspects of each test. This approach also enables the utilization of multilevel information to model the characteristics of a newly deployed diagnostic tests. HBM integrates uncertainty caused by the lack of specific data available for a new test at a given time across multiple levels thereby improving accuracy of parameter estimates for the tests with limited data availability.

The proposed hierarchical Beta-Binomial model can be formally described as follows (see Albert and Hu, 2020 for additional details). We now assume that data sampling for test characteristics (sensitivity or specificity) for each manufacturer, j, follows a binomial distribution

$$y_{ij} \sim Binomial\left(n_{j,}\theta_j\right). \tag{9}$$

Priors beliefs about the uknown parameters are constructed in two stages. In stage 1, priors on the individual characteristics θ_j are elicited using a Beta distribution, $Beta(\alpha, \beta)$ where α and β are shared among all θ_j.

$$\theta_j|\alpha, \beta \sim Beta(\alpha, \beta). \tag{10}$$

This prior does not imply that all θ_j values are the same but rather that they are related and come from the same distribution. If the variance of the prior distribution is large, it suggests that the θ_j

values can be very different from each other apriori. Conversly, if the variance is small, it indicates that θ_j are believed to be very similar in magnitude.

Note, that unlike in a simple beta-binomial model, the values of hyperparameters α and β are considered uknown in (10). In stage 2 of the hierarchical model, these hyperparameters are therefore assigned a joint prior density π:

$$\alpha, \beta \sim \pi(\alpha, \beta). \tag{11}$$

Prior in (11) provides information about the location and variability of test characteristics across the entire population of diagnostic tests that are in use, i. e. across all manufactures. The choice of $\pi(\alpha, \beta)$ is problem-specific; however uninformative (flat) or only weakly informative priors are commonly used in practice. For a more thorough discussion on the formulation of the second-stage priors, see also Albert and Hu (2020) and Gelman et al. (2013).

3.6 Implementation and Estimation of the Model

In our application, we formulate the joint prior distribution (11) as follows. Since little is known about the variability across tests, specifically at the outbreak of the pandemic, we prefer to stay rather agnostic apriori about the overall specificity or sensitivity across all diagnostic tests and opt for fully non-informative prior.

As the stage 2 prior should provide information on the average performance across the population of all tests, it is more natural to reparametrize the prior in terms of its mean value and the sample size which is directly linked to the degree of uncertainty in the prior distribution. Given that the mean μ of the Beta distribution equals $\mu = \frac{\alpha}{\alpha + \beta}$ and the prior sample size corresponds to $\eta = \alpha + \beta$, we can reformulate prior distribution (11) in terms of μ, $\eta \sim \pi(\mu, \eta)$ where the original parameters can be recovered as $\alpha = \mu\,\eta$ and $\beta = (1 - \mu)\eta$. Using this reparametrisation, we can now elicit non-informative prior on μ as:

$$\mu \sim Beta(1, 1). \tag{12}$$

which is equivalent to a uniform distribution over the interval (0,1). This reflects a prior belief that all values of the overall mean across manufacturers are equally likely a priori.

The prior on parameter η governs the extent to which the means of individual manufacturers' tests are pulled toward the overall ("grand") mean across all manufacturers. To remain agnostic, we specify the strength of this pulling force to be regulated by a uniform distribution over the interval (0,1). With some algebra, it can be shown (see Albert and Hu, 2020 for the derivation) that this leads to the following formulation of the prior:

$$log\ \eta \sim \text{Logistic}(\log\ n^*, 1), \tag{13}$$

i.e. the logarithm of η has logistic distribution with location parameter $\log n^*$ and scale one. In our model we set $\log n^*$ to 1,4624 ($n^* = 29$).

