Hellenic Complex Systems Laboratory

A Software Tool for Calculating the Uncertainty of Diagnostic Accuracy Measures

Technical Report XXVIII

Theodora Chatzimichail and Aristides T. Hatjimihail

2024

Abstract

Screening and diagnostic tests classify people into diseased and nondiseased populations. Although diagnostic accuracy measures are used to evaluate the correctness of a classification in clinical research and practice, there has been limited research on their uncertainty. This work aimed to develop a tool for calculating the uncertainty of diagnostic accuracy measures, as diagnostic accuracy is fundamental to clinical decision-making. For this reason, the freely available interactive program Diagnostic Uncertainty has been developed in the Wolfram Language. The program provides six modules with nine submodules for calculating and plotting the standard combined, measurement and sampling uncertainty and the resultant confidence intervals of various diagnostic accuracy measures of screening or diagnostic tests, which measure a normally distributed measurand applied at a single point in time to samples of nondiseased and diseased populations. This is done for differing sample sizes, mean and standard deviation of the measurand, diagnostic threshold, and standard measurement uncertainty of the test. The program's application is demonstrated with an illustrative example of glucose measurements in samples of diabetic and non-diabetic populations that shows the calculation of the uncertainty of diagnostic accuracy measures. The interactive program presented is user-friendly and can be used as a flexible educational and research tool in medical decision-making to calculate and explore the uncertainty of diagnostic accuracy measures.

Keywords: diagnostic accuracy measures; uncertainty; measurement uncertainty; sampling uncertainty; diagnostic tests; screening tests;

1. Introduction

An increasing number of in vitro screening and diagnostic tests are extensively used as binary classifiers in medicine to classify people in the non-overlapping classes of populations with and without a disease, categorized as quantitative and qualitative. Quantitative and many qualitative screening or diagnostic tests are based on measurements. There is a joint probability distribution of the measurements in the diseased and nondiseased populations. A diagnostic threshold or cutoff point is defined to classify patients with and without a disease using a test based on a measurement. If the measurement is above the threshold, the patient is classified as test-positive; otherwise, the patient is classified as test-negative (or vice versa) (Fig. 1).

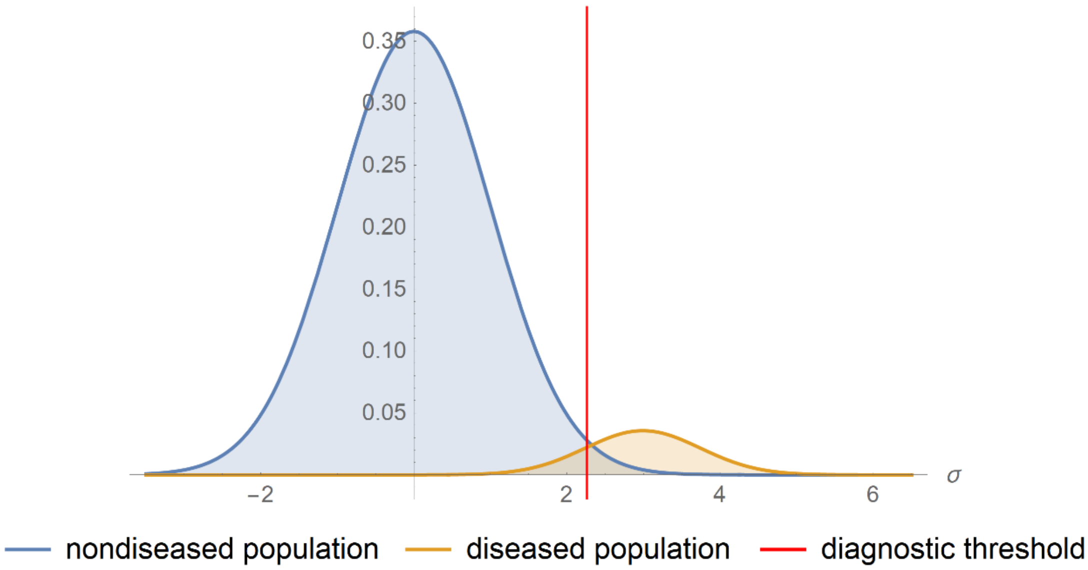


Figure 1. Probability density function plots. The probability density functions plots of a measurand in a nondiseased and diseased population.

The possible test results are summarized in Table 1.

Table 1: A2x2 Contingency Table

		populations	
		nondiseased	**diseased**
test results	**negative**	true negative (TN)	false negative (FN)
	positive	false positive (FP)	true positive (TP)

Diagnostic accuracy measures

Among the various diagnostic accuracy measures (DAM) presented in the literature, only a select few are consistently applied in clinical research and practice (1). These primary measures include:

1. Sensitivity (Se), specificity (Sp), overall diagnostic accuracy (ODA), diagnostic odds ratio (DOR), and likelihood ratios for positive or negative result ($LR+$ and $LR-$ respectively) which are defined conditionally on the true disease status (2) and are prevalence invariant.
2. Overall diagnostic accuracy (ODA), which is also defined conditionally on the true disease status and is prevalence dependent.
3. Positive predictive and negative predictive values (PPV and NPV), which are defined conditionally on the test outcome and are prevalence dependent.

The natural frequency and the equivalent probability definitions of the diagnostic accuracy measures derived from Table 1 and analyzed by the program, are presented in Table 2. The symbols are explained in the Notation Section.

Table 1: Natural frequency and probability definitions of diagnostic accuracy measures

measure	natural frequency definition	probability definition
Se	$\frac{TP}{FN+TP}$	$Pr(T\|D)$
Sp	$\frac{TN}{TN+FP}$	$Pr(\overline{T}\|\overline{D})$
PPV	$\frac{TP}{FP+TP}$	$Pr(D\|T)$
NPV	$\frac{TN}{TN+FN}$	$Pr(\overline{D}\|\overline{T})$
ODA	$\frac{TN+TP}{TN+FN+TP+FP}$	$Pr(D)\ Pr(T\|D) + Pr(\overline{D})\ \Pr(\overline{T}\|\overline{D})$
DOR	$\frac{TN\ TP}{FN\ FP}$	$\frac{\frac{Pr(T\|D)}{Pr(\overline{T}\|D)}}{\frac{Pr(T\|\overline{D})}{Pr(\overline{T}\|\overline{D})}}$
LR+	$\frac{TP(FP+TN)}{FP(FN+TP)}$	$\frac{Pr(T\|D)}{Pr(T\|\overline{D})}$
LR-	$\frac{FN(FP+TN)}{TN(FN+TP)}$	$\frac{Pr(\overline{T}\|D)}{Pr(\overline{T}\|\overline{D})}$
J	$\frac{TN\ TP - FN\ FP}{(TN+FP)(FN+TP)}$	$Pr(T\|D) +\ Pr(\overline{T}\|\overline{D}) - 1$
ED	$\sqrt{\left(\frac{FN}{FN+TP}\right)^2 + \left(\frac{FP}{TN+FP}\right)^2}$	$\sqrt{Pr(\overline{T}\|D)^2 + Pr(T\|\overline{D})^2}$
CZ	$\frac{TN\ TP}{(TN+FP)(FN+TP)}$	$Pr(T\|D)\ Pr(\overline{T}\|\overline{D})$

Uncertainty of diagnostic accuracy measures

Uncertainty is an expression of imperfect or deficient information. When quantifiable, it can be represented by probability (3). The following components of the combined uncertainty of the diagnostic accuracy measures will be considered:

Measurement uncertainty

As there is inherent variability in any measurement process, 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" (4).

The measurement uncertainty gradually replaces the total analytical error concept (5).

Sampling uncertainty

Diagnostic accuracy measures are estimated by applying a diagnostic method to samples of populations. Therefore, sampling variability contributes to their uncertainty (6).

2. Implementation

Computational methods

For the calculation of the uncertainty of the diagnostic accuracy measures of a screening or diagnostic test based on a measurand, it is assumed that:

1. There is a reference ("gold standard") diagnostic method classifying correctly a subject as diseased or nondiseased (7).
2. The measurand's values or their transforms (8, 9) are normally distributed in the diseased and nondiseased populations.
3. Measurement uncertainty is normally distributed and homoscedastic in the diagnostic threshold's range.

After that, we use the term measurand to describe either the normally distributed value of a measurand or its normally distributed applicable transform.

Diagnostic accuracy measures

The definitions of the diagnostic accuracy measures can be expressed in terms of sensitivity (*Se*), specificity (*Sp*), and rate of prevalence (r). These definitions are derived from Table 2 and presented in Table 3.

