Hellenic Complex Systems Laboratory

A Software Tool for Estimating Uncertainty of Bayesian Posterior Probability for Disease

Technical Report XXVI

Theodora Chatzimichail and Aristides T. Hatjimihail
2024

Abstract

The role of medical diagnosis is essential in patient care and healthcare. Established diagnostic practices typically rely on predetermined clinical criteria and numerical thresholds. In contrast, Bayesian inference provides an advanced framework that supports diagnosis via in-depth probabilistic analysis. This study's aim is to introduce a software tool dedicated to the quantification of uncertainty in Bayesian diagnosis, a field that has seen minimal exploration to date. The presented tool, a freely available specialized software program, utilizes uncertainty propagation techniques to estimate the sampling, measurement, and combined uncertainty of the posterior probability for disease. It features two primary modules and fifteen submodules, all designed to facilitate the estimation and graphical representation of the standard uncertainty of the posterior probability estimates for diseased and non-diseased population samples, incorporating parameters such as the mean and standard deviation of the test measurand, the size of the samples, and the standard measurement uncertainty inherent in screening and diagnostic tests. Our study showcases the practical application of the program by examining the fasting plasma glucose data sourced from the National Health and Nutrition Examination Survey. Parametric distribution models are explored to assess the uncertainty of Bayesian posterior probability for diabetes mellitus, using the oral glucose tolerance test as the reference diagnostic method.

Keywords

Bayesian diagnosis; Bayesian inference; prior probability; posterior probability; likelihood; parametric distribution; probability density function; uncertainty; combined uncertainty; measurement uncertainty; sampling uncertainty; combined uncertainty; confidence intervals; diabetes mellitus; fasting plasma glucose; oral glucose tolerance test

1. Introduction

1.1. Diagnosis in Medicine

Diagnosis in medicine fundamentally involves identifying the unique characteristics of a disease and distinguishing it from other conditions with similar presentations. The term "diagnosis", originating from the Greek word "διάγνωσις" meaning "discernment" (Weiner, Simpson, and Oxford University Press 1989 2004), emphasizes the critical role of distinguishing between healthy and diseased states in individuals. Diagnostic tests are essential in classifying individuals based on their health status. However, the reliance on a singular threshold for diagnosis across a range of data points introduces uncertainty, owing to the overlapping probability distributions of a measurand in both healthy and diseased groups (Chatzimichail and Hatjimihail 2023). While traditional diagnostic methods have been broadly effective, they may not fully encompass the diversity of disease manifestations, particularly in varied populations (Choi, Johnson, and Thurmond 2006).

As underlined previously (Chatzimichail and Hatjimihail 2023), Bayesian inference represents a paradigm shift in the field of medical diagnosis, offering a robust framework for integrating various sources of information to make probabilistic assessments. At its core, Bayesian inference relies on the Bayes' theorem for updating beliefs in light of new evidence, integrating prior disease probabilities with the distribution of diagnostic measurands to calculate posterior probabilities for disease (Bours 2021; Gelman et al. 2013; van de Schoot et al. 2021; Viana and Ramakrishnan 1992). This approach enables a more comprehensive probabilistic assessment, the evaluation of the information conveyed by diagnostic measurements, and a personalized patient approach (Choi, Johnson, and Thurmond 2006; Topol 2014).

Historically, the application of Bayesian methods in medicine has undergone significant evolution. Despite facing several challenges and being met with skepticism, these methods have gradually gained acceptance.

1.1.1. Bayes' Theorem in Medical Diagnostics

Bayes' theorem, a fundamental principle in probability theory (Gelman et al. 2013), forms a connection between the direct probability $P(H|E)$ of a hypothesis H given specific data E, and the

inverse probability P(E|H) of data E given the hypothesis H (Joyce 2021). In medical diagnostics, Bayes' theorem is instrumental in transforming the prior probability for disease into a posterior probability following diagnostic tests (Bours 2021).

1.1.2. Challenges in Applying Bayesian Inference

The application of Bayesian inference in diagnostics, however, faces significant challenges.

Computational Complexity

The computational complexity of Bayesian inference requires considerable resources.

Statistical Distributions in Diagnostics

A major challenge involves comprehensively understanding the statistical distributions of diagnostic test measurands in both diseased and nondiseased populations (Lehmann and Romano 2008). Calculation of posterior probabilities requires probability density functions (PDF) for the measurands in these populations. The normal distribution, often used for its simplicity, may not be suitable for measurands with non-normal characteristics like skewness or multimodality. Critical evaluation and potential adoption of alternative distributions are necessary for more accurate Bayesian diagnostic methods (Lehmann and Romano 2008; Box and Cox 1964; D'Agostino and Pearson 1973). Bayesian Diagnosis, our previously published software, addresses this challenge (Chatzimichail and Hatjimihail 2023).

Uncertainty of Bayesian Posterior Probabilities

Another significant challenge involves estimating the uncertainty associated with Bayesian posterior probabilities in disease diagnosis. This uncertainty can substantially affect their clinical utility. Despite its critical importance, the task of estimating, evaluating, and mitigating uncertainty in Bayesian diagnostic test interpretation has seldom been addressed in medical literature (Srinivasan, Westover, and Bianchi 2012). To confront this issue, we have developed Bayesian Diagnostic Uncertainty, a software tool for the estimation of uncertainty in Bayesian diagnosis, which is presented in detail in this study.

Both Bayesian Diagnostic Uncertainty and Bayesian Diagnosis, enhance the applicability of Bayesian methods in medical diagnostics.

1.1.3. Quantifying Uncertainty in Diagnostics

Uncertainty can be quantified and is often expressed probabilistically (Ayyub and Klir 2006).

Combined Uncertainty

In the context of Bayesian posterior probability for disease, we consider two main components of combined uncertainty:

Measurement Uncertainty

This reflects the inherent variability in measurement processes and is defined as a parameter characterizing the dispersion of values that could reasonably be attributed to the measurand (Joint Committee for Guides in Metrology 2011). While crucial for laboratory quality assurance, the impact of measurement uncertainty on clinical decision-making and outcomes is often underexplored and rarely quantified (Kallner et al. 2012; Smith et al. 2019). Emerging research focuses on its effects on misclassification (Ceriotti et al. 2017) and on diagnostic accuracy measures (Chatzimichail and Hatjimihail 2021).

Sampling Uncertainty

The variability in sampling contributes to the uncertainty of posterior probability for disease (Rostron, Fearn, and Ramsey 2020), and it is essential in evaluating diagnostic methods.

2. Methods

2.1. Computational Methods

2.1.1. Bayes' Theorem

Bayes' theorem calculates the posterior probability $P(D|T)$ of a disease D given a test result $T = x$ and a parameter vector $\boldsymbol{\theta}$, as follows:

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

Here r denotes the prior probability for disease, $f_D(x|\boldsymbol{\theta})$ the PDF in disease presence, while $f_{\bar{D}}(x;\boldsymbol{\theta})$ denotes the PDF in its absence (refer to Appendix A.1 for details).

2.1.2. Parametric Distributions

Parametric statistics operate under the assumption that data from a population can be accurately represented by a probability distribution with a fixed set of parameters (Geisser and Johnson 2006). The program supports the following parametric distributions:

1. Normal distribution
2. Lognormal distribution
3. Gamma distribution.

2.1.3. Uncertainty Quantification

Uncertainty of input parameters can manifest as standard uncertainty $u(x)$, the standard deviation of x, and expanded uncertainty $U(x)$, a range around x encompassing x with a probability p (Kallner et al. 2012).

Measurement Uncertainty

Measurement uncertainty is computed following guidelines in the "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 $u(x)$ to the value of the measurand x, is generally expressed as:

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

where b_0 is a constant and $b_1$is a proportionality constant.

If needed, it is approximated linearly as:

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

The general approach to estimating the coefficients of the above equations is delineated in Appendix A5 of "Quantifying Uncertainty in Analytical Measurement" (Ellison and Williams 2012).

Sampling Uncertainties of Means and Standard Deviations

If m_P *and* s_P are the mean and standard deviation of a measurand in a population sample of size n_P, then the standard sampling uncertainties of m_P and s_P are estimated as:

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

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

using the central limit theorem and the chi-square distribution (Agresti, Franklin, and Klingenberg 2023; Miller and Miller 2018; J. Aitchison 1957).

Sampling Uncertainty of Prior Probability for Disease

If n_D and $n_{\bar{D}}$ are the respective numbers of diseased and nondiseased in a population sample, then the standard uncertainty of the prior probability for disease $r = \frac{n_D}{n_{\bar{D}}+n_D}$ is estimated as:

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

using the Agresti–Coull adjustment of the Waldo interval (Agresti and Coull 1998).

Combined Uncertainty of Posterior Probability for Disease

The standard combined uncertainty $u_c(x)$ of posterior probability for disease is computed via uncertainty propagation rules, employing a first-order Taylor series approximation (B. M. Wilson and Smith 2013) (refer to Supplementary File II).

When there are l components of uncertainty, with standard uncertainties $u_i(x)$, then:

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

Obviously, $0 \leq u_c(x) \leq 1.0$.

2.1.4. Expanded Uncertainty of Posterior Probability for Disease

When there are l components of uncertainty, with standard uncertainties $u_i(x)$ and v_i degrees of freedom, then the effective degrees of freedom v_{eff} of the combined uncertainty $u_c(x)$ are obtained from the Welch–Satterthwaite formula (Welch 1947; Satterthwaite 1946):

$$v_{eff}(x) \cong \frac{u_c(x)^4}{\sum_{i=1}^{l} \frac{u_i(x)^4}{v_i}}$$

If v_{min} the minimum of $v_1, v_2, \ldots, v_l$, then:

$$v_{min} \leq v_{eff}(x) \leq \sum_{i=1}^{l} v_i$$

If $F_v(z)$ is the Student's t-distribution cumulative distribution function with v degrees of freedom and $u_c(x)$ is the standard combined uncertainty of posterior probability for disease, its expanded combined uncertainty $U_c(x)$ at a confidence level p is:

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

The confidence interval of x at the same confidence level p is approximated as:

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

The confidence intervals of the posterior probability for disease were truncated to the [0,1] range.

2.2. The Software

2.2.1. Program Overview

To facilitate the estimation of the uncertainty of Bayesian posterior probability for disease, the software program Bayesian Diagnostic Uncertainty was developed in Wolfram Language, using Wolfram Mathematica® Ver. 13.3 (Wolfram Research, Inc., Champaign, IL, USA). Bayesian Diagnostic Uncertainty was designed to estimate and plot the standard sampling, measurement, and combined uncertainty and the confidence intervals of the Bayesian posterior probability for disease of a screening or diagnostic test (See Figure 1).