With all priors being fully specified, Bayesian inference can be performed. However, the posterior distribution for the hierarchical model (9)-(13) cannot be obtained in a closed form in general, similarly to other complex multivariate cases and situations where the prior distribution is not naturally conjugate. In these cases, Markov Chain Monte Carlo (MCMC) simulation techniques are widely used to learn about the posterior distribution (Gelman et al, 2013). These simulation methods generate a large number of random samples through the construction of a Markov chain to approximate the posterior distribution and its characteristics. Our HBM model was implemented using the PyMC library (Salvatier et al., 2016), utilizing the No-U-Turn Sampler (NUTS) algorithm for sampling posterior distributions (Hoffman and Gellman, 2014). This algorithm is efficient for MCMC sampling and ensures effective convergence to posterior distributions of probabilistic models.

3.7 Data on diagnostic tests

The data used in this study are based on extensive research published in the article by Kliegr et al. (2022). This research covered the period from August 2021 to January 2022, during which regional public health stations collected more than 2 million diagnostic test results. To ensure accuracy and reliability, all data were subsequently verified using the RT-PCR method, providing a benchmark for evaluating the performance of individual diagnostic tests.

As demonstrated in the study by Kliegr et al. (2022), and the subsequent extension of the study to the delta and omicron virus variants by Kliegr et al. (2023), sensitivity and specificity values in practical (on-field) use differ from those in laboratory conditions. This underscores the need for precise methods that can quickly estimate the true characteristics of diagnostic tests. The following table displays the sensitivity and specificity values reported by manufacturers for the five most commonly used tests during the analyzed period (October 2020 to December 2020). Table 2 also presents the actually measured values, including confidence intervals, from the study conducted by Kliegr et al. (2022).

Table 2: Comparison of Sensitivity and Specificity between Manufacturer-Reported Values and Values Measured in the Study (Kliegr et al., 2022)

Diagnostic test	Sensitivity producer (%)	Specificity producer (%)	Sensitivity on-field (95% CI)	Specificity on-field (95% CI)	Sample in on-field study
SARS-CoV-2 Antigen Rapid Test Kit	95.06	99.62	48.7 (47.5 – 49.9)	97.8 (97.7 – 98.0)	49,505
VivaDiagTM Pro SARS-CoV-2 Ag Rapid Test	95.29	99.84	75,8 (74.7 – 76.9)	97,0 (96.8 – 97.2)	34,696

Panbio Covid-19 Ag Rapid Test	91.4	99.8	79.9 (78.9 – 80.8)	97.1 (96.8 – 97.3)	26,949
NADAL COVID-19 Ag Test	97.59	> 99.9	82,0 (80.9 – 83.1)	96.3 (96.0 – 96.6)	22,818
Humasis COVID-19 Ag Test	90.2	100	77.5 (76.4 – 78.6)	97.2 (96.9 – 97.5)	19,290

Source: Kliegr et al. (2022)

A significant deviation can be observed between the sensitivity and specificity values reported by manufacturers. For the most prevalent antigen test (n = 49,505), the measured on-field sensitivity in the study (Se = 48.7%) is even lower than a result based purely on chance. The sensitivity and specificity reported by manufacturers are based on laboratory study results with a limited number of observations and may vary in real-world conditions depending on factors such as the amount of antigen in the analyzed sample, the viral load of the tested individual, the quality of nasopharyngeal swab collection, the age and storage conditions of the test kit, and many other relevant factors.

These observations highlight the critical need for robust data-driven methods to estimate the performance of the tests in practise and guide the selection among available diagnostic tests, ensuring that resources are allocated to tests with empirically verified effectiveness.