Table 3: Definitions of diagnostic accuracy measures versus sensitivity and specificity

measure	definition
PPV	$\frac{Se\,v}{Se\,r + (1-Sp)(1-r)}$
NPV	$\frac{Sp\,(1-r)}{Sp\,(1-r) + (1-Se)r}$
ODA	$Se\,v + Sp\,(1-r)$
DOR	$\frac{\frac{Se}{1-Se}}{\frac{1-Sp}{Sp}}$
LR+	$\frac{Se}{1-Sp}$
LR-	$\frac{1-Se}{Sp}$
J	$Se + Sp - 1$
ED	$\sqrt{(1-Se)^2 + (1-Sp)^2}$
CZ	$Se\,Sp$

The functions of sensitivity (Se) and specificity (Sp), hence the functions of all the above diagnostic accuracy measures, can be expressed in terms of the cumulative distribution function of the normal distribution, consequently of the error function and the complementary error function.

The error function $erf(x)$ is defined as:

$$\boldsymbol{erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt, x \geq 0} \quad (1)$$

while the complementary error function $erfc(x)$ is defined as:

$$\boldsymbol{erfc(x) = 1 - erf(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2}\, dt, x \geq 0} \quad (2)$$

Following the definition of the sensitivity and specificity of a test (Table 2), the respective functions versus diagnostic threshold d are calculated as:

$$\boldsymbol{se(d, \mu_D, \sigma_D) = 1 - \Psi(d, \mu_D, \sigma_D) = \frac{1}{2}\left(1 + erf\left(\frac{-d+\mu_D}{\sqrt{2}\sigma_D}\right)\right)} \quad (3)$$

$$\boldsymbol{sp(d, \mu_{\bar{D}}, \sigma_{\bar{D}}) = \Psi(d, \mu_{\bar{D}}, \sigma_{\bar{D}}) = \frac{1}{2} erfc\left(\frac{-d+\mu_{\bar{D}}}{\sqrt{2}\sigma_{\bar{D}}}\right)} \quad (4)$$

Uncertainty

The uncertainty of an input parameter or a diagnostic accuracy measure x, can be expressed as standard and expanded uncertainty. The former, denoted as $u(x)$, equals the standard deviation of x. The latter, denoted as $U(x)$, is an interval around x, including x with probability p (10).

Measurement uncertainty of means and standard deviations

The standard measurement uncertainty is estimated as described in "Expression of Measurement Uncertainty in Laboratory Medicine" (11). Bias may be considered a standard measurement uncertainty component (12).

Sampling uncertainty 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 standard uncertainties of m_P and s_P are:

$$\boldsymbol{u_s(m_P) = \frac{s_P}{\sqrt{n_P}}} \quad (5)$$

$$\boldsymbol{u_s(s_P) = \frac{s_p}{\sqrt{2(n_P - 1)}}} \quad (6)$$

Combined uncertainty of means and standard deviations

If u_m the standard measurement uncertainty of a screening or diagnostic test measuring a measurand, then the standard combined uncertainties of the mean m_p and standard deviation s_p are:

$$\boldsymbol{u_c(m_P) = \sqrt{\frac{s^2}{n} + u_m^2}} \quad (7)$$

$$\boldsymbol{u_{cs}(s_P) = \sqrt{\frac{s^2}{2(n-1)} + u_m^2}} \quad (8)$$

Sampling uncertainty of prevalence

If $n_{\bar{D}}$ and n_D the respective numbers of nondiseased and diseased in a population sample, then the standard uncertainty u_r of the prevalence rate $r = \frac{n_D}{n_{\bar{D}} + n_D}$ of the disease can be approximated as:

$$\boldsymbol{u_s(r) = \sqrt{\frac{(2+n_{\bar{D}})(2+n_D)}{(4+n_{\bar{D}}+n_D)^3}}} \quad (9)$$

according to the Agresti-Coull adjustment of the Waldo interval (13).

Combined uncertainty of diagnostic accuracy measures

The standard combined uncertainty u_c of each diagnostic accuracy measure is calculated by applying the rules of uncertainty propagation from the input values to the calculated diagnostic accuracy measure, using first-order Taylor series expansion (14).

When there are l uncorrelated components of uncertainty, with standard uncertainties u_i respectively, then:

$$\boldsymbol{u_c(x)} = \sqrt{\sum_{i=1}^{l} \boldsymbol{u_i(x)^2}} \qquad 10$$

Expanded uncertainty of diagnostic accuracy measures

The effective degrees of freedom v_{eff} of the combined uncertainty u_c are calculated using Welch–Satterthwaite formula (15, 16):

$$\boldsymbol{v_{eff}} = \frac{u_c^4}{\left(\sum_{i=1}^{l} \frac{u_i^2}{v_i}\right)^2} \qquad 11$$

Consequently, if v_{min} and v_{max} the minimum and maximum of (v_1, v_2,…,v_l), it can be shown that:

$$\boldsymbol{v_{min} \leq v_{eff} \leq v_{max}} \qquad 12$$

If $F_v(z)$ is the cumulative distribution function of the Student's t-distribution with v degrees of freedom and u_c the standard combined uncertainty of a diagnostic accuracy measure, its expanded combined uncertainty U_c, at a confidence level p, is calculated as:

$$\boldsymbol{U_c(x)} = \left(\boldsymbol{F_v^{-1}}\left(\frac{1-p}{2}\right)\boldsymbol{u_c(x)}, \boldsymbol{F_v^{-1}}\left(\frac{1+p}{2}\right)\boldsymbol{u_c(x)}\right) \qquad 13$$

The resultant confidence interval of x, at the same confidence level p, is:

$$\boldsymbol{CI_p(x)} = \left(\boldsymbol{x} + \boldsymbol{F_v^{-1}}\left(\frac{1-p}{2}\right)\boldsymbol{u_c(x)}, \boldsymbol{x} + \boldsymbol{F_v^{-1}}\left(\frac{1+p}{2}\right)\boldsymbol{u_c(x)}\right) \qquad 14$$

The program

To calculate the uncertainty of the diagnostic accuracy measures, an interactive program written in Wolfram Language (17) was developed using Wolfram Mathematica® (18). This program was designed to provide six modules with nine submodules for calculating and plotting the standard uncertainty and the resultant confidence intervals of various diagnostic accuracy measures of a screening or diagnostic test applied at a single point in time in nondiseased and diseased population samples (Fig. 2). The test measures a measurand in the population samples, for varying values of their sizes, mean and standard deviation, and standard measurement uncertainty of the measurand. It is assumed that the measurands and measurement uncertainty are normally distributed and that measurement uncertainty is homoscedastic.

The program is freely available as a Wolfram Mathematica Notebook (.nb) (Supplementary file I: Uncertainty.nb). It can be run on Wolfram Player® or Wolfram Mathematica® (refer to Sections 6 and 7).

