Hellenic Complex Systems Laboratory

A Software Tool for Applying Bayes' Theorem in Medical Diagnostics

Technical Report XXVII

Theodora Chatzimichail and Aristides T. Hatjimihail
2024

1. Abstarct

Background: In medical diagnostics, determining post-test or posterior probabilities for disease and understanding the associated uncertainty and confidence intervals are essential for patient care.

Objective: This study introduces a software tool developed in the Wolfram Language for the parametric estimation, visualization, and comparison of Bayesian diagnostic measures and their uncertainty.

Methods: The tool employs Bayes' theorem to compute posterior probabilities for disease and absence of the disease, and diagnostic thresholds derived positive and negative predictive values. It also quantifies their sampling, measurement, and combined uncertainty using normal, lognormal, and gamma distributions, applying uncertainty propagation methods.

Results: The tool generates diagnostic measures, standard uncertainty, and confidence intervals estimates and provides their plots, supporting clinical decision-making. A case study using fasting plasma glucose data from the National Health and Nutrition Examination Survey in the USA showcases its application in diagnosing diabetes mellitus, highlighting the significant role of measurement uncertainty.

Conclusion: The software enhances the estimation and facilitates the comparison of Bayesian diagnostic measures, which are critical for medical practice. It provides a framework for analyzing uncertainty and assists in understanding and applying Bayes' theorem in medical diagnostics.

Keywords: Bayesian diagnosis; Bayes’ theorem; prevalence; prior probability; post-test probability; posterior probability; likelihood; positive predictive value; negative predictive value; parametric distribution; combined uncertainty; measurement uncertainty; sampling uncertainty; probability density function; disease; diabetes mellitus

2. Introduction

2.1.Medical Diagnosis

Diagnosis in medicine is fundamentally the process of identifying the unique characteristics of a disease through abduction, deduction, and induction (1). The term' diagnosis,' originating from the Greek

'διάγνωσις' meaning 'discernment' (2), underscores the critical role of distinguishing between healthy and diseased states in individuals. We can define diagnosis as the stochastic mapping of symptoms, signs, and laboratory and medical imaging findings onto a particular disease condition, derived from medical knowledge.

2.1.1. Threshold Based Diagnosis

We apply diagnostic tests or procedures to binary classification of individuals into diseased or nondiseased populations. The probability distributions of the measurands of a quantitative diagnostic test in these populations overlap. Despite this, we dichotomize the results by setting a diagnostic threshold or cut-off point (3). However, reliance on a single threshold for diagnosis across a spectrum of data points introduces uncertainty due to the overlapping probability distributions of the measurand in both nondiseased and diseased groups (4). Nevertheless, this dichotomous methodology signifies a substantial transformation in medical decision-making by correlating a continuum of evidence with binary clinical decisions, such as the decision to treat or not to treat. (5).

2.1.1.1. Diagnostic Accuracy Measures

To ensure patients' safety, the correctness of this classification must be rigorously evaluated. Among the numerous diagnostic accuracy measures (DAM) in the literature, only a few are routinely used for assessing the diagnostic accuracy of threshold-based diagnostic tests in clinical research and practice (6). These include positive and negative predictive values, which are defined conditionally on the test outcome and are prevalence-dependent.

2.1.2. Bayesian Diagnosis

In medical diagnostics, Bayes' theorem (7,8) is pivotal in transforming a disease's pre-test or prior probability into a post-test or posterior probability following diagnostic tests (4,7,9–12). This theorem connects the posterior probability P(H|E) of a hypothesis H being true given specific evidence E to the likelihood P(E|H) of observing the evidence E given that hypothesis H is true(13).

2.1.2.1. Bayesian Inference

In a purely Bayesian inference, we start with a prior distribution representing our initial beliefs about the parameters of interest before observing the evidence. This prior distribution is then updated with

likelihood functions (representing the evidence) using Bayes' theorem to obtain the posterior distribution, combining prior information with the new evidence (10).

2.1.2.1.1. Prior Distribution

Priors represent the beliefs held by researchers about parameters before seeing the evidence. They can be informative, weakly informative, or diffuse, depending on the level of certainty or uncertainty they reflect.

2.1.2.1.2. Likelihood Function

The likelihood function describes how probable the observed evidence is given different parameter values. It plays a crucial role in updating the prior distribution to form the posterior distribution.

2.1.2.1.3. Posterior Distribution

The posterior distribution is the result of combining the prior and likelihood functions. It reflects updated knowledge about the parameters after considering the observed evidence.

2.1.2.1.4. Workflow

The typical Bayesian workflow involves specifying the prior distribution, determining the likelihood function, and combining both using Bayes' theorem to obtain the posterior distribution. Model checking, refinement, and sensitivity analysis are essential steps in ensuring the robustness of Bayesian inferences provided by the evidence.

2.1.2.2. Empirical Bayesian Methods

The empirical Bayesian approach simplifies the purely Bayesian framework by using the available data to estimate the prior distribution, making it particularly practical when prior information is sparse or unavailable (14,15). Instead of specifying a fixed prior distribution, the empirical Bayesian method treats the prior as an unknown quantity to be estimated from the evidence, making it particularly suitable for medical diagnostics where real-time data integration is crucial.

2.1.2.2.1. Workflow

The typical empirical Bayesian workflow involves:

a) Collecting a large dataset and performing preliminary statistical analyses to understand the distributions and characteristics of the data,
b) Estimating prior distributions and probabilities using the empirical data through methods such as maximum likelihood,
c) Applying Bayes' theorem with the estimated prior distributions to compute the posterior probabilities, thereby incorporating the observed data.

2.2.Uncertainty

Uncertainty represents imperfect or incomplete information. When quantifiable, we can express it with probability (16). Our empirical Bayesian approach integrates frequentist methods for uncertainty quantification due to their established reliability and ease of implementation in clinical settings (17).

2.2.1. Measurement Uncertainty

Given the intrinsic variability of measurements, measurement uncertainty is defined as a 'parameter associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand'. This measurement uncertainty concept supplants the traditional notion of total analytical error (18).

2.2.2. Sampling Uncertainty

We derive diagnostic measures from screening or diagnostic tests applied to population samples. The variability within these samples contributes to their overall uncertainty (19). This intrinsic heterogeneity is present even when simple random sampling techniques are used (20).

2.2.3. Uncertainty of Diagnostic Accuracy Measures and Bayesian Posterior Probabilities

We have already explored the uncertainty of diagnostic accuracy measures and Bayes' theorem derived posterior probability for disease, which can significantly impact their clinical usefulness (21,22). Estimating, evaluating, and mitigating this uncertainty is critical in medical diagnosis.

2.3.Bayesian Diagnostic Measures

This project introduces a novel software tool designed for the parametric estimation and visualization of four diagnostic measures derived from Bayes' theorem, along with their associated uncertainty:

a) The positive predictive value and negative predictive value (11).

b) The posterior probability for disease and its complement, the posterior probability for the absence of disease.

To the best of our knowledge, this is the first publication comparing the four Bayesian diagnostic measures mentioned above and their uncertainty.

3. Methods

3.1.Calculations

3.1.1. Calculation of Bayesian Diagnostic Measures

Bayes' theorem relates the probability $P(H|E)$ of a hypothesis H being true given observed evidence E to the inverse probability $P(E|H)$ of observing E given that H is true, expressed as:

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}$$

$$= \frac{P(E|H)P(H)}{P(E|H)P(H) + P(E|\bar{H})P(\bar{H})}$$

$$= \frac{P(E|H)P(H)}{P(E|H)P(H) + P(E|\bar{H})\big(1 - P(H)\big)}$$

where $\bar{H}$ represents the negation of hypothesis H.

Bayes' theorem provides a robust framework for updating the probability of a hypothesis being true based on new evidence. In medical diagnostics, this means updating the probability for a disease given the results of diagnostic tests. By combining prior knowledge (pre-test probability) with new data (test results), Bayesian methods offer a comprehensive approach to the medical diagnostic process.

3.1.1.1. Positive and Negative Predictive Value

If D denotes the presence and $\overline{D}$ the absence of a disease, $F_D(x|\boldsymbol{\theta})$ the cumulative distribution function (CDF) of the test measurements in the presence of the disease, $F_{\overline{D}}(x|\,\boldsymbol{\theta})$ the CDF in the absence of the disease, and v the prevalence or the prior probability for disease, we can calculate the positive predictive value of a diagnostic test T for a diagnostic threshold t as:

$$P(D|T \geq t) = \frac{\big(1 - F_D(t|\boldsymbol{\theta})\big)v}{\big(1 - F_D(t|\boldsymbol{\theta})\big)v + \big(1 - F_{\overline{D}}(t|\boldsymbol{\theta})\big)(1 - v)}$$

and the negative predictive value as:

$$P(\overline{D}|T < t) = \frac{F_{\overline{D}}(t|\boldsymbol{\theta})(1 - v)}{\big(1 - F_{\overline{D}}(t|\boldsymbol{\theta})\big)(1 - v) + F_D(t|\boldsymbol{\theta})v}$$

In the above equations $1 - F_D(t|\boldsymbol{\theta})$ and $F_{\overline{D}}(t|\boldsymbol{\theta})$ are respectively the sensitivity and the specificity of the test.

3.1.1.2. Posterior Probability for Disease and Absence of Disease

Consequently, if $f_D(x|\boldsymbol{\theta})$ the probability density function (PDF) of the test measurements in the presence of the disease, $f_{\overline{D}}(x;\ \boldsymbol{\theta})$ the PDF in the absence of the disease, and v the prevalence or prior probability for disease, we calculate the posterior or post-test probability for disease of a diagnostic test T for a measurement value t as:

$$P(D|T = t) = \frac{f_D(t|\boldsymbol{\theta})v}{f_D(t|\boldsymbol{\theta})v + f_{\overline{D}}(t|\boldsymbol{\theta})(1 - v)}$$

and the posterior or post-test probability for the absence of disease as:

$$P(\overline{D}|T = t) = \frac{f_{\overline{D}}(t|\boldsymbol{\theta})(1 - v)}{f_{\overline{D}}(t|\boldsymbol{\theta})(1 - v) + f_D(t|\boldsymbol{\theta})v} = 1 - P(D|T = t)$$

3.1.2. Uncertainty Quantification

Uncertainty of input parameters can appear as standard uncertainty $u(t)$, representing the standard deviation of t, and expanded uncertainty $U(t)$, which defines a range around t with a probability p (23).

3.1.2.1. Measurement Uncertainty

Measurement uncertainty is estimated according to "Guide to the Expression of Uncertainty in Measurement" (GUM) (24) and "Expression of Measurement Uncertainty in Laboratory Medicine" (23). Bias is considered a component of this uncertainty (25). The relationship between the standard measurement uncertainty $u_m(t)$ to the value of the measurement t, is typically represented as (20):

$$u_m(t) = \sqrt{b_0^2 + b_1^2 t^2}$$

where b_0 and b_1 are constants.

For a linear approximation, it is expressed as (20):

$$u_m(t) \cong b_0 + b_1 t$$

3.1.2.2. Sampling Uncertainty of Means and Standard Deviations

Standard uncertainties in means and standard deviations are estimated utilizing the central limit theorem and the chi-square distribution (26–28) as:

$$u_s(m_P) \cong \frac{s_P}{\sqrt{n_P}}$$

$$u_s(s_P) \cong \frac{s_p}{\sqrt{2(n_P - 1)}}$$

where m_P and s_P are the mean and standard deviation of measurements in a population sample of size n_P.

3.1.2.3. Sampling Uncertainty of Prevalence or Prior Probability for Disease

Given the numbers n_D and $n_{\bar{D}}$ of diseased and nondiseased individuals in a population sample, the standard uncertainty of the prevalence or prior probability for disease $v = \frac{n_D}{n_{\bar{D}} + n_D}$ is approximated as:

$$u_s(v) \cong \sqrt{\frac{(2 + n_{\bar{D}})(2 + n_D)}{(4 + n_{\bar{D}} + n_D)^3}}$$

using the Agresti–Coull adjustment of the Waldo interval (29).

3.1.2.4. Measures Combined Uncertainty

When there are *l* independent and uncorrelated components of uncertainty, each with standard uncertainty $u_i(t)$, then their combined uncertainty ${}_l u_c(t)$ is calculated as (23):

$$_{l}u_c(t) = \sqrt{\sum_{i=1}^{l}(u_i(t))^2}$$

If the components are correlated, then (24):

$$_{l}u_c(t) = \sqrt{\sum_{i=1}^{l}\sum_{j=1}^{l}u_i(t)u_j(t)\rho_{ij}(t)}$$

where $\rho_{ij}(t)$ is the correlation coefficient between the uncertainties $u_i(t)$ and $u_j(t)$.

The standard combined uncertainty of the Bayesian diagnostic measures is computed via uncertainty propagation rules, employing a first-order Taylor series approximation (30) (refer to Supplemental File II: BayesianDiagnosticInsightsCalculations.nb). Assuming uncorrelated parameters, we use the following formula to compute uncertainty propagation (24):

$$_{l}u_c(t) = \sqrt{\sum_{i=1}^{l}\left(\frac{\partial g(t|\boldsymbol{\theta})}{\partial x_i}\right)^2 (u_i(t))^2}$$

where $g(t|\boldsymbol{\theta})$ a Bayesian diagnostic measure with a parameter vector $\boldsymbol{\theta} = (x_1, x_2, \dots, x_l)$, $_{l}u_c(t)$ the standard combined uncertainty of $g(t|\boldsymbol{\theta})$, and $u_i(t)$ the standard uncertainty of x_i at t.

3.1.2.5. Measures Expanded Uncertainty

The effective degrees of freedom $\nu_{eff}(t)$ for the combined standard uncertainty $_{l}u_c(t)$ with l components $u_i(t)$ with ν_i degrees of freedom each are determined using the Welch–Satterthwaite formula (31,32):

$$_{l}\nu_{eff}(t) \cong \frac{(_{l}u_c(t))^4}{\sum_{i=1}^{l}\frac{(u_i(t))^4}{\nu_i}}$$

where ν_i the respective degrees of freedom.

It can be shown that if ν_{min} the minimum of $\nu_1, \nu_2, \dots, \nu_l$, then :

$$\nu_{min} \leq _{l}\nu_{eff}(t) \leq \sum_{i=1}^{l}\nu_i$$

The expanded combined uncertainty $U_c(t)$ at a confidence level p is estimated as:

$$U_c(t) \cong \left(F_v^{-1}\left(\frac{1-p}{2}\right)_l u_c(t), F_v^{-1}\left(\frac{1+p}{2}\right)_l u_c(t)\right)$$

where $F_v(z)$ is the CDF of Student's *t*-distribution with v degrees of freedom and ${}_l u_c(t)$ the standard combined uncertainty of a Bayesian diagnostic measure.