3.8 Practical application: Actim SARS-CoV-2 test

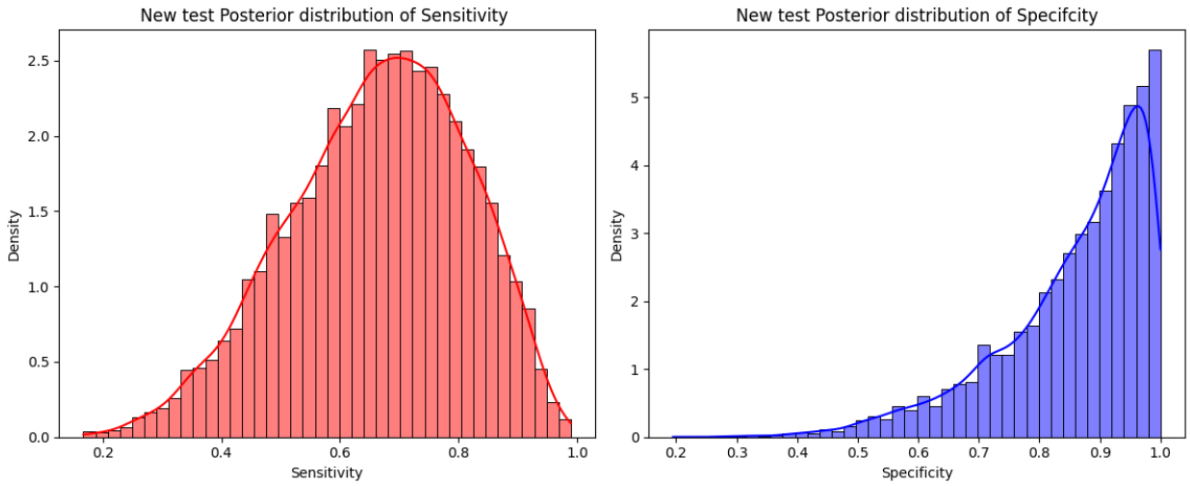
On November 10, 2021, a new diagnostic RAD test, Actim SARS-CoV-2, was introduced by the manufacturer Actim Oy. For this new test, only 29 records were available at the time of our study, leading to significant uncertainty in estimating its characteristics, specifically sensitivity and specificity. Such a low number of observations does not provide a solid ground for reliable estimates of test performance, which can affect the accuracy of subsequent analyses that rely on these estimates

4 Results

4.1 Simple Beta-Binomial model

To demonstrate the advantages of the proposed hierarchical model, we first estimate the specificity and sensitivity of the newly introduced Actim SARS-CoV-2 test only using 29 records available for this test. Under these conditions, a simple Beta-Binomial model (5)–(8) can be used to estimate the test characteristics. Its application leads to the following results: the sensitivity (Se) has a mean value of 0.663 with a 95% credible interval of 0.397 - 0.938 and a standard deviation of 0.150. The specificity (Sp) has a mean value of 0.858 with a 95% credible interval of 0.621 - 1.000 and a standard deviation of 0.123.

Figure 1: The posterior distribution of Sensitivity a Specificity of a new diagnostic test (Simple Beta-Binomial model)



Source: own calculations

Table 3: Estimates of Sensitivity and Specificity using simple Beta-Binomial model

	Mean	Standard deviation	95% CI
Sensitivity	0.663	0.150	0.397; 0.938
Specificity	0.858	0.123	0.621; 1.000

Source: own calculations

4.2 Hierarchical Bayesian Model

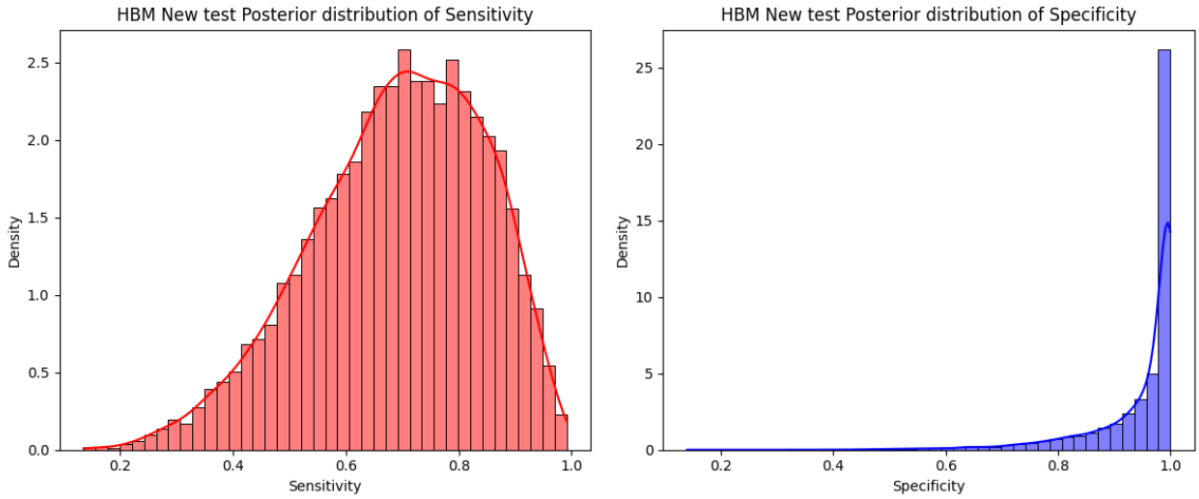
Using the hierarchical model, the information from all tests has been exploited and efficiently pooled to derive uncertainty associated with monitored characteristics. The sensitivity (Se) now increased to a mean value of 0.691 with a 95% credible interval of 0.419 - 0.963 and a standard deviation of 0.153.