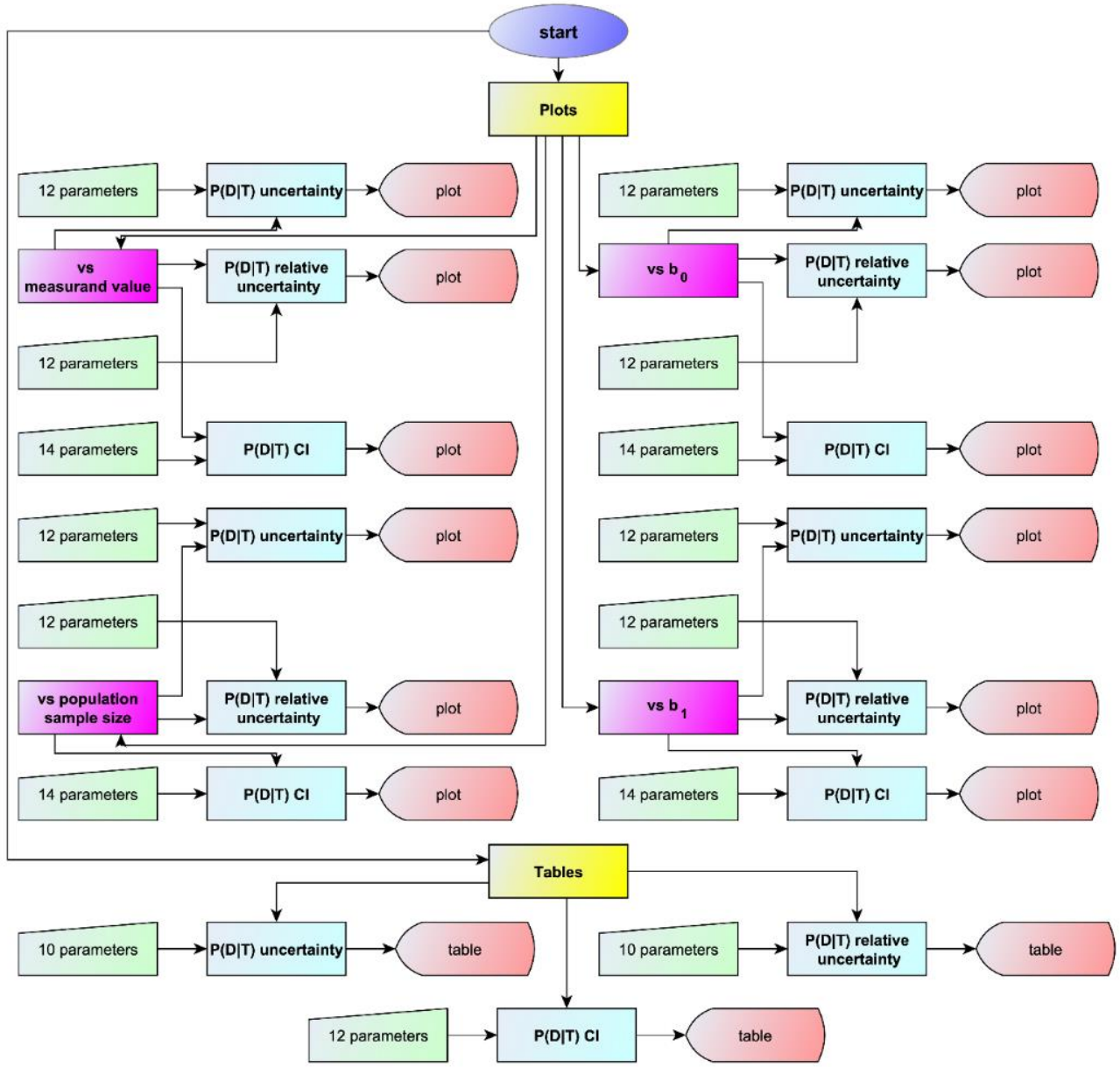

Figure 1. A simplified flowchart of the program Bayesian Diagnostic Uncertainty with the number of input parameters and of output types for each submodule.

This interactive program is freely available as a Wolfram Language notebook (.nb) (Supplementary file BayesianUncertainty.nb). It can be run on Wolfram Player® or Wolfram Mathematica® (see Appendix A.2). Due to the complexity of the calculations required, it is computationally intensive.

2.2.2. Input Parameters

The program allows for the definition of three parametric distributions of a measurand for the diseased and nondiseased populations.

Distribution Selection

The user selects the type of distribution of each population from a predefined list:

1. Normal distribution
2. Lognormal distribution
3. Gamma distribution.

Definition of Statistical Parameters

For each population, the user defines its size n, the mean μ, and the standard deviation σ of the measurand.

Measurement Uncertainty

The user selects a linear or nonlinear equation of the measurement uncertainty versuss the value x of the measurand and defines the constant contribution b_0 to the standard measurement uncertainty, the proportionality constant b_1, and the number of quality control samples that have been analyzed for its estimation.

2.2.3. Output Specifications

Visualizations

The program generates a series of plots designed to elucidate various uncertainty measures and statistics:

1. Uncertainty of posterior probability for disease: Plots are generated to show the standard sampling, measurement, and combined uncertainty of the posterior probability for disease.
2. Relative uncertainty of posterior probability for disease: Plots are generated to show the relative standard sampling, measurement, and combined uncertainty of the posterior probability for disease.
3. Confidence intervals of posterior probability for disease: Plots are generated to show the confidence intervals of the posterior probability for disease, for a user defined confidence level.

Tables

For each combination of parametric distributions of the diseased and nondiseased populations, the program tabulates for a user-defined measurand value:

1. The standard sampling, measurement, and combined uncertainty of the posterior probability for disease.
2. The relative standard sampling, measurement, and combined uncertainty of the posterior probability for disease.
3. The confidence intervals of the posterior probability for disease for a user-defined confidence level.

By providing this comprehensive set of input parameters and output specifications (see Figure 2), the program offers a robust platform for exploring the uncertainty in Bayesian diagnosis of disease using parametric distributions of medical diagnostic measurands.

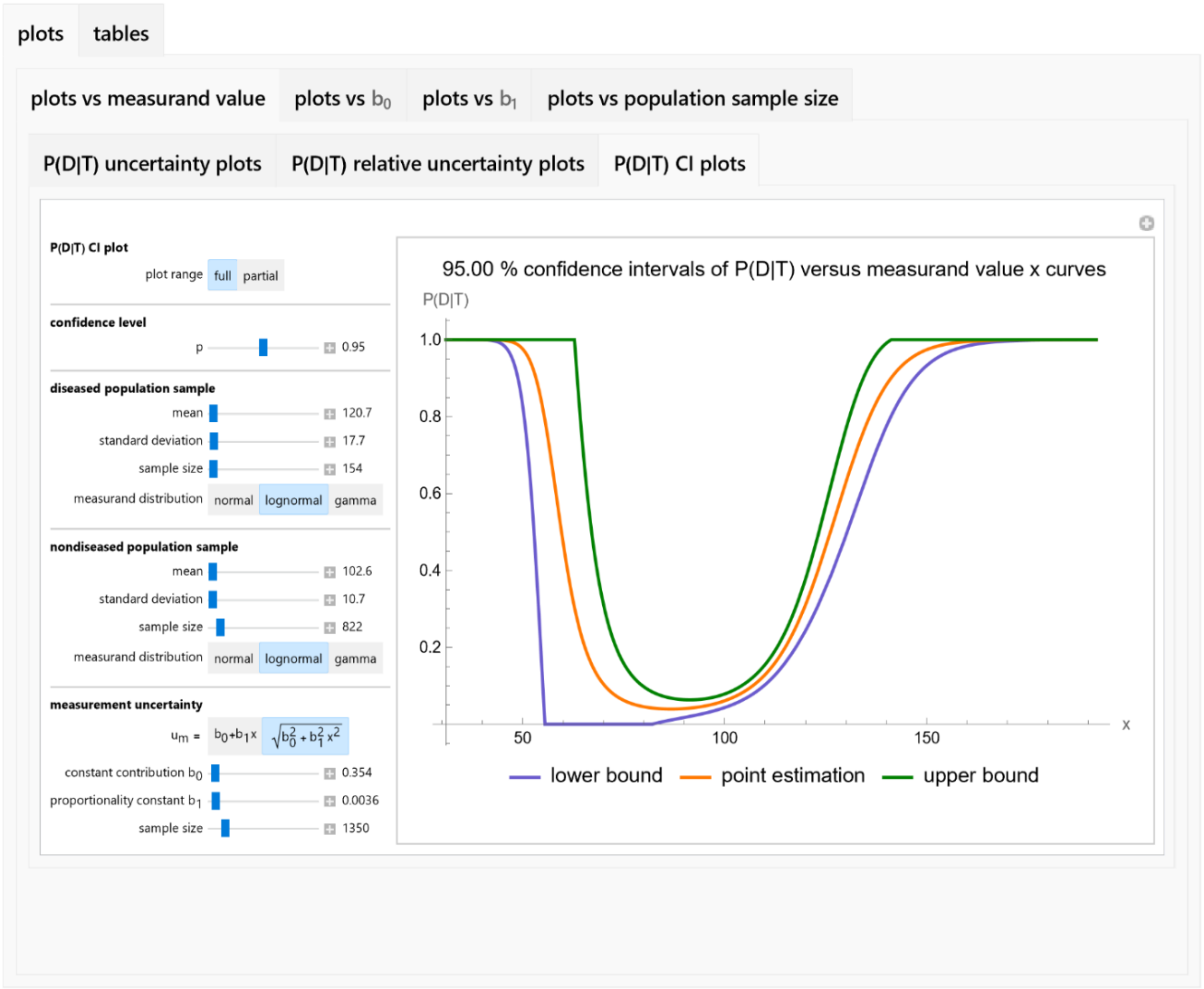

Figure 2. A screenshot of the program Bayesian Diagnostic Uncertainty.

3. Illustrative Case Study

To demonstrate the application of the program, fasting plasma glucose (FPG) was used as the diagnostic test measurand for the Bayesian diagnosis of diabetes mellitus (From now on, when mentioning "diabetes", we are referring to diabetes mellitus). The oral glucose tolerance test (OGTT) was used as the reference diagnostic method. A diagnosis of diabetes was confirmed if the plasma glucose value was equal to or greater than 200 mg/dl, measured two hours after oral administration of 75 g of glucose (ElSayed et al. 2023), during an OGTT (2-h PG). The study population was confined to individuals aged between 70 and 80 years, a decision guided by the well-documented strong correlation between age and the prevalence of diabetes (Sun et al. 2022).

National Health and Nutrition Examination Survey (NHANES) data from participants was retrieved for the period from 2005 to 2016 (n = 60,936) (National Center for Health Statistics 2005-20016). NHANES is a series of studies designed to evaluate the health and nutritional status of adults and children in the United States.

The inclusion criteria for participants were:

1. Valid fasting plasma glucose (FPG) and oral glucose tolerance test (OGTT) results (n = 13,836).
2. A negative response to NHANES question DIQ010 regarding a diabetes diagnosis (National Center for Health Statistics 2005-20016) (n = 13,465).
3. Age 70–80 years (n = 976).