Consequently, the confidence interval of t at the same confidence level p is approximated as:

$$CI_p(t) \cong \left(x + F_v^{-1}\left(\frac{1-p}{2}\right)_l u_c(t), x + F_v^{-1}\left(\frac{1+p}{2}\right)_l u_c(t)\right)$$

The confidence intervals of the Bayesian diagnostic measures are truncated to the [0,1] range.

3.2.The Software

3.2.1. Program Overview

The software program *Bayesian Diagnostic Insights* was developed in Wolfram Language, using Wolfram Mathematica® Ver 14.1 (Wolfram Research, Inc., Champaign, IL, USA), to facilitate the estimation and comparison of Bayesian diagnostic measures. This interactive program was designed to estimate and plot the values, the standard sampling, measurement, and combined uncertainty, and the confidence intervals of Bayesian diagnostic measures for a screening or diagnostic test (refer to Figures 1 and 2).

The program is freely accessible as a Wolfram Language notebook (.nb) (Supplemental File I: BayesianDiagnosticInsights.nb). It can be executed on Wolfram Player® or Wolfram Mathematica® (refer to Appendix A.3). Given the intricate nature of the required calculations, it necessitates substantial

computational resources.

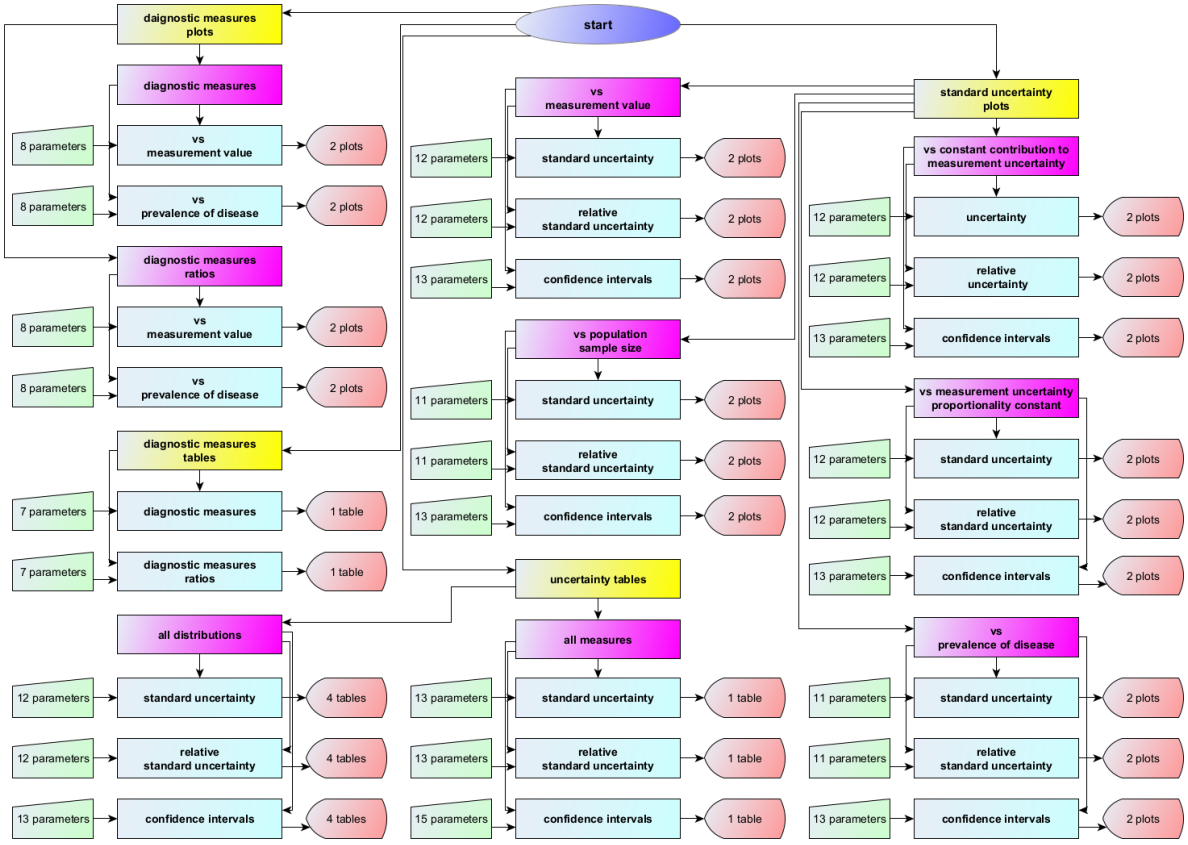

Figure 1. A simplified flowchart of the program *Bayesian Diagnostic Insights*.

3.2.2. Input Parameters

3.2.2.1. Parametric Distributions

Users select the distributions of the measurements for a diseased and nondiseased population from a predefined list of parametric distributions:

a) Normal distribution

b) Lognormal distribution

c) Gamma distribution.

3.2.2.2. Bayesian Diagnostic Measures

Users select the Bayesian diagnostic measures to be evaluated among the following:

a) The positive predictive value $P(D|T \geq t)$

b) The negative predictive value $P(\overline{D}|T < t)$

c) The posterior probability for disease $P(D|T = t)$

d) The posterior probability for the absence of disease $P(\overline{D}|T = t)$

3.2.2.3. Definition of Populations and Samples Parameters and Statistics

For each population, users define the mean μ and the standard deviation σ of the measurements, along with the prior probability or prevalence of disease v. The parameters μ and σ are specified in arbitrary units.

For each population sample, users define its size n, the mean m, and the standard deviation s of the measurements. The statistics m and s are also specified in arbitrary units.

3.2.2.4. Measurement Uncertainty

Users select a linear or nonlinear equation of the measurement uncertainty versus the value t of the measurement. They define the constant contribution b_0 to the standard measurement uncertainty, the proportionality constant b_1, and the number of quality control samples analyzed for its estimation.

For more details about the program's input, please refer to Appendix A2: Input.

3.2.3. Output

The program generates plots and tables detailing diagnostic measures, including their standard sampling, measurement, and combined uncertainty, and associated confidence intervals. By providing this extensive array of input parameters, output plots, and tables, the program presents a robust platform for exploring and comparing Bayesian diagnostic measures and their uncertainties, utilizing parametric distributions of medical diagnostic measurands.

We present more detailed documentation of the interface of the program in Supplemental file III: BayesianDiagnosticInsightsInterface.pdf

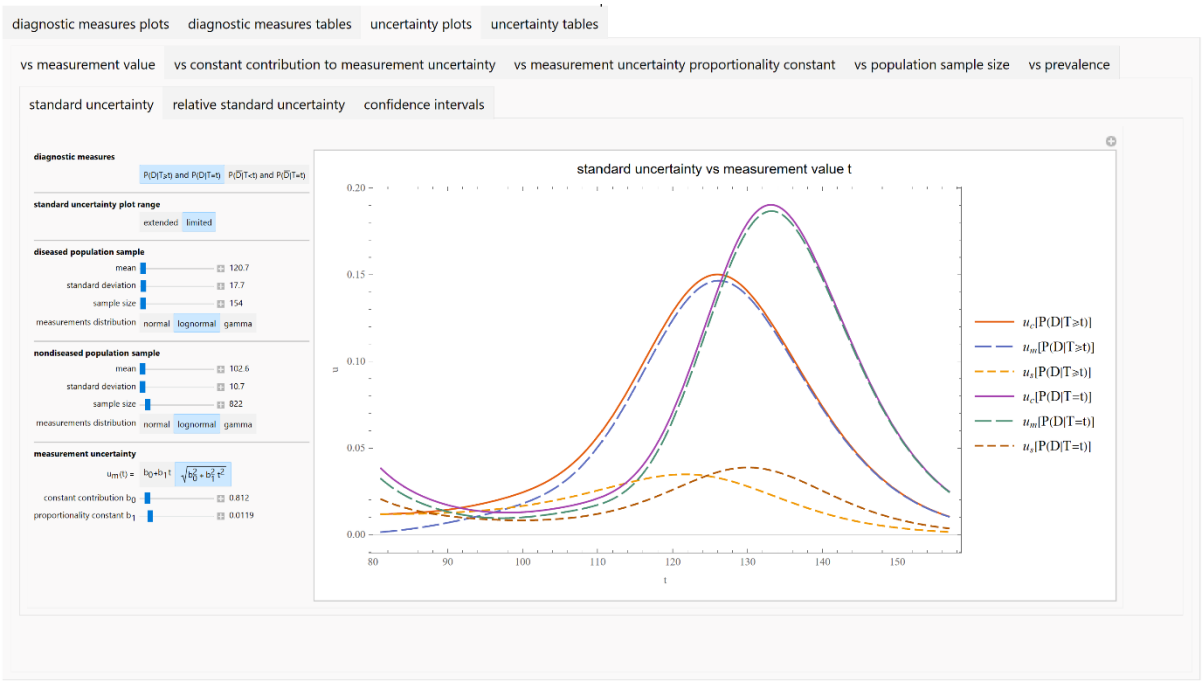

Figure 2. A screenshot of the program *Bayesian Diagnostic Insights*.

3.3.Illustrative Case Study

As previously described, we completed an illustrative case study to demonstrate the program's application (22). We used fasting plasma glucose (FPG) as the diagnostic test measurand for the Bayesian diagnosis of diabetes mellitus (hereafter referred to as "diabetes"), where the oral glucose tolerance test (OGTT) served as the reference method. Diabetes diagnosis was confirmed if the plasma glucose value was equal to or greater than 200 $\mathrm{mg/dl}$, measured two hours after 75 g of glucose oral administration (29) during an OGTT (2-h PG). The study focused on individuals aged 70 to 80 years, reflecting the significant correlation between age and diabetes prevalence (33).

Data from the National Health and Nutrition Examination Survey (NHANES) was collected from participants from 2005 to 2016 (n = 60,936), as described previously (22). NHANES is a comprehensive survey assessing adults' and children's health and nutritional status in the United States (34).

Inclusion criteria were valid FPG and OGTT results (n = 13,836), no prior diabetes diagnosis (35) (n = 13,465), and age 70–80 years (n = 976).

Participants with a 2-h PG measurement ≥200 $\mathrm{mg/dl}$ were classified as diabetic (n = 154).

The prevalence or prior probability for diabetes, along with the probability distributions for fasting plasma glucose (FPG) in both diabetic and nondiabetic participants, were estimated using empirical Bayes' methods (36).

We estimated the prevalence or prior probability for diabetes as follows:

$$v \cong \frac{154}{976} = 0.158$$

We present the FPG datasets statistics in Table 1 (hereafter, FPG and its uncertainty are expressed in $\mathrm{mg/dl}$).

Table 1. Descriptive statistics of the datasets and the estimated lognormal distributions of the diabetic and nondiabetic participants.

	Diabetic Participants			**Nondiabetic Participants**		
	Dataset	L_D	l_D	Dataset	$L_{\bar{D}}$	$l_{\bar{D}}$
n	154	-	-	822	-	-
Mean (mg/dl)	120.7	120.7	120.7	102.6	102.6	102.6
Median (mg/dl)	117.0	119.4	118.1	102.0	102.1	101.5
Standard Deviation (mg/dl)	19.1	17.8	17.7	10.9	10.7	10.7
Mean uncertainty (mg/dl)	1.665	1.665	0	1.473	1.473	0
Skewness	1.448	0.446	0.448	0.523	0.315	0.314
Kurtosis	6.354	3.355	3.360	3.445	3.177	3.176
p-value (Cramér–von Mises test)	-	0.294	0.562	-	0.281	0.260

Lognormal distributions were employed to model FPG measurements in diabetic and nondiabetic participants using the maximum likelihood estimation method (37). Parametrized for their means m_D and $m_{\bar{D}}$, and standard deviations s_D and $s_{\bar{D}}$, were defined as:

$$L_D = Lognormal(m_D, s_D) = Lognormal(120.671, 17.791)$$

$$L_{\bar{D}} = Lognormal(m_{\bar{D}}, s_{\bar{D}}) = Lognormal(102.642, 10.747)$$

Quality control data for FPG measurements from NHANES for the same period (2005–2016) included 1350 QC samples. Nonlinear least squares regression (38,39) applied to the QC data provided the following function for standard measurement uncertainty $u_m(t)$ relative to the measurement value t:

$$u_m(t) = \sqrt{b_0^2 + b_1^2 t^2} = \sqrt{0.6600 + 0.00014t^2}$$

where $b_0 = 0.8124$ and $b_1 = 0.0119$.

We estimated the means of the standard measurement uncertainty of FPG in the diabetic and nondiabetic participants as follows:

$$\hat{u}_D \cong 1.665 \text{ mg/dl}$$

$$\hat{u}_{\bar{D}} \cong 1.473 \text{ mg/dl}$$

Consequently, we estimated the distributions of the measurements, assuming negligible measurement uncertainty, as:

$$d_D \cong Lognormal\left(m_D, \sqrt{s_D^2 - \hat{u}_D^2}\right) \cong Lognormal(120.671, 17.713)$$

$$d_{\bar{D}} \cong Lognormal\left(m_{\bar{D}}, \sqrt{s_{\bar{D}}^2 - \hat{u}_{\bar{D}}^2}\right) \cong Lognormal(102.642, 10.747)$$

Table 1 presents the descriptive statistics of the estimated lognormal distributions of the diabetic and nondiabetic participants and the respective p-values in the Cramér–von Mises goodness-of-fit test (40). This test compares the samples' empirical CDFs with the estimated distributions' CDFs. The *p*-values of the table show that the observed differences between the samples of the FPG measurements and the estimated lognormal distributions can be attributed to random sampling variability.

Figures 3 and 4 show the estimated PDFs of FPG in the diabetic and nondiabetic participants, assuming a lognormal distribution and negligible measurement uncertainty, and the histograms of the respective NHANES datasets.

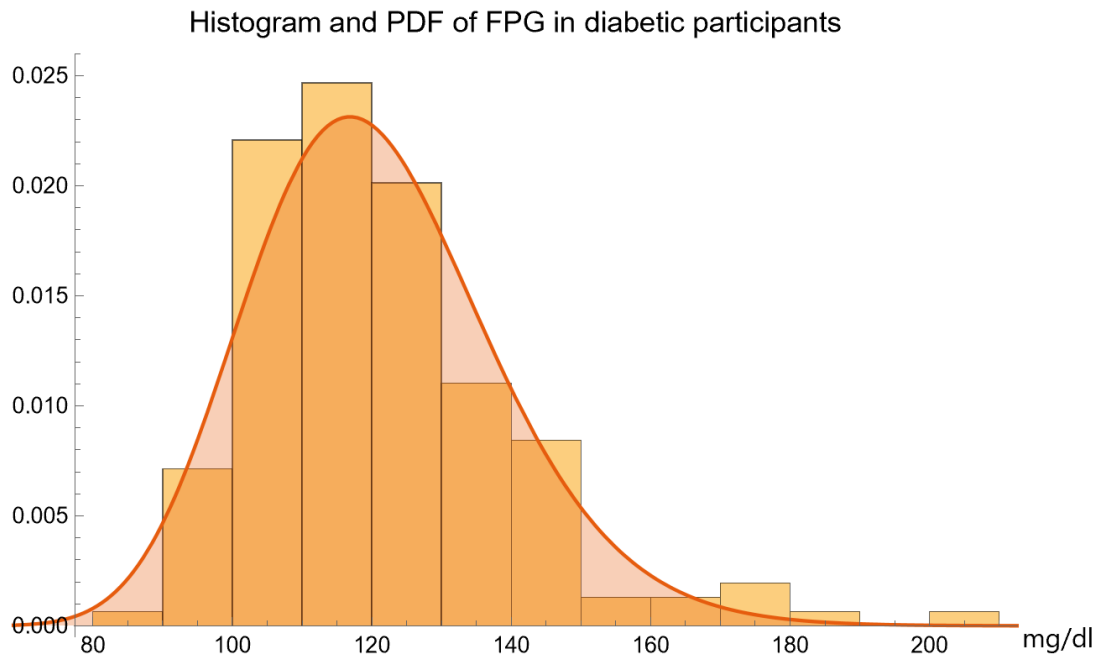

Figure 3. The estimated PDF of the FPG (mg/dl) in diabetic participants, assuming a lognormal distribution and negligible measurement uncertainty, and the histogram of the respective NHANES dataset, with the distribution parameters in Table 2.