These results suggest that the application of the HBM led to a significant reduction in the estimation uncertainty associated with the key test characteristics. This is most apparent in the narrower 95% credible intervals for specificity, where their length dropped by 41% and the standard deviation decreased from 0.123 to 0.089, pointing to a more precise estimate. This reflects the fact that the HBM makes use of additional external information and effectively propagates uncertainty from previously validated RAD tests. As we show below, the reduction in uncertainty can be crucial in practical applications, such as large-scale screening programs, where more precise estimates lead to better planning and more efficient allocation financial resources.

It is worth noting however, that the sensitivity estimates obtained from both models remain relatively similar, with only a slight increase in the mean value from 0.663 to 0.691. This can be attributed to the low number of true positive cases in the available data, which constrains the model's ability to improve sensitivity estimates under uninformative priors. In contrast, there has been a much larger

proportion of true negative cases, which explains why the hierarchical Bayesian model had a greater impact on improving the estimate for specificity than for sensitivity. As estimates of sensitivity is naturally more variable in low-prevalence settings (not enough true positive observations), the pooling of information across tests contributes less to its precision, whereas the abundance of true negative cases enables a more substantial reduction in uncertainty for specificity.

Figure 2: The posterior distribution of Sensitivity a Specificity of a new diagnostic test (Hierarchiacal Bayesian Model)



Source: own calculations

Table 4: Estimates of Sensitivity and Specificity using HBM

	Mean	Standard deviation	95% CI
Sensitivity	0.691	0.153	0.419; 0.963
Specificity	0.946	0.089	0.777; 1.000

Source: own calculations

5. Economic Analysis of False Positive Results

Obtaining more precise estimates of the test characteristics is important in itself as it may help the authorities take more informed decisions about the new acquisition of specific diagnostic test. However, there also are other practical merits that can result in direct economic gains. To illustrate this, we present a realistic quantitative example demonstrating how enhanced specificity estimates can lead to improved cost management, better economic planning and data-driven decision-making in the context of a pandemic outbreak.

We devise a realistic scenario in which the newly introduced diagnostic test Actim SARS-CoV-2 test should be used for preventive screening of 10,000 employees, who, without any clinically manifested symptoms or epidemiological contact, are required to take this diagnostic test before entering employment. Such a scenario can reflect a microeconomic level of the largest companies in the country, but it can easily be scaled to macroeconomic level as well. We further suppose that

in the sample of 10,000 tested individuals, 100 are actually infected, and the prevalence is 1% ($Prev$ = 0.01). This is a realistic assumption as the prevalence ranged between 0.025 % and 2.5 % in the Czech Republic at that time (Ministry of Health, 2022e). While modeling the overall effects of the described screening procedure may be complex and challenging in practice, assessing the costs associated with false negative results is relatively straightforward and can be crucial for efficient policy making.