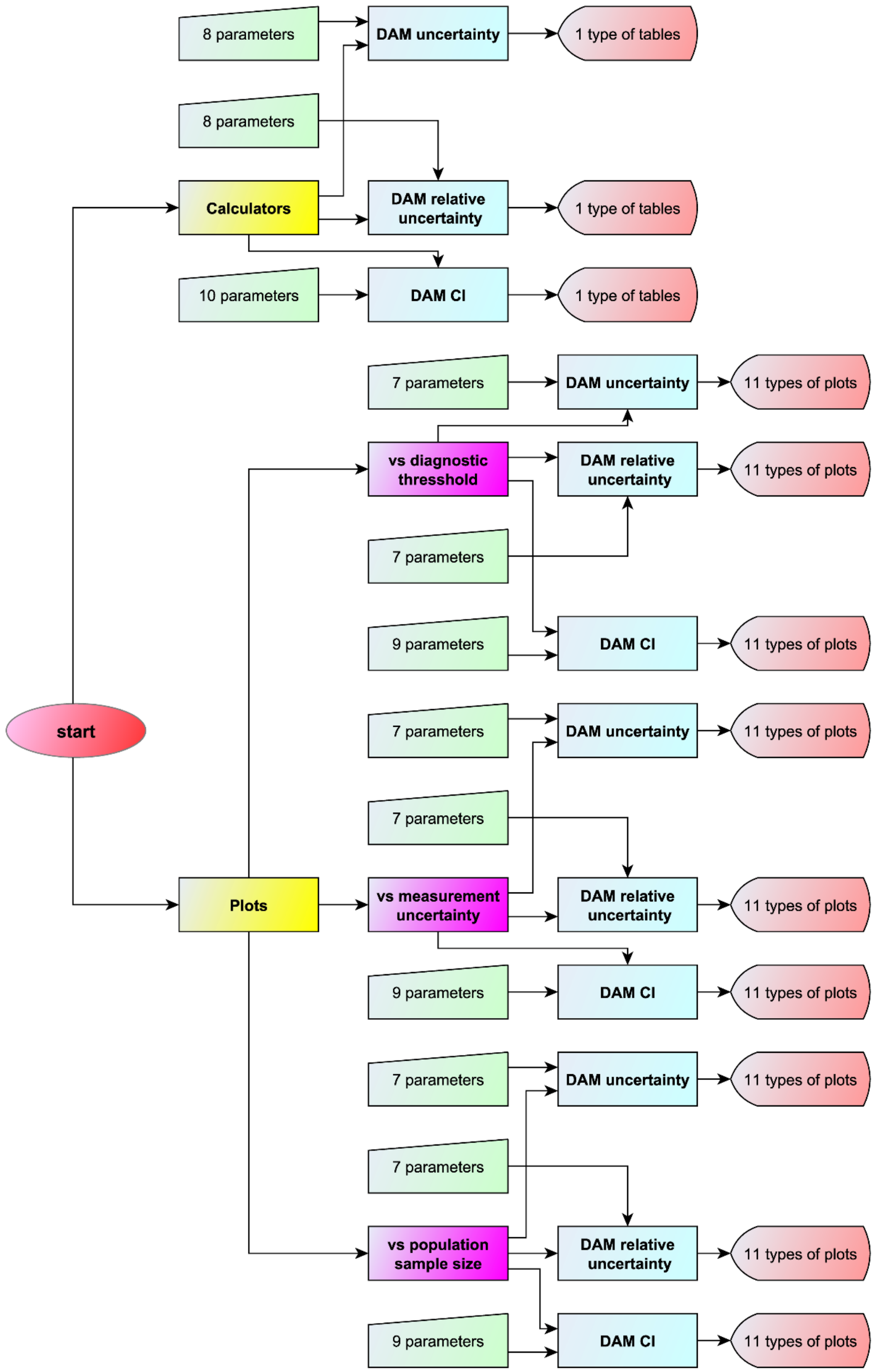


Figure 2. Program flowchart. A simplified flowchart of the program with the number of the input parameters and of the output types for each module.

3. Results

Interface of the program

The program's modules include panels with controls that allow the interactive manipulation of various parameters, as described in the Appendix. These are the following:

Plots

Plots vs. diagnostic threshold module

Diagnostic accuracy measures standard uncertainties plots submodule

The values of the standard combined, measurement, and sampling uncertainties of the diagnostic accuracy measures of a screening or diagnostic test are plotted versus the diagnostic threshold (Fig. 3).

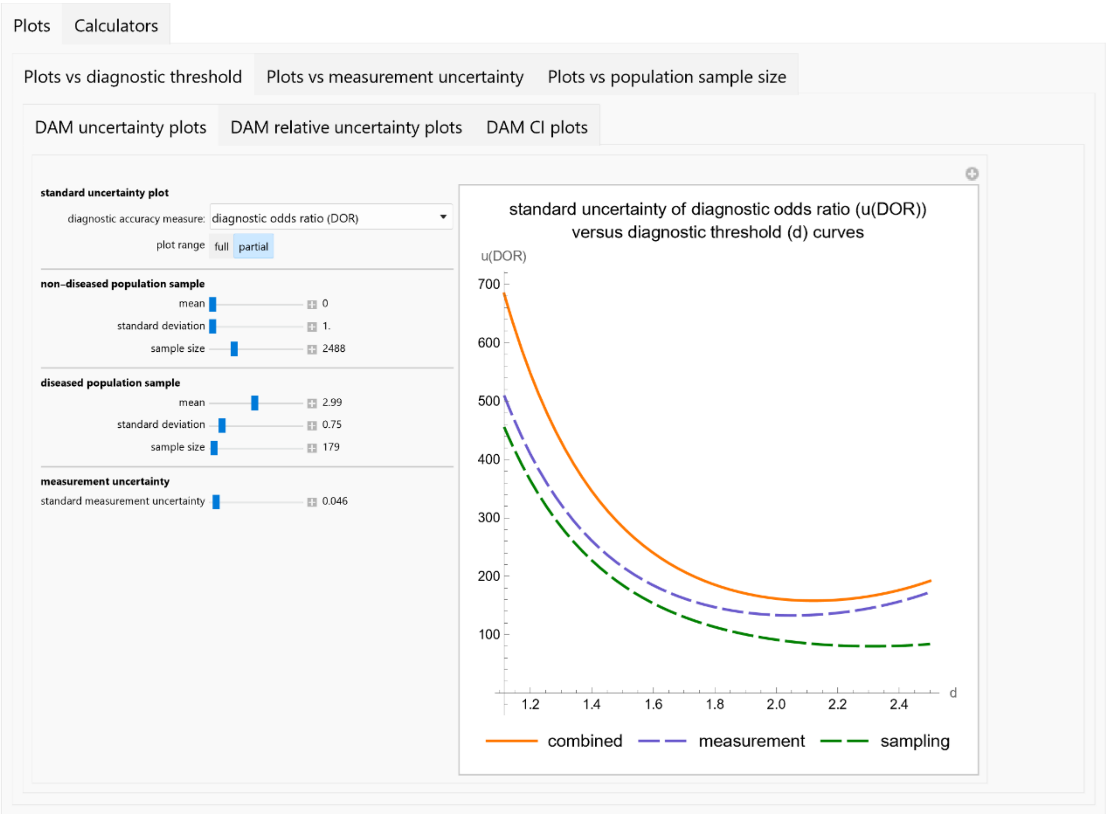


Figure 3. Plots vs. diagnostic threshold module, DAM uncertainty submodule screenshot. Standard combined, sampling, and measurement uncertainty of diagnostic odds ratio ($u(DOR)$) versus diagnostic threshold (d) curve plot, with the settings shown at the left.

Diagnostic accuracy measures relative standard uncertainties plots submodule

The values of the relative standard combined, measurement, and sampling uncertainties of the diagnostic accuracy measures of a screening or diagnostic test are plotted versus the diagnostic threshold (Fig. 4).

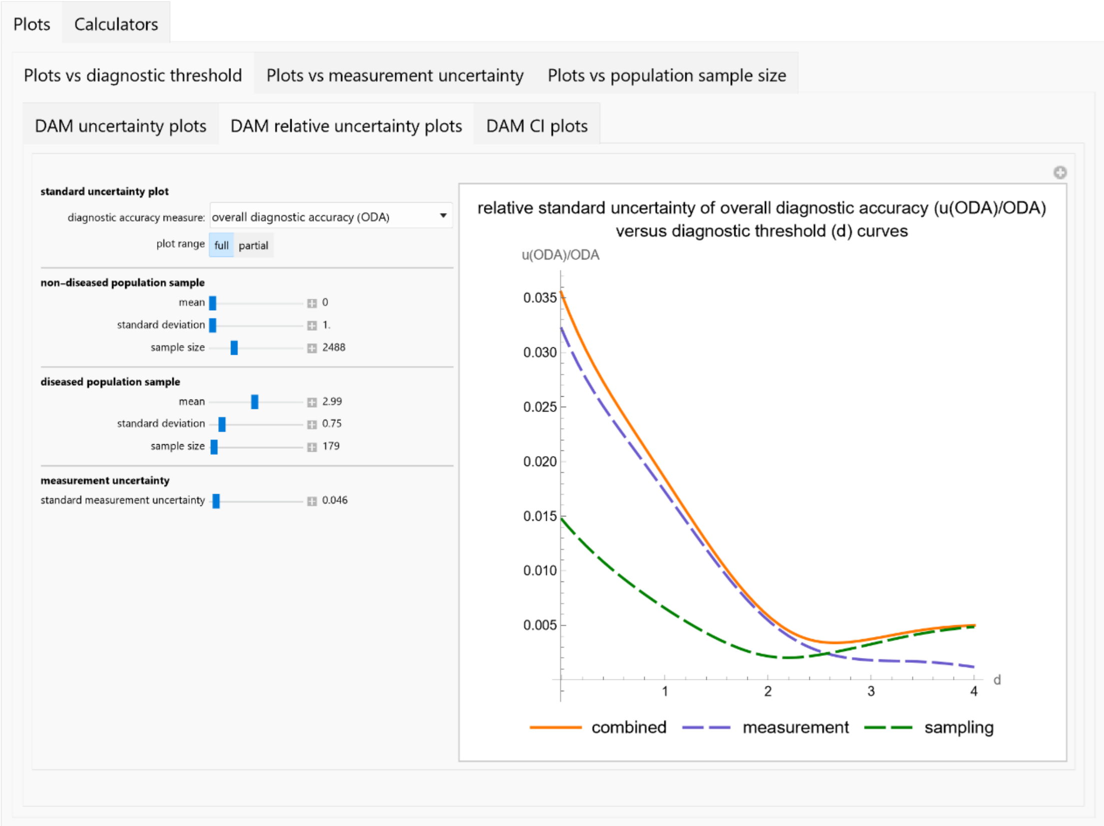


Figure 4. Plots vs. diagnostic threshold module, DAM relative uncertainty submodule screenshot. The relative standard combined, sampling, and measurement uncertainty of overall diagnostic accuracy ($u(ODA)$) versus diagnostic threshold (d) curve plot, with the settings shown at the left.

