Hellenic Complex Systems Laboratory

A Software Tool for Applying Bayes' Theorem in Medical Diagnostics

Technical Report XXVII

Theodora Chatzimichail and Aristides T. Hatjimihail
2024

A Software Tool for Applying Bayes' Theorem in Medical Diagnostics

Theodora Chatzimichail, MRCS [a], Aristides T. Hatjimihail, MD, PhD [a]

[a] Hellenic Complex Systems Laboratory

Abstract

Background: In medical diagnostics, estimating post-test or posterior probabilities for disease, positive and negative predictive values, and their associated uncertainty is essential for patient care.

Objective: To introduce 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 estimate positive and negative predictive values and posterior probabilities for the presence and absence of a disease. It estimates their standard sampling, measurement, and combined uncertainty, as well as their confidence intervals, applying uncertainty propagation methods based on first-order Taylor series approximations. It employs normal, lognormal, and gamma distributions.

Results: The software generates plots and tables of the estimates to support clinical decision-making. An illustrative case study using fasting plasma glucose data from the National Health and Nutrition Examination Survey (NHANES) demonstrates its application in diagnosing diabetes mellitus. The results highlight the significant impact of measurement uncertainty on Bayesian diagnostic measures, particularly on positive predictive value and posterior probabilities.

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 their uncertainty quantification 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

1. Introduction

1.1. Medical Diagnosis

Diagnosis in medicine is fundamentally the process of identifying a disease by analyzing its unique characteristics through abduction, deduction, and induction (Stanley and Campos 2013). The term' diagnosis,' originating from the Greek 'διάγνωσις' meaning 'discernment' (Weiner, Simpson, and Oxford University Press 1989 2004), underscores the critical role of distinguishing between healthy and diseased states in individuals. Diagnosis can be defined as the stochastic mapping of symptoms, signs, and laboratory and medical imaging findings onto a particular disease condition, based on medical knowledge.

1.1.1. Threshold Based Diagnosis

Diagnostic tests or procedures are often applied to classify individuals into diseased or nondiseased populations in a binary manner. Although the probability distributions of measurands from a quantitative diagnostic test in these populations may overlap, results are typically dichotomized by setting a diagnostic threshold or cut-off point (Zou, O'Malley, and Mauri 2007). Reliance on a single threshold for diagnosis across a spectrum of data points introduces uncertainty due to this overlap (Chatzimichail and Hatjimihail 2023). Nonetheless, this dichotomous approach represents a significant transformation in medical decision-making by correlating a continuous spectrum of evidence with binary clinical decisions, such as whether to treat or not (Djulbegovic et al. 2015).

1.1.1.1. Diagnostic Accuracy Measures

To ensure patient safety, the correctness of this classification must be rigorously evaluated. Although numerous diagnostic accuracy measures (DAMs) are described in the literature, only a few are routinely used

in clinical research and practice to assess the diagnostic accuracy of threshold-based tests (Šimundić 2009). These include the prevalence-dependent positive and negative predictive values, defined conditionally on the test outcome.

1.1.2. Bayesian Diagnosis

Bayes' theorem (Gelman et al. 2013; Bayes and Price 1763) plays a pivotal role in medical diagnostics by transforming the pre-test or prior probability for a disease into a post-test or posterior probability after considering diagnostic test results (Viana and Ramakrishnan 1992; Gelman et al. 2013; van de Schoot et al. 2021; Bours 2021; Fischer 2021; Chatzimichail and Hatjimihail 2023). 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 (Joyce 2021).

1.1.2.1. Bayesian Inference

In purely Bayesian inference, the process begins with a prior distribution representing initial beliefs about the parameters of interest before observing any evidence. This prior distribution is then updated with the likelihood function—which represents the probability of the observed evidence given different parameter values—using Bayes' theorem to obtain the posterior distribution (van de Schoot et al. 2021). After considering the observed data, the posterior distribution combines prior information with new evidence, reflecting updated parameter knowledge.

1.1.2.1.1. Prior Distribution

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

1.1.2.1.2. Likelihood Function

The likelihood function describes the probability of the observed evidence given various parameter values. It plays a crucial role in updating the prior distribution to form the posterior distribution.

1.1.2.1.3. Posterior Distribution

The posterior distribution results from combining the prior distribution and the likelihood function. It reflects the updated understanding of the parameters after incorporating the observed evidence.

1.1.2.1.4. Workflow

The typical Bayesian workflow involves:

a) *Specifying the Prior Distribution*: Defining initial beliefs about the parameters based on prior knowledge or assumptions.
b) *Determining the Likelihood Function*: Modeling how likely the observed data are given different parameter values.
c) *Computing the Posterior Distribution*: Applying Bayes' theorem to update the prior distribution with the likelihood function.
d) *Model Checking and Refinement*: Assessing the model's fit and making necessary adjustments.
e) *Sensitivity Analysis*: Evaluating how sensitive the results are to changes in the prior assumptions or model specifications.

These steps are essential for ensuring the robustness of Bayesian inferences.

1.1.2.2. Empirical Bayesian Methods

The empirical Bayesian approach simplifies the purely Bayesian framework by using available data to estimate the prior distribution, making it practical when prior information is sparse or unavailable (Casella 1985, 1992). Instead of specifying a fixed prior distribution, the empirical Bayesian method treats the prior as an unknown quantity to be estimated from the data. This approach is particularly suitable for medical diagnostics, where real-time data integration is crucial.

1.1.2.2.1. Workflow

The typical empirical Bayesian workflow involves:

a) *Data Collection and Preliminary Analysis*: Gathering a large dataset and performing statistical analyses to understand the distributions and characteristics of the data.

b) *Estimating Prior Distributions*: Using empirical data to estimate prior distributions and probabilities through methods such as maximum likelihood estimation.

c) *Applying Bayes' Theorem*: Computing posterior probabilities by combining the estimated prior distributions with the likelihood function, thereby incorporating the observed data.

This method allows for adaptive updating of beliefs based on the data, enhancing the applicability of Bayesian methods in practical settings where prior information may be limited.

1.2. Uncertainty

Uncertainty reflects imperfect or incomplete information. When quantifiable, it can be expressed using probability (Ayyub and Klir 2006). In our empirical Bayesian approach, we integrate frequentist methods for uncertainty quantification due to their established reliability and ease of implementation in clinical settings (Willink and White 2012).

1.2.1. Measurement Uncertainty

Due to 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 (Oosterhuis and Theodorsson 2016).

1.2.2. Sampling Uncertainty

Diagnostic measures are derived from screening or diagnostic tests applied to population samples. The variability within these samples contributes to the overall uncertainty of the measures(M H Ramsey S L R Ellison P Rostron 2019). This intrinsic heterogeneity is present even when simple random sampling techniques are employed (Ellison and Williams 2012).

1.2.3. Uncertainty of Diagnostic Accuracy Measures and Bayesian Posterior Probabilities

Previous studies have explored the uncertainty associated with diagnostic accuracy measures and the posterior probabilities for disease derived from Bayes' theorem, demonstrating that this uncertainty can significantly impact their clinical usefulness (Chatzimichail and Hatjimihail 2021, 2024). Estimating, evaluating, and mitigating this uncertainty are critical tasks in medical diagnosis.

1.3. 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) Positive predictive value and negative predictive value (Bours 2021).

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

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

2. Methods

2.1. Calculations

2.1.1. 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. It is expressed as:

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

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

where $\bar{H}$ represents the negation of hypothesis H. Substituting back into Bayes' theorem:

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

In medical diagnostics, Bayes' theorem provides a robust framework for updating the probability of a disease (hypothesis H) being present given new evidence E (such as test results). By combining prior knowledge (pre-test probability) with new data (test results), Bayesian methods offer a comprehensive approach to the medical diagnostic process.

2.1.1.1. Positive and Negative Predictive Value

Let D denote the presence and $\bar{D}$ the absence of a disease, $F_D(x|\boldsymbol{\theta})$ the cumulative distribution function (CDF) of the test measurements T in individuals with the disease, $F_{\bar{D}}(x|\ \boldsymbol{\theta})$ the CDF in individuals without the disease, and v the prevalence or prior probability for disease. The positive predictive value of a diagnostic test T for a diagnostic threshold t is calculated as:

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

Similarly, the negative predictive value is:

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

In these equations, $1 - F_D(t|\boldsymbol{\theta})$ represents the **sensitivity** of the test at threshold t and $F_{\bar{D}}(t|\boldsymbol{\theta})$ its specificity.

These measures assess the test's ability to correctly identify diseased and nondiseased individuals based on the threshold t.

2.1.1.2. Posterior Probability for Disease and Absence of Disease

Let $f_D(x|\boldsymbol{\theta})$ denote the probability density function (PDF) of the test measurements T in individuals with the disease, $f_{\bar{D}}(x;\ \boldsymbol{\theta})$ the PDF in individuals without the disease, and v the prevalence or prior probability for disease. The posterior or post-test probability for disease given a diagnostic test result $T = t$ is:

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

Similarly, the posterior or post-test probability for the absence of disease is:

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

These posterior probabilities provide a continuous assessment of disease likelihood based on the test measurement t, rather than dichotomizing the results using a threshold.

2.1.2. Uncertainty Quantification

Uncertainty in input parameters can be represented as standard uncertainty $u(t)$, which is the standard deviation of t, and expanded uncertainty $U(t)$, which defines a range around t with a specified probability p (Kallner et al. 2012).

2.1.2.1. Measurement Uncertainty

Measurement uncertainty is estimated according to "Guide to the Expression of Uncertainty in Measurement" (GUM) (Joint Committee for Guides in Metrology 2011) and "Expression of Measurement Uncertainty in Laboratory Medicine" (Kallner et al. 2012). Bias is considered a component of this uncertainty (White 2008). The relationship between the standard measurement uncertainty $u_m(t)$ to the value of the measurement t, is typically represented as (Ellison and Williams 2012):

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

where b_0 and b_1 are constants.

For a linear approximation, it is expressed as (Ellison and Williams 2012):

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

2.1.2.2. Sampling Uncertainty of Means and Standard Deviations

Standard uncertainty in means and standard deviations are estimated utilizing the central limit theorem and the chi-square distribution (Agresti, Franklin, and Klingenberg 2023; Miller and Miller 2018; J. Aitchison 1957) 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.

2.1.2.3. 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 (Agresti and Coull 1998).

2.1.2.4. 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 (Kallner et al. 2012):

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

If the components are correlated, then (Joint Committee for Guides in Metrology 2011):

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

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 are computed via uncertainty propagation rules, employing a first-order Taylor series approximation (B. M. Wilson and Smith 2013) (refer to Supplemental File II: BayesianDiagnosticInsightsCalculations.nb). Assuming uncorrelated parameters, we use the following formula to compute uncertainty propagation (Joint Committee for Guides in Metrology 2011):

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

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

The estimated standard uncertainty of the Bayesian diagnostic measures is truncated to the [0,1] range.

2.1.2.5. 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 (Satterthwaite 1946; Welch 1947):

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

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:

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

where $F_\nu^{-1}(z)$ is the inverse CDF of the Student's t-distribution with ν degrees of freedom and ${}_l u_c(t)$ is the standard combined uncertainty of the Bayesian diagnostic measure.

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

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

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

2.2. The Software

2.2.1. Program Overview

The software program *Bayesian Diagnostic Insights* was developed using the Wolfram Language with Wolfram Mathematica® Ver 14.1 (Wolfram Research, Inc., Champaign, IL, USA). It facilitates the estimation and comparison of Bayesian diagnostic measures. This interactive program is designed to estimate and plot the values, standard sampling uncertainty, measurement uncertainty, combined uncertainty, and 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 using Wolfram Player® or Wolfram Mathematica® (refer to Appendix A.3). The intricate nature of the required computations necessitates substantial computational resources.