Specifically, the ability of the test to identify negative cases when individuals are not infected has a significant impact on estimating the economic costs caused by false positives and the unnecessary quarantine of the tested individual and their epidemiological contacts.Considering the anti-epidemic measures in place in the Czech Republic in summer 2021, false positive results from diagnostic tests led to significant economic and social impacts, such as loss of productivity, loss of government revenue in the form of social and health insurance, unnecessary healthcare expenses, psychological stress, and increased pressure on healthcare and tracking systems (see Karel et al. 2022 for details).

Given the uncertainty surrounding the magnitude of these economic costs, it is crucial to assess their potential scale once the new test is introduced, to help better allocate financial resources from both the government and corporate sectors and reduce inefficiencies. To proceed, we use the methology proposed in Karel et al. (2022) to determine the total costs associated with the isolation of falsely negative individual. It covers the direct impact of lost wages of these individuals on insurance and income tax, additional losses of employer required to pay their employees wage compensation for the first 14 days amounting to approximately 9,600 CZK for a gross salary of 35,000 CZK in 2022 (Covid Portal, 2022b). Moreover, if an employer needed to replace an isolated or quarantined employee with another fully-fledged employee, they had to pay the corresponding wage. On the other hand, in terms of budgetary revenues, there would be no loss of income from insurance and income tax. In sum, the overal cost per one individual with a false negative result was estimated to amount to approximately CZK 10,350 in 2022 for 14 days staying in quaranteen or isolation (Karel et al., 2022).

5.1. Comparison of results using simple and the Hierarchical Beta-Binomial Model

Armed with the estimate of total cost per one individual, we can estimate the total implied (14day) cost associated with the proposed mass screening. We start again with the case of the simple Beta-Binomial Bayesian model which only uses information on the results obtained specifically for this rapid antigen test. Given that the estimated specificity (Sp) under this model is 0.858 (mean of the posterior distribution, see Table 3), this results in a point estimate of 1,406 false positive individuals in the screening sample and the estimate of total cost for false positive individuals amounts to 14,552,100 CZK. The number of false positive individual can be obtained using the relation:

$$FP = (1 - Sp)\ (1 - Prev)\ N, \tag{14}$$

where N =10,000 is the number of individuals tested, Sp is the estimated specificity and $Prev$ denotes prevalence. Accounting further for the uncertainty associated with the point estimate under the simple Beta-Binomial model, the total cost is estimated to lie between (13,848,300; 15,255,900) with 95% probability.

Using the hierarchical Bayesian model, the estimated specificity (Sp) is 0.946, leading to an estimated expected number of 535 false positive individuals. The total cost for false positive individuals is 5,589,000 CZK, with a 95% CI of (5,123,250; 5,909,850).