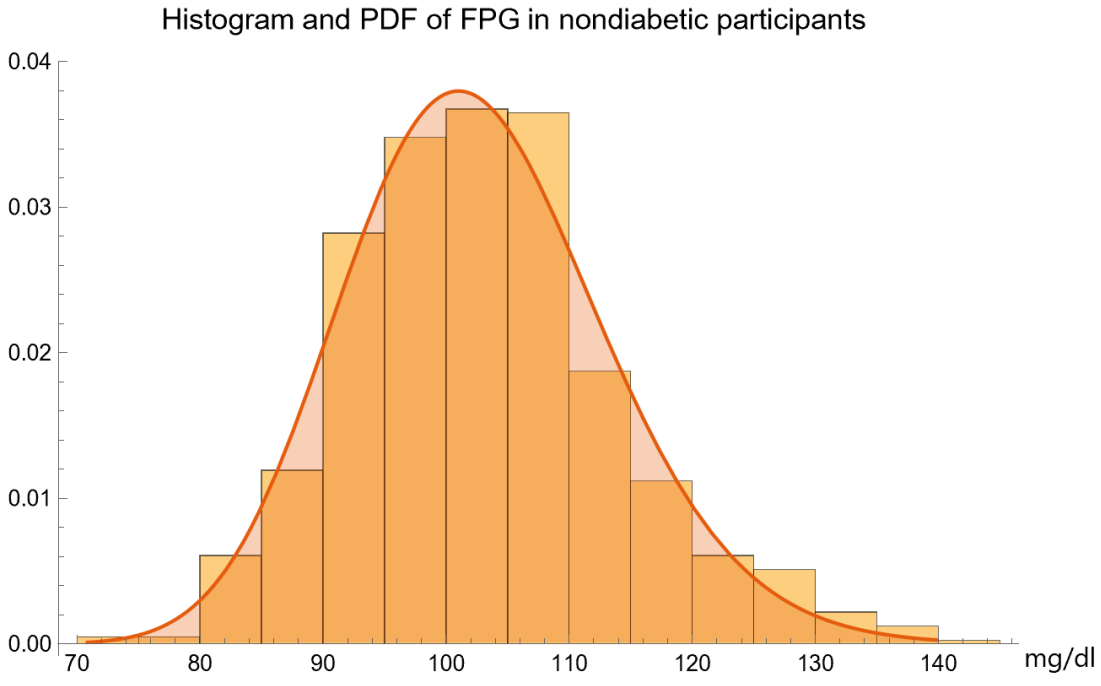

Figure 4. The estimated PDF of the FPG (mg/dl) in nondiabetic participants, assuming a lognormal distribution and negligible measurement uncertainty, and the histogram of the respective NHANES dataset, with the parameters of the distribution in Table 2.

Likelihoods and posterior probabilities were estimated accordingly.

4. Results

The results of the program's application on are presented in Figures 5-19, and the program settings are detailed in Tables 2 and 3.

4.1.Measures

Table 2. The settings of the program *Bayesian Diagnostic Insights* for the Figures 5-9

	Units	**Figures 5-6**	**Figures 7-8**	**Figure 9**
t	mg/dl	32.0– 210.0	126	126
μ_D	mg/dl	120.7	120.7	120.7
σ_D	mg/dl	17.7	17.7	17.7
$\mu_{\overline{D}}$	mg/dl	102.6	102.6	102.6
$\sigma_{\overline{D}}$	mg/dl	10.7	10.7	10.7
v		0.158	0.001-0.999	0.158
d_D		lognormal	lognormal	normal lognormal gamma
$d_{\overline{D}}$		lognormal	lognormal	normal lognormal gamma

Figure 5 shows the plots of:

a) The positive predictive value $P(D|T \geq t)$ of FPG for diabetes versus threshold value t $(\mathrm{mg/dl})$, (orange curve). The curve is smooth, increasing monotonically, and approximately sigmoidal. $P(D|T \geq t)$ is asymptotically equal to the prevalence of diabetes for lower values of t, then rises rapidly to become asymptotically equal to 1.00.

b) The posterior probability for diabetes versus FPG value t $(\mathrm{mg/dl})$. The curve is smooth, approximately double sigmoidal. For $t = 86.8$ mg/dl $P(D|T = t)$ has a minimum value of 0.04. $P(D|T = t)$ is asymptotically equal to 1.00 for lower values of t, then decreases rapidly to its minimum before rising rapidly again to become asymptotically equal to 1.00.

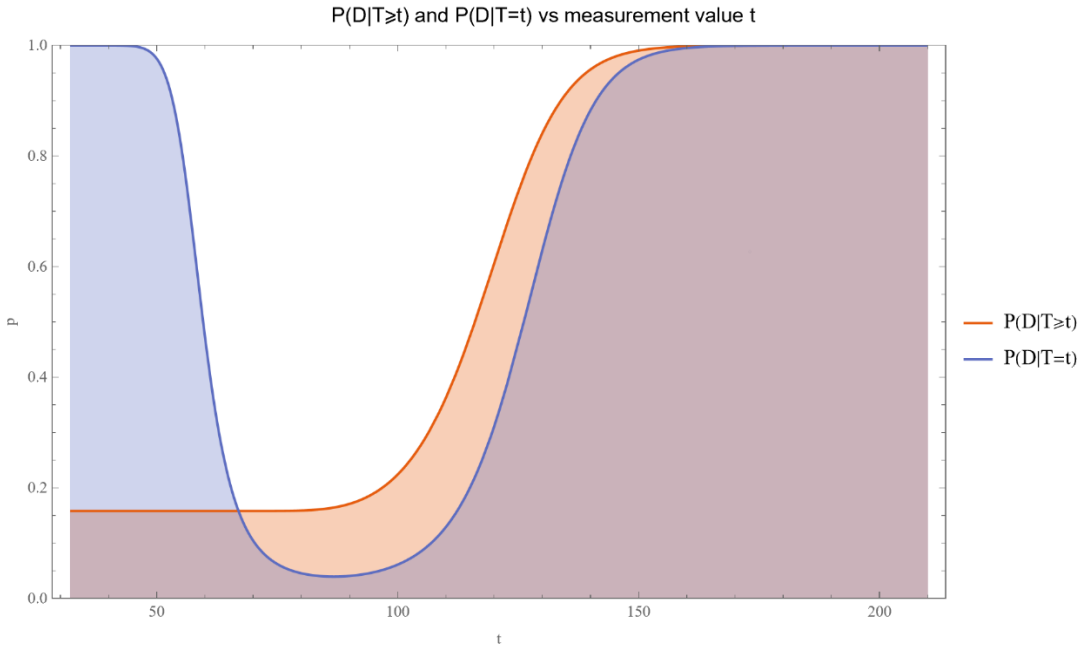

Figure 5. Positive predictive value and posterior probability for diabetes versus FPG value t $(\mathrm{mg/dl})$ curves plot, with the program's settings in Table 2.

Figure 6 shows the plots of:

a) The negative predictive value $P(\bar{D}|T < t)$ of FPG for diabetes versus threshold value t $(\mathrm{mg/dl})$ (orange curve). The curve is smooth and unimodal, with a maximum value of 0.96 at $t = 91.3$ $\mathrm{mg/dl}$. $P(\bar{D}|T < t)$ is asymptotically equal to 0.00 for lower values of t, then rises rapidly to its maximum and becomes asymptotically equal to $1.00 - v$, where v the prevalence of diabetes.

b) The posterior probability $P(\bar{D}|T = t)$ for the absence of diabetes versus FPG value t $(\mathrm{mg/dl})$ (orange curve). The curve is smooth, unimodal, and approximately double sigmoidal. For an FPG value $t = 86.8$ $\mathrm{mg/dl}$, $P(\bar{D}|T = t)$ has a maximum value of 0.96. $P(\bar{D}|T = t)$ is asymptotically equal to 0.00 for lower and higher values of t.

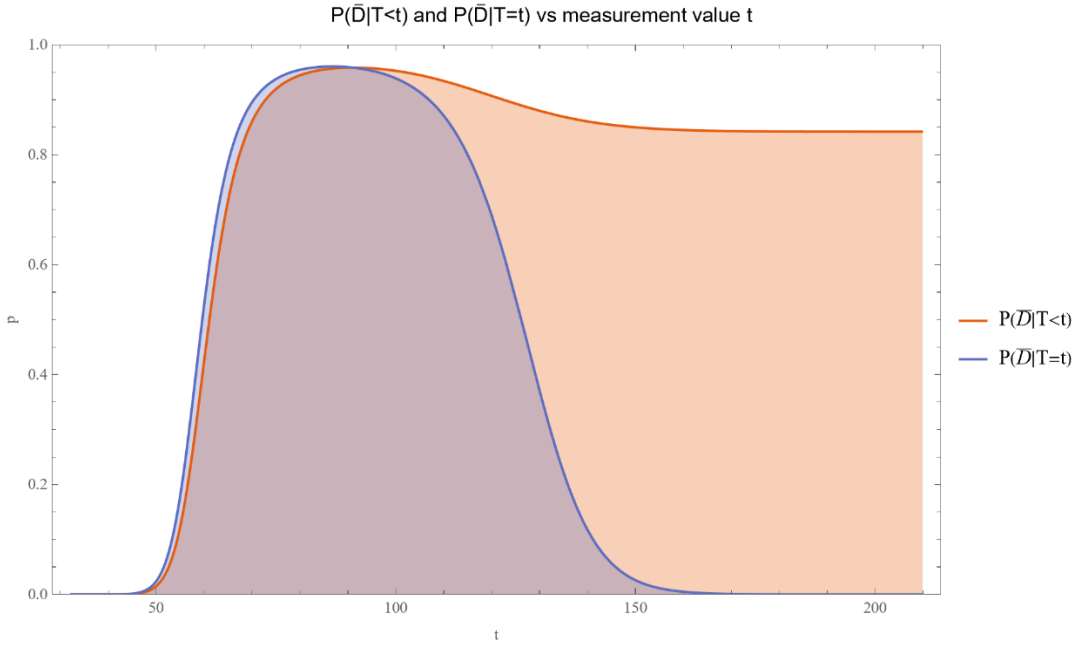

Figure 6. Negative predictive value for diabetes and posterior probability for the absence of diabetes versus FPG value t $(\mathrm{mg/dl})$ curves plot, with the program's settings in Table 2.

Moreover:

a) For $t = 67.4\ \mathrm{mg/dl}$, we have $P(D|T \geq t) = P(D|T = t) = 0.158 = v$

b) For $t < 67.4\ \mathrm{mg/dl}$, we have $P(D|T \geq t) < P(D|T = t)$,

c) For $t > 67.4\ \mathrm{mg/dl}$, we have $P(D|T \geq t) > P(D|T = t)$.

d) For $t = 91.0\ \mathrm{mg/dl}$, we have $P(\bar{D}|T < t) = P(\bar{D}|T = t) = 0.96$.

e) For $t < 91.0\ \mathrm{mg/dl}$, we have $P(\bar{D}|T < t) < P(\bar{D}|T = t)$

f) For $t > 91.0\ \mathrm{mg/dl}$, we have $P(\bar{D}|T < t) > P(\bar{D}|T = t)$.

Additionally, as Figures 7 and 8 show, for an FPG value $t = 126.0\ \mathrm{mg/dl}$ and for prevalence $0.0 < v < 1.0$:

a) Both $P(D|T \geq t)$ and $P(D|T = t)$ curves are smooth, starting from a probability asymptotically equal to 0.00, monotonically increasing as prevalence increases.

b) Both $P(\bar{D}|T < t)$ and $P(\bar{D}|T = t)$ curves are smooth, starting from a probability asymptotically equal to 1.00, monotonically decreasing as prevalence increases.

c) $P(D|T \geq t) > P(D|T = t)$, and

d) $P(\bar{D}|T < t) > P(\bar{D}|T = t)$.

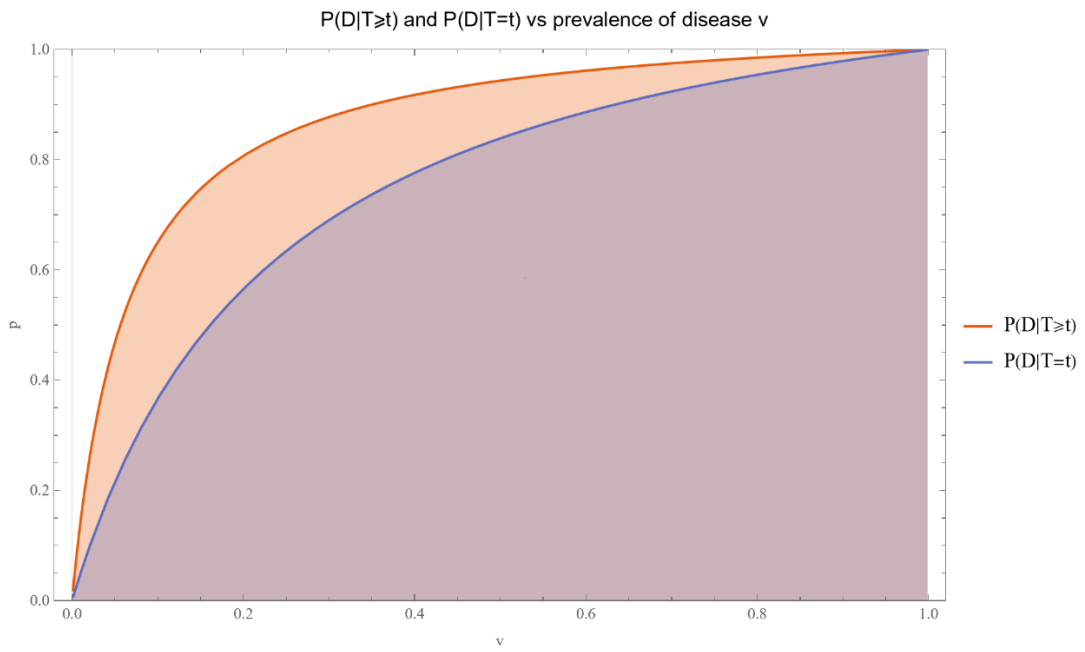

Figure 7. Positive predictive value and posterior probability for diabetes versus prior probability for diabetes v curves plot for an FPG value $t = 126$ mg/dl, with the other program settings in Table 2.