Confidence intervals of diagnostic accuracy measures plots submodule

The values of the lower and upper bounds of the confidence intervals of a diagnostic accuracy measure of a screening or diagnostic test at a selected confidence level are plotted versus the diagnostic threshold of the test (Fig. 5).

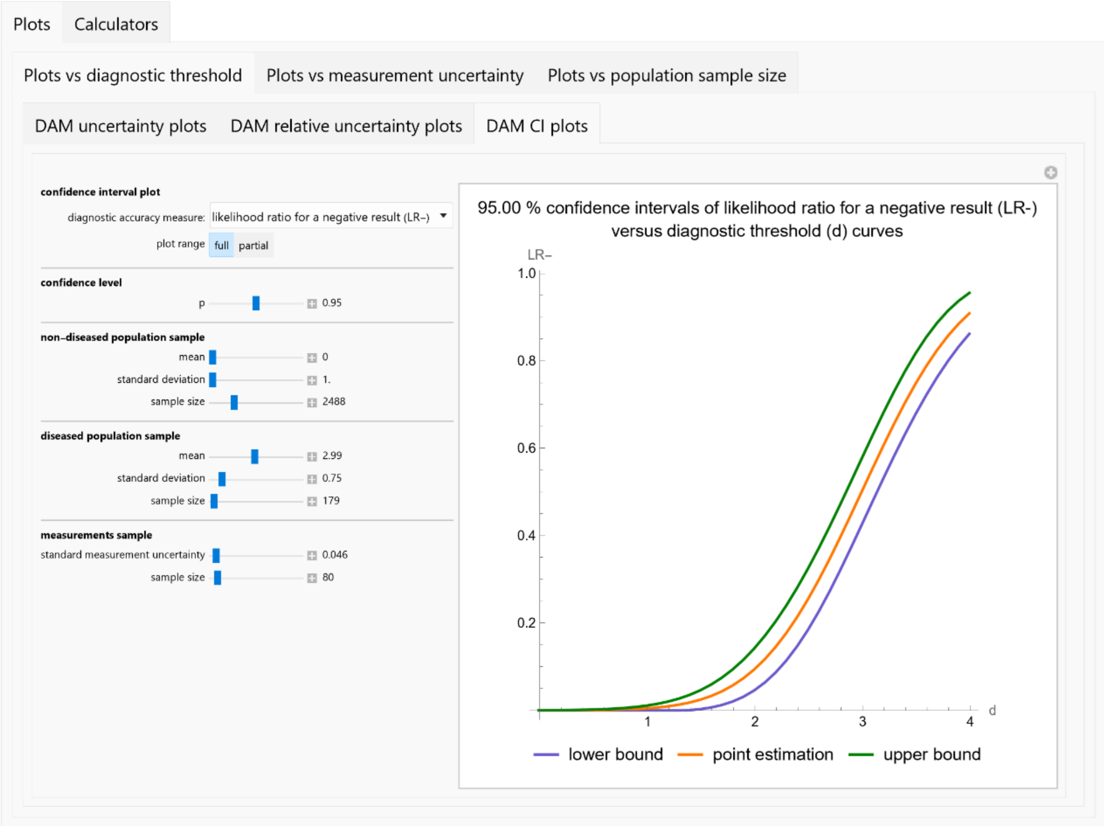


Figure 5. Plots vs diagnostic threshold module, DAM CI submodule screenshot. Confidence intervals of the likelihood ratio for a negative test result $(u(LR\text{-}))$ versus diagnostic threshold (d) curves plot, with the settings shown at the left.

Plots vs. measurement uncertainty module

Diagnostic accuracy measures standard uncertainties plots submodule

The values of the standard combined, measurement, and sampling uncertainties of the diagnostic accuracy measures of a screening or diagnostic test are plotted versus the measurement uncertainty of the test (Fig. 6).

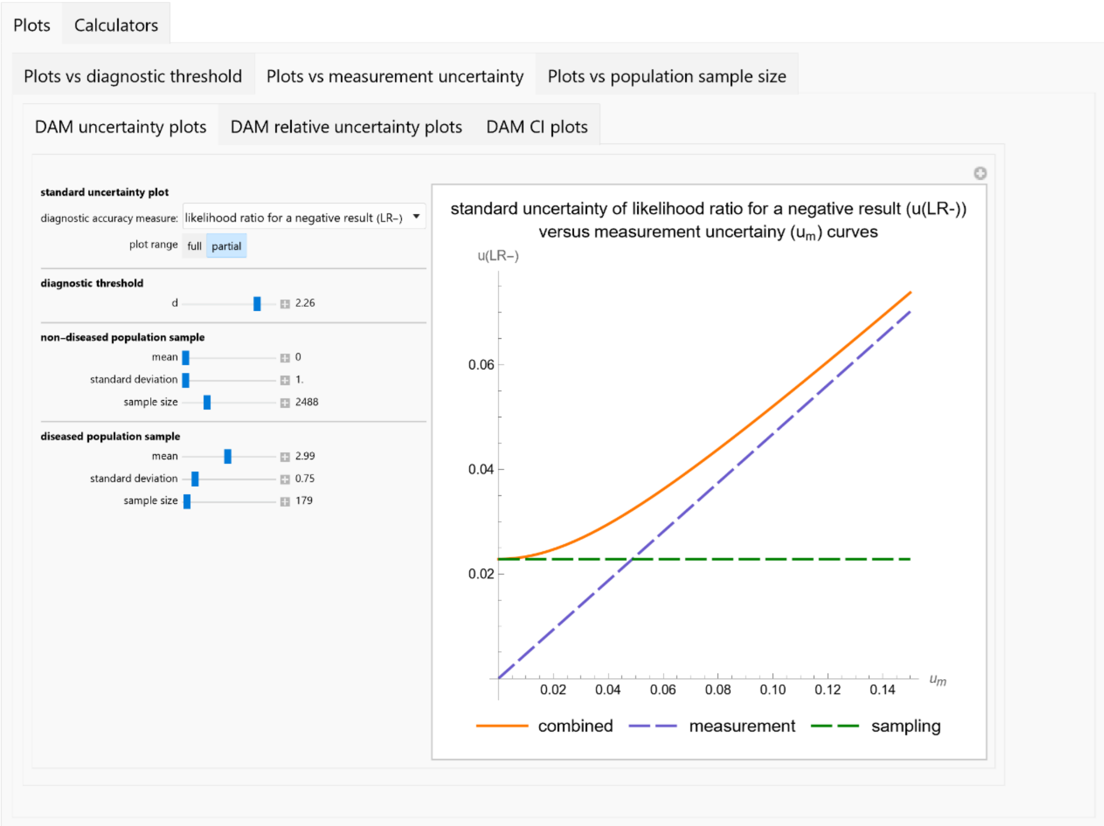


Figure 6. Plots vs. measurement uncertainty module, DAM uncertainty submodule screenshot. Standard combined, sampling, and measurement uncertainty of likelihood ratio for a negative test result $(u(LR\text{-}))$ versus standard measurement uncertainty (u_m) curve plot, with the settings shown at the left.

Diagnostic accuracy measures relative standard uncertainties plots submodule

The values of the relative standard combined, measurement, and sampling uncertainties of the diagnostic accuracy measures of a screening or diagnostic test are plotted versus the measurement uncertainty of the test (Fig. 7).