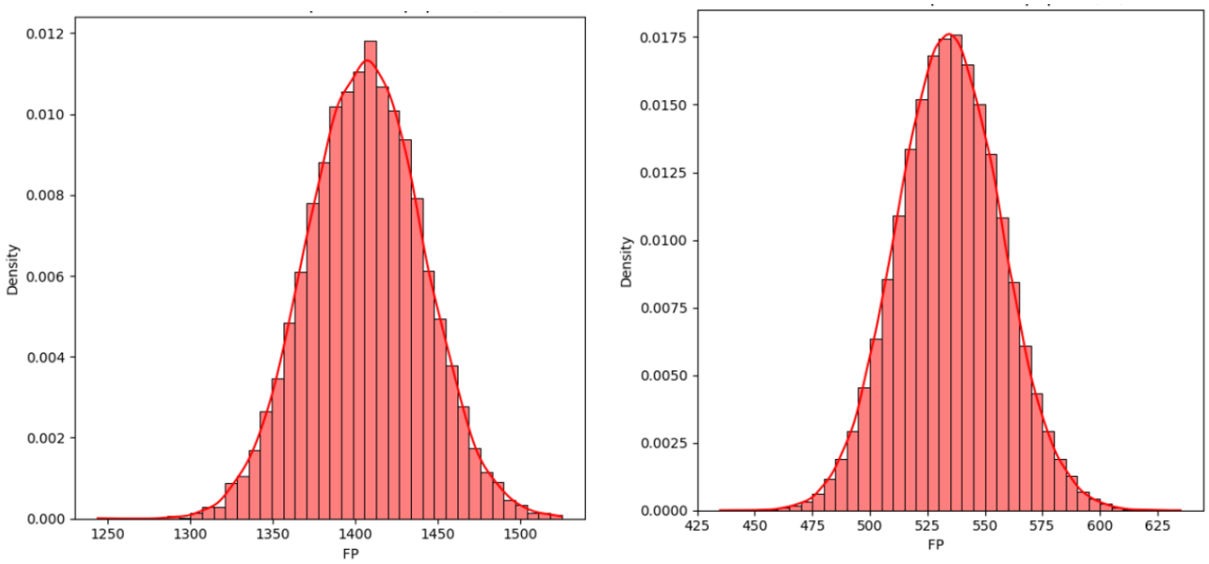
Again, the reduction in uncertainty is significant when comparing the credible intervals for economic cost across the two models. The enhanced precision and reduced uncertainty in cost estimation has notable economic benefits. With reduced uncertainty and more accurate forecasts, public health authorities as well as corporations should be able to allocate financial resources more effectively as they no loger have to keep them untouched to cover potential losses implied by false negative results. At the government level, improved accuracy in cost predictions allows for better financial planning, ensuring that limited public funds are allocated where they are needed most.

Table 5: Estimates of Specificity, FP and Total Costs, Simple Beta-Binomial model and HBM

	Classical approach	HBM
Specificity	0.858	0.946
False positivity cases	1,406	535
Total Costs	14,552,100	5,537,250

Source: own calculations

Figure 3: The posterior distribution of False positive estimates (Simple Beta-Binomial model vs. HBM)



Source: own calculations

5 Conclusions

Fast, cheap, and scalable diagnostic testing is essential for controlling the spread of global pandemics and balancing public health with social and economic costs. In practice, the demand for rapid and affordable testing often comes at the expense of test quality, highlighting the need for rigorous, data-driven approaches to assess their field performance. Precise estimates can support decision-making by authorities, helping to conserve public funds while keeping the disease spread under control.

In this paper, we employed advanced modeling techniques to obtain reliable estimates of key test characteristics, utilizing the Bayesian paradigm and a hierarchical structure of the available data. Hierarchical Bayesian models provide flexible tools for pooling available information, accounting for both the commonalities and differences among the tests. As such they can lead to substantial improvements vis-à-vis other competing models. We showed that the application of HBM led to a clear reduction in uncertainty associated with the estimate of specificity of a newly deployed RAD test, Actim SARS-CoV-2, where only limited information on its field performance was available at the beginning of the pandemic. In both the estimate of the test's specificity and the economic costs associated with false positive results, the reduction in uncertainty reached several tens of percent.

The benefits of these results are at least twofold. First, obtaining precise estimates for all available tests can assist governments in selecting tests that minimize false positive and false negative results in real-world conditions. Second, these estimates can be used for more efficient economic planning, as demonstrated in our devised yet realistic scenario. Reducing the estimated number of false positive results, particularly by minimizing the uncertainty represented by 95% credible intervals, also alleviates the burden on healthcare and surveillance systems. Forecasted false positive outcomes necessitate additional confirmatory testing, contact tracing, and often unnecessary medical interventions. In scenarios where the resources are limited, such misallocation represents an inefficiency that could otherwise be diverted to more critical needs. A more efficient allocation of healthcare resources, made possible through enhanced forecasting, allows for a greater focus on truly infected individuals, thereby improving the efficiency and effectiveness of healthcare interventions.

The hierarchical Bayesian model, along with other advanced data-driven methods, highlights the broader importance of sophisticated statistical methodologies in epidemiology and public health. The thorough study of the COVID-19 pandemic serves as a valuable testing ground for these methods, enabling us to build on this experience when future pandemics emerge.