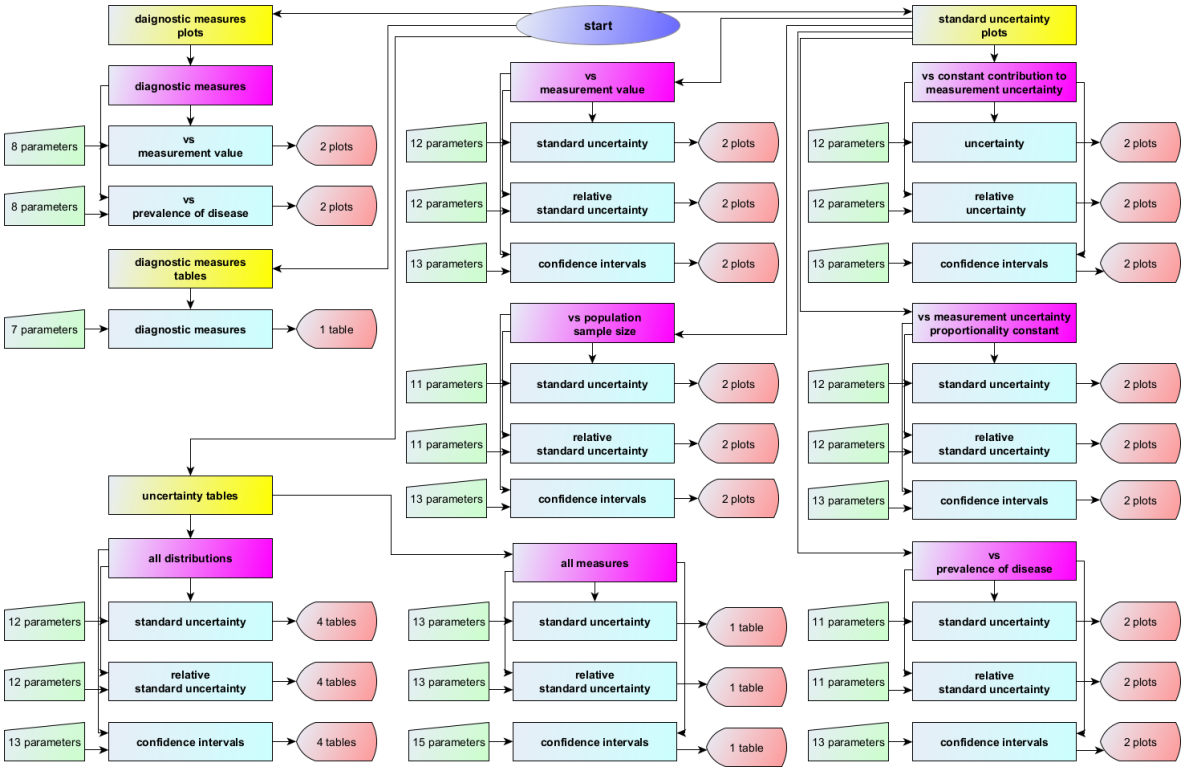

Figure 1. A simplified flowchart of the program *Bayesian Diagnostic Insights*.

2.2.2. Input Parameters

2.2.2.1. Parametric Distributions

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

a) Normal distribution
b) Lognormal distribution
c) Gamma distribution.

2.2.2.2. Bayesian Diagnostic Measures

Users select the Bayesian diagnostic measures to be evaluated from the following options:

a) The positive predictive value $P(D|T \geq t)$
b) The negative predictive value $P(\bar{D}|T < t)$
c) The posterior probability for disease $P(D|T = t)$
d) The posterior probability for the absence of disease $P(\bar{D}|T = t)$

2.2.2.3. Definition of Populations and Samples Parameters and Statistics

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

For each population sample, users define its size n, the mean m, and the standard deviation s of the measurements (in arbitrary units).

2.2.2.4. Measurement Uncertainty

Users select a linear or a nonlinear equation to describe the measurement uncertainty as a function of the measurement value t. 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.

2.2.3. Output

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

More detailed documentation of the program's interface is provided in Supplemental file III: BayesianDiagnosticInsightsInterface.pdf

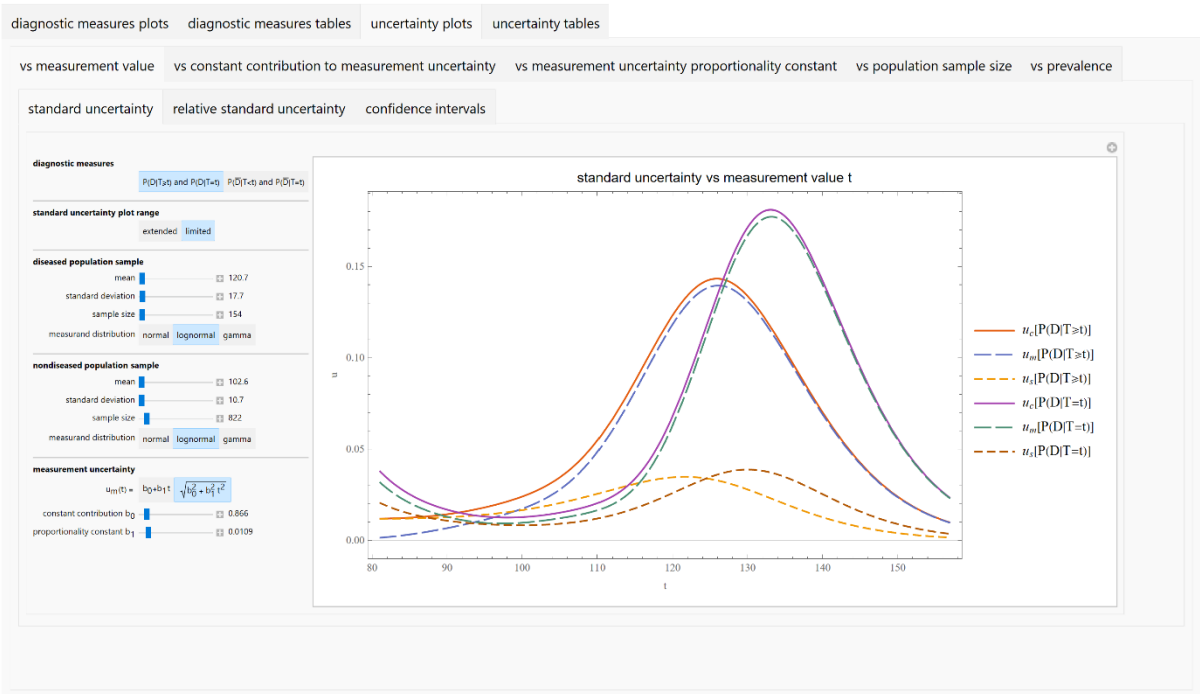

Figure 2. A screenshot of the program *Bayesian Diagnostic Insights*.

2.3. Illustrative Case Study

As previously described (Chatzimichail and Hatjimihail 2024), we conducted an illustrative case study to demonstrate the program's application. We used fasting plasma glucose (FPG) as the diagnostic test measurand for the Bayesian diagnosis of diabetes mellitus (hereafter referred to as "diabetes"), with the oral glucose tolerance test (OGTT) serving as the reference method. Diabetes was diagnosed if the plasma glucose value was equal to or greater than 200 mg/dL, measured two hours after the oral administration of 75 g of glucose during an OGTT (2-hour PG) (American Diabetes Association 2021). The study focused on individuals aged 70 to 80 years, reflecting the significant correlation between age and diabetes prevalence (Sun et al. 2022).

Data were collected from the National Health and Nutrition Examination Survey (NHANES) participants from 2005 to 2016 (n = 60,936), as previously described (Chatzimichail and Hatjimihail 2024). NHANES is a comprehensive survey assessing the health and nutritional status of adults and children in the United States (National Center for Health Statistics 2005-20016).

The inclusion criteria were valid FPG and OGTT results (n = 13,836), no prior diabetes diagnosis (National Center for Health Statistics 2005-20016) (n = 13,465), and age 70–80 years (n = 976).

Participants with a 2-h PG measurement ≥200 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 (Petrone, Rousseau, and Scricciolo 2014). We estimated the prevalence or prior probability for diabetes as follows:

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

The FPG datasets statistics are presented in Table 1 (hereafter, FPG and its uncertainty are expressed in 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 (Myung 2003). 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 (Johnson 2008; Bates and Watts 1988) 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 for diabetic and nondiabetic participants and the respective p-values from the Cramér–von Mises goodness-of-fit test (Darling 1957). This test assesses the goodness-of-fit by comparing the empirical cumulative distribution functions (CDFs) of the measurement samples with those of the estimated distributions. The calculated p-values indicate that any observed differences between the empirical data and the estimated distributions can be attributed to random sampling variability, suggesting that the lognormal distributions provide an acceptable fit to the FPG measurements in both groups.

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

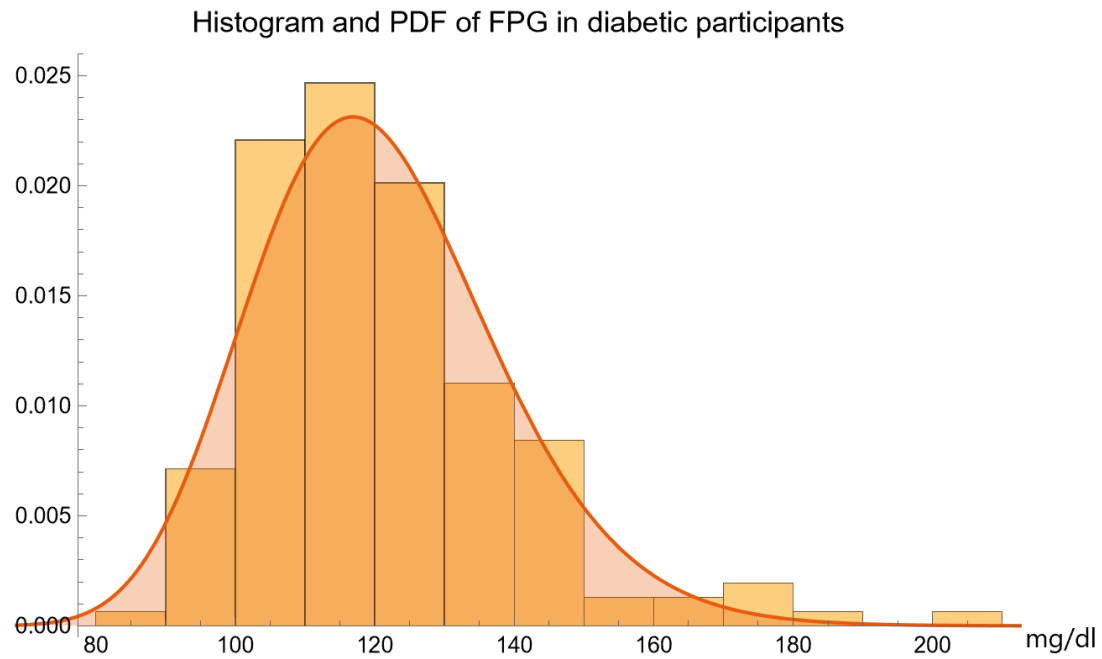

Figure 3. The estimated PDF of the FPG (mg/dl) in diabetic participants, assuming a lognormal distribution and negligible measurement uncertainty, and the histogram of the respective NHANES dataset, with the distribution parameters in Table 2.

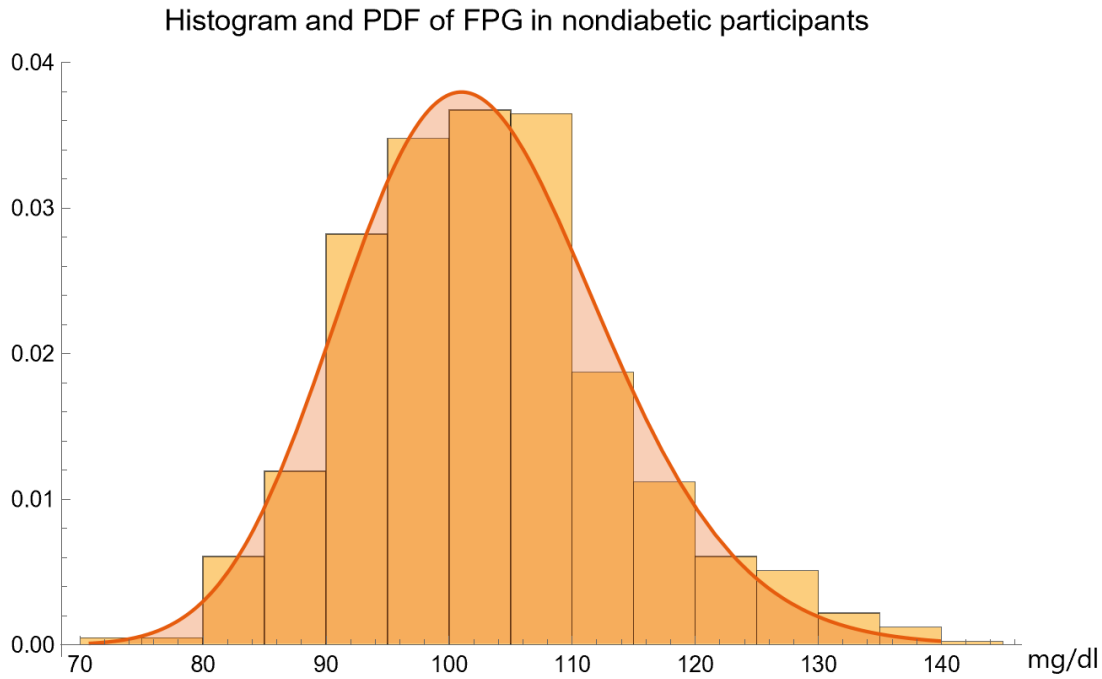

Figure 4. The estimated PDF of the FPG (mg/dl) in nondiabetic participants, assuming a lognormal distribution and negligible measurement uncertainty, and the histogram of the respective NHANES dataset, with the parameters of the distribution in Table 2.