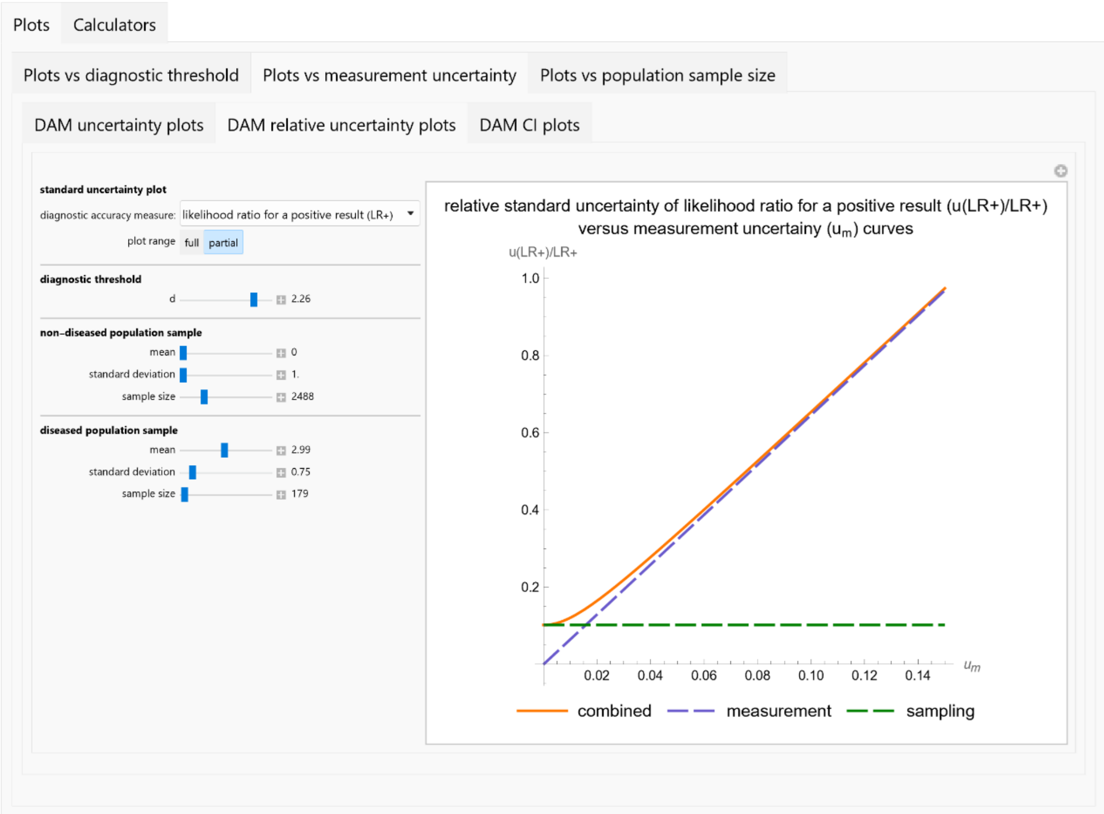


Figure 7. Plots vs. measurement uncertainty module, DAM relative uncertainty submodule screenshot. The relative standard combined, sampling, and measurement uncertainty of likelihood ratio for a positive test result $(u(LR+))$ vs. measurement uncertainty (u_m) curves plot, with the settings shown at the left.

Confidence intervals of diagnostic accuracy measures plots submodule

The values of the lower and upper bounds of the confidence intervals of a diagnostic accuracy measure of a screening or diagnostic test at a selected confidence level are plotted versus the measurement uncertainty of the test (Fig. 8).

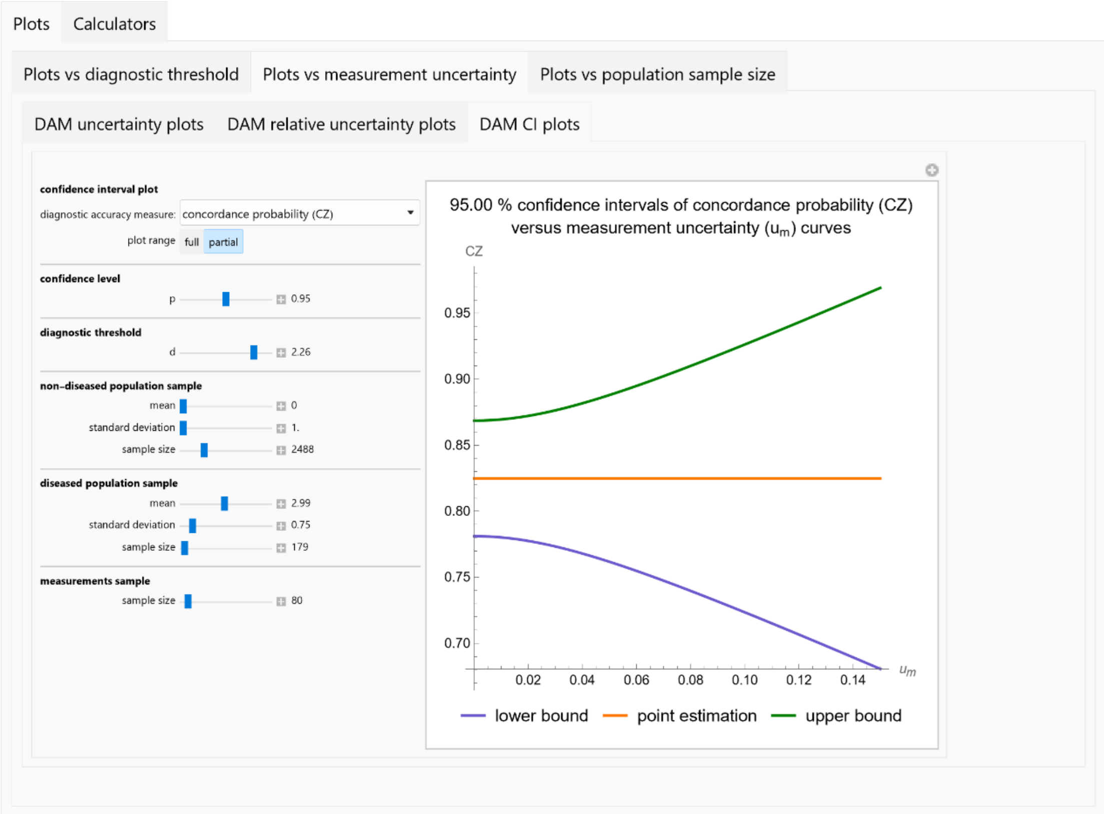


Figure 8. Plots vs. measurement uncertainty module, DAM CI submodule screenshot. Confidence intervals of concordance probability (CZ) versus standard measurement uncertainty (u_m) curves plot, with the settings at the left.

Plots vs population sample size

Diagnostic accuracy measures standard uncertainties plots submodule

The values of the standard combined measurement and sampling uncertainties of the diagnostic accuracy measures of a screening or diagnostic test are plotted versus the total population sample size (Fig. 9).