The potential applications of the HBM for estimating test characteristics, and in epidemiology more broadly, extend far beyond what is presented in this paper. For instance, HBM can be employed to design an optimal strategy for mass test procurement by governments, based on estimated test qualities. The strategy can be updated daily as new data on individual tests become available. Although not demonstrated here, such a strategy can be derived directly from the posterior distribution samples, assigning probabilities to each manufacturer that their test is the best-performing at any given time. We leave the full development of this strategy for our future work.

References

ALBERT, J. (2009). *Bayesian Computation with R (2nd ed.)*. Dordrecht: Springer. Available at: https://doi.org/10.1007/978-0-387-92297-3

ALBERT, J., and HU, J. (2020). *Probability and Bayesian Modeling*. Boca Raton: CRC Press. Available at: https://doi.org/10.1201/9780429273875

ALSHEIKH, H., ABDULLAH, A., and NASSAR, M. (2021). False positive results in SARS-CoV-2 testing and the impact on public health. *Journal of Clinical Medicine*, 10(3), 654-660. https://doi.org/10.3390/jcm10030654.

ANDRIEU, C., DE FREITAS, N., DOUCET, A., and JORDAN, M. I. (2003). An Introduction to MCMC for Machine Learning. *Machine Learning*, 50(1-2), pp. 5-43. Available at: https://doi.org/10.1023/A:1020281327116

BASILE, K., MADDOCKS, S., KOK, J., and DWYER, D.E. (2020). Accuracy amidst ambiguity: false positive SARS-CoV-2 nucleic acid tests when COVID-19 prevalence is low. *Pathology*, 52(7), pp. 809-811.

BOLSTAD, W.M. (2007). *Introduction to Bayesian Statistics*. John Wiley & Sons.

GELMAN, A., CARLIN, J.B., STERN, H.S., DUNSON, D.B., VEHTARI, A., and RUBIN, D.B. (2013). *Bayesian Data Analysis* (3rd ed.). CRC Press.

GEMAN, S., and GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6), pp. 721-741. Available at: https://doi.org/10.1109/TPAMI.1984.4767596

HEBÁK, P., HUSTOPECKÝ, J., and MALÝ, I. (2013). *Vícerozměrné statistické metody: Metody a aplikace (Multivariate Statistical Methods: Methods and Applications)*. Praha: Informatorium.

HEALY, B., KHAN, A., METEZAI, H., BLYTH, I., and ASAD, H. (2021). The impact of false positive COVID-19 results in an area of low prevalence. *Clinical Medicine (London)*, 21(1).

HOFFMAN, M. D., and GELMAN, A. (2014). The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(1), 1593–1623.

KAREL, T., MAZOUCH, P., and FISCHER, J. (2022). The economic impact of false positivity of COVID-19 PCR testing in the Czech Republic. *International Journal of Economic Sciences*, 11(1), pp. 37-46.

KLIERGR, T., JARKOVSKÝ, J., JIŘINCOVÁ, H., KUCHAŘ, J., KAREL, T., TACHEZY, R. (2022). Role of population and test characteristics in antigen-based SARS-CoV-2 diagnosis, Czechia, August to November 2021. *Euro Surveillence*, 27(33), 2200070. https://doi.org/10.2807/1560-7917.ES.2022.27.33.2200070.

KLIERGR, T., JARKOVSKÝ, J., JIŘINCOVÁ, H., KUCHAŘ, J., KAREL, T., CHUDÁN, D., VOJÍŘ, S., ZAVŘEL, M., & ŠANCA, O., & TACHEZY, R. (2023). Can variants, reinfection, symptoms and test types affect COVID-19 diagnostic performance? A large-scale retrospective study of AG-RDTs during circulation of Delta and Omicron variants, Czechia, December 2021 to February 2022. *Euro Surveillence*, 28(38), 2200938. https://doi.org/10.2807/1560-7917.ES.2023.28.38.2200938.