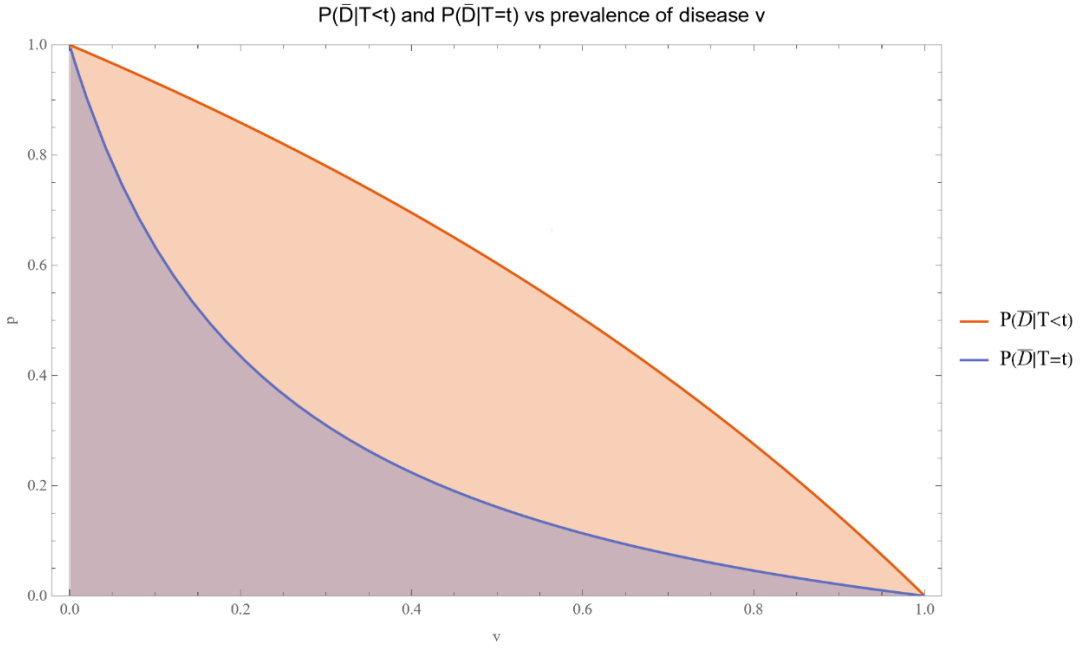

Figure 8. Negative predictive value for diabetes and posterior probability for the absence of diabetes versus prior probability for diabetes v curves plot, for an FPG value $t = 126$ mg/dl, with the other settings of the program in Table 2.

Figure 9 shows a table of the Bayesian diagnostic measures for an FPG value $t = 126$ mg/dl, the established threshold for the diagnosis of diabetes (41), assuming normal, lognormal, and gamma distributions of FPG.

diagnostic measures					
measurements distribution		measure			
diseased	nondiseased	P(D|T≥t)	P(D|T=t)	P($\bar{D}$|T<t)	P($\bar{D}$|T=t)
normal	normal	0.833	0.542	0.895	0.458
	lognormal	0.771	0.509	0.894	0.491
	gamma	0.790	0.516	0.894	0.484
lognormal	normal	0.823	0.527	0.891	0.473
	lognormal	0.758	0.494	0.890	0.506
	gamma	0.778	0.501	0.890	0.499
gamma	normal	0.827	0.532	0.892	0.468
	lognormal	0.763	0.498	0.892	0.502
	gamma	0.783	0.505	0.892	0.495

Figure 9. Table of positive predictive value, posterior probability, and negative predictive value for diabetes, and posterior probability for the absence of diabetes, for an FPG value $t = 126$ mg/dl, with the other settings of the program in Table 2.

4.2.Uncertainty

Table 3. The settings of the program *Bayesian Diagnostic Insights* for the Figures 10-19

	Units	Figures 10-11	Figures 12-13	Figures 14-15	Figures 16-17	Figure 18	Figure 19
p		-	0.95	-	0.95	-	0.95
t	mg/dl	32.0– 210.0	32.0– 210.0	126.0	126.0	126.0	126.0
m_D	mg/dl	120.7	120.7	120.7	120.7	120.7	120.7
s_D	mg/dl	17.7	17.7	17.7	17.7	17.7	17.7
n_D		154	154	-	-	154	154
$m_{\bar{D}}$	mg/dl	102.6	102.6	102.6	102.6	102.6	102.6
$s_{\bar{D}}$	mg/dl	10.7	10.7	10.7	10.7	10.7	10.7
$n_{\bar{D}}$		822	822	-	-	822	822
n		976	976	976	976	976	976
v		0.158	0.158	0.001-0.999	0.001-0.999	0.158	0.158
b_0		0.812	0.812	0.812	0.812	0.812	0.812

b_1	0.0119	0.0119	0.0119	0.0119	0.0119	0.0119
n_U	-	1350	-	1350	-	1350
d_D	lognormal	lognormal	lognormal	lognormal	lognormal	lognormal
$d_{\overline{D}}$	lognormal	lognormal	lognormal	lognormal	lognormal	lognormal

Figure 10 shows the plots of:

a) The standard sampling, measurement, and combined uncertainty of the positive predictive value for diabetes versus FPG value t $(\mathrm{mg/dl})$. The curves are smooth and unimodal.

b) The standard sampling, measurement, and combined uncertainty of the posterior probability for diabetes versus FPG value t $(\mathrm{mg/dl})$. The curves are smooth and bimodal.

Figure 11 shows the plots of:

a) The standard sampling, measurement, and combined uncertainty of the negative predictive value for diabetes versus FPG value t $(\mathrm{mg/dl})$. The curves are smooth and unimodal.

b) The standard sampling, measurement, and combined uncertainty of the posterior probability for the absence of diabetes versus FPG value t $(\mathrm{mg/dl})$. The curves are smooth and bimodal.

In the assessment of the combined standard uncertainty of posterior probability for diabetes $u_c[P(D|T=t)]$ and absence of diabetes $u_c[P(\overline{D}|T=t)]$:

a) They are equal.

b) They are substantially affected by the measurement uncertainty of FPG.

c) Two local maxima are observed, corresponding to the regions near the steepest segments of the posterior probability curves, which display an approximately double sigmoidal configuration. The maxima are quantitatively approximated as follows:

   a. At an FPG value of $t = 58.5\ \mathrm{mg/dl}$, the combined standard uncertainty is 0.898, for $P(D|T=t) = 0.581$, where $P(\overline{D}|T=t) = 0.419$.

   b. At an FPG value of $t = 133.1\ \mathrm{mg/dl}$, the combined standard uncertainty is 0.190, where $P(D|T=t) = 0.726$ and $P(\overline{D}|T=t) = 0.274$.

   c. The standard combined uncertainty $u_c[P(D|T \geq t)]$ of the positive predictive value for diabetes of FPG has a maximum value of 0.150 for $t = 126.0\ \mathrm{mg/dl}$, where $P(D|T \geq t) =$

0.758, while the standard combined uncertainty $u_c[P(\overline{D}|T < t)]$ of the negative predictive value for diabetes has a maximum value of 0.900 for $t = 58.5$ mg/dl, where $P(\overline{D}|T < t) = 0.321$. This local maxima pattern indicates heightened uncertainty in the regions where the diagnostic measures curves demonstrate their most pronounced inflections (refer to Figures 5 and 6).

In addition:

a) For $t = 91.6$mg/dl, we have $u_c[P(D|T \geq t)] = u_c[P(D|T = t)] = 0.015$, while $P(D|T \geq t) = 0.175$ and $P(D|T = t) = 0.042$.

b) For $t = 126.8$ mg/dl, we have $u_c[P(D|T \geq t)] = u_c[P(D|T = t)] = 0.150$, while $P(D|T \geq t) = 0.776$ and $P(D|T = t) = 0.520$.

c) For $0 < t < 91.6$ mg/dl and $t > 126.8$ we have $u_c[P(D|T \geq t)] < u_c[P(D|T = t)]$.

d) For 91.6 mg/dl $< t < 126.8$ mg/dl we have $u_c[P(D|T = t)] < u_c[P(D|T \geq t)]$

e) For $t = 59.1$ mg/dl, we have $u_c[P(\overline{D}|T < t)] = u_c[P(\overline{D}|T = t)] = 0.887$, while $P(\overline{D}|T < t) = 0.362$ and $P(\overline{D}|T = t) = 0.463$.

f) For $t = 103.8$ mg/dl, we have $u_c[P(\overline{D}|T < t)] = u_c[P(\overline{D}|T = t)] = 0.015$, while $P(\overline{D}|T < t) = 0.947$ and $P(\overline{D}|T = t) = 0.921$.

g) For $0 < t < 59.1$ mg/dl and $103.8 < t$ we have $u_c[P(\overline{D}|T < t)] < u_c[P(\overline{D}|T = t)]$.

h) For 59.1 mg/dl $< t < 103.8$ mg/dl we have $u_c[P(\overline{D}|T = t)] < u_c[P(\overline{D}|T < t)]$.

The confidence intervals are affected accordingly (refer to Figures 12 and 13):

a) The confidence intervals of positive predictive value $P(D|T = t)$ (blue curves) are narrower for lower and higher values of t.

b) The confidence intervals of Bayesian posterior probability $P(D|T \geq t)$ (orange curves) narrow considerably for lower values of t.

c) The confidence intervals of Bayesian posterior probability $P(\overline{D}|T = t)$ (blue curves) are wider at the extremes of the t spectrum.

d) The confidence intervals of negative predictive value $P(\overline{D}|T < t)$ (orange curves) are wide at lower t values, to become considerably narrower at higher values.

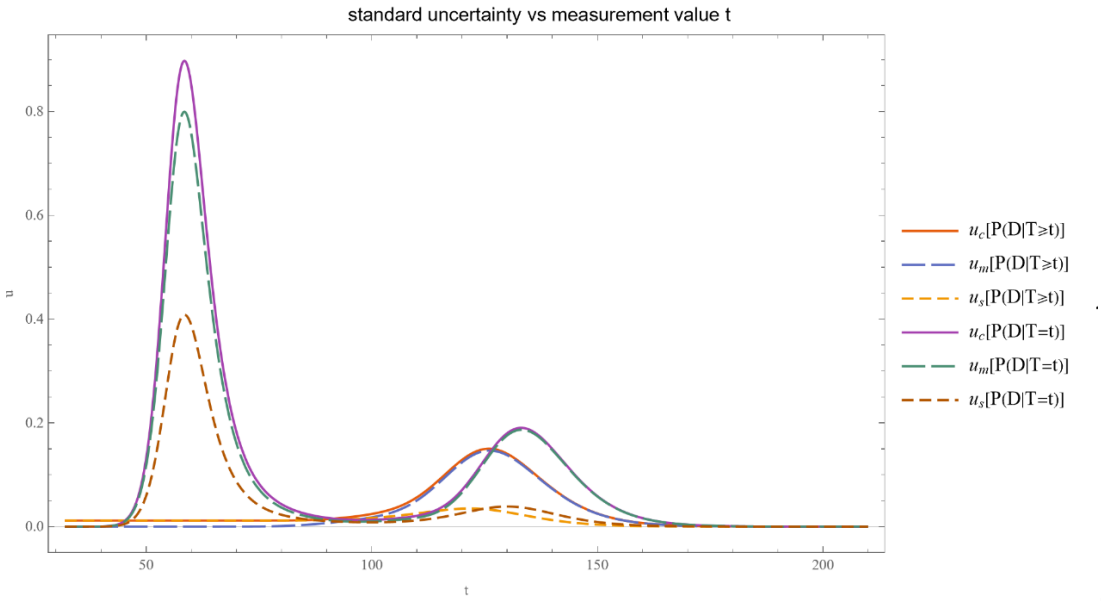

Figure 10. Standard sampling, measurement, and combined uncertainty of the positive predictive value and posterior probability for diabetes versus FPG value t $(\mathrm{mg/dl})$ curves plot, with the program's settings in Table 3.

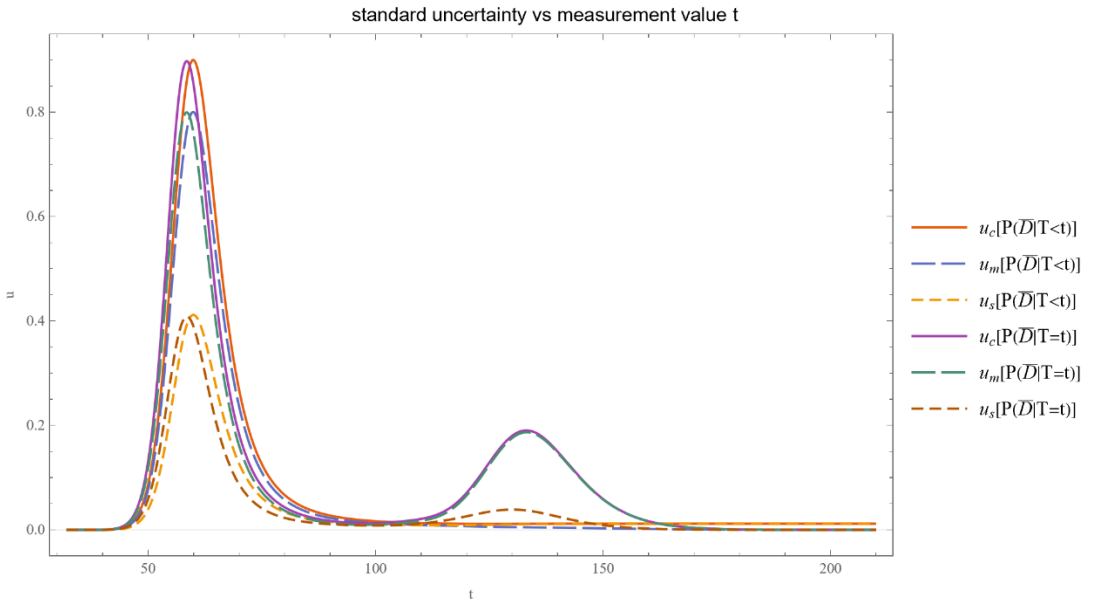

Figure 11. Standard sampling, measurement, and combined uncertainty of the negative predictive value for diabetes and posterior probability for the absence of diabetes versus FPG value t $(\mathrm{mg/dl})$ curves plot, with the program's settings in Table 3.