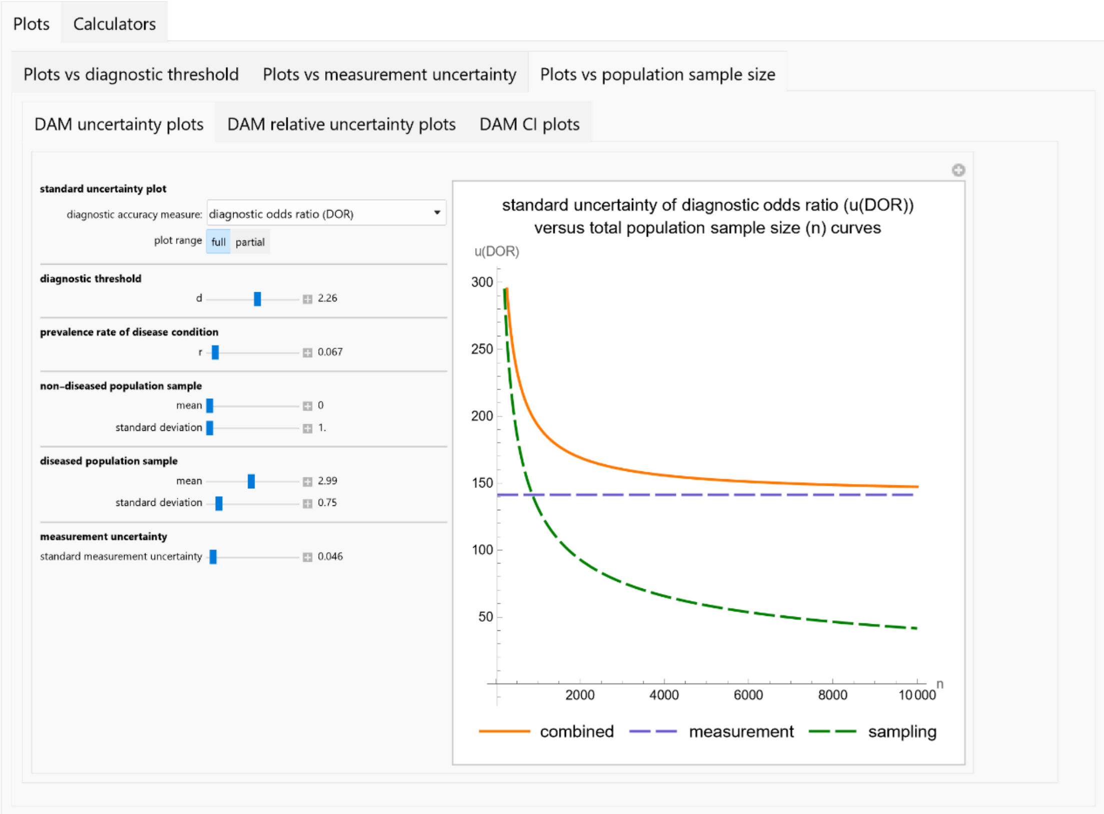


Figure 9. Plots vs. total population sample size module, DAM uncertainty submodule screenshot. Standard combined, sampling, and measurement uncertainty of diagnostic odds ratio ($u(DOR)$) versus total population sample size (n) curves plot, with the settings shown at the left.

Diagnostic accuracy measures relative standard uncertainties plots submodule

The values of the relative standard combined, measurement, and sampling uncertainties of the diagnostic accuracy measures of a screening or diagnostic test are plotted versus the total population sample size (Fig. 10).

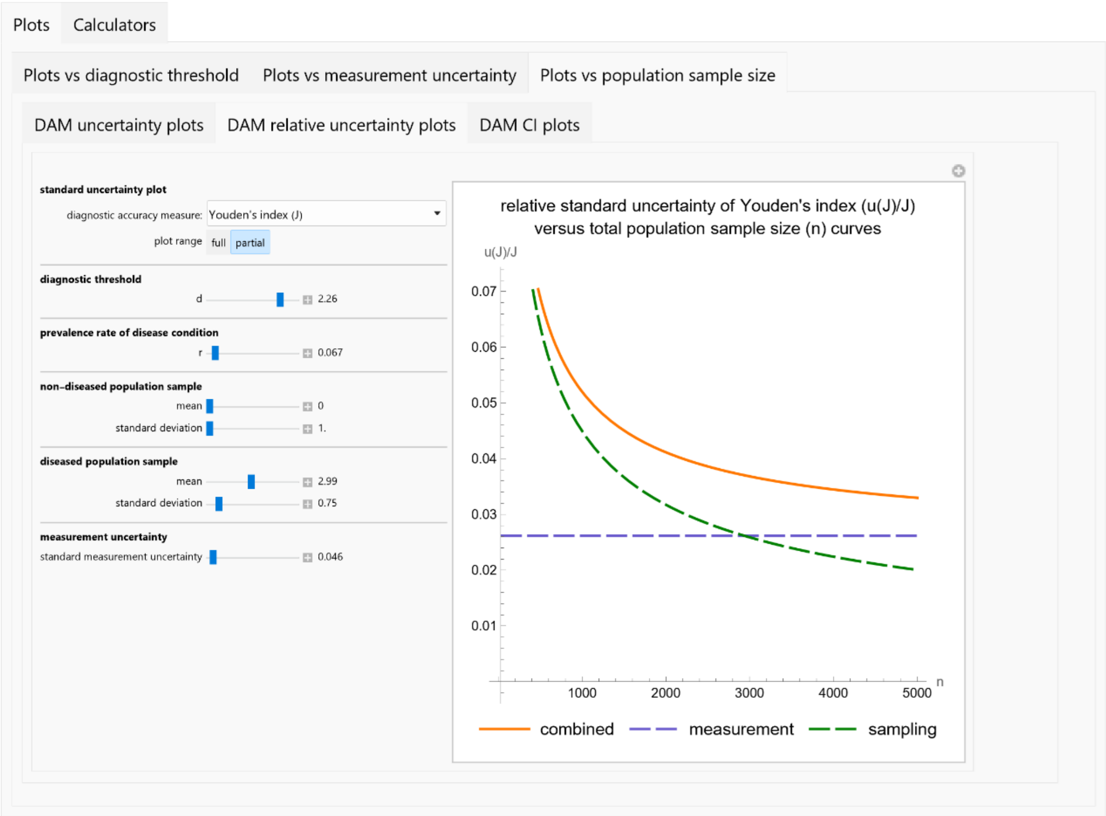


Figure 10. Plots vs. population sample size module, DAM relative uncertainty submodule screenshot. The relative standard combined, sampling, and measurement uncertainty of Youden's index ($u(J)$) versus total population sample size (n) curves plot, with the settings shown at the left.

Confidence intervals of diagnostic accuracy measures plots submodule

The values of the lower and upper bounds of the confidence intervals of a diagnostic accuracy measure of a screening or diagnostic test are plotted at a selected confidence level versus the total population sample size (Fig. 11).

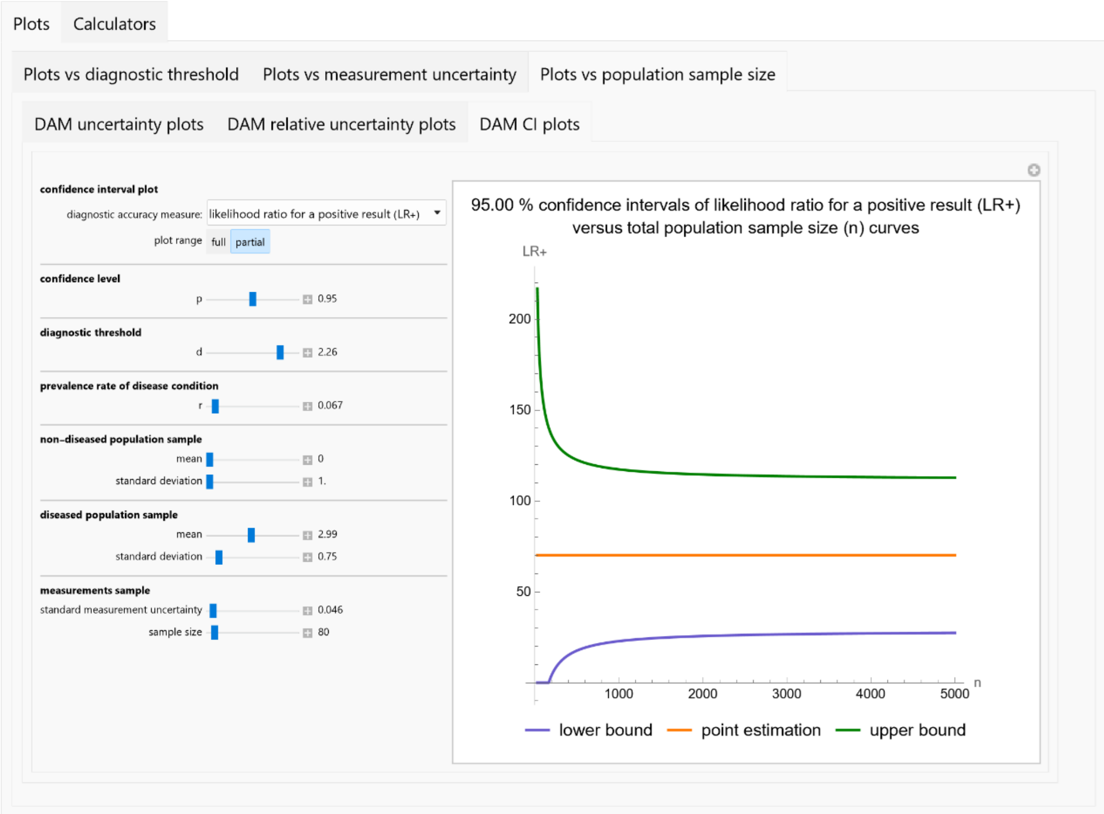
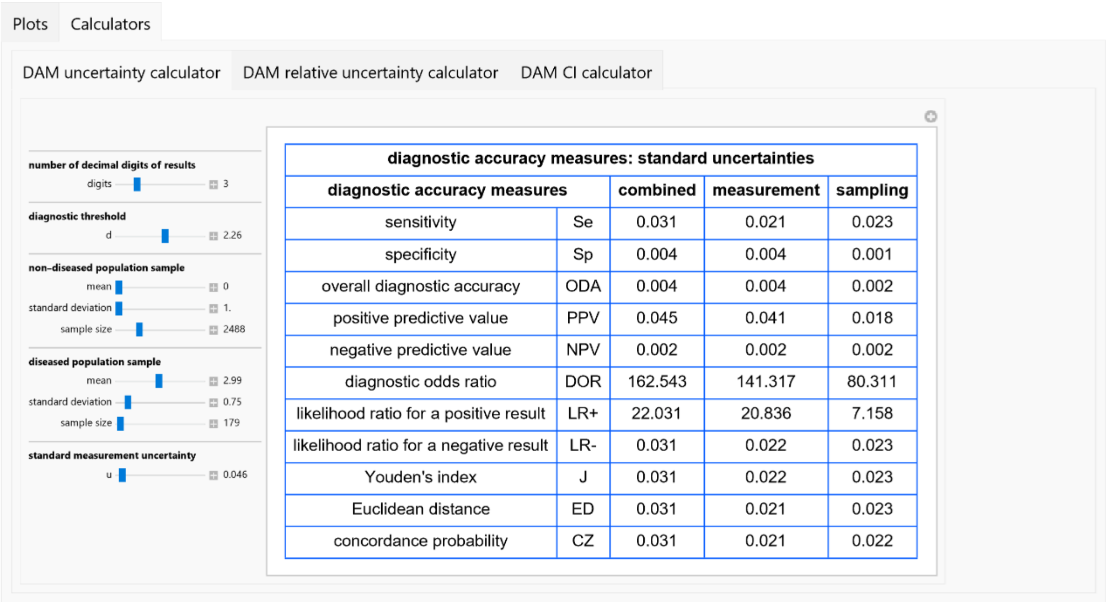