Participants with a 2-h PG measurement ≥200 mg/dl were considered diabetic (n = 154).

The prior probability for diabetes was estimated as:

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

The statistics of the FPG datasets are presented in Table 1 (Hereafter, FPG and its uncertainty are expressed in mg/dl).

Table 1. Descriptive statistics of the fasting plasma glucose datasets.

	Diabetic Patients	**Nondiabetic Patients**
n	154	822
Mean	120.7	102.6
Median	117.0	102.0
Standard Deviation	19.1	10.9
Skewness	1.448	0.523
Kurtosis	6.354	3.445

Lognormal distributions were estimated to model FPG measurements in diabetic and nondiabetic participants, using the maximum likelihood estimation method (Myung 2003). The respective distributions, parametrized for their means μ_D and $\mu_{\bar{D}}$, and standard deviations σ_D and $\sigma_{\bar{D}}$, were the following:

$$L_D = Lognormal(\mu_D, \sigma_D) = Lognormal(120.671, 17.720)$$
$$L_{\bar{D}} = Lognormal(\mu_{\bar{D}}, \sigma_{\bar{D}}) = Lognormal(102.642, 10.653)$$

NHANES quality control data of the FPG measurements was retrieved for the same period (2005–2016). 1350 QC samples had been analyzed. The weighted nonlinear least squares analysis (Nielsen 2007) yielded the following function relating the standard measurement uncertainty $u_m(x)$ to the measurement value x:

$$u_m(x) = \sqrt{b_0^2 + b_1^2 x^2} = \sqrt{0.7501 + 0.00012x^2}$$

where $b_0 = 0.866$ and $b_1 = 0.109$.

The means of the standard measurement uncertainty of the FPG of the included diabetic and nondiabetic participants were estimated as:

$$\hat{u}_D \cong 1.586 \text{ mg/dl}$$
$$\hat{u}_{\bar{D}} \cong 1.028 \text{ mg/dl}$$

Consequently, the distributions of the measurands, assuming negligible uncertainty, were estimated as:

$$l_D \cong Lognormal\left(\mu_D, \sqrt{\sigma_D^2 - \hat{u}_D^2}\right) \cong Lognormal(120.671, 17.720)$$

$$l_{\bar{D}} \cong Lognormal\left(\mu_{\bar{D}}, \sqrt{\sigma_{\bar{D}}^2 - \hat{u}_{\bar{D}}^2}\right) \cong Lognormal(102.642, 10.653)$$

Table 2 displays the descriptive statistics of the estimated lognormal distributions of the diabetic and nondiabetic populations, including the respective p -values of the Cramér–von Mises goodness-of-fit test (Darling 1957).

Table 2. Descriptive statistics of the estimated lognormal distributions of the diabetic and nondiabetic populations.

	Diabetic Participants		Nondiabetic Participants	
Estimated Distribution	L_D	l_D	$L_{\bar{D}}$	$l_{\bar{D}}$
Mean Uncertainty	1.586	0	1.028	0
Mean	120.7	120.7	102.6	102.6
Median	119.4	119.4	102.1	102.1
Standard Deviation	17.8	17.7	10.9	10.7
Skewness	0.446	0.444	0.315	0.312
Kurtosis	3.355	3.352	3.177	3.174
p-value (Cramér–von Mises test)	0.294	0.295	0.281	0.299

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

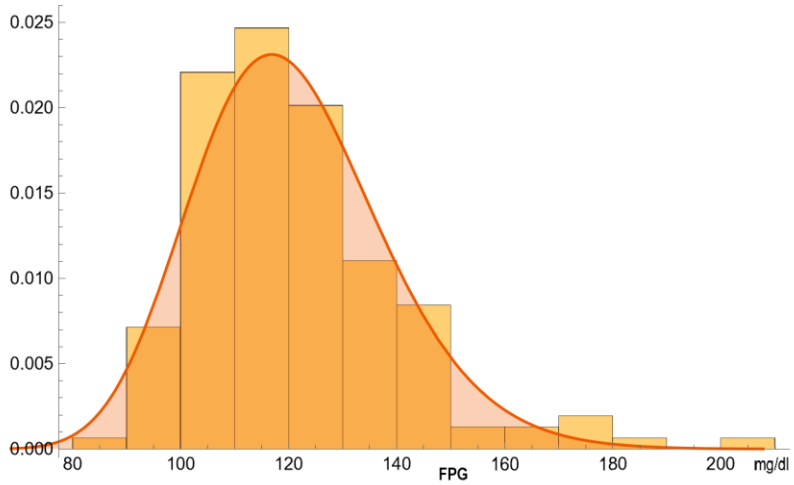

Figure 3. The estimated PDF of the FPG (mg/dl) in diabetic participants, assuming a lognormal distribution and negligible measurement uncertainty, and the histogram of the respective NHANES dataset, with the parameters of the distribution in Table 2.

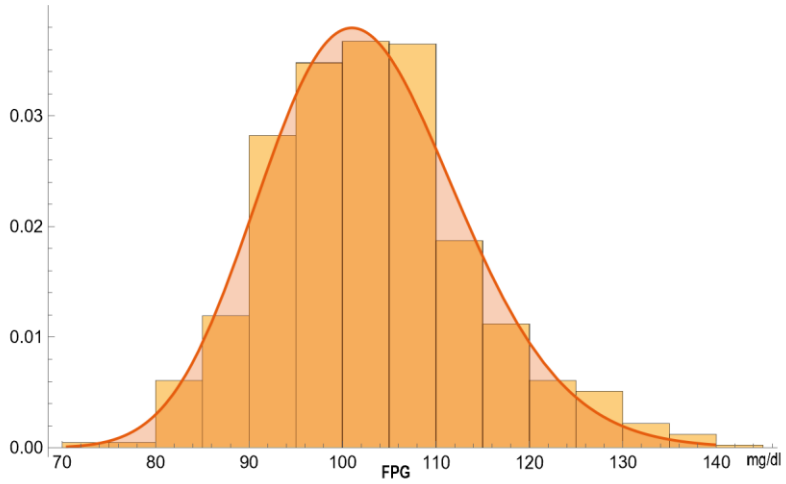

Figure 4. The estimated PDF of the FPG (mg/dl) in nondiabetic participants, assuming a lognormal distribution and negligible measurement uncertainty, and the histogram of the respective NHANES dataset, with the parameters of the distribution in Table 2.

Likelihoods and posterior probabilities were estimated accordingly.

4. Results

Using the settings of Table 3, the program generated the plots of Figures 5–16 and the tables of Figures 17–19.

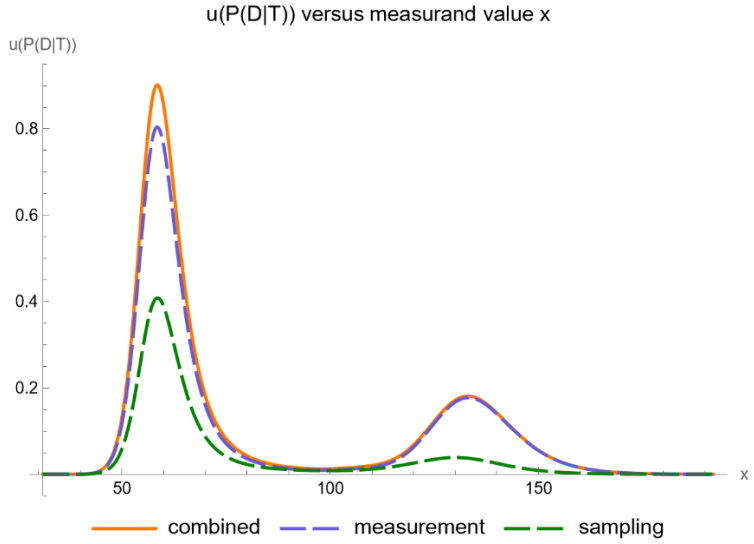

Figure 5. Standard sampling, measurement, and combined uncertainty of the posterior probability for diabetes versus FPG curve plot, with the settings of the program in Table 2.

Table 3. The settings of the program Bayesian Diagnostic Uncertainty for Figures 5–19.

Settings	Figures 5 and 6	Figure 7	Figures 8 and 9	Figure 10	Figures 11 and 12	Figure 13	Figures 14 and 15	Figure 16	Figures 17 and 18	Figure 19
p	-	0.95	-	0.95	-	0.95	-	0.95	-	0.95
x	31.0–192.0	31.0–192.0	126.0	126.0	126.0	126.0	126.0	126.0	126.0	126.0
μ_D	120.7	120.7	120.7	120.7	120.7	120.7	120.7	120.7	120.7	120.7
σ_D	17.7	17.7	17.7	17.7	17.7	17.7	17.7	17.7	17.7	17.7
n_D	154	154	154	154	154	154	-	-	154	154
$\mu_{\bar{D}}$	102.7	102.7	102.7	102.7	102.7	102.7	102.7	102.7	102.7	102.7
$\sigma_{\bar{D}}$	10.7	10.7	10.7	10.7	10.7	10.7	10.7	10.7	10.7	10.7
$n_{\bar{D}}$	822	822	822	822	822	822	-	-	822	822
n	-	-	-	-	-	-	65–5000	65–5000	-	-
r	-	-	-	-	-	-	0.158	0.158	-	-
b_0	0.866	0.866	0.0–0.161	0.0–0.161	0.866	0.866	0.866	0.866	0.866	0.866
b_1	0.0109	0.0109	0.0109	0.0109	0.0–0.1	0.0–0.1	0.0109	0.0109	0.0109	0.0109
n_U	-	1350	-	1350	-	1350	-	1350	-	1350
l_D	lognormal	lognormal	lognormal	lognormal	lognormal	lognormal	lognormal	lognormal	normal lognormal gamma	normal lognormal gamma
$l_{\bar{D}}$	lognormal	lognormal	lognormal	lognormal	lognormal	lognormal	lognormal	lognormal	normal lognormal gamma	normal lognormal gamma

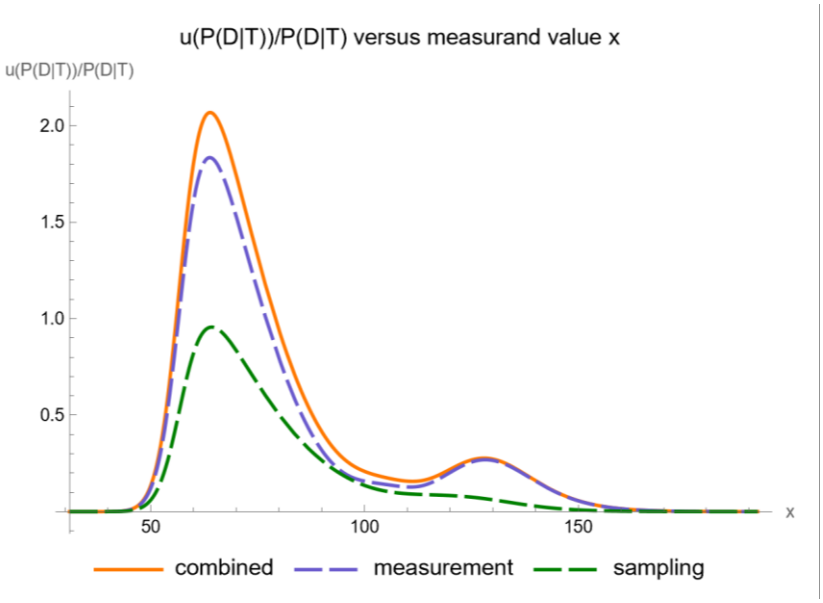

Figure 6. Relative standard sampling, measurement, and combined uncertainty of the posterior probability for diabetes versus FPG curve plot, with the settings of the program in Table 2.