Likelihoods and posterior probabilities were estimated accordingly.

3. Results

The results of applying the program to the illustrative case study data are presented in Figures 5-19, and the program settings are detailed in Tables 2 and 3.

3.1. Measures

Table 2. The settings of the program *Bayesian Diagnostic Insights* for 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 displays the plots of:

a) Positive predictive value $P(D|T \geq t)$ of FPG for diabetes versus threshold value t (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 approach an asymptote at 1.00.

b) Posterior probability for diabetes versus FPG value t (mg/dl)(blue curve). 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 very low and very high values of t, decreasing rapidly to its minimum before increasing rapidly again.

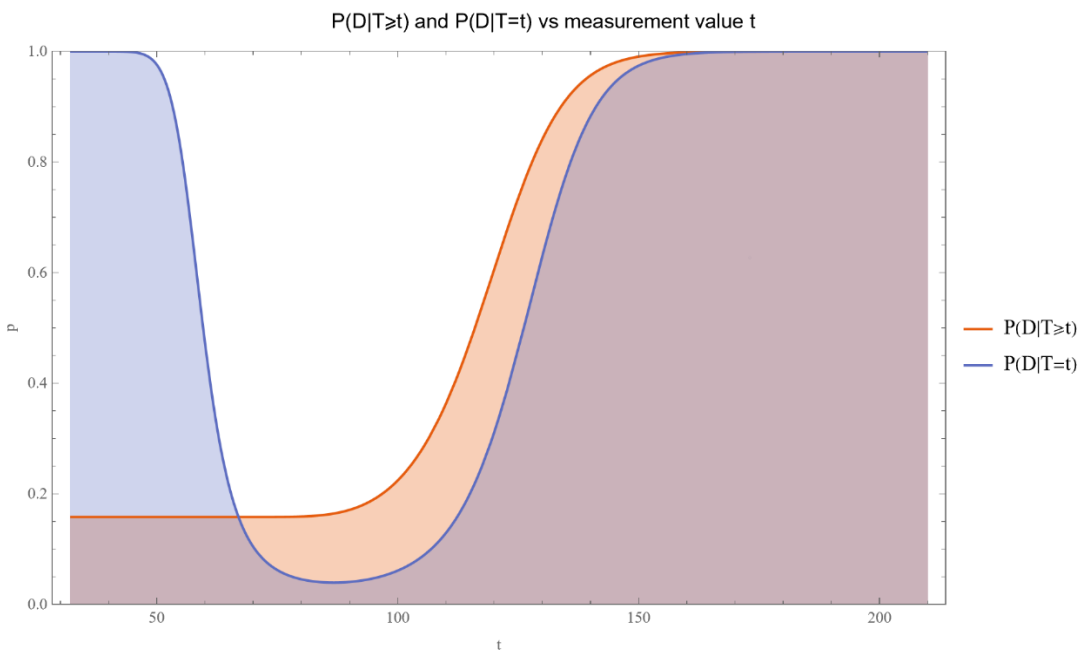

Figure 5. Positive predictive value and posterior probability for diabetes versus FPG value t (mg/dl) curves plot, with the program's settings in Table 2.

Figure 6 presents the plots of:

a) The negative predictive value $P(\bar{D}|T < t)$ of FPG for diabetes versus threshold value t (mg/dl) (orange curve). The curve is smooth and unimodal, with a maximum value of 0.96 at $t = 91.3$ 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 (mg/dl) (orange curve). The curve is smooth, unimodal, and approximately double sigmoidal. For an FPG value $t = 86.8$ 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.

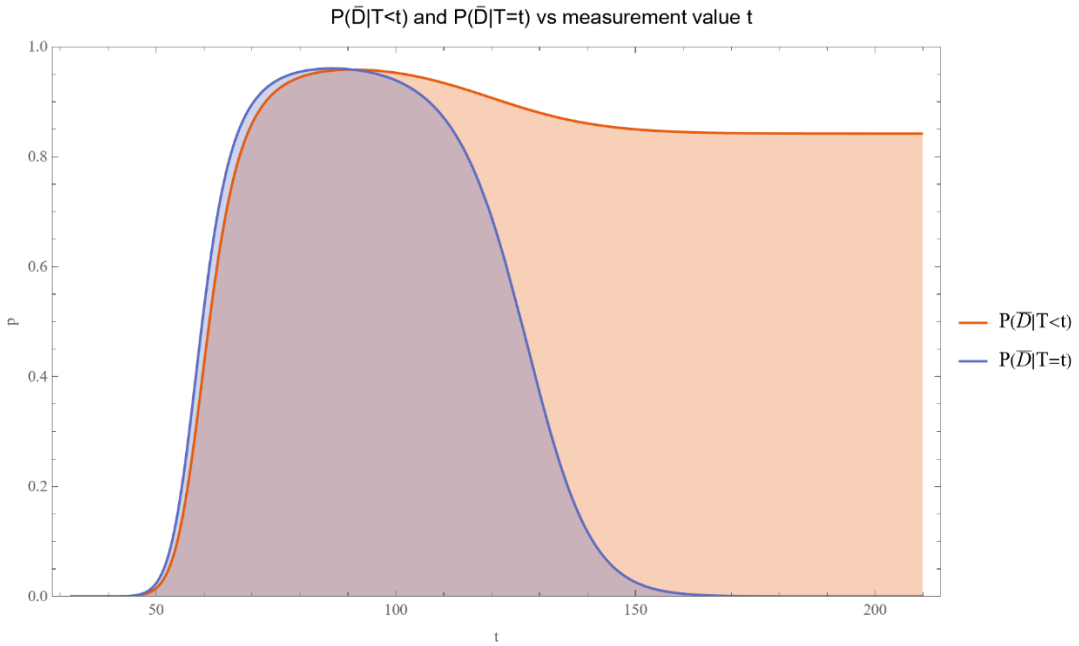

Figure 6. Negative predictive value for diabetes and posterior probability for the absence of diabetes versus FPG value t (mg/dl) curves plot, with the program's settings in Table 2.

Additionally:

a) For $t = 67.4$ mg/dl, we have $P(D|T \geq t) = P(D|T = t) = 0.158 = v$
b) For $t < 67.4$ mg/dl, we have $P(D|T \geq t) < P(D|T = t)$,
c) For $t > 67.4$ mg/dl, we have $P(D|T \geq t) > P(D|T = t)$.
d) For $t = 91.0$ mg/dl, we have $P(\bar{D}|T < t) = P(\bar{D}|T = t) = 0.96$.
e) For $t < 91.0$ mg/dl, we have $P(\bar{D}|T < t) < P(\bar{D}|T = t)$
f) For $t > 91.0$ mg/dl, we have $P(\bar{D}|T < t) > P(\bar{D}|T = t)$.

As shown in Figures 7 and 8, for an FPG value $t = 126.0$ mg/dl and varying 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) It is observed that $P(D|T \geq t) > P(D|T = t)$ and $P(\bar{D}|T < t) > P(\bar{D}|T = t)$.

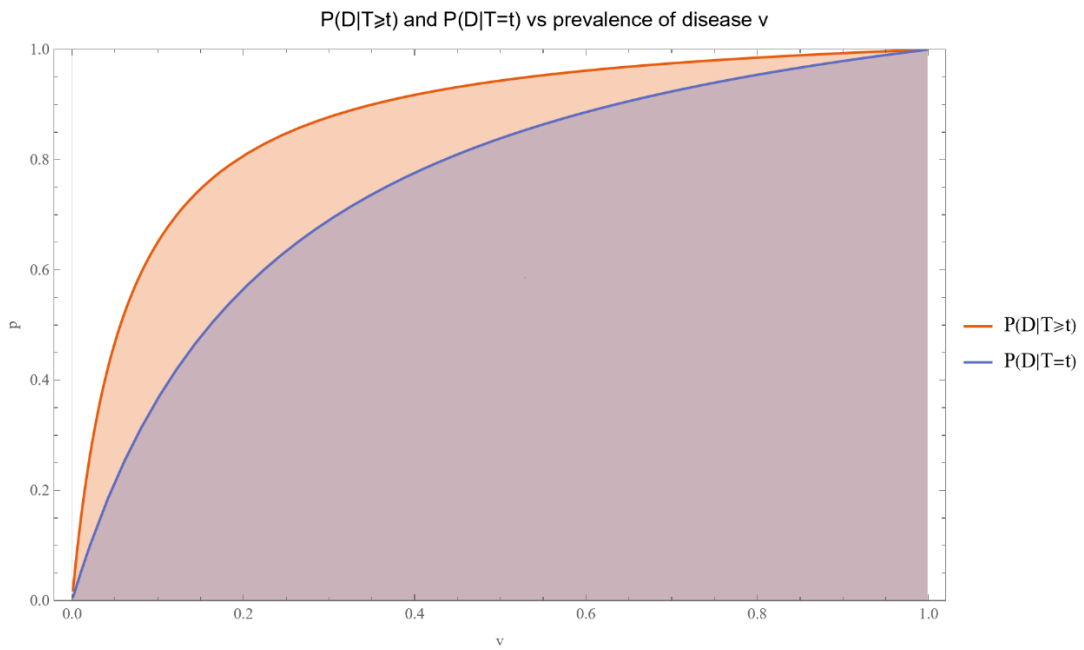

Figure 7. Positive predictive value and posterior probability for diabetes versus prior probability or prevalence of diabetes$v$ curves plot for an FPG value $t = 126$ mg/dl, with the other program settings in Table 2.

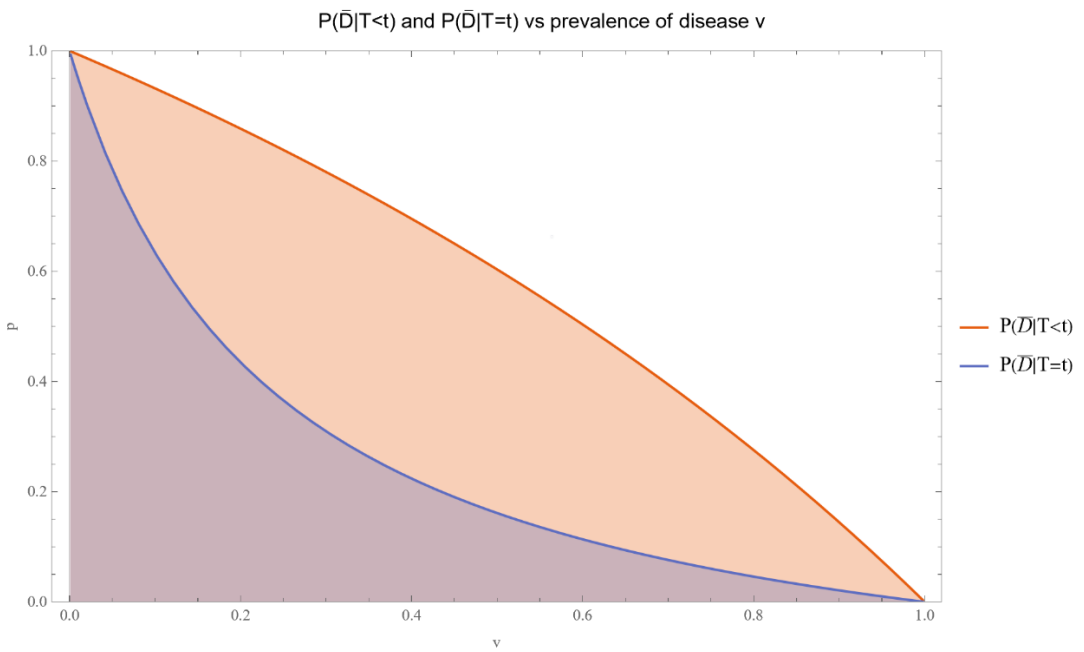

Figure 8. Negative predictive value for diabetes and posterior probability for the absence of diabetes versus prior probability or prevalence of diabetes$v$ curves plot, for an FPG value $t = 126$ mg/dl, with the other settings of the program in Table 2.