Figure 11. Plots vs population sample size module, DAM CI submodule screenshot. Confidence intervals of the likelihood ratio for a positive test result ($LR+$) versus total population sample size (n) curves plot, with the settings at the left.

Calculators

Diagnostic accuracy measures standard uncertainties calculator module

The standard combined, measurement and sampling uncertainties of the diagnostic accuracy measures of a screening or diagnostic test at a selected diagnostic threshold are calculated and presented in a table (Fig. 12).

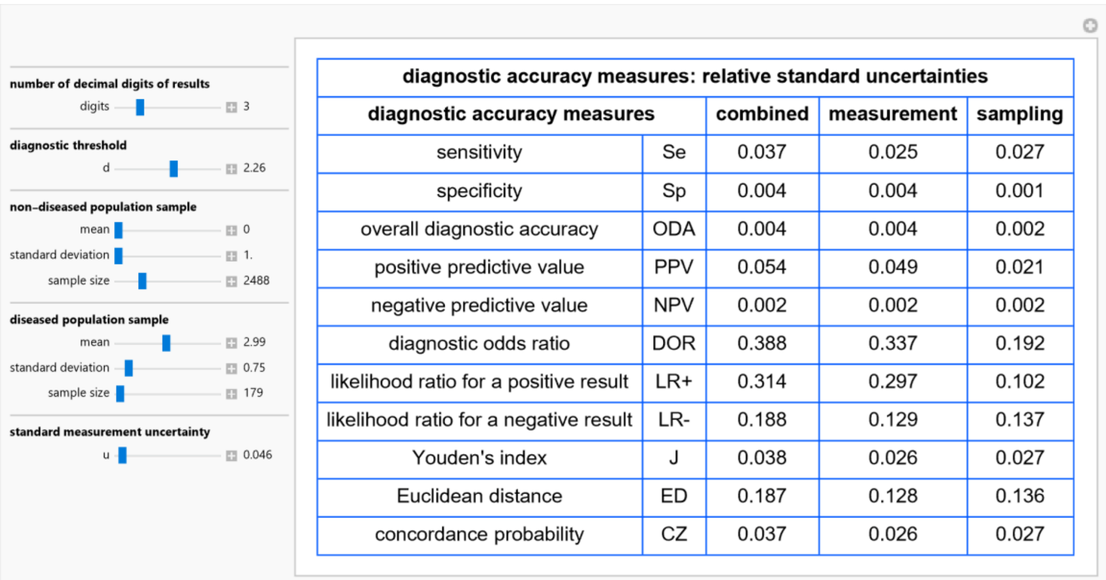


diagnostic accuracy measures: standard uncertainties				
diagnostic accuracy measures		combined	measurement	sampling
sensitivity	Se	0.031	0.021	0.023
specificity	Sp	0.004	0.004	0.001
overall diagnostic accuracy	ODA	0.004	0.004	0.002
positive predictive value	PPV	0.045	0.041	0.018
negative predictive value	NPV	0.002	0.002	0.002
diagnostic odds ratio	DOR	162.543	141.317	80.311
likelihood ratio for a positive result	LR+	22.031	20.836	7.158
likelihood ratio for a negative result	LR-	0.031	0.022	0.023
Youden's index	J	0.031	0.022	0.023
Euclidean distance	ED	0.031	0.021	0.023
concordance probability	CZ	0.031	0.021	0.022

Figure 12. ***Calculators module, DAM uncertainty calculator submodule screenshot.*** The calculated standard combined, measurement and sampling uncertainty of diagnostic accuracy measures, with the settings shown at the left.

Diagnostic accuracy measures relative standard uncertainties calculator module

The values of the relative standard combined, measurement, and sampling uncertainties of the diagnostic accuracy measures of a screening or diagnostic test at a selected diagnostic threshold are calculated and presented in a table (Fig. 13).

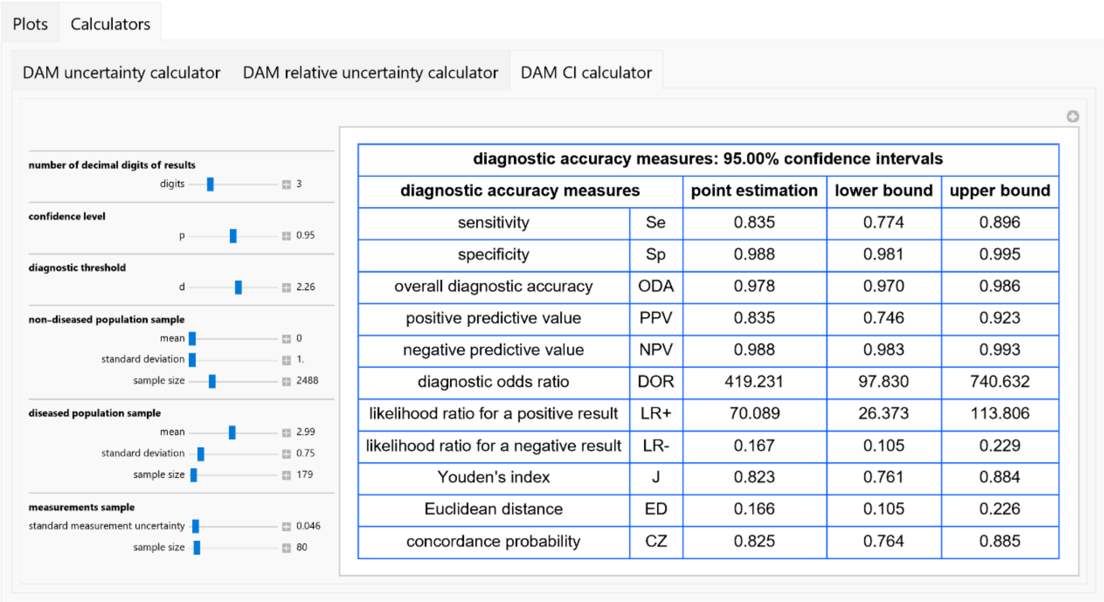


diagnostic accuracy measures: relative standard uncertainties				
diagnostic accuracy measures		combined	measurement	sampling
sensitivity	Se	0.037	0.025	0.027
specificity	Sp	0.004	0.004	0.001
overall diagnostic accuracy	ODA	0.004	0.004	0.002
positive predictive value	PPV	0.054	0.049	0.021
negative predictive value	NPV	0.002	0.002	0.002
diagnostic odds ratio	DOR	0.388	0.337	0.192
likelihood ratio for a positive result	LR+	0.314	0.297	0.102
likelihood ratio for a negative result	LR-	0.188	0.129	0.137
Youden's index	J	0.038	0.026	0.027
Euclidean distance	ED	0.187	0.128	0.136
concordance probability	CZ	0.037	0.026	0.027

Figure 13. ***Calculators module, DAM relative uncertainty calculator submodule screenshot.*** The calculated relative standard combined, measurement, and sampling uncertainty of diagnostic accuracy measures, with the settings shown at the left.