Figure 5 shows the plots of the standard sampling, measurement, and combined uncertainty of posterior probability for diabetes versus FPG, while Figure 6 shows the respective plots of the relative standard uncertainty.

Figure 7 shows the plots of the confidence intervals of posterior probability for diabetes versus FPG for a confidence level $p = 0.95$.

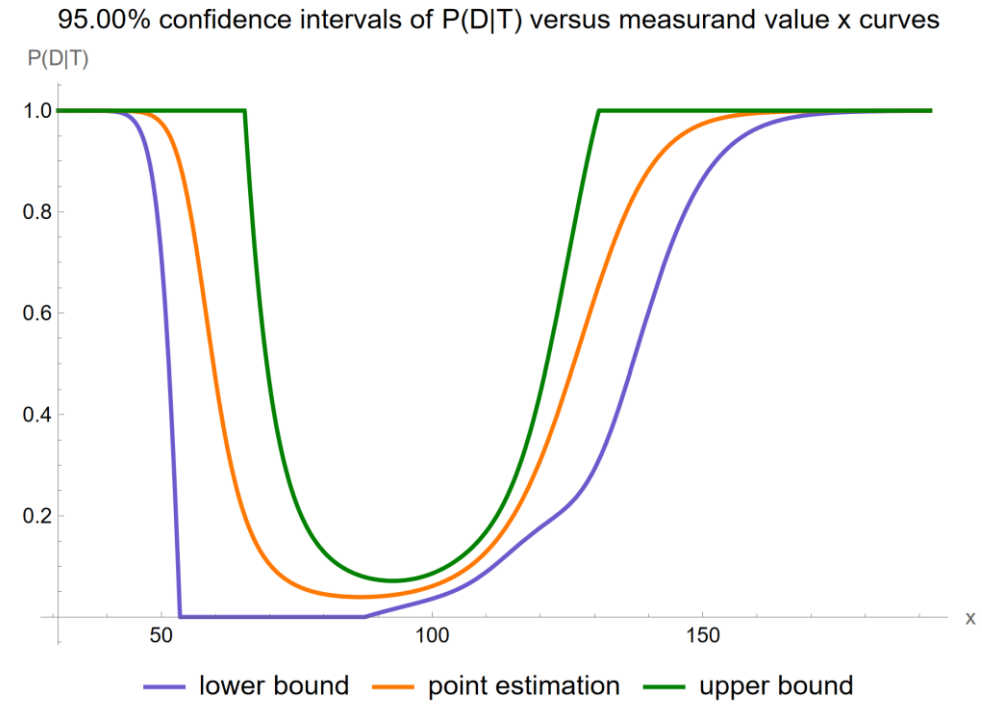

Figure 7. Confidence intervals of the posterior probability for diabetes versus FPG curves plot, with the settings of the program in Table 2.

Assessing the combined standard uncertainty of the posterior probability for diabetes, we note the following:

1. It is substantially affected by measurement uncertainty of FPG.

2. Two local maxima are observed, corresponding to the regions near the steepest segments of the posterior probability curve, which exhibits an approximately double sigmoidal configuration. These maxima are quantitatively defined as following:

   2.1. At an FPG value of 58.7 mg/dl, the posterior probability for disease is equal to 0.585, while the combined standard uncertainty is equal to 0.893.

   2.2. At an FPG value of 133.2 mg/dl, the posterior probability for disease is equal to 0.725, while the combined standard uncertainty is equal to 0.182.

This pattern of local maxima is indicative of heightened uncertainty in the regions where the posterior probability curve demonstrates its most pronounced inflections. The confidence intervals are affected accordingly.

Assessing the relative combined standard uncertainty of the posterior probability for diabetes, we note that two local maxima are observed as well, quantitatively defined as following:

1. At an FPG value of 64.1 mg/dl, the posterior probability for disease is equal to 0.257, while the relative combined standard uncertainty is equal to 2.044.

2. At an FPG value of 128.1 mg/dl, the posterior probability for disease is equal to 0.561, while the relative combined standard uncertainty is equal to 0.278.

Figure 8 shows the plots of the standard sampling, measurement, and combined uncertainty of posterior probability for diabetes versus the constant contribution b_0 of measurement uncertainty of FPG, while Figure 9 shows the respective plots of the relative standard uncertainty.

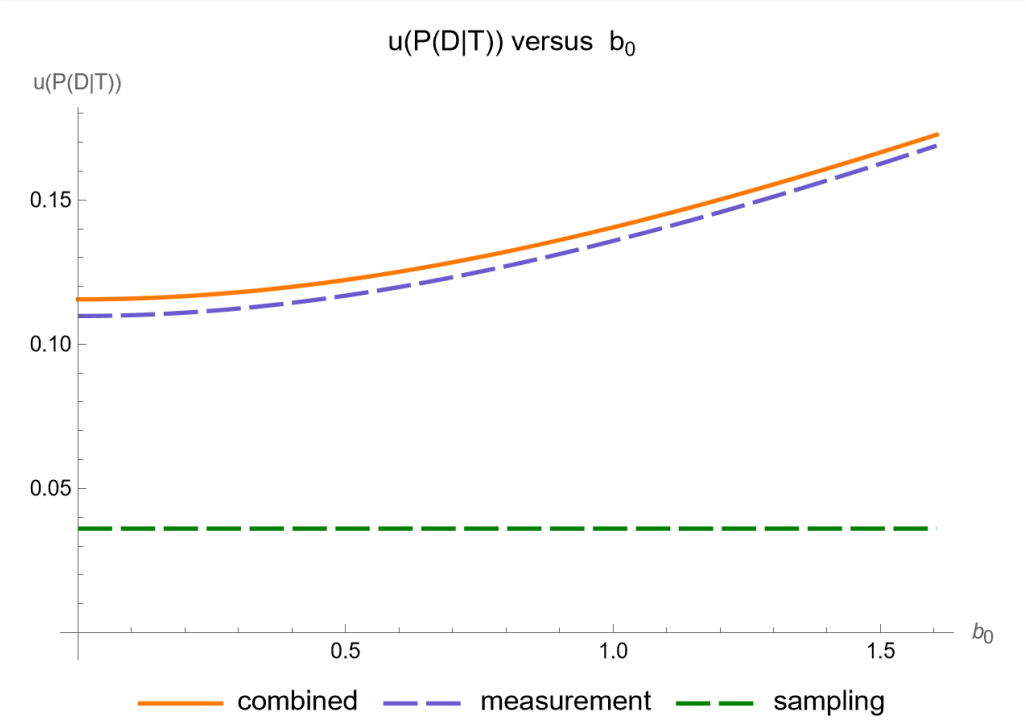

Figure 8. Standard sampling, measurement, and combined uncertainty of the posterior probability for diabetes versus measurement uncertainty constant contribution b_0 curve plot, with the settings of the program in Table 2.

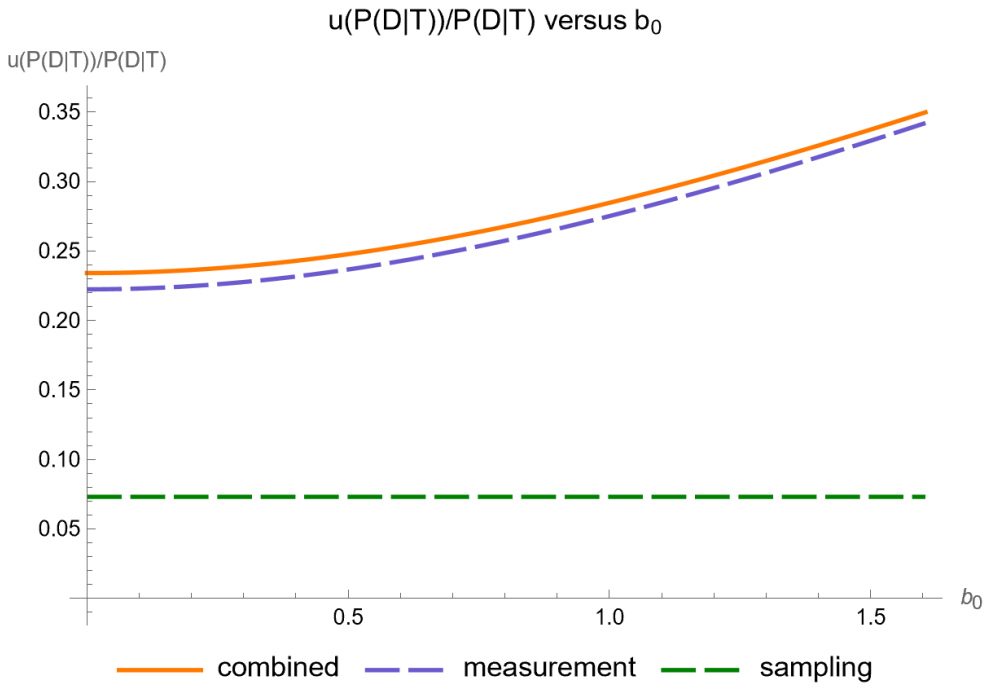

Figure 9. Relative standard sampling, measurement, and combined uncertainty of the posterior probability for diabetes versus measurement uncertainty constant contribution b_0 curve plot, with the settings of the program in Table 2.

Figure 10 shows the plots of the confidence intervals of posterior probability for diabetes versus the constant contribution b_0 of measurement uncertainty of FPG, for a confidence level $p = 0.95$.