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 (ElSayed et al. 2023), 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.

3.2. Uncertainty

Table 3. The settings of the program *Bayesian Diagnostic Insights* for 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
$d_{\bar{D}}$		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 (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 (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 (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 (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 for the absence of diabetes $u_c[P(\bar{D}|T=t)]$:

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 exhibit an approximately double sigmoidal configuration. The maxima are quantitatively approximated as follows:
    a. At an FPG value of $t = 58.5$ mg/dl, the combined standard uncertainty is 0.898, for $P(D|T=t) = 0.581$and $P(\bar{D}|T=t) = 0.419$.
    b. At an FPG value of $t = 133.1$ mg/dl, the combined standard uncertainty is 0.190, where $P(D|T=t) = 0.726$ and $P(\bar{D}|T=t) = 0.274$.
    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 $t = 126.0$ mg/dl, where $P(D|T \geq t) = 0.758$.
    d. The standard combined uncertainty $u_c[P(\bar{D}|T < t)]$ of the negative predictive value for diabetes has a maximum value of 0.900 for $t = 58.5$ mg/dl, where $P(\bar{D}|T < t) = 0.321$.
    e. This pattern indicates heightened uncertainty in the regions where the diagnostic measures curves have their most pronounced inflections (Figures 5 and 6).

In addition:

a) For $t = 95.7$mg/dl, we have $u_c[P(D|T \geq t)] = u_c[P(D|T=t)] = 0.013$, while $P(D|T \geq t) = 0.193$ and $P(D|T=t) = 0.049$.
b) For $t = 126.7$ mg/dl, we have $u_c[P(D|T \geq t)] = u_c[P(D|T=t)] = 0.149$, while $P(D|T \geq t) = 0.774$ and $P(D|T=t) = 0.517$.
c) For $0 < t < 95.7$ mg/dl and $t > 126.7$ we have $u_c[P(D|T \geq t)] < u_c[P(D|T=t)]$.
d) For 95.7 mg/dl $< t < 1267$ 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 $u_c[P(\bar{D}|T < t)] = u_c[P(\bar{D}|T=t)] = 0.887$, while $P(\bar{D}|T < t) = 0.362$ and $P(\bar{D}|T=t) = 0.463$.
f) For $t = 103.8$ mg/dl, we have $u_c[P(\bar{D}|T < t)] = u_c[P(\bar{D}|T=t)] = 0.015$, while $P(\bar{D}|T < t) = 0.947$ and $P(\bar{D}|T=t) = 0.921$.
g) For $0 < t < 59.1$ mg/dl and $103.8 < t$ we have $u_c[P(\bar{D}|T < t)] < u_c[P(\bar{D}|T=t)]$.
h) For 59.1 mg/dl $< t <$ 103.8 mg/dl we have $u_c[P(\bar{D}|T=t)] < u_c[P(\bar{D}|T < t)]$.

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 both 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 $P(\bar{D}|T=t)$ (blue curves) are wider at the extremes of the t spectrum.
d) The confidence intervals of negative predictive value $P(\bar{D}|T < t)$ (orange curves) are wide at lower t values, to become considerably narrower at higher values.

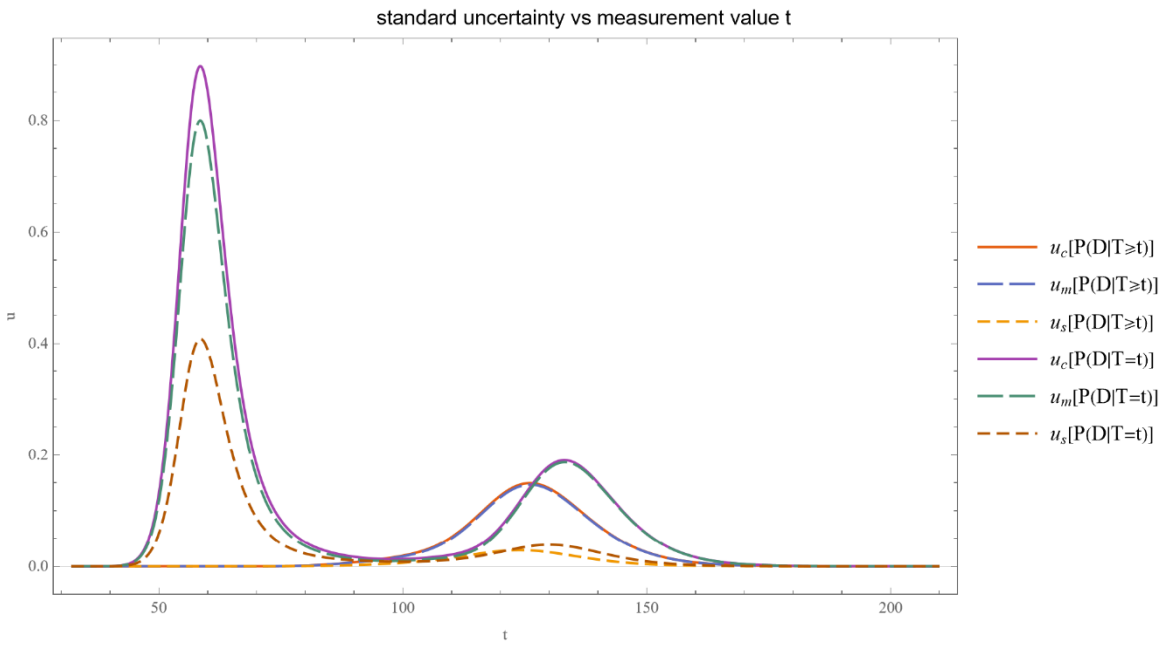

Figure 10. Standard sampling, measurement, and combined uncertainty of the positive predictive value and posterior probability for diabetes versus FPG value t (mg/dl) curves plot, with the program's settings in Table 3.

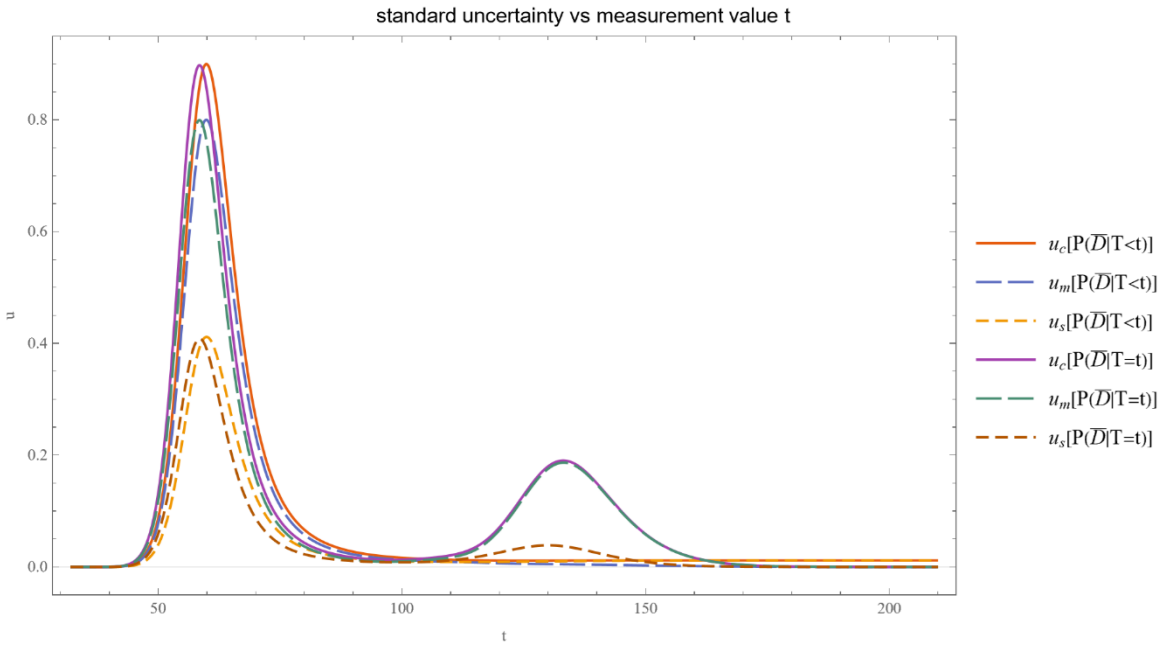

Figure 11. Standard sampling, measurement, and combined uncertainty of the negative predictive value for diabetes and posterior probability for the absence of diabetes versus FPG value t (mg/dl) curves plot, with the program's settings in Table 3.

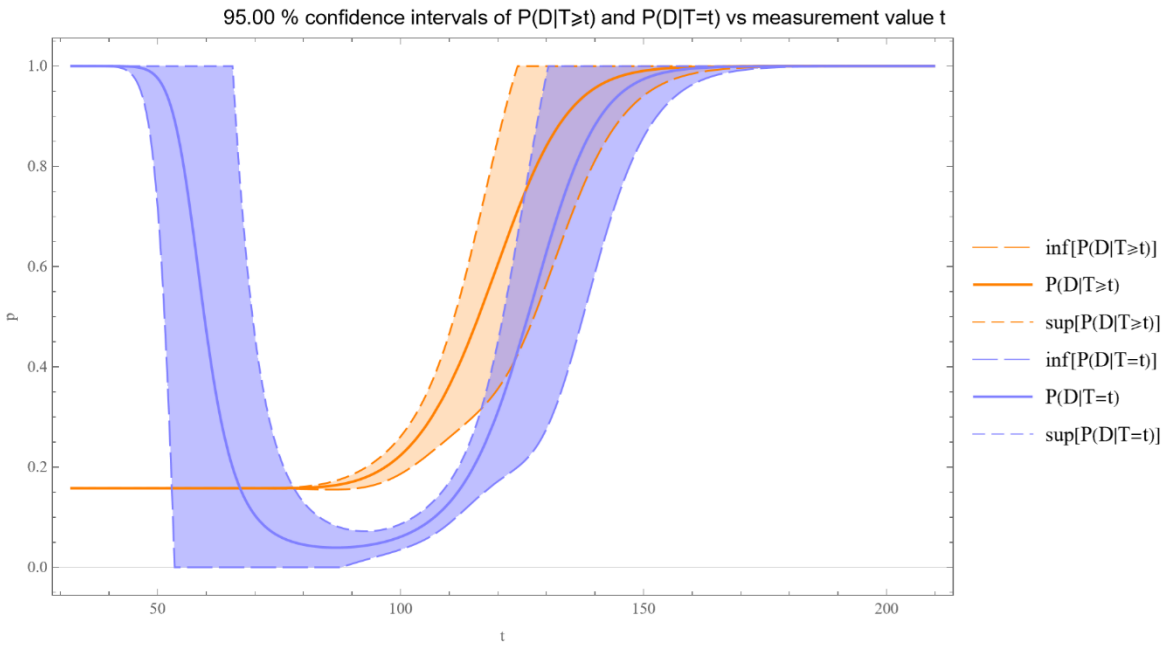

Figure 12. Confidence intervals of the positive predictive value and posterior probability for diabetes versus FPG value t (mg/dl) curves plot, with the program's settings in Table 3.