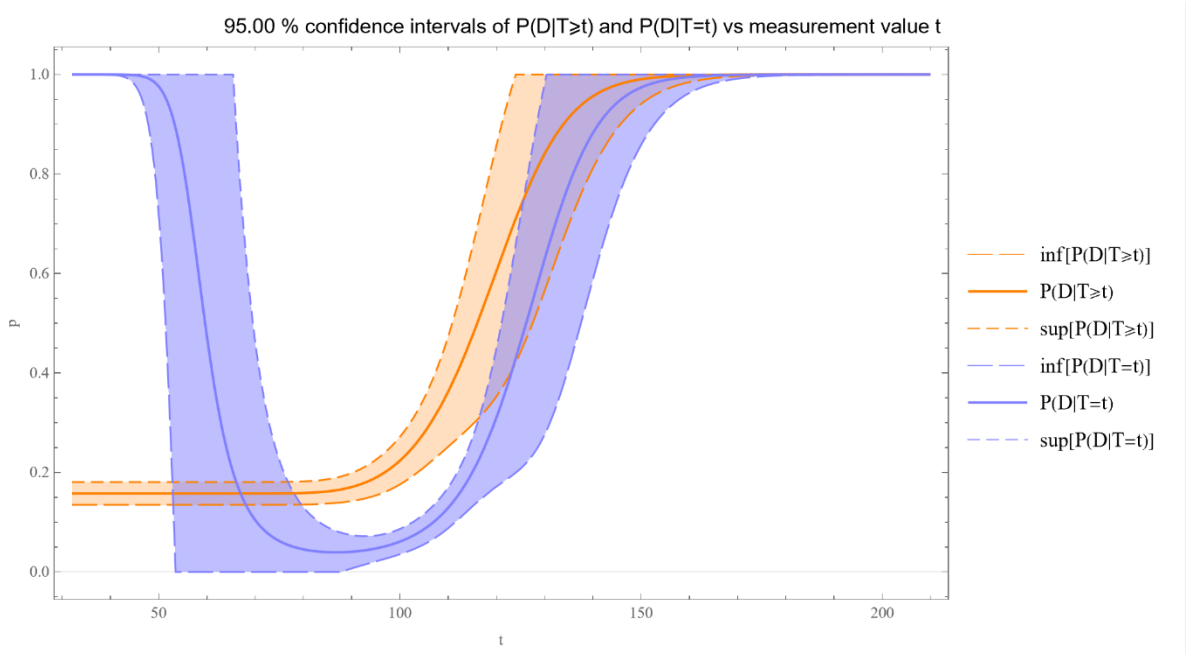

Figure 12. Confidence intervals of the positive predictive value and posterior probability for diabetes versus FPG value t $(\mathrm{mg/dl})$ curves plot, with the program's settings in Table 3.

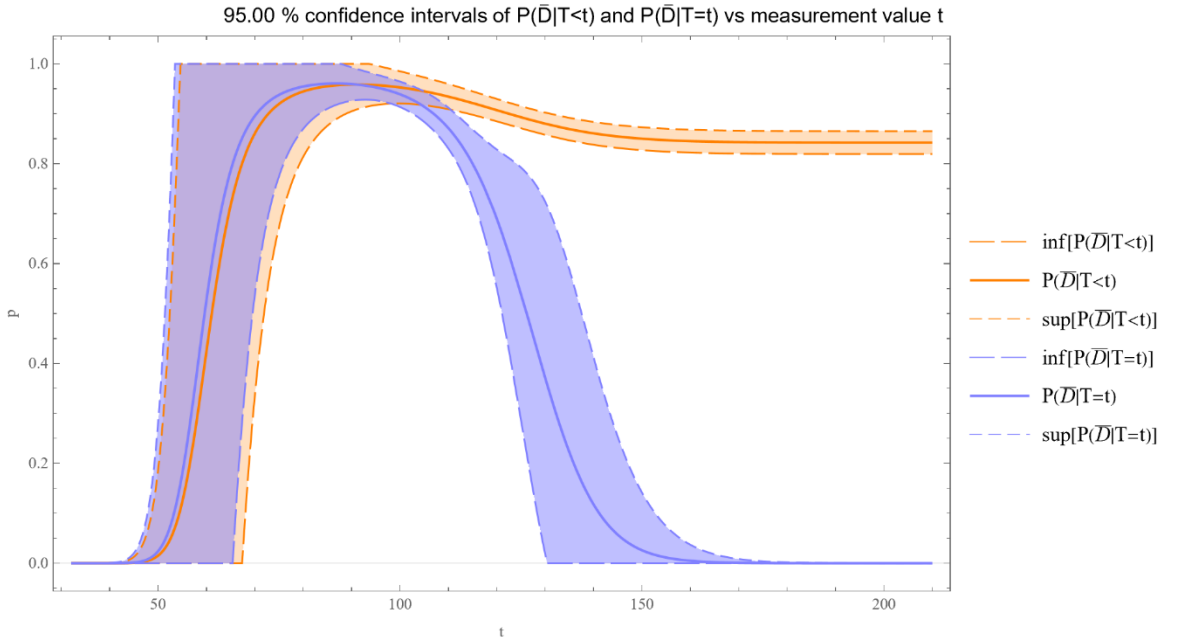

Figure 13. Confidence intervals of the negative predictive value and posterior probability for the absence of diabetes versus FPG value t $(\mathrm{mg/dl})$ curves plot, with the program's settings in Table 3.

For an FPG value $t = 126$ mg/dl, Figures 14 and 15 show the plots of the standard sampling, measurement, and combined uncertainty of positive predictive value, the posterior probability for

diabetes, the negative predictive value, and the posterior probability for the absence of diabetes versus prevalence or prior probability for diabetes v. The combined uncertainty of the diagnostic measures is substantially affected by the measurement uncertainty of FPG. The curves are unimodal, with the respective maxima quantitatively approximated as follows:

a) For $v = 0.054,\ u_c[P(D|T \geq t)] = 0.208$ where $P(D|T \geq t) = 0.488$.

b) For $v = 0.158,\ u_c[P(D|T = t)] = 0.141$ where $P(D|T = t) = 0.494$.

c) For $v = 0.631,\ u_c[P(\overline{D}|T < t)] = 0.023$ where $P(\overline{D}|T < t) = 0.471$.

d) For $v = 0.158,\ u_c[P(\overline{D}|T = t)] = 0.141$ where $P(\overline{D}|T = t) = 0.506$.

The local maxima indicate heightened uncertainty in the regions where the respective diagnostic measures curves demonstrate their most pronounced inflections (refer to Figures 7 and 8).

Additionally:

a) For $v = 0.175$ we have $u_c[P(D|T \geq t)] = u_c[P(D|T = t)] = 0.140$, and $P(D|T \geq t) = 0.780$ and $P(D|T = t) = 0.525$.

b) For $0 < v < 0.175$ we have $u_c[P(D|T \geq t)] > u_c[P(D|T = t)]$.

c) For $0.175 < v < 1.0$ we have $u_c[P(D|T \geq t)] < u_c[P(D|T = t)]$.

d) For $0 < v < 1.0$ we have $u_c[P(\overline{D}|T < t)] < u_c[P(\overline{D}|T = t)]$.

Remarkably, the combined uncertainty of the negative predictive value is considerably less than the combined uncertainty of the posterior probability for the absence of diabetes.

The confidence intervals are adjusted accordingly (refer to Figures 16-17):

a) The confidence intervals of Bayesian posterior probability $P(D|T = t)$ for diabetes (Figure 16, blue curves), positive predictive value $P(D|T \geq t)$ (Figure 16, blue curves), Bayesian posterior probability $P(\overline{D}|T = t)$ for the absence of diabetes (Figure 17, blue curves) and negative predictive value $P(\overline{D}|T < t)$ (Figure 17, orange curves) are narrowest at both lower and higher prevalences.

b) The confidence intervals of $P(D|T \geq t)$ (Figure 16, orange curves) are generally narrower than the confidence intervals of $P(D|T = t)$ (Figure 16, blue curves).

c) The confidence intervals of $P(\overline{D}|T < t)$(Figure 17, orange curves) are considerably narrower than the confidence intervals of $P(\overline{D}|T = t)$ (Figure 17, blue curves).

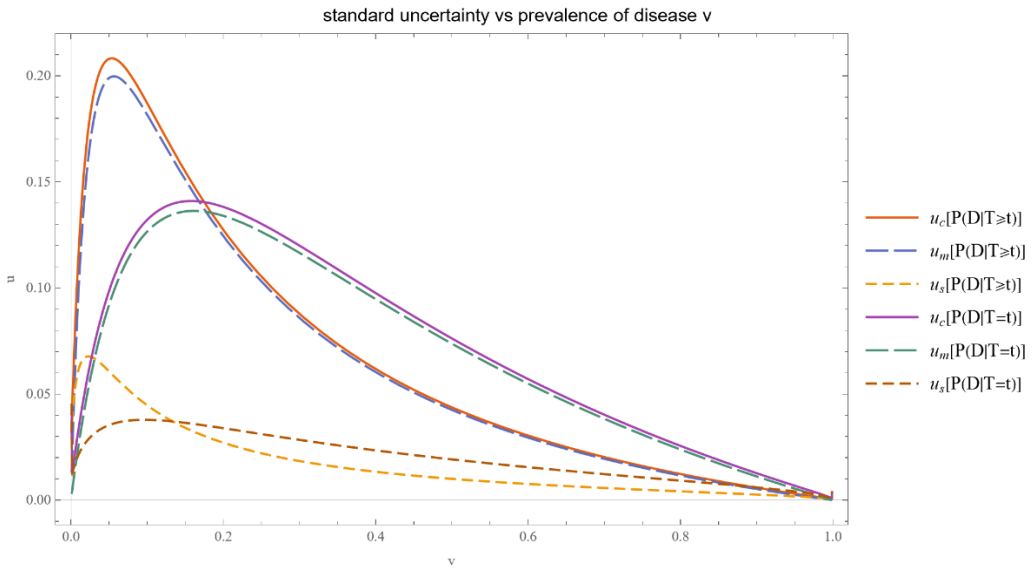

Figure 14. Standard sampling, measurement, and combined uncertainty of the positive predictive value and posterior probability for diabetes versus prior probability for diabetes v curves plot, for an FPG value $t = 126$ mg/dl, with the other settings of the program in Table 3.

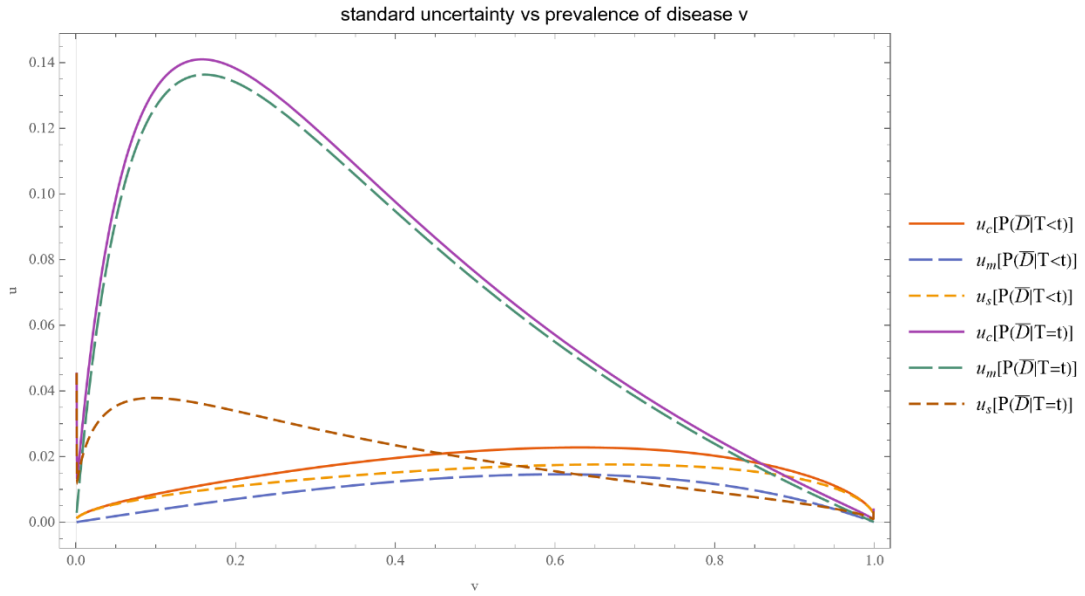

Figure 15. Standard sampling, measurement, and combined uncertainty of the negative predictive value for diabetes, and posterior probability for the absence of diabetes versus prior probability for diabetes v curves plot, for an FPG value $t = 126$ mg/dl, with the other settings of the program in Table 3.

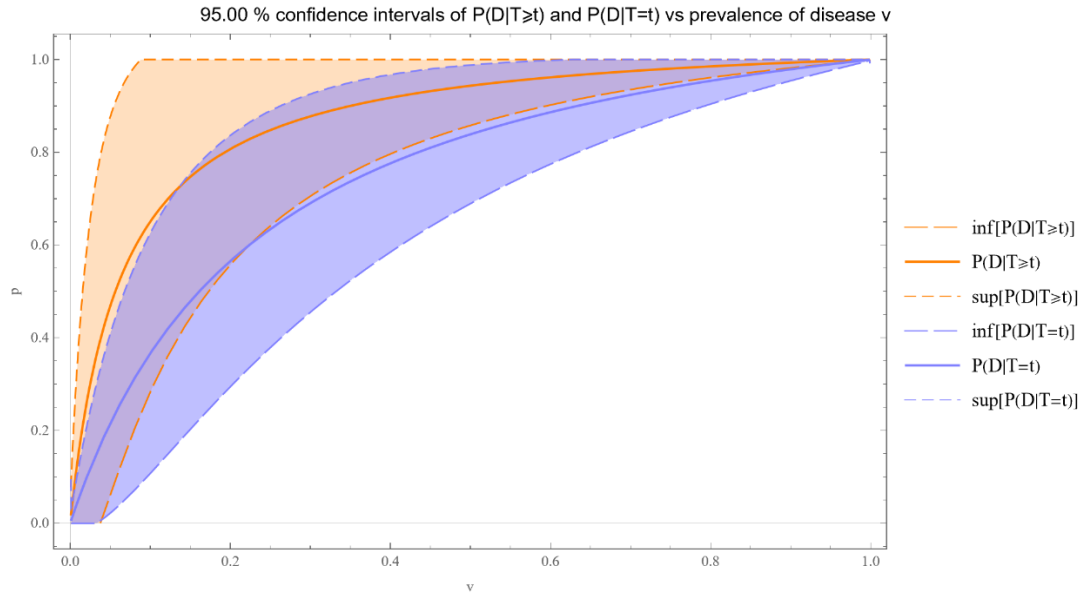

Figure 16. Confidence intervals of the positive predictive value and posterior probability for diabetes versus prior probability for diabetes v curves plot, for an FPG value $t = 126$ mg/dl, with the other settings of the program in Table 3.