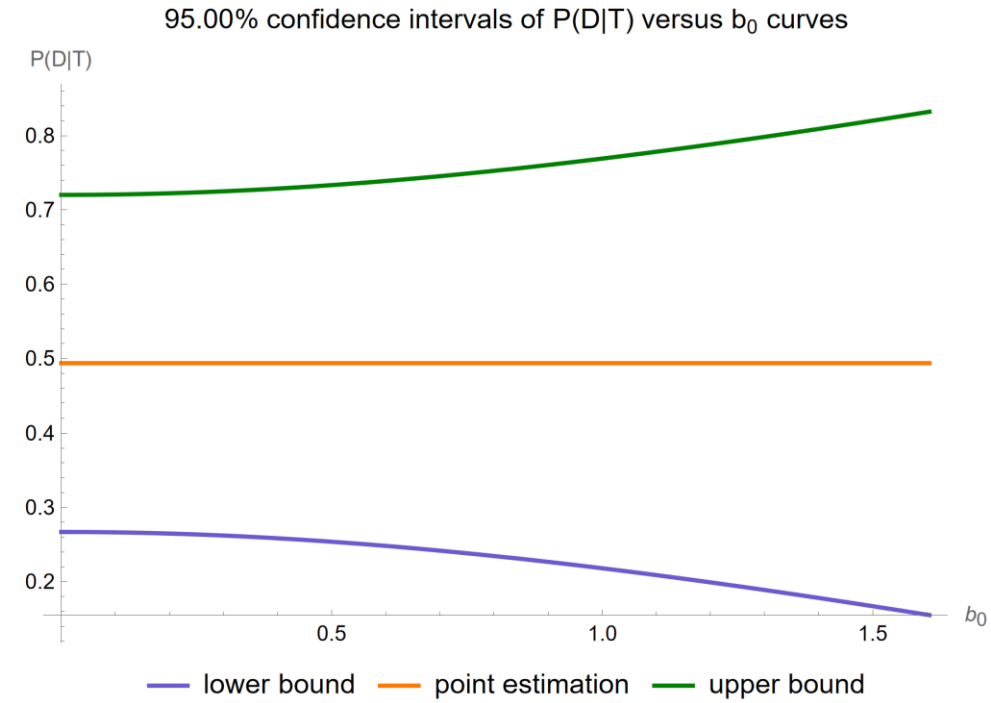

Figure 10. Confidence intervals of the posterior probability for diabetes versus measurement uncertainty constant contribution b_0 curves plot, with the settings of the program in Table 2.

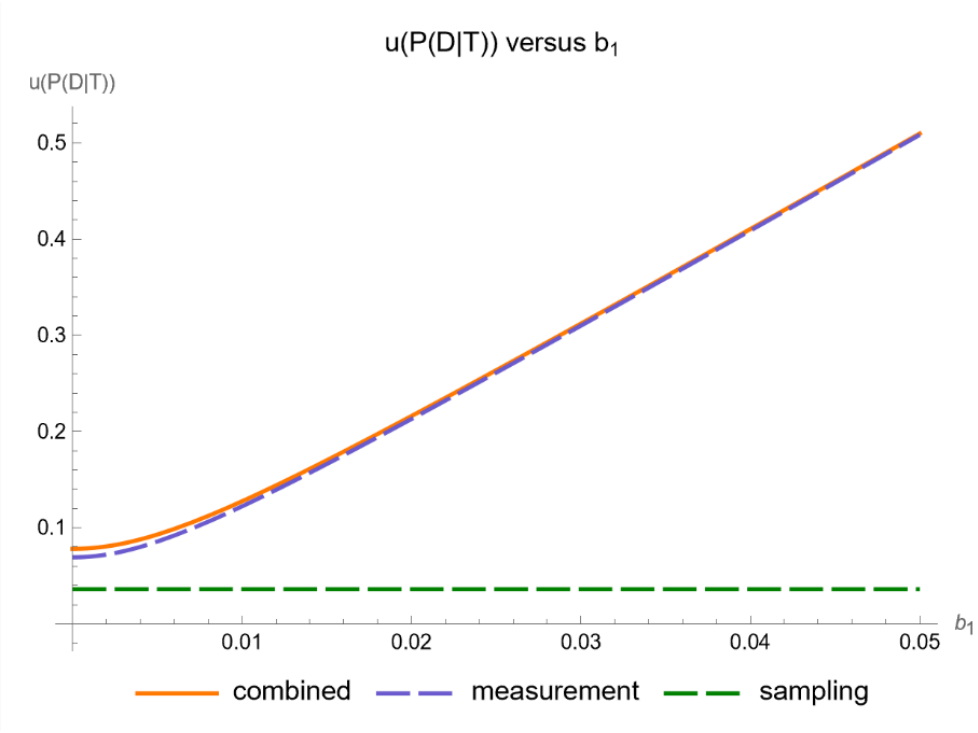

Figure 11. Standard sampling, measurement, and combined uncertainty of the posterior probability for diabetes versus measurement uncertainty proportionality constant b_1 curve plot, with the settings of the program in Table 2.

Figure 11 shows the plots of the standard sampling, measurement, and combined uncertainty of posterior probability for diabetes versus the proportionality constant b_1 of measurement uncertainty of FPG, while Figure 12 shows the respective plots of the relative standard uncertainty.

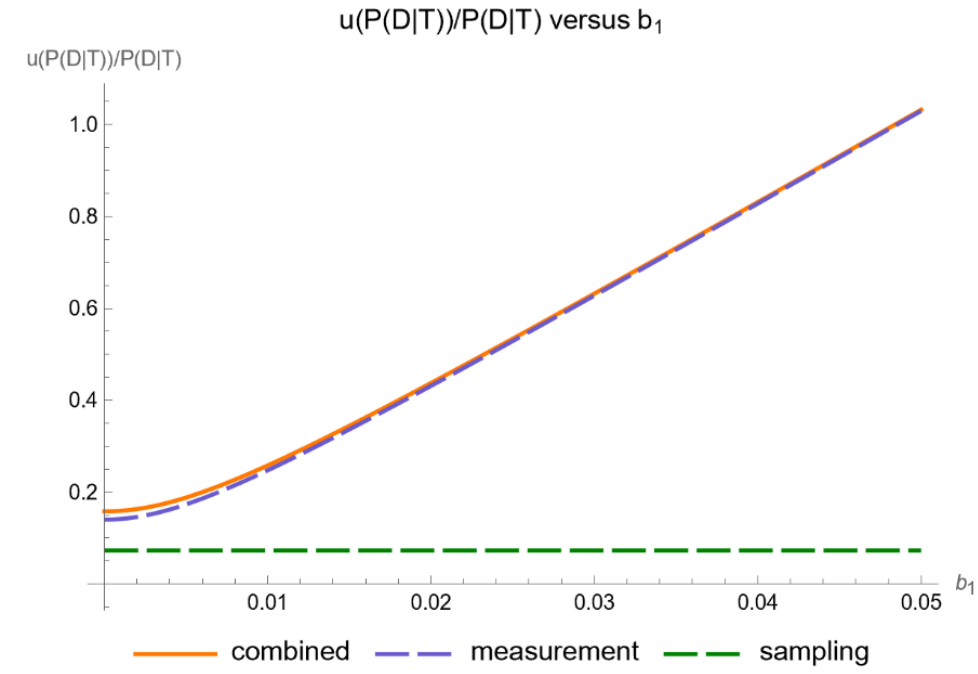

Figure 12. Relative standard sampling, measurement, and combined uncertainty of the posterior probability for diabetes versus measurement uncertainty proportionality constant $b_1$curve plot, with the settings of the program in Table 2.

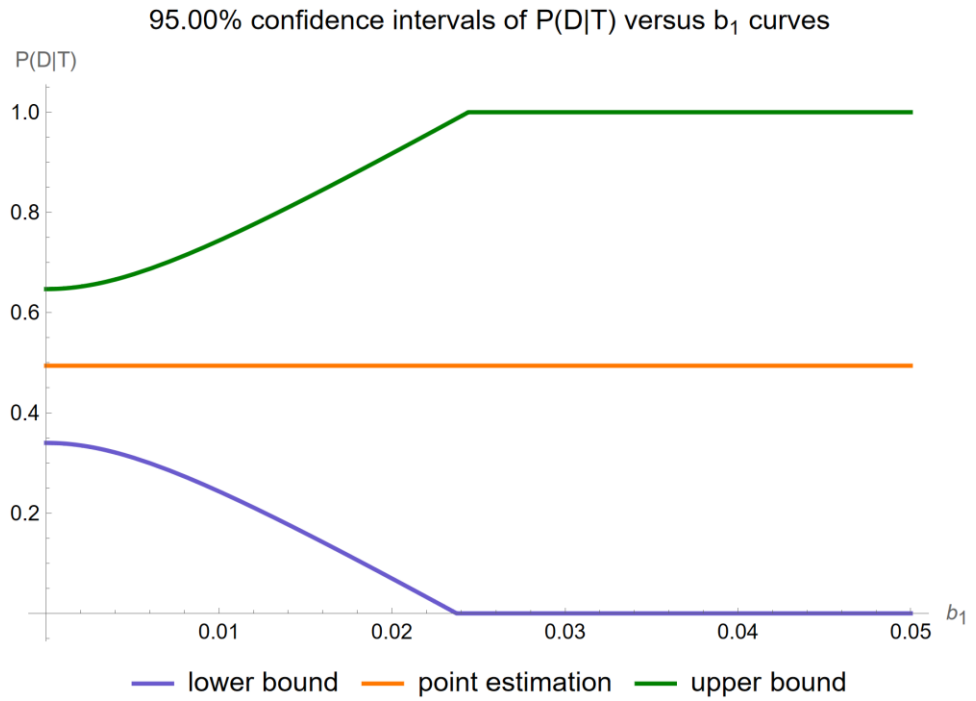

Figure 13. Confidence intervals of the posterior probability for diabetes versus measurement uncertainty proportionality constant b_1 curves plot, with the settings of the program in Table 2.

Figure 13 shows the plots of the confidence intervals of posterior probability for diabetes versus the proportionality constant b_1 of measurement uncertainty of FPG for a confidence level $p = 0.95$.

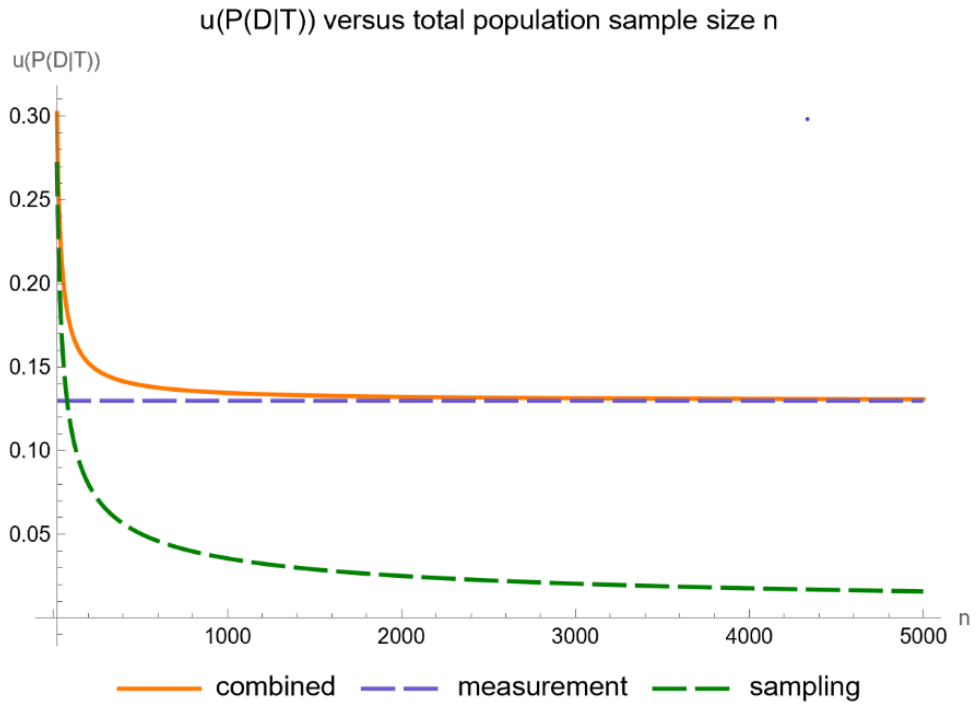

Figure 14. Standard sampling, measurement, and combined uncertainty of the posterior probability for diabetes versus total population sample size n curve plot, with the settings of the program in Table 2.