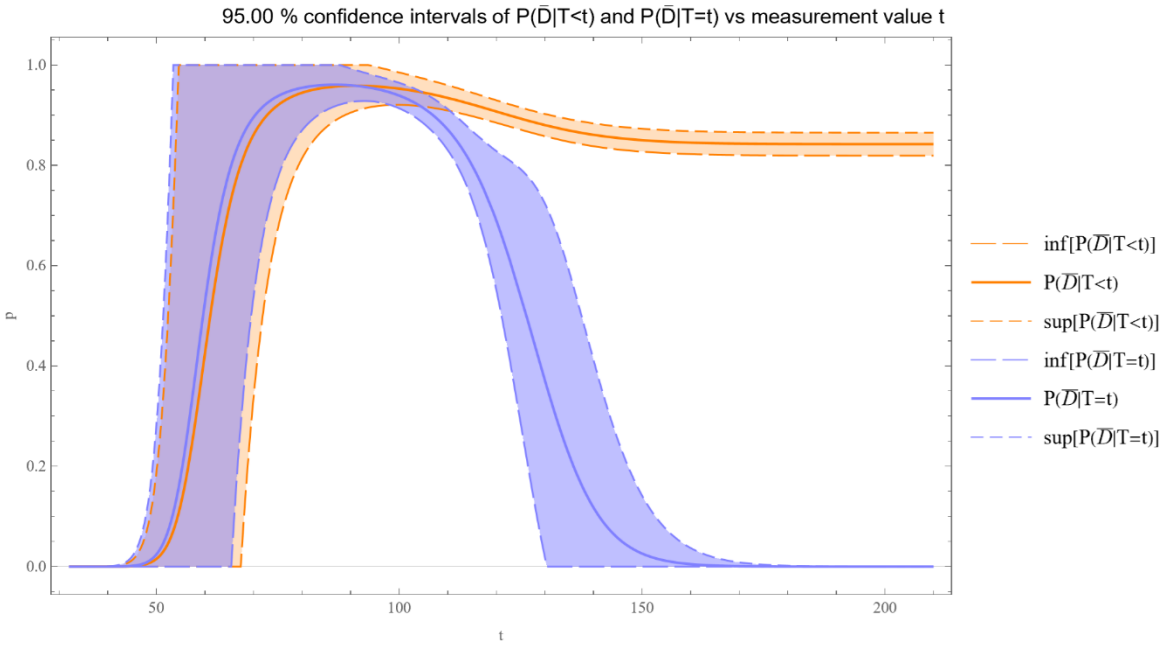

Figure 13. Confidence intervals of the negative predictive value and posterior probability for the absence of diabetes versus FPG value t (mg/dl) curves plot, with the program's settings in Table 3.

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 prior probability or prevalence of diabetes v. The combined uncertainty of the diagnostic measures is substantially affected by the measurement uncertainty of FPG. The curves are unimodal, with maxima approximately:

a) For $v = 0.055,\ u_c[P(D|T \geq t)] = 0.205$ where $P(D|T \geq t) = 0.493$.
b) For $v = 0.158,\ u_c[P(D|T = t)] = 0.141$ where $P(D|T = t) = 0.494$.
c) For $v = 0.631,\ u_c[P(\bar{D}|T < t)] = 0.023$ where $P(\bar{D}|T < t) = 0.471$.
d) For $v = 0.158,\ u_c[P(\bar{D}|T = t)] = 0.141$ where $P(\bar{D}|T = t) = 0.506$.

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

Additionally:

a) For $v = 0.173$ we have $u_c[P(D|T \geq t)] = u_c[P(D|T = t)] = 0.141$, $P(D|T \geq t) = 0.777$ and $P(D|T = t) = 0.521$.
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 $u_c[P(\bar{D}|T < t)] < u_c[P(\bar{D}|T = t)]$.

Notably, 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 $P(\bar{D}|T = t)$ for the absence of diabetes (Figure 17, blue curves) and negative predictive value $P(\bar{D}|T < t)$ (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 those of $P(D|T = t)$ (Figure 16, blue curves).
c) The confidence intervals of $P(\bar{D}|T < t)$(Figure 17, orange curves) are considerably narrower than those of $P(\bar{D}|T = t)$ (Figure 17, blue curves).

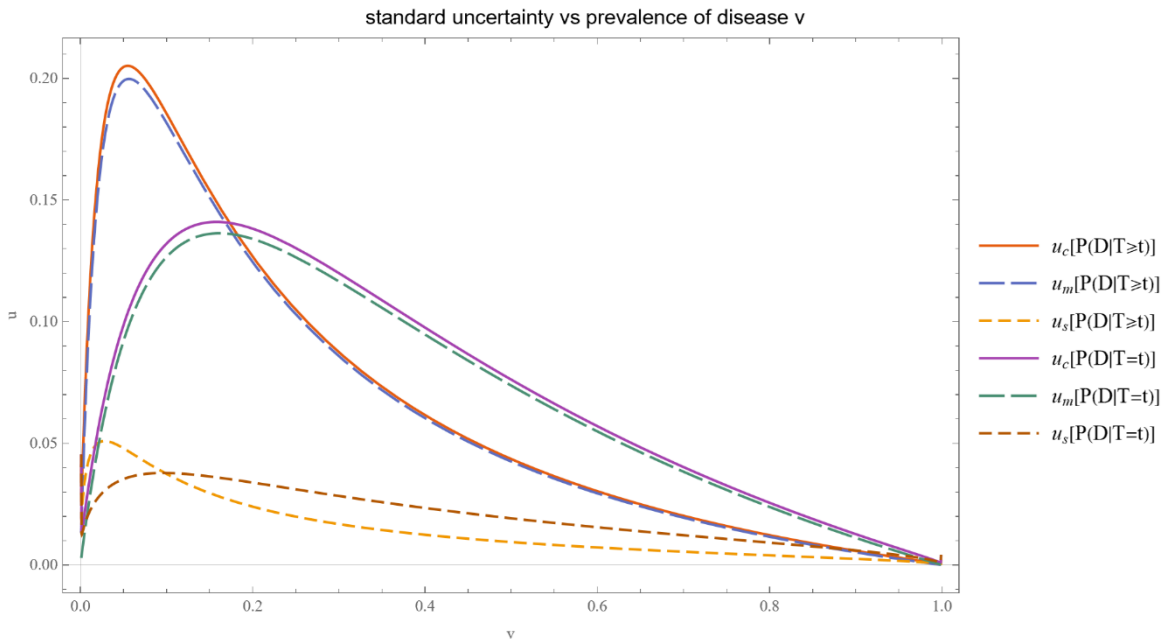

Figure 14. Standard sampling, measurement, and combined uncertainty of the positive predictive value and posterior probability for diabetes versus prior probability or prevalence of diabetes v curves plot, for an FPG value $t = 126$ mg/dl, with the other settings of the program in Table 3.

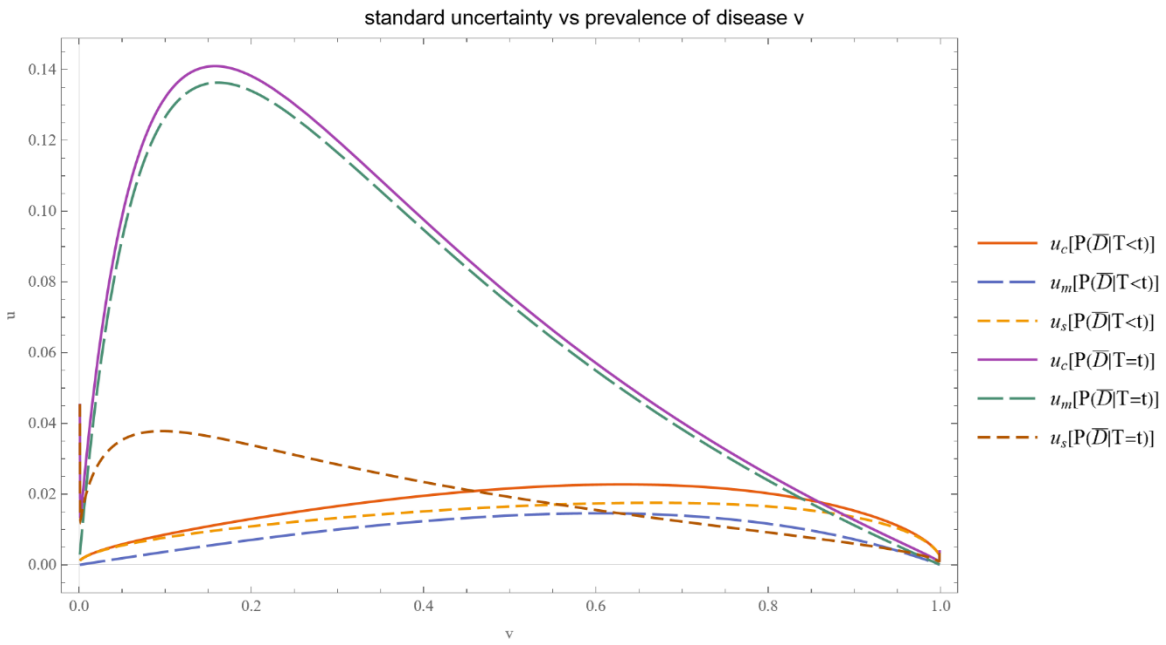

Figure 15. Standard sampling, measurement, and combined uncertainty of the negative predictive value for diabetes, and posterior probability for the absence of diabetes versus prior probability or prevalence of diabetes v curves plot, for an FPG value $t = 126$ mg/dl, with the other settings of the program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

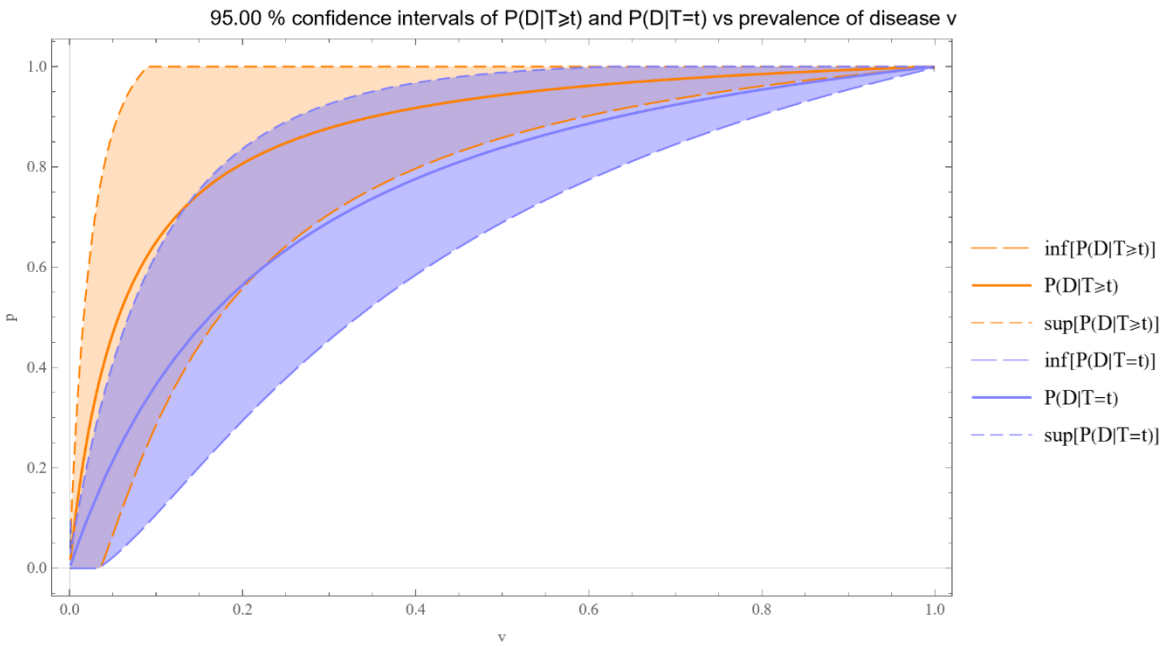

Figure 16. Confidence intervals of the positive predictive value and posterior probability for diabetes versus prior probability or prevalence of diabetes v curves plot, for an FPG value $t = 126$ mg/dl, with the other settings of the program in Table 3.