KOOP, G. (2003). *Bayesian Econometrics*. Chichester: John Wiley & Sons Ltd. Available at: https://doi.org/10.1002/0470867156

LOPES, L. C., BOMFIM, E., SILVEIRA, D. S. C. D., et al. (2020). Effectiveness of mass testing for control of COVID-19: A systematic review protocol. *BMJ Open*, 10. https://doi.org/10.1136/bmjopen-2020-040413.

MAYERS, C., and BAKER, K. (2020). *Impact of false-positives and false-negatives in the UK's COVID-19 RT-PCR testing programme*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/895843/S0519_Impact_of_false_positives_and_negatives.pdf.

MINISTRY OF HEALTH (2022a). Covid 19: *Otevrene datove sady a sady ke stazeni*. Available at: https://onemocneni-aktualne.mzcr.cz/api/v2/covid-19 (Accessed: 1.7.2024).

MINISTRY OF HEALTH (2022b). Covid 19: *Přehled aktuální situace*. Available at: https://onemocneni-aktualne.mzcr.cz/covid-19 (Accessed: 1.7.2024).

MINISTRY OF HEALTH (2022c). Covid 19: *Přehled vykázaných očkování*. Available at: https://onemocneni-aktualne.mzcr.cz/vakcinace-cr (Accessed: 1.7.2024).

MINISTRY OF HEALTH (2022d). *Price of the Covid test*. Available at: https://koronavirus.mzcr.cz/ministerstvo-zdravotnictvi-snizuje-maximalni-cenu-testu-na-covid-19-pro-samoplatce/ (Accessed: 8.7.2024).

MINISTRY OF HEALTH (2022e). *Covid-19: Přehled aktuální situace v ČR*. Získáno z https://onemocneni-aktualne.mzcr.cz/covid-19 (Accessed: 5.8.2024).

NĚMEC, J., and ŠPAČEK, D. (2020). The Covid-19 pandemic and local government finance: Czechia and Slovakia. *Journal of Public Budgeting, Accounting & Financial Management*. 32(5). doi: 10.1108/JPBAFM-06-2020-0087.

ROBERT, C.P. (2007). *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation* (2nd ed.).

SALVATIER, J., WIECKI, T. V., & FONNESBECK, C. (2016). Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*, 2, e55. https://doi.org/10.7717/peerj-cs.55

SURKOVA, E., NIKOLAYEVSKYY, V., and DROBNIEWSKI, F. (2020). False-positive Covid-19 results: hidden problems and costs. *The Lancet Respiratory Medicine*, 8(12), pp. 1167-1168.

PUGH, S., LEVIN, A. T., MEYEROWITZ-KATZ, G., SOMAN, S., OWUSU-BOAITEY, N., ZWI, A. B., MALANI, A., WILSON, A., and FOSDICK, B. K. (2023). *A hierarchical Bayesian model for estimating age-specific COVID-19 infection fatality rates in developing countries*. arXiv. https://doi.org/10.48550/arXiv.2308.04012¨

REHMS, L., BIEG, T., and MALLICK, B. (2024). Predicting pandemic outbreaks with Bayesian hierarchical models. *Journal of Infectious Disease Modeling*, 23(1), 77-93. https://doi.org/10.1016/j.idm.2024.01.005.

SIERRA, P., LEVIN, A. T., MEYEROWITZ-KATZ, G., SOMAN, S., OWUSU-BOAITEY, N., ZWI, A. B., MALANI, A., and FOSDICK, B. K. (2023). A hierarchical Bayesian model for estimating age-specific COVID-19 infection fatality rates in developing countries. *arXiv*. https://doi.org/10.48550/arXiv.2308.04012.

WAGENMAKERS, E.-J., and MATZKE, D. (2023). *Bayesian inference from the ground up: The theory of common sense*. Available at: https://www.bayesianspectacles.org/free-course-book/

WATSON, J., RICHTER, A., and DEEKS, J. J. (2020). Testing for SARS-CoV-2 antibodies. *BMJ*, 370. https://doi.org/10.1136/bmj.m3325.