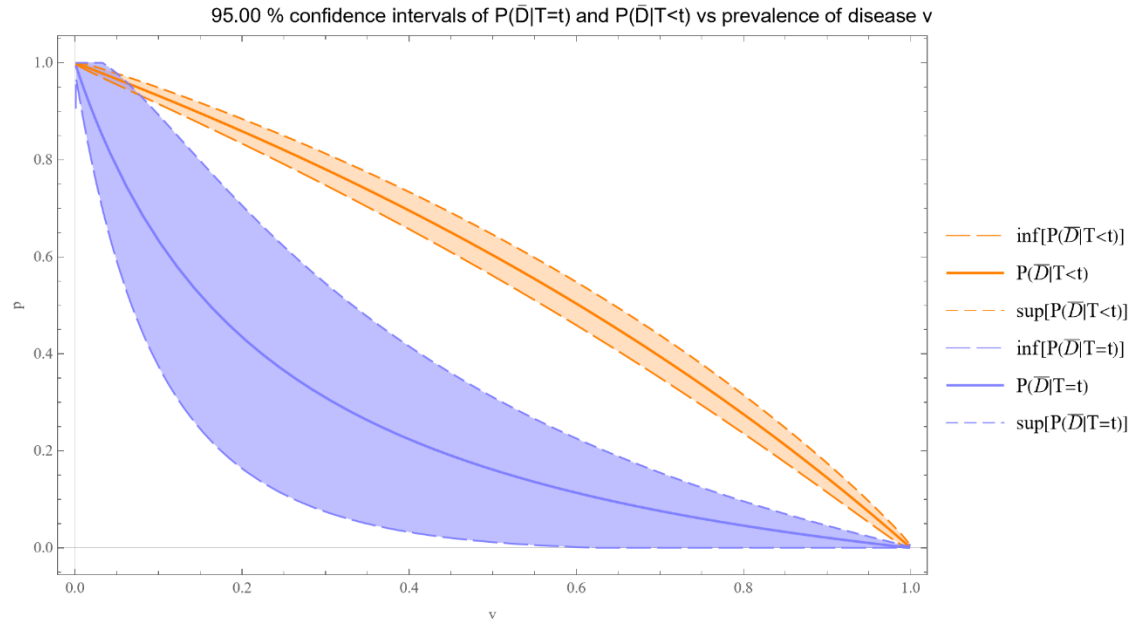

Figure 17. Confidence intervals of the negative predictive value for diabetes and posterior probability for the absence of diabetes versus prior probability for diabetes v curves plot for an FPG value $t = 126$ mg/dl, with the other settings of the program in Table 3.

standard uncertainty				
prevalence of disease v = 0.158				
measure	point estimation	standard uncertainty		
		combined	measurement	sampling
P(D|T≥t)	0.758	0.150	0.147	0.033
P(D|T=t)	0.494	0.141	0.136	0.036
P($\bar{D}$|T<t)	0.890	0.011	0.006	0.010
P($\bar{D}$|T=t)	0.506	0.141	0.136	0.036

Figure 18. Table of the sampling, measurement, and combined uncertainty of the Bayesian diagnostic measures for an FPG value $t = 126\ \mathrm{mg/dl}$, with the other program settings in Table 3.

95.00% confidence intervals			
prevalence of disease v = 0.158			
measure	point estimation	lower bound	upper bound
P(D|T≥t)	0.758	0.464	1.000
P(D|T=t)	0.494	0.217	0.770
P($\bar{D}$|T<t)	0.890	0.868	0.912
P($\bar{D}$|T=t)	0.506	0.230	0.783

Figure 19. Table of the confidence intervals of the Bayesian diagnostic measures for an FPG value $t = 126\ \mathrm{mg/dl}$, with the other settings of the program in Table 3.

The tables of Figures 18 and19 present Bayesian diagnostic measures for FPG measurements at the diabetes diagnostic threshold $t = 126\ \mathrm{mg/dl}$, in accordance with the American Diabetes Association (ADA) guidelines. The standard for diagnosing diabetes used in this study is the oral glucose tolerance test (OGTT) with a threshold of $200\ \mathrm{mg/dl}$. The limited concordance between these two diagnostic thresholds is evident from the point estimations and their associated uncertainties. For an FPG diagnostic threshold $t = 126\ \mathrm{mg/dl}$:

a) $P(D|T \geq t) = 0.758$, with a wide confidence interval (0.479 - 1.000), indicating substantial uncertainty.

b) $P(D|T = t) = 0.494$, with a wide confidence interval (0.229 - 0.758), showing lower certainty.

c) $P(\bar{D}|T < t) = 0.890$, with a confidence interval (0.242 - 0.771), showing lower certainty.

d) $P(\bar{D}|T = t) = 0.506$, with a narrow confidence interval (0.877 - 0.904), indicating high certainty.

Therefore:

a) $P(D|T = t) < P(D|T \geq t)$

b) The sizes of the confidence intervals of $P(D|T \geq t)$ and $P(D|T = t)$ are comparable.

c) There is a considerable overlap between the confidence intervals of $P(D|T \geq t)$ and $P(D|T = t)$.

d) $P(\overline{D}|T = t) < P(\overline{D}|T < t)$

e) The size of the confidence intervals of $P(\overline{D}|T < t)$ are considerably less than the size of the confidence intervals of $P(\overline{D}|T = t)$.

f) There is no overlap between the confidence intervals of $P(\overline{D}|T < t)$ and $P(\overline{D}|T = t)$.

In addition, the table with the standard uncertainty of the Bayesian diagnostic measures of Figure 18 shows that for $t = 126\ \mathrm{mg/dl}$, measurement uncertainty is the main component of their combined uncertainty.

All the figures provided by the program about the *Illustrative Case Study* data are presented in Supplemental file IV: BayesianDiagnosticInsightsFigures.pdf.

5. Discussion

There is a persistent need to estimate diagnostic measures and their uncertainty, especially regarding screening and diagnostic tests for potentially life-threatening diseases. The COVID-19 pandemic has convincingly exposed this need (42–47).

Conventional diagnostic approaches typically rely on set thresholds, often overlooking certain aspects of disease pathology. While historically influential, these methods may lack the comprehensive perspective required in modern patient-centered medicine. The continuous evolution of disease progression and changing patient demographics further complicate the diagnostic process, challenging the limits of traditional methods. In this context, Bayesian inference emerges as a viable alternative, offering probabilistic assessments tailored to individual patient profiles (4,48). Bayes' theorem provides a statistical framework to update the probability estimate of a disease as new information or test results become available. This approach enables healthcare professionals to refine disease probability estimates based on new data and prior knowledge.

To facilitate the application of Bayes' theorem in medical diagnosis, we have developed the software tool introduced in this study to explore and compare two pairs of Bayesian diagnostic measures of screening or diagnostic tests, assuming parametric distributions of the measurements:

a) The positive predictive value with the posterior probability for disease and

b) The negative predictive value for disease with the posterior probability for the absence of disease.

Academic publications that thoroughly explore the statistical distributions of diagnostic test measurements in diseased and nondiseased populations are limited (49). Therefore, exploratory data analysis and fitting of statistical distributions to diagnostic measurement data may be needed to apply the software tool (50). Recently, we have made available the *Bayesian Inference* program, which might be helpful in this regard (4).

Their broad applicability in medical diagnostic measurements motivated our parametric distribution choices:

a) *Normal distribution*

   A normal distribution is suited for data symmetric around the mean, indicating minimal skewness. This distribution assumes that data points are equally likely to occur on either side of the mean, forming the well-known bell curve.

b) *Lognormal distribution*

   In contrast, a lognormal distribution is appropriate for modeling positively skewed data, where the variable's logarithm follows a normal distribution. The parameters of the underlying normal distribution of the logarithm of the variable, a location parameter and a scale parameter, define this distribution. It can model data that cannot assume negative values and exhibits a long right tail, such as biological measurements.

c) *Gamma distribution*

   The gamma distribution is suitable for data with varying skewness and kurtosis that a lognormal distribution cannot adequately model. This distribution is characterized by two parameters: a shape parameter and a scale parameter. The flexibility of these parameters allows the gamma distribution to model a wide range of data behaviors, including varying degrees of skewness and kurtosis.

For our Illustrative case study, we have implemented an empirical Bayesian approach, as it is advantageous in several ways:

a) *Adaptability*

   It can adapt to the specific characteristics of the dataset, making it more flexible and applicable to diverse clinical settings.

b) *Robustness*

   Using empirical data to inform the prior mitigates the risk of bias introduced by subjective prior choices.

c) *Computational efficiency*

   Estimating the prior from data reduces the computational burden compared to fully Bayesian methods that require specifying and integrating complex prior distributions.

Estimating the uncertainty inherent in diagnostic measures is a considerable challenge in medical diagnostics (21,22,51). This challenge is particularly pronounced in medical decision-making for potentially life-threatening conditions. Assessing uncertainty is vital for ensuring reliable diagnoses and appropriate clinical interventions. Several notable examples of diagnostic measures where uncertainty estimation is critical include:

a) Cardiac troponin for diagnosing myocardial injury and infarction

   Cardiac troponin is a crucial biomarker for diagnosing myocardial injury and infarction (52).

b) Natriuretic peptides for diagnosing heart failure

   Natriuretic peptides, such as B-type natriuretic peptide (BNP) and N-terminal pro-b-type natriuretic peptide (NT-proBNP), are essential in diagnosing heart failure (53).

c) D-dimer for diagnosing thromboembolic events

   The measurement of D-dimer levels plays a crucial role in diagnosing thromboembolic events, such as deep vein thrombosis and pulmonary embolism (54).

d) Fasting plasma glucose (FPG), oral glucose tolerance test (OGTT), and glycated hemoglobin (HbA1c) for diagnosing diabetes

   Diagnosing diabetes relies on measuring blood glucose levels through tests like FPG, OGTT, and HbA1c (41).

e) *OGTT for diagnosing gestational diabetes*

The oral glucose tolerance test (OGTT) is the standard diagnostic tool for gestational diabetes and is vital for the health of both the mother and the developing fetus (55).

f) *Thyroid stimulating hormone (TSH), free serum triiodothyronine (T3), and free serum thyroxine (T4) for diagnosing thyroid dysfunction*

Measurement of thyroid function tests, including TSH, free T3, and free T4, is essential for diagnosing thyroid dysfunctions (56).

Our software allows the computation and plotting of the sampling, measurement, and combined uncertainty of Bayesian diagnostic measures as well as their confidence intervals.

Confidence interval plots serve multiple purposes:

a) *Precision assessment*

They provide insights into the precision of probability estimates at different measurement levels (57).

b) *Decision-making support*

For clinical decision-making, these plots can highlight the measurement thresholds where the probability for disease shifts significantly, guiding interventions or further testing.

c) *Epidemiological insights*

In epidemiological studies, understanding how disease probability varies across a population's measurement spectrum helps identify risk factors and inform public health strategies.

This exploration is imperative in quality and risk management in laboratory medicine and may contribute to the design and implementation of test accuracy studies (58). Despite extensive research on Bayesian diagnosis and uncertainty as separate areas, their intersection remains relatively unexplored (59,60).

The illustrative case study, focusing on individuals aged 70 to 80 years, aimed to minimize age-related variations in disease prevalence. This focus demonstrates the considerations required in modern diagnostics, where factors such as age, genetics, and lifestyle choices must be accounted for in the diagnostic equation. The case study underscores the substantial impact of combined uncertainty on the diagnostic process, highlighting the predominant role of measurement uncertainty and the challenging path toward enhancing diagnostic accuracy. Improving the analytical methods of screening and

diagnostic tests could enable the medical community to achieve more accurate diagnoses, facilitating more effective and personalized patient care.

Analyzing in more detail the Figures 5-8, 12,13, 16, and 17 of the illustrative case study described above, we may note the following clinical implications:

a) The positive predictive value $P(D|T \geq t)$ is highly influenced by the chosen threshold and the prevalence of diabetes, emphasizing the importance of selecting the appropriate cut-off for accurate diagnosis.
b) The double-threshold pattern in the Bayesian posterior probability $P(D|T = t)$ for diabetes suggests the need to understand the pathological implications of different FPG levels for tailored diagnostic strategies.
c) The variability in confidence intervals of both $P(D|T \geq t)$ and $P(D|T = t)$ at intermediate FPG levels suggests an increased risk of false positives or false negatives. This variability could result in unnecessary treatments or missed diagnoses, highlighting the importance of carefully interpreting test results within this range.
d) The differing trends in negative predictive value $P(\overline{D}|T < t)$ highlight the significance of selecting the appropriate threshold for excluding diabetes.
e) The unique behavior of Bayesian posterior probability $P(\overline{D}|T = t)$ for the absence of diabetes at lower FPG values, and the variability in its confidence intervals at both lower and higher FPG values impact diagnostic decisions, necessitating careful interpretation.
f) Despite the interpretative challenges of $P(\overline{D}|T < t)$ at lower FPG values, it is generally more robust than $P(\overline{D}|T = t)$ at higher FPG values.

The tables in Figures 18 and 19:

a) Indicate limited concordance between the diabetes classification criteria derived from the OGTT and FPG tests, consistent with findings previously reported in the literature (61,62).
b) Show that for FPG and diabetes, the point estimation of each Bayesian posterior probability is substantially less than the respective predictive value.

The discrepancies between FPG and OGTT thresholds for diagnosing diabetes highlight the need for a careful and comprehensive approach in clinical practice. By implementing combined testing strategies,

repeat testing protocols, and informed clinical judgment, healthcare providers can improve diagnostic accuracy and patient outcomes. Further research and patient education are also necessary in addressing the challenges posed by the limited concordance between these diagnostic methods and their considerable uncertainty.

Our approach integrates frequentist methods for uncertainty quantification due to their established reliability and ease of implementation in clinical settings. This empirical Bayesian framework allows for the practical application of the Bayes' theorem while leveraging the robustness of frequentist techniques for estimating sampling and measurement uncertainties.

Future research should focus on improving the estimations of the uncertainty of Bayesian diagnostic measures of different measurands under a diverse array of clinically and laboratory-relevant parameter settings. Furthermore, the full implementation of Bayesian methods for all aspects of uncertainty quantification could be explored, including utilizing Bayesian hierarchical models (7,63). Additionally, applying Bayes' factors to compare the evidence provided by different diagnostic measures represents a promising area for further investigation (64,65). These advancements could enhance the robustness and applicability of Bayesian methods in medical diagnostics, overcoming their current limitations (17,66).

To transition from research to practical application, clinical decision analysis, cost-effectiveness studies, and research on risk assessment and quality of care, including implementing studies, are required (67). These efforts are essential for addressing the complex issues in diagnostic medicine and developing new and effective strategies to overcome ongoing challenges.