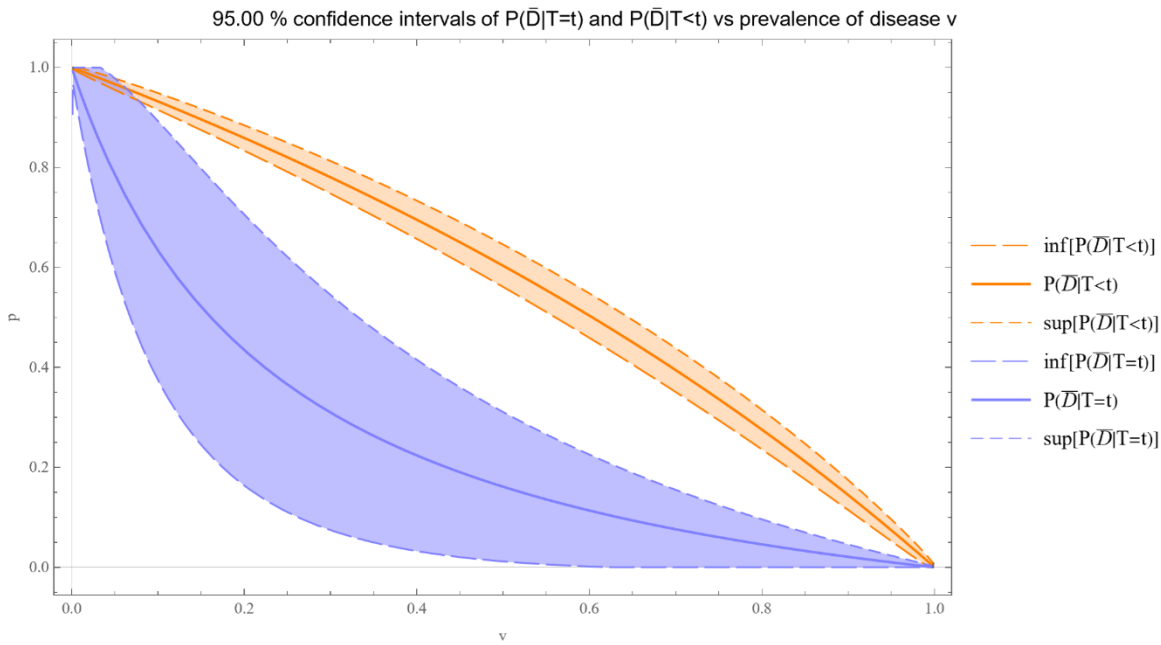

Figure 17. Confidence intervals of the negative predictive value for diabetes and posterior probability for the absence of diabetes versus prior probability or prevalence of diabetes v curves plot for an FPG value $t = 126\ \mathrm{mg/dl}$, with the other settings of the program in Table 3.

standard uncertainty				
prevalence of disease v = 0.158				
measure	point estimation	standard uncertainty		
		combined	measurement	sampling
P(D|T≥t)	0.758	0.149	0.147	0.029
P(D|T=t)	0.494	0.141	0.136	0.036
P(D̄|T<t)	0.890	0.011	0.006	0.010
P(D̄|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	0.465	1.000
P(D|T=t)	0.494	0.217	0.770
P(D̄|T<t)	0.890	0.868	0.912
P(D̄|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.

Figures 18 and 19 present tables of Bayesian diagnostic measures for FPG measurements at the diabetes diagnostic threshold $t = 126$ mg/dl, following the American Diabetes Association (ADA) guidelines. The standard for diagnosing diabetes used in this study is the oral glucose tolerance test (OGTT) with a 200 mg/dl threshold. The limited concordance between these two diagnostic thresholds is evident from the point estimations and their associated uncertainty. For an FPG diagnostic threshold $t = 126$ mg/dl:

a) $P(D|T \geq t) = 0.758$, with a confidence interval of (0.465 - 1.000).
b) $P(D|T = t) = 0.494$, with a confidence interval of (0.217 - 0.770).
c) $P(\bar{D}|T < t) = 0.890$, with a confidence interval of (0.868 - 0.912).
d) $P(\bar{D}|T = t) = 0.506$, with a confidence interval of (0.230 - 0.783).

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) $P(\bar{D}|T = t) < P(\bar{D}|T < t)$
e) The size of the confidence interval of $P(\bar{D}|T < t)$ is considerably less than the size of the confidence interval of $P(\bar{D}|T = t)$.
f) There is no overlap between the confidence intervals of $P(\bar{D}|T < t)$ and $P(\bar{D}|T = t)$.

In addition, the table with the standard uncertainty of the Bayesian diagnostic measures of Figure 18 shows that for $t = 126$ 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.

4. Discussion

There is a persistent need to estimate diagnostic measures and their uncertainty, especially concerning screening and diagnostic tests for potentially life-threatening diseases. The COVID-19 pandemic has highlighted this necessity (Lippi, Simundic, and Plebani 2020; 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 2011; Tang et al. 2020; Deeks et al. 2020; Infantino et al. 2020; Mahase 2020).

Traditional diagnostic approaches often rely on fixed thresholds, which may overlook 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(Choi, Johnson, and Thurmond 2006; Chatzimichail and Hatjimihail 2023). Bayes' theorem provides a statistical framework to update the probability estimate of a disease as new information or test results become available, enabling healthcare professionals to refine disease probability estimates based on new data and prior knowledge.

We developed the software tool introduced in this study to facilitate the application of Bayes' theorem in medical diagnosis. It allows for the exploration and comparison of two pairs of Bayesian diagnostic measures for screening or diagnostic tests, assuming parametric distributions of the measurements:

a) The positive predictive value and the posterior probability for disease and
b) The negative predictive value and 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 (Smith and Gelfand 1992). Therefore, exploratory data analysis and fitting of statistical distributions to diagnostic measurement data may be necessary to apply the software tool effectively (Forbes et al. 2011). Our previously developed Bayesian Inference program may be helpful in this regard (Chatzimichail and Hatjimihail 2023).

Our choice of parametric distributions was motivated by their broad applicability in modeling medical diagnostic measurements:

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*
A lognormal distribution is appropriate for modeling positively skewed data, where the logarithm of the variable follows a normal distribution. Defined by a location parameter and a scale parameter of the underlying normal distribution, it can model data that cannot assume negative values and exhibit a long right tail, such as many biological measurements.

c) *Gamma distribution*
The gamma distribution is suitable for data with varying skewness and kurtosis that a lognormal distribution cannot adequately model. It is characterized by 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.

In our illustrative case study, we implemented an empirical Bayesian approach due to several advantages:

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 (Srinivasan, Westover, and Bianchi 2012; Chatzimichail and Hatjimihail 2021, 2024). 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 (Wereski et al. 2021).

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 (Roberts et al. 2015).

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 (Freund et al. 2021).

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 (ElSayed et al. 2023).

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 (Rani and Begum 2016).

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 (Reyes Domingo, Avey, and Doull 2019).

Our software allows the estimation and plotting of the sampling, measurement, and combined uncertainty of Bayesian diagnostic measures and 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 (Greenland et al. 2016).

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.

Diagnostic uncertainty quantification is imperative in quality and risk management in laboratory medicine and may contribute to the design and implementation of test accuracy studies (Horvath et al. 2024). Despite extensive research on Bayesian diagnosis and uncertainty as separate areas, their intersection remains relatively unexplored (Baron 1994; Ashby and Smith 2000).

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 challenges in 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.

A detailed analysis of Figures 5-8, 12,13, 16, and 17 from the illustrative case study reveals several clinical implications:

a) *Influence of Threshold and Prevalence on Positive Predictive Value*: 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) *Double-Threshold Pattern in Posterior Probability*: The double-threshold pattern observed 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) *Variability in Confidence Intervals at Intermediate FPG Levels*: 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) *Significance of Threshold Selection for Negative Predictive Value*: The differing trends in negative predictive value $P(\bar{D}|T < t)$ highlight the significance of selecting the appropriate threshold for excluding diabetes.
e) *Unique Behavior of Posterior Probability for Absence of Disease*: The unique behavior of Bayesian posterior probability $P(\bar{D}|T = t)$ 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) *Robustness of Negative Predictive Values*: Despite the interpretative challenges of $P(\bar{D}|T < t)$ at lower FPG values, it is generally more robust than $P(\bar{D}|T = t)$ 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 (Tucker 2020; Sacks et al. 2023).
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 Bayes' theorem while leveraging the robustness of frequentist techniques for estimating sampling and measurement uncertainty.

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 (Gelman et al. 2013; Congdon 2021). Additionally, applying Bayes' factors to compare the evidence provided by different diagnostic measures represents a promising area for further investigation (Kass and Raftery 1995; Bozza, Taroni, and Biedermann 2022). These advancements could enhance the robustness and applicability of Bayesian methods in medical diagnostics, overcoming their current limitations (Willink and White 2012; Willink 2013).

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 (J. Andre Knottnerus and Buntinx 2011). 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. 0.19.1, Mathematica® ver. 14.1, Matlab® ver. R2024a, MedCalc® ver. 23.0.2, metRology ver. 1.1-3, NCSS® ver. 24.0.3, NIST Uncertainty Machine ver. 1.6.2, OpenBUGS ver. 3.2.3, R ver. 4.4.1, SAS Viya® ver. 2024.09, SPSS® ver. 30.0.0, Stan ver. 2.35, Stata® ver. 19, and UQLab ver. 2.0) include routines for calculating and plotting various diagnostic measures and their confidence intervals. However, the program presented in this work provides 34 types of plots and 16 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 (Chatzimichail and Hatjimihail 2018, 2021, 2023, 2024), facilitating their comparison.

4.1. Limitations of the Program

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

a) Underlying assumptions

a. Existence of "Gold Standards" in Diagnostics: The program assumes the availability of a "gold standard" for disease classification. Without a "gold standard", alternative approaches like latent class models or expert consensus methods may be necessary (J. A. Knottnerus and Dinant 1997; Pfeiffer and Castle 2005; Nair, Aggarwal, and Khanna 2011; van Smeden et al. 2014).
b. Assumption of Specific Distributions: The tool assumes that the measurements or their transformations follow normal, lognormal, or gamma distributions. While these distributions are often used in biomedical data, they may not accurately represent the underlying data characteristics. Literature on reference intervals, diagnostic thresholds, and clinical decision limits provides alternative distribution models that could be considered (Solberg 1987; Pavlov, Wilson, and Delgado 2012; Sikaris 2012; Daly et al. 2013; Ozarda et al. 2018).
c. Assumption of Bimodality**:** The program generally accepts that the measurements are bimodally distributed, corresponding to diseased and nondiseased populations. However, in some cases, an unimodal distribution might be more appropriate (J. M. G. Wilson and Jungner 1968; Petersen and Horder 1992).

b) Approximations used for the estimations

a. Uncertainty Approximation in Disease Prevalence: The uncertainty associated with a disease's prevalence is approximated using the Agresti–Coull-adjusted Wald interval. Although this method is widely used, more accurate techniques are available, especially for small sample sizes or extreme probabilities (Pires and Amado 2008).
b. Sampling Uncertainty Approximations: The program approximations of the sampling uncertainty for sample means and standard deviations may be less reliable for small sample sizes or when the data exhibit significant skewness, as is often the case with lognormal and gamma distributions (Schmoyeri et al. 1996; Bhaumik, Kapur, and Gibbons 2009).

c. First-Order Taylor Series Approximations: The program employs first-order Taylor series approximations for uncertainty propagation. While this method simplifies calculations, it may not capture the complexity of uncertainty in nonlinear functions. Higher-order approximations or Monte Carlo simulations could provide more accurate results (Joint Committee for Guides in Metrology 2008, 2011).
d. Confidence Intervals Based on the t-Distribution: Confidence intervals are derived using the t-distribution, which, despite the high relative uncertainty (Williams 2020), is a practical choice in selected scenarios, particularly in metrology (Willink and White 2012; Willink 2013; Gelman et al. 2013; Stephens 2023). Alternatives like credible intervals in a Bayesian framework could provide more accurate uncertainty quantification of nonlinear functions, especially for small samples.
e. Truncation to the [0,1] Range: Truncation of the estimated standard uncertainty and the confidence intervals to the [0,1] range is implemented since probabilities cannot logically assume values less than zero or greater than one. However, this approach may distort the uncertainty representation. Quantile-derived credible intervals inherently avoid truncation by constructing intervals within the [0,1] range.

While addressing these limitations would considerably increase computational complexity, they represent critical areas for future enhancement (Joint Committee for Guides in Metrology 2008, 2020). We should, however, keep in mind that "all models will be based on assumptions and can only approach complex reality" (Oosterhuis 2017), as "all models are wrong, but some models are useful" (Box 1979).

4.2. 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 (Rao, Disraeli, and McGregor 2004; Meneilly and Elliott 1999; Geer and Shen 2009; Van Cauter, Polonsky, and Scheen 1997; Colberg et al. 2010; Salmerón et al. 1997; Surwit et al. 2002; Pandit et al. 1993; Dupuis et al. 2010).
b) Approximation of the FPG measurements distributions from NHANES datasets by lognormal distributions.
c) The implied assumption of simple random sampling.

5. Conclusion

Bayesian Diagnostic Insights provides modules for estimating, visualizing, and comparing Bayesian diagnostic measures, including their associated uncertainty. Exploring the uncertainty of disease probability estimates can assist in the clinical decision-making process. The illustrative case study using fasting plasma glucose (FPG) for diabetes diagnosis demonstrates the impact of measurement uncertainty on diagnostic measures, highlighting its relevance in clinical and laboratory practices. While the software offers a framework for applying Bayes' theorem in medical diagnostics, further research is needed to fully assess its utility in diagnosing various health conditions.

6. Supplemental Material

The following supplemental files are available at https://www.hcsl.com/Supplements/SBDI.zip (accessed on November 7, 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.