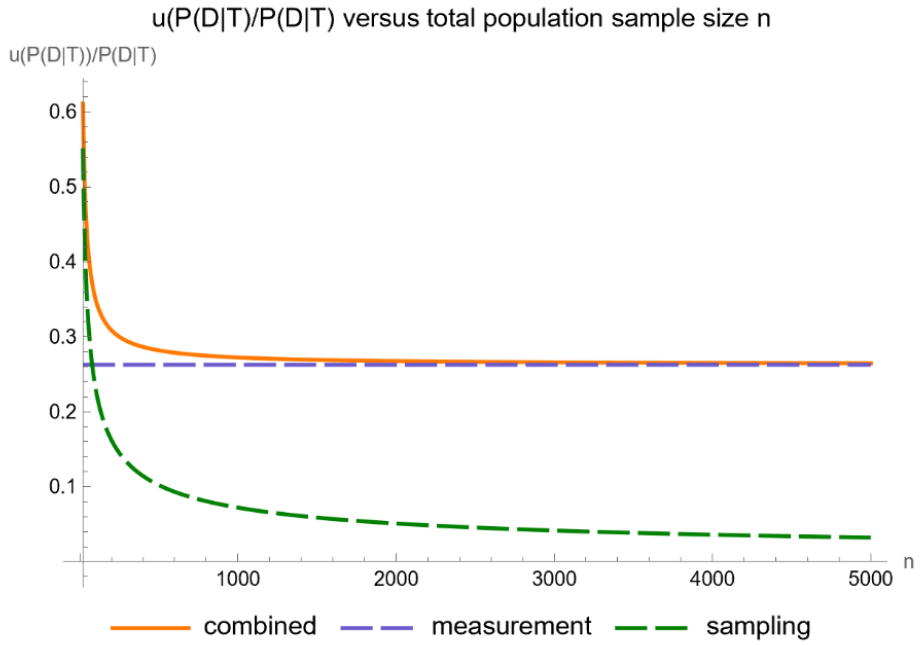

Figure 15. Relative standard sampling, measurement, and combined uncertainty of the posterior probability for diabetes versus total population sample size n curve plot, with the settings of the program in Table 2.

Figure 14 shows the plots of the standard sampling, measurement, and combined uncertainty of posterior probability for diabetes versus the total population size n, while Figure 15 shows the respective plots of the relative standard uncertainty.

Figure 16 shows the plots of the confidence intervals of posterior probability for diabetes versus the total population size n, for a confidence level $p = 0.95$.

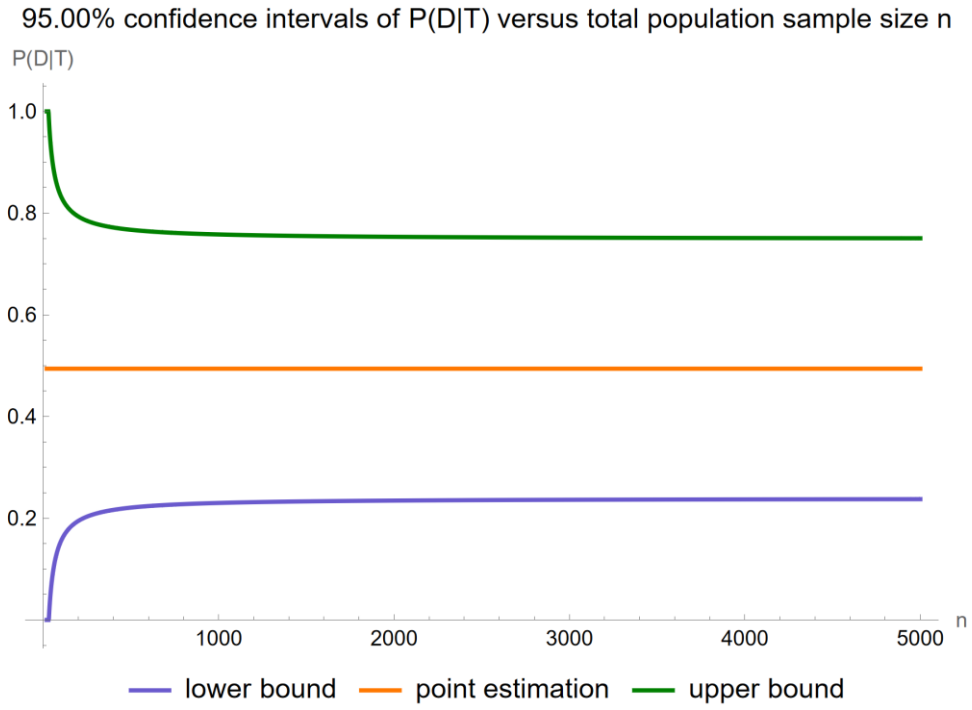

Figure 16. Confidence intervals of the posterior probability for diabetes versus total population sample size n curve plot curves plot, with the settings of the program in Table 2.

As anticipated, the impact of sampling uncertainty decreases markedly as the size of the population sample increases.

Figure 17 shows a table of the standard sampling, measurement, and combined standard uncertainty of posterior probability for diabetes for FPG value equal to 126 mg/dl, while Figure 18 shows a table of the respective values of relative standard uncertainty.

standard uncertainty of posterior probability for disease					
measurand distribution		probability	standard uncertainty		
diseased	nondiseased		combined	measurement	sampling
normal	nornal	0.541991	0.170704	0.166022	0.039708
	lognormal	0.508692	0.134128	0.129443	0.035140
	gamma	0.515938	0.145254	0.140599	0.036475
lognormal	nornal	0.526974	0.134704	0.129786	0.036066
	lognormal	0.493604	0.134704	0.129786	0.036066
	gamma	0.500853	0.145916	0.141041	0.037400
gamma	nornal	0.531176	0.171538	0.166723	0.040358
	lognormal	0.497819	0.134538	0.129702	0.035746
	gamma	0.505069	0.145730	0.140933	0.037083

Figure 17. Table of the standard sampling, measurement, and combined uncertainty of the posterior probability for diabetes, with the settings of the program in Table 2.

relative standard uncertainty of posterior probability for disease					
measurand distribution		probability	relative standard uncertainty		
diseased	nondiseased		combined	measurement	sampling
normal	nornal	0.541991	0.314958	0.306318	0.073264
	lognormal	0.508692	0.263672	0.254462	0.069079
	gamma	0.515938	0.281533	0.272512	0.070697
lognormal	nornal	0.526974	0.255618	0.246285	0.068441
	lognormal	0.493604	0.272900	0.262936	0.073068
	gamma	0.500853	0.291335	0.281602	0.074673
gamma	nornal	0.531176	0.322940	0.313875	0.075979
	lognormal	0.497819	0.270255	0.260541	0.071806
	gamma	0.505069	0.288535	154.000000	0.073422

Figure 18. Table of the relative standard sampling, measurement, and combined uncertainty of the posterior probability for diabetes, with the settings of the program in Table 2.

posterior probability for disease: 95.00% confidence intervals				
diseased	nondiseased	point estimation	lower bound	upper bound
normal	nornal	0.541991	0.207147	0.876836
	lognormal	0.508692	0.245600	0.771784
	gamma	0.515938	0.231020	0.800856
lognormal	nornal	0.526974	0.262753	0.791196
	lognormal	0.493604	0.229382	0.757825
	gamma	0.500853	0.214638	0.787069
gamma	nornal	0.531176	0.194697	0.867655
	lognormal	0.497819	0.233923	0.761715
	gamma	0.505069	0.219218	0.790920

Figure 19. Confidence intervals of the posterior probability for diabetes, with the settings of the program in Table 2.

Figure 18 shows the confidence intervals of posterior probability for diabetes for FPG value equal to 126 mg/dl and confidence level $p = 0.95$.

The tables distinctly demonstrate the considerable magnitude of uncertainty and relative uncertainty associated with the posterior probability for diabetes at a FPG level of 126 mg/dl, the established threshold for the diagnosis of diabetes. Furthermore the posterior probabilities delineated in the tables suggest a limited concordance between the classification criteria of diabetes derived from the OGTT and FPG tests (ElSayed et al. 2023), as found previously in the existing literature (Tucker 2020).

5. Discussion

5.1. Reevaluation of Traditional Diagnostic Methods

Traditional diagnostic methods rely on the use of predetermined thresholds; however, this often fails to consider the complexities of disease pathology. While this has been historically effective, it may lack the ability to offer a holistic approach in today's patient-centered medicine, where personalized care is paramount (Obermeyer and Emanuel 2016). The evolving nature of diseases and shifts in patient demographics increase the complexity of the diagnostic process, pushing the boundaries of conventional methodologies. In this challenging context, Bayesian inference emerges as an alternative approach, offering probabilistic evaluations that can adapt to the individual patient profiles (Choi, Johnson, and Thurmond 2006; Chatzimichail and Hatjimihail 2023).

Nevertheless, estimating the uncertainty of posterior probabilities within Bayesian inference remains a pivotal challenge (Srinivasan, Westover, and Bianchi 2012). This issue is critically important in the context of diagnostic and screening tests for life-threatening conditions or those associated with considerable morbidity risk. It underscores the need for well-informed clinical judgments and comprehensive uncertainty evaluation in medical decision-making. Key examples include:

1. Cardiac troponin for diagnosing myocardial injury and infarction (Wereski et al. 2021).
2. Natriuretic peptides for the diagnosis of heart failure (Roberts et al. 2015).
3. D-dimer for diagnosing thromboembolic events (Freund et al. 2021).
4. FPG, OGTT, and glycated hemoglobin (HbA1c) for diagnosing diabetes (ElSayed et al. 2023).
5. OGTT for the diagnosis of gestational diabetes (Rani and Begum 2016).
6. Thyroid stimulating hormone (TSH), free serum triiodothyronine (T_3), and free serum thyroxine (T_4) for diagnosing thyroid dysfunction (Reyes Domingo, Avey, and Doull 2019).