All major general or medical statistical software packages (JASP® ver. 0.20.0, Mathematica® ver. 14.0, Matlab® ver. R2023b, MedCalc® ver. 22.008, metRology ver. 1.1-3, NCSS® ver. 24.0.0, NIST Uncertainty Machine ver. 2.0.0, OpenBUGS ver. 3.3.0, R ver. 4.3.1, SAS® ver. 9.5, SPSS® ver. 29, Stan ver. 2.33.0, Stata® ver. 19, and UQLab ver. 2.0.0) include routines for calculating and plotting various diagnostic measures and their confidence intervals. The program presented in this work provides 38 types of plots and 17 types of comprehensive tables of the four Bayesian diagnostic measures, their uncertainty, and the associated confidence intervals (Figure 1), many of which are novel. To the best of our knowledge, neither the programs mentioned above, nor any other software offers this extensive range of plots and tables without requiring advanced statistical programming.

The program complements our previously published tools for exploring diagnostic measures and posterior probability for disease and their uncertainty (4,21,22,68), facilitating their comparison.

5.1.Limitations of the Program

This program's limitations, which provide paths for further research, include:

a) Underlying assumptions

a. The existence of "gold standards" in diagnostics: In the absence of a "gold standard," alternative approaches for classification are available (69–71).
b. The hypothesis that measurements or their transforms follow a normal, lognormal, or gamma distribution. There is relevant literature concerning reference intervals, diagnostic thresholds, and clinical decision limits (72–76).
c. The generally accepted bimodality of the measurements, although unimodal distributions, could be considered (77,78).

b) Approximations used for the estimations

a. Utilization of first-order Taylor series approximations: First-order Taylor series approximations are employed in the propagation of uncertainty calculations. While this method provides a baseline estimation, higher-order approximations or Monte Carlo simulations may yield more precise results (24,79).
b. Uncertainty approximation in disease prevalence: The uncertainty associated with the prevalence or prior probability for a disease is approximated using the Agresti–Coull-adjusted Waldo interval. Although this method is widely used, more accurate techniques are available (80).
c. Approximations of the sampling uncertainty for both the sample means and standard deviations: These approximations can be refined for smaller sample sizes or in the presence of pronounced skewness in lognormal and gamma distributions (81,82).
d. Confidence intervals based on the t-distribution: Confidence intervals are derived using the t-distribution, which, despite the high relative uncertainty (83), is a practical alternative to credible intervals in selected scenarios, particularly in metrology (7,17,66,84).

While addressing these limitations would considerably increase computational complexity, they represent critical areas for future enhancement (79,85). We should, however, keep in mind that "all models will be based on assumptions and can only approach complex reality" (86), as "all models are wrong, but some models are useful" (87).

5.2.Limitations of the Case Study

The primary limitations of the case study are:

a) Dependence on the OGTT as the reference method for diagnosing diabetes mellitus, despite various factors affecting glucose tolerance (88–96).
b) Approximation of the FPG measurements distributions from NHANES datasets by lognormal distributions.
c) The implied assumption of simple random sampling.

6. Conclusion

Bayesian Diagnostic Insights enhances the estimation, visualization, and comparison of Bayesian diagnostic measures, including their associated uncertainty. It facilitates better clinical decision-making by providing insights into the uncertainty of disease probabilities. The illustrative case study, using FPG to diagnose diabetes, demonstrates the impact of measurement uncertainty on diagnostic measures, underlining its importance in improving clinical and laboratory diagnostic practices. Overall, the software provides a comprehensive framework for understanding and applying Bayes' theorem in medical diagnostics, fostering improved assessment and diagnosis of various health conditions.

7. Supplemental Material

The following supplemental files are available at https://www.hcsl.com/Supplements/SBDI.zip (accessed on August 4, 2024):

a) Supplemental File I:

   BayesianDiagnosticInsights.nb: The program as a Wolfram Notebook.

b) Supplemental File II:

BayesianDiagnosticInsightsCalculations.nb: The calculations for estimating Bayesian diagnostic measures and their standard uncertainty in a Wolfram Notebook

c) Supplemental File III:

BayesianDiagnosticInsightsInterface.pdf: A brief interface documentation of the program.

d) Supplemental File IV:

BayesianDiagnosticInsightsFigures.pdf: The figures of the program's output for the illustrative case study.

8. Declarations

Institutional Review Board Statement: Data collection was carried out following the rules of the Declaration of Helsinki. The National Center for Health Statistics Ethics Review Board approved data collection and posting of the data online for public use. The National Center for Health Statistics NHANES—NCHS Research Ethics Review Board Approval (Protocols #2005-06 and #2011-17) is available online at: https://www.cdc.gov/nchs/nhanes/irba98.htm (accessed on May 18, 2024).

Informed Consent Statement: Written consent was obtained from each subject participating in the survey.

Data Availability Statement: The data presented in this study are available at https://wwwn.cdc.gov/nchs/nhanes/default.aspx (accessed on August 4, 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

9. Appendix A

A.1. Notation

A.1.1. Acronyms

CDF: cumulative distribution function

PDF: probability density function

FPG: fasting plasma glucose

ADA: American Diabetes Association

A.1.2. Abbreviations

D: disease

$\overline{D}$: absence of disease

T: diagnostic test result

A.1.3. Parameters

t: diagnostic threshold

n_D : size of diseased population sample

μ_D: mean of the measurements of the diseased population

σ_D: standard deviation of the measurements of the diseased population

d_D: distribution of the measurements of the diseased population

$\mu_{\overline{D}}$: mean of the measurements of the nondiseased population

$\sigma_{\overline{D}}$: standard deviation of the measurements of the nondiseased population

$d_{\overline{D}}$: distribution of the measurements of the nondiseased population

m_D: mean of the measurements of the diseased population sample

s_D: standard deviation the measurements of the of diseased population sample

$n_{\overline{D}}$: size of nondiseased population sample

$m_{\overline{D}}$: mean of the measurements of the nondiseased population sample

$s_{\overline{D}}$: standard deviation of the measurements of the nondiseased population sample

v : prior probability for disease (prevalence rate)

n_U : number of quality control measurements

b_0 : constant contribution to measurement uncertainty

b_1: measurement uncertainty proportionality constant

p : confidence level

$\boldsymbol{\theta}$: Parameter vector

A.1.4. Bayesian Diagnostic Measures

$P(D|T \geq t)$: positive predictive value

$P(\overline{D}|T < t)$: negative predictive value

$P(D|T = t)$: posterior probability for disease

$P(\overline{D}|T = t)$: posterior probability for the absence of disease

A.1.5. Functions

$f(x)$: probability density function

$F(x)$: cumulative distribution function

$u_m(x)$: standard measurement uncertainty

$u_s(x)$: standard sampling uncertainty

${}_{i}u_c(x)$: standard combined uncertainty

${}_{i}\nu_{eff}(x)$: effective degrees of freedom

$inf(f)$: lower bound of f

$sup(f)$: upper bound of f

A.2. Input

A.2.1. Range of input parameters

t: $maximum(0, minimum(m_{\overline{D}} - 5s_{\overline{D}}, m_D - 5s_{\overline{D}})) - maximum(m_{\overline{D}} + 5s_{\overline{D}}, m_D + 5s_{\overline{D}})$

n_D : 2 – 10,000

m_D: 0.1 – 10,000

s_D: 0.01 – 1,000

$n_{\overline{D}}$: 2 – 10,000

$m_{\overline{D}}$: 0.1 – 10,000

$s_{\bar{D}}$: 0.01 – 1,000

v : 0.001 – 0.999

n_U : 20 – 10,000

b_0 : 0 – $\sigma_{\bar{D}}$

b_1 : 0 – 0.1000

p: 0.900 – 0.999

$t, m_D, s_D, m_{\bar{D}}$, and $s_{\bar{D}}$ are defined in arbitrary units.

A.2.2. Additional Input Options

A.2.2.1. Plots

Users can select between an extended and limited plot range.

A.2.2.2.2. Tables

Users can define the number of decimal digits for results, ranging from 1 to 10.

A.3. Software Availability and Requirements

Program name: Bayesian Diagnostic Insights

Version: 1.2.4

Project home page: https://www.hcsl.com/Tools/BayesianDiagnosticInsights/ (accessed on September 23, 2024)

Program source: BayesianDiagnosticInsights.nb . Available at: https://www.hcsl.com/Tools/BayesianDiagnosticInsights/BayesianDiagnosticInsights.nb (accessed on September 23, 2024).

Operating systems: Microsoft Windows 10+, Linux 3.15+, Apple macOS 11+

Programming language: Wolfram Language

Other software requirements: To run the program and read the BayesianDiagnosticInsightsCalculations.nb file Wolfram Player® ver. 14.0 is required, freely available at

https://www.wolfram.com/player/ (accessed on September 23, 2024) or Wolfram Mathematica® ver. 14.0.

System requirements: Intel® i9™ or equivalent CPU and 32 GB of RAM

License: Attribution—Noncommercial—ShareAlike 4.0 International Creative Commons License

A.4. A Note about the Program

About the Program Controls

The program features an intuitive tabbed user interface to streamline user interaction and facilitate effortless navigation across multiple modules and submodules.

Users may define the numerical settings with menus or sliders. Sliders are finely manipulated by pressing the alt or opt key while dragging the mouse. Pressing the shift or ctrl keys can even more finely manipulate them.

Dragging with the mouse while pressing the ctrl, alt, or opt keys zooms plots in or out. When the mouse cursor is positioned over a point on a curve in a plot, the coordinates of that point are displayed, and vertical drop lines are drawn to the respective axes.

10. References

1. Stanley DE, Campos DG. The logic of medical diagnosis. Perspect Biol Med. 2013 Spring;56(2):300–15.

2. Weiner ESC, Simpson JA, Oxford University Press. The Oxford English dictionary. Oxford, Oxford: Clarendon Press ; Melbourne; 1989 2004.

3. Zou KH, O'Malley AJ, Mauri L. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. Circulation. 2007 Feb 6;115(5):654–7.

4. Chatzimichail T, Hatjimihail AT. A Bayesian Inference Based Computational Tool for Parametric and Nonparametric Medical Diagnosis. Diagnostics. 2023 Oct 5;13(19):3135.

5. Djulbegovic B, van den Ende J, Hamm RM, Mayrhofer T, Hozo I, Pauker SG, et al. When is rational to order a diagnostic test, or prescribe treatment: the threshold model as an explanation of practice variation. Eur J Clin Invest. 2015 May;45(5):485–93.

6. Šimundić A-M. Measures of Diagnostic Accuracy: Basic Definitions. EJIFCC. 2009 Jan;19(4):203–11.

7. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian Data Analysis. CRC Press; 2013. 675 p.

8. Bayes M, Price M. LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. Philos Trans R Soc Lond. 1763 Dec 31;53(0):370–418.

9. Viana MAG, Ramakrishnan V. Bayesian estimates of predictive value and related parameters of a diagnostic test. Can J Stat. 1992 Sep;20(3):311–21.

10. van de Schoot R, Depaoli S, King R, Kramer B, Märtens K, Tadesse MG, et al. Bayesian statistics and modelling. Nature Reviews Methods Primers. 2021 Jan 14;1(1):1–26.

11. Bours MJ. Bayes' rule in diagnosis. J Clin Epidemiol. 2021 Mar;131:158–60.

12. Fischer F. Using Bayes theorem to estimate positive and negative predictive values for continuously and ordinally scaled diagnostic tests. Int J Methods Psychiatr Res. 2021 Jun;30(2):e1868.

13. Joyce J. Bayes' Theorem. In: Zalta EN, editor. The Stanford Encyclopedia of Philosophy [Internet]. Fall 2021. Metaphysics Research Lab, Stanford University; 2021. Available from: https://plato.stanford.edu/archives/fall2021/entries/bayes-theorem/

14. Casella G. An Introduction to Empirical Bayes Data Analysis. Am Stat. 1985;39(2):83–7.

15. Casella G. Illustrating empirical Bayes methods. Chemometrics Intellig Lab Syst. 1992 Oct 1;16(2):107–25.

16. Ayyub BM, Klir GJ. Uncertainty Modeling and Analysis in Engineering and the Sciences. Chapman and Hall/CRC; 2006.

17. Willink R, White R. Disentangling Classical and Bayesian Approaches to Uncertainty Analysis. New Zeland: Measurement Standards Laboratory; 2012 Nov.

18. Oosterhuis WP, Theodorsson E. Total error vs. measurement uncertainty: revolution or evolution? Clin Chem Lab Med. 2016 Feb;54(2):235–9.

19. M H Ramsey S L R Ellison P Rostron. Measurement uncertainty arising from sampling - A guide to methods and approaches. 2nd ed. EURACHEM/CITAC; 2019.

20. Ellison SLR, Williams A. Quantifying Uncertainty in Analytical Measurement. 3rd ed. EURACHEM/CITAC; 2012. Report No.: CG 4.

21. Chatzimichail T, Hatjimihail AT. A Software Tool for Calculating the Uncertainty of Diagnostic Accuracy Measures. Diagnostics (Basel) [Internet]. 2021 Feb 27;11(3). Available from: http://dx.doi.org/10.3390/diagnostics11030406

22. Chatzimichail T, Hatjimihail AT. A Software Tool for Estimating Uncertainty of Bayesian Posterior Probability for Disease. Diagnostics (Basel) [Internet]. 2024 Feb 12;14(4). Available from: http://dx.doi.org/10.3390/diagnostics14040402

23. Kallner A, Boyd JC, Duewer DL, Giroud C, Hatjimihail AT, Klee GG, et al. Expression of Measurement Uncertainty in Laboratory Medicine; Approved Guideline. Clinical and Laboratory Standards Institute; 2012.

24. Joint Committee for Guides in Metrology. Evaluation of measurement data – Supplement 2 to the "Guide to the expression of uncertainty in measurement" – Extension to any number of output quantities [Internet]. Pavillon de Breteuil, F-92312 Sèvres, Cedex, France: BIPM; 2011 Oct. Report No.: JCGM 102:2011. Available from: https://www.bipm.org/documents/20126/2071204/JCGM_102_2011_E.pdf/6a3281aa-1397-d703-d7a1-a8d58c9bf2a5

25. White GH. Basics of estimating measurement uncertainty. Clin Biochem Rev. 2008 Aug;29 Suppl 1:S53-60.

26. Agresti A, Franklin C, Klingenberg B. Statistics: The art and science of learning from data, global edition. 4th ed. London, England: Pearson Education; 2023.

27. Miller J, Miller JC. Statistics and Chemometrics for Analytical Chemistry. 7th ed. London, England: Pearson Education; 2018. 312 p.

28. J. Aitchison JACB. The Lognormal Distribution with special reference to its uses in econometrics. Cambridge: Cambridge University Press; 1957.

29. Agresti A, Coull BA. Approximate is Better than "Exact" for Interval Estimation of Binomial Proportions. Am Stat. 1998 May 1;52(2):119–26.

30. Wilson BM, Smith BL. Taylor-series and Monte-Carlo-method uncertainty estimation of the width of a probability distribution based on varying bias and random error. Meas Sci Technol. 2013 Jan 24;24(3):035301.