7. Declarations

Funding: This research received no external funding.

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.

8. 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

$\bar{D}$: absence of disease

T: diagnostic test result

A.1.3. Parameters

t: diagnostic threshold

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

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

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

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

n_D : size of the diseased population sample

m_D: mean of the measurements of the diseased population sample

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

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

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

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

v : prior probability for disease or 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

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

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

$P(\bar{D}|T = t)$: 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

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

${}_{l}v_{eff}(x)$: 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: $maximum(0, minimum(m_{\bar{D}} - 5s_{\bar{D}}, m_D - 5s_{\bar{D}})) - maximum(m_{\bar{D}} + 5s_{\bar{D}}, m_D + 5s_{\bar{D}})$

n_D : 2 – 10,000

m_D: 0.1 – 10,000

s_D: 0.01 – 1,000

$n_{\bar{D}}$: 2 – 10,000

$m_{\bar{D}}$: 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: 2.1.0

Project home page: https://www.hcsl.com/Tools/BayesianDiagnosticInsights/ (accessed on October 4, 2024)

Program source: BayesianDiagnosticInsights.nb . Available at: https://www.hcsl.com/Tools/BayesianDiagnosticInsights/BayesianDiagnosticInsights.nb (accessed on Npvember 7, 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.

9. References

Agresti, Alan, and Brent A. Coull. 1998. "Approximate Is Better than 'Exact' for Interval Estimation of Binomial Proportions." *The American Statistician* 52 (2): 119–26.

Agresti, Alan, Christine Franklin, and Bernhard Klingenberg. 2023. *Statistics: The Art and Science of Learning from Data, Global Edition*. 4th ed. London, England: Pearson Education.

American Diabetes Association. 2021. "2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2021." *Diabetes Care* 44 (Suppl 1): S15–33.

Ashby, D., and A. F. Smith. 2000. "Evidence-Based Medicine as Bayesian Decision-Making." *Statistics in Medicine* 19 (23): 3291–3305.

Ayyub, Bilal M., and George J. Klir. 2006. *Uncertainty Modeling and Analysis in Engineering and the Sciences*. Chapman and Hall/CRC.

Baron, J. A. 1994. "Uncertainty in Bayes." *Medical Decision Making: An International Journal of the Society for Medical Decision Making* 14 (1): 46–51.

Bates, Douglas M., and Donald G. Watts. 1988. *Nonlinear Regression Analysis and Its Applications*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Bayes, Mr, and Mr Price. 1763. "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." *Philosophical transactions of the Royal Society of London* 53 (0): 370–418.

Bhaumik, Dulal K., Kush Kapur, and Robert D. Gibbons. 2009. "Testing Parameters of a Gamma Distribution for Small Samples." *Technometrics: A Journal of Statistics for the Physical, Chemical, and Engineering Sciences* 51 (3): 326–34.

Bours, Martijn Jl. 2021. "Bayes' Rule in Diagnosis." *Journal of Clinical Epidemiology* 131 (March): 158–60.

Box, G. E. P. 1979. “Robustness in the Strategy of Scientific Model Building.” In *Robustness in Statistics*, 201–36. Elsevier.

Bozza, Silvia, Franco Taroni, and Alex Biedermann. 2022. *Bayes Factors for Forensic Decision Analyses with R*. 1st ed. Springer Texts in Statistics. Cham, Switzerland: Springer International Publishing.

Casella, George. 1985. “An Introduction to Empirical Bayes Data Analysis.” *The American Statistician* 39 (2): 83–87.

———. 1992. “Illustrating Empirical Bayes Methods.” *Chemometrics and Intelligent Laboratory Systems* 16 (2): 107–25.

Chatzimichail, Theodora, and Aristides T. Hatjimihail. 2018. “Relation of Diagnostic Accuracy Measures.” XIV. Hellenic Complex Systems Laboratory. https://heracleitos.github.io/TR/hcsltr14/index.html.

———. 2021. “A Software Tool for Calculating the Uncertainty of Diagnostic Accuracy Measures.” *Diagnostics (Basel, Switzerland)* 11 (3). https://doi.org/10.3390/diagnostics11030406.

———. 2023. “A Bayesian Inference Based Computational Tool for Parametric and Nonparametric Medical Diagnosis.” *Diagnostics* 13 (19): 3135.

———. 2024. “A Software Tool for Estimating Uncertainty of Bayesian Posterior Probability for Disease.” *Diagnostics (Basel, Switzerland)* 14 (4). https://doi.org/10.3390/diagnostics14040402.

Choi, Young-Ku, Wesley O. Johnson, and Mark C. Thurmond. 2006. “Diagnosis Using Predictive Probabilities without Cut-Offs.” *Statistics in Medicine* 25 (4): 699–717.

Colberg, Sheri R., Ronald J. Sigal, Bo Fernhall, Judith G. Regensteiner, Bryan J. Blissmer, Richard R. Rubin, Lisa Chasan-Taber, et al. 2010. “Exercise and Type 2 Diabetes: The American College of Sports Medicine and the American Diabetes Association: Joint Position Statement.” *Diabetes Care* 33 (12): e147-67.

Congdon, Peter D. 2021. *Bayesian Hierarchical Models: With Applications Using R, Second Edition*. 2nd ed. Philadelphia, PA: Chapman & Hall/CRC. https://books.google.com/books/about/Bayesian_Hierarchical_Models.html?hl=el&id=hlivDwAAQBAJ.

Daly, Caitlin H., Xiaofeng Liu, Vijay L. Grey, and Jemila S. Hamid. 2013. “A Systematic Review of Statistical Methods Used in Constructing Pediatric Reference Intervals.” *Clinical Biochemistry* 46 (13–14): 1220–27.

Darling, D. A. 1957. “The Kolmogorov-Smirnov, Cramer-von Mises Tests.” *Annals of Mathematical Statistics* 28 (4): 823–38.

Deeks, Jonathan J., Jacqueline Dinnes, Yemisi Takwoingi, Clare Davenport, Mariska M. G. Leeflang, René Spijker, Lotty Hooft, Ann Van den Bruel, Devy Emperador, and Sabine Dittrich. 2020. “Diagnosis of SARS-CoV-2 Infection and COVID-19: Accuracy of Signs and Symptoms; Molecular, Antigen, and Antibody Tests; and Routine Laboratory Markers.” Edited by Cochrane Infectious Diseases Group. *Cochrane Database of Systematic Reviews* 26 (April): 1896.

Djulbegovic, Benjamin, Jef van den Ende, Robert M. Hamm, Thomas Mayrhofer, Iztok Hozo, Stephen G. Pauker, and International Threshold Working Group (ITWG). 2015. “When Is Rational to Order a Diagnostic Test, or Prescribe Treatment: The Threshold Model as an Explanation of Practice Variation.” *European Journal of Clinical Investigation* 45 (5): 485–93.

Dupuis, Josée, Claudia Langenberg, Inga Prokopenko, Richa Saxena, Nicole Soranzo, Anne U. Jackson, Eleanor Wheeler, et al. 2010. “New Genetic Loci Implicated in Fasting Glucose Homeostasis and Their Impact on Type 2 Diabetes Risk.” *Nature Genetics* 42 (2): 105–16.

Ellison, Stephen L. R., and A. Williams. 2012. “Quantifying Uncertainty in Analytical Measurement.” CG 4. 3rd ed. EURACHEM/CITAC.

ElSayed, Nuha A., Grazia Aleppo, Vanita R. Aroda, Raveendhara R. Bannuru, Florence M. Brown, Dennis Bruemmer, Billy S. Collins, et al. 2023. "2. Classification and Diagnosis of Diabetes: Standards of Care in Diabetes-2023." *Diabetes Care* 46 (Suppl 1): S19–40.

Fischer, Felix. 2021. "Using Bayes Theorem to Estimate Positive and Negative Predictive Values for Continuously and Ordinally Scaled Diagnostic Tests." *International Journal of Methods in Psychiatric Research* 30 (2): e1868.

Forbes, Catherine, Merran Evans, Nicholas Hastings, and Brian Peacock. 2011. *Statistical Distributions*. John Wiley & Sons.

Freund, Yonathan, Anthony Chauvin, Sonia Jimenez, Anne-Laure Philippon, Sonja Curac, Florent Fémy, Judith Gorlicki, et al. 2021. "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: The Journal of the American Medical Association* 326 (21): 2141–49.

Geer, Eliza B., and Wei Shen. 2009. "Gender Differences in Insulin Resistance, Body Composition, and Energy Balance." *Gender Medicine* 6 Suppl 1 (Suppl 1): 60–75.

Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. 2013. *Bayesian Data Analysis*. CRC Press.

Greenland, Sander, Stephen J. Senn, Kenneth J. Rothman, John B. Carlin, Charles Poole, Steven N. Goodman, and Douglas G. Altman. 2016. "Statistical Tests, P Values, Confidence Intervals, and Power: A Guide to Misinterpretations." *European Journal of Epidemiology* 31 (4): 337–50.

Horvath, Andrea Rita, Katy J. L. Bell, Ferruccio Ceriotti, Graham R. D. Jones, Tze Ping Loh, Sally Lord, Sverre Sandberg, and Task Group Analytical Performance Specifications based on Outcomes of the European Federation of Clinical Chemistry and Laboratory Medicine. 2024. "Outcome-Based Analytical Performance Specifications: Current Status and Future Challenges." *Clinical Chemistry and Laboratory Medicine: CCLM / FESCC*, June. https://doi.org/10.1515/cclm-2024-0125.

Infantino, Maria, Valentina Grossi, Barbara Lari, Riccardo Bambi, Alessandro Perri, Matteo Manneschi, Giovanni Terenzi, et al. 2020. "Diagnostic Accuracy of an Automated Chemiluminescent Immunoassay for Anti-SARS-CoV-2 IgM and IgG Antibodies: An Italian Experience." *Journal of Medical Virology*, April. https://doi.org/10.1002/jmv.25932.

J. Aitchison, J. A. C. Brown. 1957. *The Lognormal Distribution with Special Reference to Its Uses in Econometrics*. Cambridge: Cambridge University Press.

Johnson, Michael L. 2008. "Nonlinear Least-Squares Fitting Methods." In *Methods Cell Biol.*, 84:781–805. Academic Press.

Joint Committee for Guides in Metrology. 2008. "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." JCGM 101:2008. Pavillon de Breteuil, F-92312 Sèvres, Cedex, France: BIPM. https://www.bipm.org/documents/20126/2071204/JCGM_101_2008_E.pdf/325dcaad-c15a-407c-1105-8b7f322d651c.

———. 2011. "Evaluation of Measurement Data – Supplement 2 to the 'Guide to the Expression of Uncertainty in Measurement' – Extension to Any Number of Output Quantities." JCGM 102:2011. Pavillon de Breteuil, F-92312 Sèvres, Cedex, France: BIPM. https://www.bipm.org/documents/20126/2071204/JCGM_102_2011_E.pdf/6a3281aa-1397-d703-d7a1-a8d58c9bf2a5.

———. 2020. "Guide to the Expression of Uncertainty in Measurement — Part 6: Developing and Using Measurement Models." JCGM GUM-6:2020. Pavillon de Breteuil, F-92312 Sèvres, Cedex, France: BIPM.

https://www.bipm.org/documents/20126/2071204/JCGM_GUM_6_2020.pdf/d4e77d99-3870-0908-ff37-c1b6a230a337.

Joyce, James. 2021. “Bayes’ Theorem.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Fall 2021. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2021/entries/bayes-theorem/.

Kallner, Anders, James C. Boyd, David L. Duewer, Claude Giroud, Aristides T. Hatjimihail, George G. Klee, Stanley F. Lo, et al. 2012. *Expression of Measurement Uncertainty in Laboratory Medicine; Approved Guideline*. Clinical and Laboratory Standards Institute.

Kass, Robert E., and Adrian E. Raftery. 1995. “Bayes Factors.” *Journal of the American Statistical Association* 90 (430): 773–95.

Knottnerus, J. A., and G. J. Dinant. 1997. “Medicine Based Evidence, a Prerequisite for Evidence Based Medicine.” *BMJ* 315 (7116): 1109–10.

Knottnerus, J. Andre, and Frank Buntinx, eds. 2011. *The Evidence Base of Clinical Diagnosis*. EPUB. 2nd ed. Evidence-Based Medicine. BMJ Books.