7. Protein-to-creatinine ratio for the diagnosis of preeclampsia (Rodriguez-Thompson and Lieberman 2001).
8. Creatinine or cystatin C derived glomerular filtration rate (GFR), and albuminuria for diagnosing chronic kidney disease (Moynihan, Glassock, and Doust 2013).

The ability to quantify this uncertainty is not only an academic concern but also a practical necessity in improving diagnosis and patient outcomes.

To address this, our software explores the sampling, measurement, and combined uncertainty of Bayesian posterior probabilities. This exploration is not only vital for enhancing clinical decision-making but also plays a significant role in the fields of quality and risk management in laboratory medicine (Haeckel et al. 2016). Additionally, it may contribute to the design and implementation of test accuracy studies (J. Andre Knottnerus and Buntinx 2011; Hajian-Tilaki 2014). As mentioned in the Introduction section, despite the extensive body of research on Bayesian diagnosis and uncertainty as separate entities, the intersection of these two areas remains relatively unexplored (Baron 1994; Ashby and Smith 2000).

The illustrative case study, focusing on individuals aged 70 to 80 years, was designed to mitigate age-related variations in disease prevalence. This focus exemplifies the considerations required in modern diagnostics, where factors such as age, genetics, and lifestyle choices should be accounted for in the diagnostic equation.

Our software manages through its analysis of sampling, measurement, and combined uncertainty (as illustrated in Figures 5, 8, 11, 14, and 17), relative uncertainty (Figures 6, 9, 12, 15, and 18) and the corresponding confidence limits (Figures 7, 10, 13, 16, and 19), to display its versatility in addressing these diagnostic challenges. Although the software's calculations are highly sophisticated, its user-friendly interface renders it an effective tool for medical researchers and professionals.

The case study from Section 4 highlights the substantial impact of combined uncertainty on the diagnostic process. This finding emphasizes the predominant role of measurement uncertainty, and thus stresses the demanding path toward enhancing diagnostic accuracy. By improving the analytical methods of screening and diagnostic tests, the medical community could achieve more accurate diagnosis, leading to more effective and tailored patient care.

Looking ahead, future research should focus on improving the estimations of the uncertainty of posterior probabilities under a diverse array of clinically relevant parameter settings. To transition from research into practical application, it is necessary to focus on clinical decision analysis, studies on cost-effectiveness, and research on quality of care, which includes conducting implementation studies (J. Andre Knottnerus and Buntinx 2011). Such efforts are necessary in addressing the complex issues in diagnostic medicine and finding new and effective approaches to tackle ongoing challenges..

5.2. Limitations and Future Research Directions

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

1. Underlying assumptions:
   1.1. The existence of "gold standards" in diagnostics. If a "gold standard" does not exist, there are alternative approaches for classification (J. A. Knottnerus and Dinant 1997; Pfeiffer and Castle 2005; van Smeden et al. 2014).

   1.2. The hypothesis of parametric distribution of measurements or their transformations. However, existing literature underlines the robustness of nonparametric techniques in capturing complex data distributions (Wasserman 2006).

   1.3. The generally accepted bimodality of the measurands, although unimodal distributions could be considered (J. M. G. Wilson and Jungner 1968; Petersen and Horder 1992).

If these assumptions are not valid, the program may underestimate the standard uncertainty of the posterior probability for disease.

2. The use of first-order Taylor series approximations in uncertainty propagation calculations, where higher-order approximations may provide more accurate estimations (Joint Committee for Guides in Metrology 2011).

3. The approximation of the uncertainty of the prior probability for disease using the Agresti–Coull-adjusted Waldo interval, despite more accurate methods being available (Pires and Amado 2008).

4. The approximations of the sampling uncertainties for both the sample means and standard deviations, which can be improved for smaller samples or pronounced skewness observed in lognormal and gamma distributions (Schmoyeri et al. 1996; Bhaumik, Kapur, and Gibbons 2009).

5. The use of confidence intervals derived from the *t*-distribution despite the high relative uncertainty (Williams 2020). Though not typical in a Bayesian context, this can be employed instead of credible intervals as a practical tool under certain circumstances (Gelman et al. 2013; Stephens 2023).

While addressing these limitations would increase considerably computational complexity, they represent key areas for future enhancement (Joint Committee for Guides in Metrology 2008, 2020).

5.3. Case Study Shortcomings

The case study's main limitations include reliance on the OGTT as the reference method for diagnosing diabetes mellitus, despite several factors influencing glucose tolerance (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). Additionally, the lognormal distributions used only approximate the distributions of the FPG measurements from NHANES datasets, highlighting the need for more flexible statistical models.

5.4. Challenges in Bayesian Analysis for Disease Diagnosis

While Bayesian analysis may be beneficial in medical diagnostics, it presents certain challenges. For instance, the substantial uncertainty of the posterior probability for disease revealed in our study could lead to clinical indecision. Additionally, there is a notable lack of comprehensive statistical research on the distribution of measurands in both diseased and nondiseased populations, hindering further advancements in Bayesian analysis in this field.

5.5. Implications of Incomplete Data

1. Over-reliance on prior probabilities: Limited empirical data may cause an overdependence on prior probabilities, leading to distorted posterior probabilities and potentially flawed clinical decisions (O'Hagan et al. 2006).

1. Increased uncertainty: Insufficient data amplifies the uncertainty of computed posterior probabilities, which in turn could exacerbate clinical indecision (Berger 1985).

2. Bias risks: Unrepresentative datasets could introduce systemic bias, increasing the uncertainty in Bayesian computations (Gelman et al. 2013).

5.6. Analysis of the Double Sigmoidal Curve in Posterior Probability Estimation and Its Impact on Uncertainty

The posterior probability for disease curve, characterized by a double sigmoidal shape featuring two symmetrical sigmoid functions, presents compelling analytical perspectives in the field of medical diagnostic statistics. This configuration implies that the risk associated with the disease may escalate at both the lower and upper extremes of a given measurand, while a zone of relative safety exists in the intermediate range. Notably, the uncertainty associated with the posterior probability for disease becomes markedly pronounced along the steep segments of the double sigmoidal curve. This heightened uncertainty is attributable to the fact that minor variations in the measurand value can lead to significant alterations in the computed posterior probability.

5.7. Software Comparison

Our software easily generates a wide array of parametric plots and comprehensive tables for the analysis of posterior probability uncertainty. To the best of our knowledge, no software, including all major general or medical or Bayesian statistical and uncertainty quantification software packages (JASP® ver. 0.20.0, Mathematica® ver. 14.0, Matlab® ver. R2023b, MedCalc® ver. 20.2.1, metRology ver. 2023, NCSS® ver. 24.0.0, NIST Uncertainty Machine ver. 2.0.0, OpenBUGS ver. 3.3.0, R ver. 4.3.1, SAS® ver. 9.5, SPSS® ver. 29, Stan ver. 2.33.0, Stata® ver. 19, and UQLab ver. 2.0.0) provides this range of plots and tables without advanced programming.

6. Conclusions

The program we have developed represents a novel approach to estimating and analyzing the uncertainty of Bayesian posterior probabilities in disease diagnosis. This tool stands out not only for its innovative capabilities in the field of medical diagnostics but also as a significant educational and research asset. Considering the difficulties and complexities we have outlined, this software offers essential assistance in applying Bayesian methods and dealing with diagnostic uncertainties, thereby enhancing well-informed decision-making.

Looking forward, it seems imperative that future research should focus on improving this method with advanced statistical concepts and empirically validating it with comprehensive test accuracy studies. Such studies are essential to verify the efficacy and reliability of the program in real clinical settings. Additionally, it is necessary to expand its application across a diverse range of diagnostic modalities. Doing so could enable the program to address a broader spectrum of diagnostic challenges, further enhancing its utility and impact in the medical field.

Our research, undertaken alongside our prior work on the uncertainty of diagnostic accuracy measures (Chatzimichail and Hatjimihail 2021), creates a foundation for understanding uncertainties in diagnostic tests. With this consideration, we would recommend employing our approach in diagnostic accuracy research, aiming at formulating clear guidelines and establishing best practices to effectively integrate such information into clinical practice (J. Andre Knottnerus and Buntinx 2011; Whiting et al. 2013; Salameh et al. 2020; Schlattmann 2023).

Regarding regulatory issues, it is necessary to ensure that the application of the software adheres to the standards set forth by local regulatory authorities.

The potential of this program seems to be extending beyond its practical implications in medical diagnostics. As an educational resource, it could offer significant opportunities for training in medical statistics, particularly in the understanding of the uncertainty of Bayesian posterior probabilities. Its user-friendly interface, coupled with the depth of its analytical capabilities, makes it an effective learning tool for both aspiring and experienced professionals in the medical community.

In conclusion, the development and refinement of the Bayesian Diagnostic Uncertainty program are pivotal steps towards navigating the complexities of modern medical diagnostics. Its role in enhancing Bayesian diagnostic methods, coupled with its educational benefits, highlights its capability as a supporting tool in the ongoing evolution of medical practice and research.

Appendix A

Appendix A.1. Formalisms and Notation

Acronyms

PDF: probability density function
FPG: fasting plasma glucose
OGTT: oral glucose tolerance test
NHANES: National Health and Nutrition Examination Survey

Notation

Parameters

n_D: size of diseased population
μ_D: mean of diseased population
σ_D: standard deviation of diseased population
$n_{\bar{D}}$: size of nondiseased population
$\mu_{\bar{D}}$: mean of nondiseased population
$\sigma_{\bar{D}}$: standard deviation of nondiseased population
r: prior probability for disease (prevalence rate)
$u_s(x)$: standard sampling uncertainty of x
$u_m(x)$: standard measurement uncertainty of x
$u_c(x)$: standard combined uncertainty of x
n_U: number of quality control measurements

b_0: constant contribution to measurement uncertainty
b_1: measurement uncertainty proportionality constant
$\hat{u}$: mean standard measurement uncertainty
p: confidence level
v_{eff} : effective degrees of freedom

Functions

$P(A)$: probability of the event A
$P(A|B)$: conditional probability of the event A given the event B
$L(\theta|z)$: likelihood function
$F^{-1}(.)$: the inverse function $F(.)$

Bayes' Theorem

For the purposes of our study, Bayes' theorem is formulated as:

$$P(D|T) = \frac{P(T|D)P(D)}{P(T)} = \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|\bar{D})\big(1 - P(D)\big)}$$

where
$P(D|T)$ denotes the posterior probability of having a disease D given a test result T.
$P(T|D)$ denotes the likelihood of obtaining the result T given the presence of the disease D.
$P(T|\bar{D})$ denotes the likelihood of obtaining the result T given the absence of the disease D.
$P(D)$ is the prior probability or prevalence r of the disease D.
$P(T)$ is the overall probability of the result T.