31. Satterthwaite FE. An approximate distribution of estimates of variance components. Biometrics. 1946 Dec;2(6):110–4.

32. Welch BL. The Generalization of `Student's' Problem when Several Different Population Variances are Involved. Biometrika. 1947;34(1/2):28–35.

33. Sun H, Saeedi P, Karuranga S, Pinkepank M, Ogurtsova K, Duncan BB, et al. IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Res Clin Pract. 2022 Jan;183:109119.

34. National Center for Health Statistics. National Health and Nutrition Examination Survey Data [Internet]. Centers for Disease Control and Prevention. 2005-20016 [cited 2023 Sep 4]. Available from: https://wwwn.cdc.gov/nchs/nhanes/default.aspx

35. National Center for Health Statistics. National Health and Nutrition Examination Survey Questionnaire [Internet]. Centers for Disease Control and Prevention. 2005-20016 [cited 2023 Sep 4]. Available from: https://wwwn.cdc.gov/nchs/nhanes/Search/variablelist.aspx?Component=Questionnaire

36. Petrone S, Rousseau J, Scricciolo C. Bayes and empirical Bayes: do they merge? Biometrika. 2014 Jun 1;101(2):285–302.

37. Myung IJ. Tutorial on maximum likelihood estimation. J Math Psychol. 2003 Feb 1;47(1):90–100.

38. Johnson ML. Nonlinear least-squares fitting methods. In: Methods Cell Biol. Academic Press; 2008. p. 781–805.

39. Bates DM, Watts DG. Nonlinear Regression Analysis and Its Applications. Hoboken, New Jersey: John Wiley & Sons, Inc.; 1988.

40. Darling DA. The Kolmogorov-Smirnov, Cramer-von Mises Tests. Ann Math Stat. 1957;28(4):823–38.

41. ElSayed NA, Aleppo G, Aroda VR, Bannuru RR, Brown FM, Bruemmer D, et al. 2. Classification and Diagnosis of Diabetes: Standards of Care in Diabetes-2023. Diabetes Care. 2023 Jan 1;46(Suppl 1):S19–40.

42. Lippi G, Simundic A-M, Plebani M. Potential preanalytical and analytical vulnerabilities in the laboratory diagnosis of coronavirus disease 2019 (COVID-19). Clin Chem Lab Med [Internet]. 2020 Mar 16; Available from: http://dx.doi.org/10.1515/cclm-2020-0285

43. Martin H. Kroll, MD Bipasa Biswas Jeffrey R. Budd, PhD Paul Durham, MA Robert T. Gorman, PhD Thomas E. Gwise, PhD Abdel-Baset Halim, PharmD, PhD, DABCC Aristides T. Hatjimihail, MD, PhD Jørgen Hilden, MD Kyunghee Song. Assessment of the Diagnostic Accuracy of Laboratory Tests Using Receiver Operating Characteristic Curves; Approved Guideline—Second Edition. Clinical and Laboratory Standards Institute; 2011.

44. Tang Y-W, Schmitz JE, Persing DH, Stratton CW. The Laboratory Diagnosis of COVID-19 Infection: Current Issues and Challenges. J Clin Microbiol [Internet]. 2020 Apr 3; Available from: http://dx.doi.org/10.1128/JCM.00512-20

45. Deeks JJ, Dinnes J, Takwoingi Y, Davenport C, Leeflang MMG, Spijker R, et al. Diagnosis of SARS-CoV-2 infection and COVID-19: accuracy of signs and symptoms; molecular, antigen, and antibody

tests; and routine laboratory markers. Cochrane Infectious Diseases Group, editor. Cochrane Database Syst Rev. 2020 Apr 24;26:1896.

46. Infantino M, Grossi V, Lari B, Bambi R, Perri A, Manneschi M, et al. Diagnostic accuracy of an automated chemiluminescent immunoassay for anti-SARS-CoV-2 IgM and IgG antibodies: an Italian experience. J Med Virol [Internet]. 2020 Apr 24; Available from: http://dx.doi.org/10.1002/jmv.25932

47. Mahase E. Covid-19: "Unacceptable" that antibody test claims cannot be scrutinised, say experts. BMJ. 2020 May 18;369:m2000.

48. Choi Y-K, Johnson WO, Thurmond MC. Diagnosis using predictive probabilities without cut-offs. Stat Med. 2006 Feb 28;25(4):699–717.

49. Smith AFM, Gelfand AE. Bayesian Statistics without Tears: A Sampling-Resampling Perspective. Am Stat. 1992;46(2):84–8.

50. Forbes C, Evans M, Hastings N, Peacock B. Statistical Distributions. John Wiley & Sons; 2011. 230 p.

51. Srinivasan P, Westover MB, Bianchi MT. Propagation of uncertainty in Bayesian diagnostic test interpretation. South Med J. 2012 Sep;105(9):452–9.

52. Wereski R, Kimenai DM, Taggart C, Doudesis D, Lee KK, Lowry MTH, et al. Cardiac Troponin Thresholds and Kinetics to Differentiate Myocardial Injury and Myocardial Infarction. Circulation. 2021 Aug 17;144(7):528–38.

53. Roberts E, Ludman AJ, Dworzynski K, Al-Mohammad A, Cowie MR, McMurray JJV, et al. The diagnostic accuracy of the natriuretic peptides in heart failure: systematic review and diagnostic meta-analysis in the acute care setting. BMJ. 2015 Mar 4;350:h910.

54. Freund Y, Chauvin A, Jimenez S, Philippon A-L, Curac S, Fémy F, et al. Effect of a Diagnostic Strategy Using an Elevated and Age-Adjusted D-Dimer Threshold on Thromboembolic Events in Emergency Department Patients With Suspected Pulmonary Embolism: A Randomized Clinical Trial. JAMA. 2021 Dec 7;326(21):2141–9.

55. Rani PR, Begum J. Screening and Diagnosis of Gestational Diabetes Mellitus, Where Do We Stand. J Clin Diagn Res. 2016 Apr;10(4):QE01-4.

56. Reyes Domingo F, Avey MT, Doull M. Screening for thyroid dysfunction and treatment of screen-detected thyroid dysfunction in asymptomatic, community-dwelling adults: a systematic review. Syst Rev. 2019 Nov 18;8(1):260.

57. Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol. 2016 Apr;31(4):337–50.

58. Horvath AR, Bell KJL, Ceriotti F, Jones GRD, Loh TP, Lord S, et al. Outcome-based analytical performance specifications: current status and future challenges. Clin Chem Lab Med [Internet]. 2024 Jun 6; Available from: http://dx.doi.org/10.1515/cclm-2024-0125

59. Baron JA. Uncertainty in Bayes. Med Decis Making. 1994 Jan-Mar;14(1):46–51.

60. Ashby D, Smith AF. Evidence-based medicine as Bayesian decision-making. Stat Med. 2000 Dec 15;19(23):3291–305.

61. Tucker LA. Limited Agreement between Classifications of Diabetes and Prediabetes Resulting from the OGTT, Hemoglobin A1c, and Fasting Glucose Tests in 7412 U.S. Adults. J Clin Med Res [Internet]. 2020 Jul 13;9(7). Available from: http://dx.doi.org/10.3390/jcm9072207

62. Sacks DB, Arnold M, Bakris GL, Bruns DE, Horvath AR, Lernmark Å, et al. Guidelines and Recommendations for Laboratory Analysis in the Diagnosis and Management of Diabetes Mellitus. Clin Chem. 2023 Aug 2;69(8):808–68.

63. Congdon PD. Bayesian hierarchical models: With applications using R, second edition [Internet]. 2nd ed. Philadelphia, PA: Chapman & Hall/CRC; 2021 [cited 2024 Aug 4]. 592 p. Available from: https://books.google.com/books/about/Bayesian_Hierarchical_Models.html?hl=el&id=hlivDwAAQBAJ

64. Kass RE, Raftery AE. Bayes Factors. J Am Stat Assoc. 1995;90(430):773–95.

65. Bozza S, Taroni F, Biedermann A. Bayes factors for forensic decision analyses with R. 1st ed. Cham, Switzerland: Springer International Publishing; 2022. 187 p. (Springer texts in statistics).

66. Willink R. Measurement Uncertainty and Probability. Cambridge University Press; 2013.

67. Knottnerus JA, Buntinx F, editors. The evidence base of clinical diagnosis. 2nd ed. BMJ Books; 2011. 320 p. (Evidence-Based Medicine).

68. Chatzimichail T, Hatjimihail AT. Relation of diagnostic accuracy measures [Internet]. Hellenic Complex Systems Laboratory; 2018 [cited 2023 Nov 5] p. 1–7. Report No.: XIV. Available from: https://heracleitos.github.io/TR/hcsltr14/index.html

69. Knottnerus JA, Dinant GJ. Medicine based evidence, a prerequisite for evidence based medicine. BMJ. 1997 Nov 1;315(7116):1109–10.

70. Pfeiffer RM, Castle PE. With or without a gold standard. Epidemiology. 2005 Sep;16(5):595–7.

71. van Smeden M, Naaktgeboren CA, Reitsma JB, Moons KGM, de Groot JAH. Latent class models in diagnostic studies when there is no reference standard--a systematic review. Am J Epidemiol. 2014 Feb 15;179(4):423–31.

72. Solberg HE. Approved recommendation (1987) on the theory of reference values. Part 5. Statistical treatment of collected reference values. Determination of reference limits. Clin Chim Acta. 1987 Dec 1;170(2):S13–32.

73. Pavlov IY, Wilson AR, Delgado JC. Reference interval computation: which method (not) to choose? Clin Chim Acta. 2012 Jul 11;413(13–14):1107–14.

74. Sikaris K. Application of the stockholm hierarchy to defining the quality of reference intervals and clinical decision limits. Clin Biochem Rev. 2012 Nov;33(4):141–8.

75. Daly CH, Liu X, Grey VL, Hamid JS. A systematic review of statistical methods used in constructing pediatric reference intervals. Clin Biochem. 2013 Sep;46(13–14):1220–7.

76. Ozarda Y, Sikaris K, Streichert T, Macri J, IFCC Committee on Reference intervals and Decision Limits (C-RIDL). Distinguishing reference intervals and clinical decision limits - A review by the IFCC Committee on Reference Intervals and Decision Limits. Crit Rev Clin Lab Sci. 2018 Sep;55(6):420–31.

77. Wilson JMG, Jungner G. Principles and practice of screening for disease. Geneva: World Health Organization; 1968. 163 p. (Public health papers; vol. 34).

78. Petersen PH, Horder M. 2.3 Clinical test evaluation. Unimodal and bimodal approaches. Scand J Clin Lab Invest. 1992 Jan 1;52(sup208):51–7.

79. Joint Committee for Guides in Metrology. Evaluation of measurement data — Supplement 1 to the "Guide to the expression of uncertainty in measurement"— Propagation of distributions using a Monte Carlo method Joint Committee for Guides in Metrology [Internet]. Pavillon de Breteuil, F-92312 Sèvres, Cedex, France: BIPM; 2008. Report No.: JCGM 101:2008. Available from: https://www.bipm.org/documents/20126/2071204/JCGM_101_2008_E.pdf/325dcaad-c15a-407c-1105-8b7f322d651c

80. Pires AM, Amado C. Interval Estimators for a Binomial Proportion: Comparison of Twenty Methods. Revstat Stat J. 2008 Jun 24;6(2):165–97.

81. Schmoyeri RL, Beauchamp JJ, Brandt CC, Hoffman FO. Difficulties with the lognormal model in mean estimation and testing. Environ Ecol Stat. 1996 Mar 1;3(1):81–97.

82. Bhaumik DK, Kapur K, Gibbons RD. Testing Parameters of a Gamma Distribution for Small Samples. Technometrics. 2009 Aug 1;51(3):326–34.

83. Williams A. Calculation of the expanded uncertainty for large uncertainties using the lognormal distribution. Accredit Qual Assur. 2020 Dec 1;25(5):335–8.

84. Stephens M. The Bayesian lens and Bayesian blinkers. Philos Trans A Math Phys Eng Sci. 2023 May 15;381(2247):20220144.

85. Joint Committee for Guides in Metrology. Guide to the expression of uncertainty in measurement — Part 6: Developing and using measurement models [Internet]. Pavillon de Breteuil, F-92312 Sèvres, Cedex, France: BIPM; 2020. Report No.: JCGM GUM-6:2020. Available from: https://www.bipm.org/documents/20126/2071204/JCGM_GUM_6_2020.pdf/d4e77d99-3870-0908-ff37-c1b6a230a337

86. Oosterhuis WP. Analytical performance specifications in clinical chemistry: the holy grail? J Lab Precis Med. 2017 Sep;2:78–78.

87. Box GEP. Robustness in the strategy of scientific model building. In: Robustness in Statistics. Elsevier; 1979. p. 201–36.

88. Rao SS, Disraeli P, McGregor T. Impaired glucose tolerance and impaired fasting glucose. Am Fam Physician. 2004 Apr 15;69(8):1961–8.

89. Meneilly GS, Elliott T. Metabolic alterations in middle-aged and elderly obese patients with type 2 diabetes. Diabetes Care. 1999 Jan;22(1):112–8.

90. Geer EB, Shen W. Gender differences in insulin resistance, body composition, and energy balance. Gend Med. 2009;6 Suppl 1(Suppl 1):60–75.

91. Van Cauter E, Polonsky KS, Scheen AJ. Roles of circadian rhythmicity and sleep in human glucose regulation. Endocr Rev. 1997 Oct;18(5):716–38.

92. Colberg SR, Sigal RJ, Fernhall B, Regensteiner JG, Blissmer BJ, Rubin RR, et al. Exercise and type 2 diabetes: the American College of Sports Medicine and the American Diabetes Association: joint position statement. Diabetes Care. 2010 Dec;33(12):e147-67.

93. Salmerón J, Manson JE, Stampfer MJ, Colditz GA, Wing AL, Willett WC. Dietary fiber, glycemic load, and risk of non-insulin-dependent diabetes mellitus in women. JAMA. 1997 Feb 12;277(6):472–7.

94. Surwit RS, van Tilburg MAL, Zucker N, McCaskill CC, Parekh P, Feinglos MN, et al. Stress management improves long-term glycemic control in type 2 diabetes. Diabetes Care. 2002 Jan;25(1):30–4.

95. Pandit MK, Burke J, Gustafson AB, Minocha A, Peiris AN. Drug-induced disorders of glucose tolerance. Ann Intern Med. 1993 Apr 1;118(7):529–39.

96. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet. 2010 Feb;42(2):105–16.

11. Permanent Citation:

Chatzimichail T, Hatjimihail AT. *A Software Tool for Applying Bayes' Theorem in Medical Diagnostics*. Technical Report XXVII. Hellenic Complex Systems Laboratory; 2024. https://www.hcsl.com/TR/hcsltr27/hcsltr27.pdf

12. License

First Published: June 10, 2024

Revised: September 23, 2024