Lippi, Giuseppe, Ana-Maria Simundic, and Mario Plebani. 2020. “Potential Preanalytical and Analytical Vulnerabilities in the Laboratory Diagnosis of Coronavirus Disease 2019 (COVID-19).” *Clinical Chemistry and Laboratory Medicine: CCLM / FESCC*, March. https://doi.org/10.1515/cclm-2020-0285.

M H Ramsey S L R Ellison P Rostron. 2019. “Measurement Uncertainty Arising from Sampling - A Guide to Methods and Approaches.” 2nd ed. EURACHEM/CITAC.

Mahase, Elisabeth. 2020. “Covid-19: ‘Unacceptable’ That Antibody Test Claims Cannot Be Scrutinised, Say Experts.” *BMJ* 369 (May): m2000.

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. 2011. *Assessment of the Diagnostic Accuracy of Laboratory Tests Using Receiver Operating Characteristic Curves; Approved Guideline—Second Edition*. Clinical and Laboratory Standards Institute.

Meneilly, G. S., and T. Elliott. 1999. “Metabolic Alterations in Middle-Aged and Elderly Obese Patients with Type 2 Diabetes.” *Diabetes Care* 22 (1): 112–18.

Miller, James, and Jane C. Miller. 2018. *Statistics and Chemometrics for Analytical Chemistry*. 7th ed. London, England: Pearson Education.

Myung, In Jae. 2003. “Tutorial on Maximum Likelihood Estimation.” *Journal of Mathematical Psychology* 47 (1): 90–100.

Nair, Raj, Rohit Aggarwal, and Dinesh Khanna. 2011. “Methods of Formal Consensus in Classification/Diagnostic Criteria and Guideline Development.” *Seminars in Arthritis and Rheumatism* 41 (2): 95–105.

National Center for Health Statistics. 2005-20016. “National Health and Nutrition Examination Survey Data.” Centers for Disease Control and Prevention. 2005-20016. https://wwwn.cdc.gov/nchs/nhanes/default.aspx.

———. 2005-20016. “National Health and Nutrition Examination Survey Questionnaire.” Centers for Disease Control and Prevention. 2005-20016. https://wwwn.cdc.gov/nchs/nhanes/Search/variablelist.aspx?Component=Questionnaire.

Oosterhuis, Wytze P. 2017. “Analytical Performance Specifications in Clinical Chemistry: The Holy Grail?” *Journal of Laboratory and Precision Medicine* 2 (September): 78–78.

Oosterhuis, Wytze P., and Elvar Theodorsson. 2016. “Total Error vs. Measurement Uncertainty: Revolution or Evolution?” *Clinical Chemistry and Laboratory Medicine: CCLM / FESCC* 54 (2): 235–39.

Ozarda, Yesim, Ken Sikaris, Thomas Streichert, Joseph Macri, and IFCC Committee on Reference intervals and Decision Limits (C-RIDL). 2018. “Distinguishing Reference Intervals and Clinical Decision Limits - A Review by the IFCC Committee on Reference Intervals and Decision Limits.” *Critical Reviews in Clinical Laboratory Sciences* 55 (6): 420–31.

Pandit, M. K., J. Burke, A. B. Gustafson, A. Minocha, and A. N. Peiris. 1993. “Drug-Induced Disorders of Glucose Tolerance.” *Annals of Internal Medicine* 118 (7): 529–39.

Pavlov, Igor Y., Andrew R. Wilson, and Julio C. Delgado. 2012. “Reference Interval Computation: Which Method (Not) to Choose?” *Clinica Chimica Acta; International Journal of Clinical Chemistry* 413 (13–14): 1107–14.

Petersen, Per Hyltoft, and Morgens Horder. 1992. “2.3 Clinical Test Evaluation. Unimodal and Bimodal Approaches.” *Scandinavian Journal of Clinical and Laboratory Investigation* 52 (sup208): 51–57.

Petrone, S., J. Rousseau, and C. Scricciolo. 2014. “Bayes and Empirical Bayes: Do They Merge?” *Biometrika* 101 (2): 285–302.

Pfeiffer, Ruth M., and Philip E. Castle. 2005. “With or without a Gold Standard.” *Epidemiology* .

Pires, Ana M., and Conceição Amado. 2008. “Interval Estimators for a Binomial Proportion: Comparison of Twenty Methods.” *Revstat Statistical Journal* 6 (2): 165–97.

Rani, P. Reddi, and Jasmina Begum. 2016. “Screening and Diagnosis of Gestational Diabetes Mellitus, Where Do We Stand.” *Journal of Clinical and Diagnostic Research: JCDR* 10 (4): QE01-4.

Rao, Shobha S., Phillip Disraeli, and Tamara McGregor. 2004. “Impaired Glucose Tolerance and Impaired Fasting Glucose.” *American Family Physician* 69 (8): 1961–68.

Reyes Domingo, Francesca, Marc T. Avey, and Marion Doull. 2019. “Screening for Thyroid Dysfunction and Treatment of Screen-Detected Thyroid Dysfunction in Asymptomatic, Community-Dwelling Adults: A Systematic Review.” *Systematic Reviews* 8 (1): 260.

Roberts, Emmert, Andrew J. Ludman, Katharina Dworzynski, Abdallah Al-Mohammad, Martin R. Cowie, John J. V. McMurray, Jonathan Mant, and NICE Guideline Development Group for Acute Heart Failure. 2015. “The Diagnostic Accuracy of the Natriuretic Peptides in Heart Failure: Systematic Review and Diagnostic Meta-Analysis in the Acute Care Setting.” *BMJ* 350 (March): h910.

Sacks, David B., Mark Arnold, George L. Bakris, David E. Bruns, Andrea R. Horvath, Åke Lernmark, Boyd E. Metzger, David M. Nathan, and M. Sue Kirkman. 2023. “Guidelines and Recommendations for Laboratory Analysis in the Diagnosis and Management of Diabetes Mellitus.” *Clinical Chemistry* 69 (8): 808–68.

Salmerón, J., J. E. Manson, M. J. Stampfer, G. A. Colditz, A. L. Wing, and W. C. Willett. 1997. “Dietary Fiber, Glycemic Load, and Risk of Non-Insulin-Dependent Diabetes Mellitus in Women.” *JAMA: The Journal of the American Medical Association* 277 (6): 472–77.

Satterthwaite, F. E. 1946. “An Approximate Distribution of Estimates of Variance Components.” *Biometrics* 2 (6): 110–14.

Schmoyeri, R. L., J. J. Beauchamp, C. C. Brandt, and F. O. Hoffman. 1996. “Difficulties with the Lognormal Model in Mean Estimation and Testing.” *Environmental and Ecological Statistics* 3 (1): 81–97.

Schoot, Rens van de, Sarah Depaoli, Ruth King, Bianca Kramer, Kaspar Märtens, Mahlet G. Tadesse, Marina Vannucci, et al. 2021. “Bayesian Statistics and Modelling.” *Nature Reviews Methods Primers* 1 (1): 1–26.

Sikaris, Ken. 2012. “Application of the Stockholm Hierarchy to Defining the Quality of Reference Intervals and Clinical Decision Limits.” *The Clinical Biochemist. Reviews / Australian Association of Clinical Biochemists* 33 (4): 141–48.

Šimundić, Ana-Maria. 2009. “Measures of Diagnostic Accuracy: Basic Definitions.” *EJIFCC* 19 (4): 203–11.

Smeden, Maarten van, Christiana A. Naaktgeboren, Johannes B. Reitsma, Karel G. M. Moons, and Joris A. H. de Groot. 2014. “Latent Class Models in Diagnostic Studies When There Is No Reference Standard--a Systematic Review.” *American Journal of Epidemiology* 179 (4): 423–31.

Smith, A. F. M., and A. E. Gelfand. 1992. “Bayesian Statistics without Tears: A Sampling-Resampling Perspective.” *The American Statistician* 46 (2): 84–88.

Solberg, H. E. 1987. “Approved Recommendation (1987) on the Theory of Reference Values. Part 5. Statistical Treatment of Collected Reference Values. Determination of Reference Limits.” *Clinica Chimica Acta; International Journal of Clinical Chemistry* 170 (2): S13–32.

Srinivasan, Preethi, M. Brandon Westover, and Matt T. Bianchi. 2012. “Propagation of Uncertainty in Bayesian Diagnostic Test Interpretation.” *Southern Medical Journal* 105 (9): 452–59.

Stanley, Donald E., and Daniel G. Campos. 2013. “The Logic of Medical Diagnosis.” *Perspectives in Biology and Medicine* 56 (2): 300–315.

Stephens, Matthew. 2023. “The Bayesian Lens and Bayesian Blinkers.” *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences* 381 (2247): 20220144.

Sun, Hong, Pouya Saeedi, Suvi Karuranga, Moritz Pinkepank, Katherine Ogurtsova, Bruce B. Duncan, Caroline Stein, et al. 2022. “IDF Diabetes Atlas: Global, Regional and Country-Level Diabetes Prevalence Estimates for 2021 and Projections for 2045.” *Diabetes Research and Clinical Practice* 183 (January): 109119.

Surwit, Richard S., Miranda A. L. van Tilburg, Nancy Zucker, Cynthia C. McCaskill, Priti Parekh, Mark N. Feinglos, Christopher L. Edwards, Paula Williams, and James D. Lane. 2002. “Stress Management Improves Long-Term Glycemic Control in Type 2 Diabetes.” *Diabetes Care* 25 (1): 30–34.

Tang, Yi-Wei, Jonathan E. Schmitz, David H. Persing, and Charles W. Stratton. 2020. “The Laboratory Diagnosis of COVID-19 Infection: Current Issues and Challenges.” *Journal of Clinical Microbiology*, April. https://doi.org/10.1128/JCM.00512-20.

Tucker, Larry A. 2020. “Limited Agreement between Classifications of Diabetes and Prediabetes Resulting from the OGTT, Hemoglobin A1c, and Fasting Glucose Tests in 7412 U.S. Adults.” *Journal of Clinical Medicine Research* 9 (7). https://doi.org/10.3390/jcm9072207.

Van Cauter, E., K. S. Polonsky, and A. J. Scheen. 1997. “Roles of Circadian Rhythmicity and Sleep in Human Glucose Regulation.” *Endocrine Reviews* 18 (5): 716–38.

Viana, M. A. G., and V. Ramakrishnan. 1992. “Bayesian Estimates of Predictive Value and Related Parameters of a Diagnostic Test.” *The Canadian Journal of Statistics = Revue Canadienne de Statistique* 20 (3): 311–21.

Weiner, E. S. C., J. A. Simpson, and Oxford University Press. 1989 2004. *The Oxford English Dictionary*. Oxford, Oxford: Clarendon Press ; Melbourne.

Welch, B. L. 1947. “The Generalization of `Student’s’ Problem When Several Different Population Variances Are Involved.” *Biometrika* 34 (1/2): 28–35.

Wereski, Ryan, Dorien M. Kimenai, Caelan Taggart, Dimitrios Doudesis, Kuan Ken Lee, Matthew T. H. Lowry, Anda Bularga, et al. 2021. “Cardiac Troponin Thresholds and Kinetics to Differentiate Myocardial Injury and Myocardial Infarction.” *Circulation* 144 (7): 528–38.

White, Graham H. 2008. “Basics of Estimating Measurement Uncertainty.” *The Clinical Biochemist. Reviews / Australian Association of Clinical Biochemists* 29 Suppl 1 (August): S53-60.

Williams, Alex. 2020. “Calculation of the Expanded Uncertainty for Large Uncertainties Using the Lognormal Distribution.” *Accreditation and Quality Assurance* 25 (5): 335–38.

Willink, Robin. 2013. *Measurement Uncertainty and Probability*. Cambridge University Press.

Willink, Robin, and Rod White. 2012. "Disentangling Classical and Bayesian Approaches to Uncertainty Analysis." New Zeland: Measurement Standards Laboratory.

Wilson, Brandon M., and Barton L. Smith. 2013. "Taylor-Series and Monte-Carlo-Method Uncertainty Estimation of the Width of a Probability Distribution Based on Varying Bias and Random Error." *Measurement Science & Technology* 24 (3): 035301.

Wilson, James Maxwell Glover, and Gunnar Jungner. 1968. *Principles and Practice of Screening for Disease*. Vol. 34. Public Health Papers. Geneva: World Health Organization.

Zou, Kelly H., A. James O'Malley, and Laura Mauri. 2007. "Receiver-Operating Characteristic Analysis for Evaluating Diagnostic Tests and Predictive Models." *Circulation* 115 (5): 654–57.

10. 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

11. License

First Published: June 10, 2024

Revised: November 7, 2024