According to Bayes' theorem, the posterior probability for a disease D given a test result $T = x$ and a parameter vector $\boldsymbol{\theta}$ is calculated as:

$$P(D|T) = \frac{L_D(\boldsymbol{\theta}|x)r}{L_D(x|\boldsymbol{\theta})r + L_{\bar{D}}(x|\boldsymbol{\theta})(1-r)} = \frac{f_D(x|\boldsymbol{\theta})r}{f_D(x|\boldsymbol{\theta})r + f_{\bar{D}}(x|\boldsymbol{\theta})(1-r)}$$

where r denotes the prior probability for disease, $L_D(\boldsymbol{\theta}|x)$ and $f_D(x|\boldsymbol{\theta})$ denote the likelihood function and the PDF of the test measurand in the presence of the disease, respectively, while $L_{\bar{D}}(x|\boldsymbol{\theta})$ and $f_{\bar{D}}(x|\,\boldsymbol{\theta})$ are the respective functions in the absence of the disease.

Appendix A.1.1. Parametric Distributions

It is assumed that the test measurands of the diseased or nondiseased populations follow the normal, lognormal or gamma distribution. The domains of random variables for the respective distributions are defined as follows:

1. The domain of a random variable X following a normal distribution is the set of all real numbers, denoting $-\infty < X < \infty$.

2. The domain of a random variable X following a lognormal distribution is the set of all positive real numbers, denoting $0 < X < \infty$.

3. The domain of a random variable X following a gamma distribution is the set of all positive real numbers, denoting $0 < X < \infty$.

Appendix A.1.2. Calculations of the Posterior Probability for Disease and Its Uncertainty

These calculations are detailed in Supplementary File II (Refer to Supplementary Files).

Appendix A.2. Software Availability and Requirements

Program name: Bayesian Diagnostic Uncertainty
Version: 1.0.0
Project home page: https://www.hcsl.com/Tools/BayesianDiagnosticUncertainty/ (accessed on 4 January 2024). Available at: https://www.hcsl.com/Tools/BayesianDiagnosticUncertainty/BayesianDiagnosticUncertainty.nb (accessed on 4 January 2024)
Operating systems: Microsoft Windows 10+, Linux 3.15+, Apple macOS 11+

Programming language: Wolfram Language

Other software requirements:

For running the program and reading the BayesianDiagnosticUncertaintyCalculations.nb file Wolfram Player® ver. 12.0+ is required, freely available at: https://www.wolfram.com/player/ (accessed 18 December 2023) or Wolfram Mathematica® ver. 12.0+

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

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

Appendix A.3. A Note about the Program

About the Program Controls

The program features an intuitive tabbed user interface, designed to streamline user interaction and facilitate effortless navigation across its multiple modules and sub-modules.

The numerical settings are defined by the user with menus or sliders. Sliders can be finely manipulated by holding down the *alt* key or *opt* key while dragging the mouse. They be even more finely manipulated by also holding the *shift* and/or *ctrl* keys.

Dragging with the mouse while pressing the *ctrl*, *alt*, or *opt* keys zooms plots in or out.

Range of input parameters

x: $maximum(\mu_{\bar{D}} - 5\sigma_{\bar{D}}, 0) - \mu_D + 5\sigma_{\bar{D}}$
n_D: 2–10,000
μ_D: 0.1–10,000
σ_D: 0.01–1000
$n_{\bar{D}}$: 2–10,000
$\mu_{\bar{D}}$: 0.1–10,000
$\sigma_{\bar{D}}$: 0.01–1000
r: 0.010–0.500
n_U: 20–10,000
b_0: 0–$\sigma_{\bar{D}}$
b_1: 0–0.1
p: 0.900–0.999

7.2. Appendix II: Software Availability and Requirements

Program name: Bayesian Diagnostic Uncertainty

Version: 1.0.0

Project home page: https://www.hcsl.com/Tools/BayesianDiagnosticUncertainty/ (accessed 4 January 2024).

7.3. Appendix III: A Note about the Program

About the program controls

The program features an intuitive tabbed user interface, designed to streamline user interaction and facilitate effortless navigation across its multiple modules and sub-modules.

The numerical settings are defined by the user with menus or sliders. Sliders can be finely manipulated by holding down the *alt* key or *opt* key while dragging the mouse. They be even more finely manipulated by also holding the *shift* and/or *ctrl* keys.

Dragging with the mouse while pressing the *ctrl*, *alt*, or *opt* keys zooms plots in or out.

Range of input parameters

x: $maximum(\mu_{\bar{D}} - 5\sigma_{\bar{D}}, 0) - \mu_D + 5\sigma_{\bar{D}}$

n_D : 2 – 10,000

μ_D: 0.1 – 10,000

σ_D: 0.01 – 1,000

$n_{\bar{D}}$: 2 – 10,000

$\mu_{\bar{D}}$: 0.1 – 10,000

$\sigma_{\bar{D}}$: 0.01 – 1,000

r: 0.010 – 0.500

n_U: 20 – 10,000

b_0 : 0 – $\sigma_{\bar{D}}$

b_1: 0 – 0.1

p: 0.900 – 0.999

8. Supplementary Files

8.1 Supplementary File I

BayesianDiagnosticUncertainty.nb: The program as a Wolfram Mathematica Notebook. Available at https://www.hcsl.com/Tools/BayesianDiagnosticUncertainty/BayesianDiagnosticUncertainty.nb

8.2 Supplementary File II

BayesianUncertaintyCalculations.nb: The calculations for the estimation Bayesian posterior probability for disease and its standard uncertainty, in a Wolfram Mathematica Notebook. Available at https://www.hcsl.com/Supplements/ SBDU.zip

8.3 Supplementary File III

BayesianDiagnosticUncertaintyInterface.pdf: A brief description of the interface of the program. Available at: https://www.hcsl.com/Documents/BayesianDiagnosticUncertaintyInterface.pdf

9. Statements

9.1 Institutional Review Board Statement

Data collection was carried out following the rules of the Declaration of Helsinki. The Ethics Review Board of the National Center for Health Statistics approved data collection and posting the data online for public use. National Center for Health Statistics NHANES—NCHS Research Ethics Review Board Approval (Protocols #2005-06 and #2011-17), available online at: https://www.cdc.gov/nchs/nhanes/irba98.htm (accessed on 20 December 2023).

9.2 Informed Consent Statement

Written consent was obtained from each subject participating in the survey.

9.3 Data Availability Statement

The data presented in this study are available at https://wwwn.cdc.gov/nchs/nhanes/default.aspx (Accessed at 20/12/2023).

9.4 Conflicts of Interest

The authors declare no conflict of interest.

10. 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.

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.

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

Box, G. E. P., and D. R. Cox. 1964. “An Analysis of Transformations.” *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 26 (2): 211–43.

Ceriotti, Ferruccio, Pilar Fernandez-Calle, George G. Klee, Gunnar Nordin, Sverre Sandberg, Thomas Streichert, Joan-Lluis Vives-Corrons, and Mauro Panteghini. 2017. “Criteria for Assigning Laboratory Measurands to Models for Analytical Performance Specifications Defined in the 1st EFLM Strategic Conference.” *Clinical Chemistry and Laboratory Medicine: CCLM / FESCC* 55 (2): 189–94.

Chatzimichail, Theodora, and Aristides T. Hatjimihail. 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.

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.

D’Agostino, Ralph, and E. S. Pearson. 1973. “Tests for Departure from Normality. Empirical Results for the Distributions of b_2 and $\sqrt{b_1}$.” *Biometrika* 60 (3): 613–22.

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

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.

Geisser, Seymour, and Wesley O. Johnson. 2006. *Modes of Parametric Statistical Inference*. John Wiley & Sons.

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

Haeckel, Rainer, Werner Wosniok, Eberhard Gurr, Burkhard Peil, and on behalf of Arbeitsgruppe Richtwer. 2016. “Supplements to a Recent Proposal for Permissible Uncertainty of Measurements in Laboratory Medicine,” January. https://doi.org/10.1515/labmed-2015-0112.

Hajian-Tilaki, Karimollah. 2014. “Sample Size Estimation in Diagnostic Test Studies of Biomedical Informatics.” *Journal of Biomedical Informatics* 48 (April): 193–204.

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

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.

Lehmann, Erich L., and Joseph P. Romano. 2008. *Testing Statistical Hypotheses*. Springer New York.

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.

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.

Nielsen, Allan Aasbjerg. 2007. “Least Squares Adjustment: Linear and Nonlinear Weighted Regression Analysis.” Technical University of Denmark. http://www2.imm.dtu.dk/pubdb/edoc/imm2804.pdf.

Obermeyer, Ziad, and Ezekiel J. Emanuel. 2016. “Predicting the Future - Big Data, Machine Learning, and Clinical Medicine.” *The New England Journal of Medicine* 375 (13): 1216–19.

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.

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

Rostron, Peter D., Tom Fearn, and Michael H. Ramsey. 2020. “Confidence Intervals for Robust Estimates of Measurement Uncertainty.” *Accreditation and Quality Assurance* 25 (2): 107–19.

Salameh, Jean-Paul, Patrick M. Bossuyt, Trevor A. McGrath, Brett D. Thombs, Christopher J. Hyde, Petra Macaskill, Jonathan J. Deeks, et al. 2020. “Preferred Reporting Items for Systematic Review and Meta-Analysis of Diagnostic Test Accuracy Studies (PRISMA-DTA): Explanation, Elaboration, and Checklist.” *BMJ* 370 (August): m2632.

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

Schlattmann, Peter. 2023. “Tutorial: Statistical Methods for the Meta-Analysis of Diagnostic Test Accuracy Studies.” *Clinical Chemistry and Laboratory Medicine: CCLM / FESCC* 61 (5): 777–94.

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.

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, Alison F., Bethany Shinkins, Peter S. Hall, Claire T. Hulme, and Mike P. Messenger. 2019. “Toward a Framework for Outcome-Based Analytical Performance Specifications: A Methodology Review of Indirect Methods for Evaluating the Impact of Measurement Uncertainty on Clinical Outcomes.” *Clinical Chemistry* 65 (11): 1363–74.

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.

Topol, Eric J. 2014. "Individualized Medicine from Prewomb to Tomb." *Cell* 157 (1): 241–53.

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.

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.

Wasserman, Larry. 2006. *All of Nonparametric Statistics*. Springer Science & Business Media.

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.

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

Whiting, Penny F., Anne W. S. Rutjes, Marie E. Westwood, Susan Mallett, and QUADAS-2 Steering Group. 2013. "A Systematic Review Classifies Sources of Bias and Variation in Diagnostic Test Accuracy Studies." *Journal of Clinical Epidemiology* 66 (10): 1093–1104.

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.

11. Permanent Citation:

Chatzimichail T, Hatjimihail AT. A Software Tool for Estimating Uncertainty of Bayesian Posterior Probability for Disease. Technical Report XXVI. Drama: Hellenic Complex Systems Laboratory, 2024. https://www.hcsl.com/TR/hcsltr26/hcsltr26.pdf

12. License

First Published: January 4, 2024

Revised: September 27, 2024