Diagnostic accuracy measures confidence intervals calculator module

The values of the lower and upper bounds of the confidence intervals of various diagnostic accuracy measures of a screening or diagnostic at a selected confidence level and diagnostic threshold are calculated and presented in a table (**Error! Reference source not found.** 14).



diagnostic accuracy measures: 95.00% confidence intervals				
diagnostic accuracy measures		point estimation	lower bound	upper bound
sensitivity	Se	0.835	0.774	0.896
specificity	Sp	0.988	0.981	0.995
overall diagnostic accuracy	ODA	0.978	0.970	0.986
positive predictive value	PPV	0.835	0.746	0.923
negative predictive value	NPV	0.988	0.983	0.993
diagnostic odds ratio	DOR	419.231	97.830	740.632
likelihood ratio for a positive result	LR+	70.089	26.373	113.806
likelihood ratio for a negative result	LR-	0.167	0.105	0.229
Youden's index	J	0.823	0.761	0.884
Euclidean distance	ED	0.166	0.105	0.226
concordance probability	CZ	0.825	0.764	0.885

Figure 14. Calculators *module, DAM CI calculator module screenshot.* Calculated confidence intervals of diagnostic accuracy measures, with the settings shown at the left.

Brief interface documentation of the program is available as Supplemental File II: DiagnosticUncertaintyCalculations.pdf. Supplemental File III: DiagnosticUncertaintyCalculations.nb presents the calculations for estimating diagnostic accuracy measures and their standard uncertainty (refer to Section 7)

Illustrative case study

The program was applied to a bimodal joint distribution of log-transformed blood glucose measurements in non-diabetic and diabetic Malay populations during an oral glucose tolerance test (OGTT) (19). Briefly, after ingesting 75 g glucose monohydrate, the 2-h postprandial blood glucose of 2667 Malay adults aged 40 - 49 years was measured with reflectance photometry. The estimated prevalence of diabetes was 0.067; therefore, the respective numbers of diseased and nondiseased were 179 and 2488. To estimate the distribution of the measurand in the diabetic and non-diabetic populations, it was assumed that the measurement coefficient of variation and bias were equal to 4% and 2%, respectively. The log-transformed measurands of each population were normally distributed, as shown in Fig. 1. In this case study, the normalized log-transformed measurand means and standard deviations in the diseased and nondiseased populations, the standard measurement uncertainty and the diagnostic threshold were expressed in units equal to the standard deviation of the log-transformed measurand in the nondiseased population. The normalized log-transformed diagnostic threshold of 2.26 corresponds to the American Diabetes Association (ADA) diagnostic threshold for diabetes 2-h postprandial glucose during OGTT, equal to 11.1 mmol/l (20). The normalized log-transformed standard measurement uncertainty 0.046 of the test corresponds to standard measurement uncertainty equal to 2% of the mean of the measurand in the non-diabetic population or equivalently to the coefficient of variation equal to 2%.

The results of the illustrative case study are presented:

1. In the plots of Fig. 3-11 and 15-21.
2. In the chart of Fig. 22
3. In the tables of Fig. 12-14.

The parameter settings of Fig. 15-22 are presented in Table 4.

Table 4. The parameter settings of Fig. 15 - 22

settings	Fig. 15, 16	Fig. 17	Fig. 18	Fig. 19	Fig. 20	Fig. 21	Fig. 22
p	-	0.95	-	0.95	-	0.95	-
d	0.0– 4.0	0.0 – 4.0	2.26	-	2.26	-	2.26
r	-	-	-	-	0.067	0.067	-
μ_D	2.99	2.99	2.99	2.99	2.99	2.99	2.99
σ_D	0.75	0.75	0.75	0.75	0.75	0.75	0.75
n_D	179	179	179	179	-	-	179
$\mu_{\overline{D}}$	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$\sigma_{\overline{D}}$	1.0	1.0	1.0	1.0	1.0	1.0	1.0
$n_{\overline{D}}$	2488	2488	2488	2488	-	-	2488
n	-	-	-	-	30-5000	30-5000	-
u_m	0.046	0.046	0-0.15	0-0.15	0.046	0.046	0.046
n_U	-	80	-	80	-	80	-

The symbols of the settings column are explained in the Notation Section.

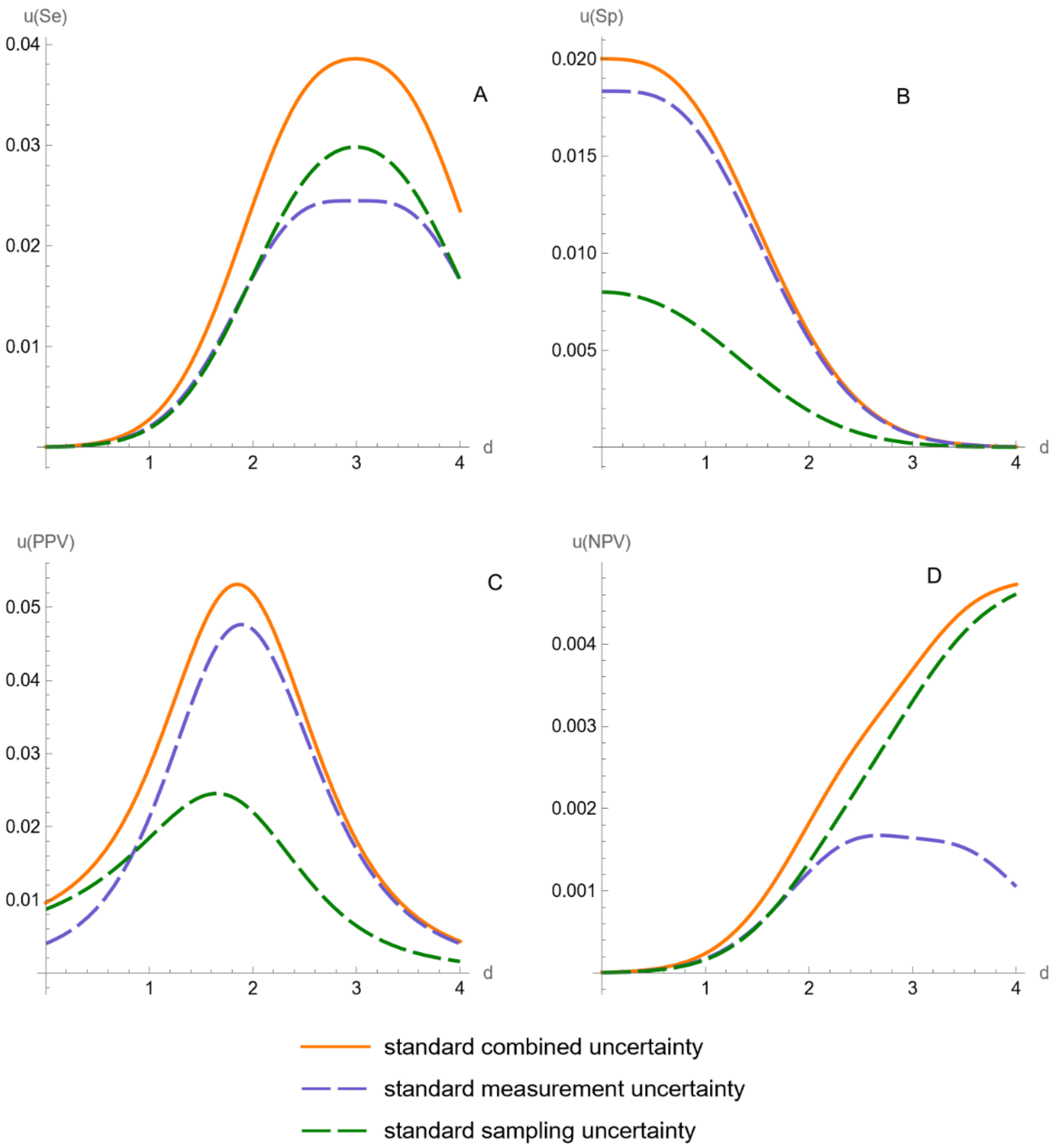


Figure 15. DAM standard uncertainties versus diagnostic threshold plots. Plots of standard combined, sampling, and measurement uncertainties of (A) sensitivity (*Se*), (B) specificity (*Sp*), (C) positive predictive value (*PPV*), and (D) negative predictive value (*NPV*) versus diagnostic threshold (*d*) curves, with the respective parameters in Table 4**.**

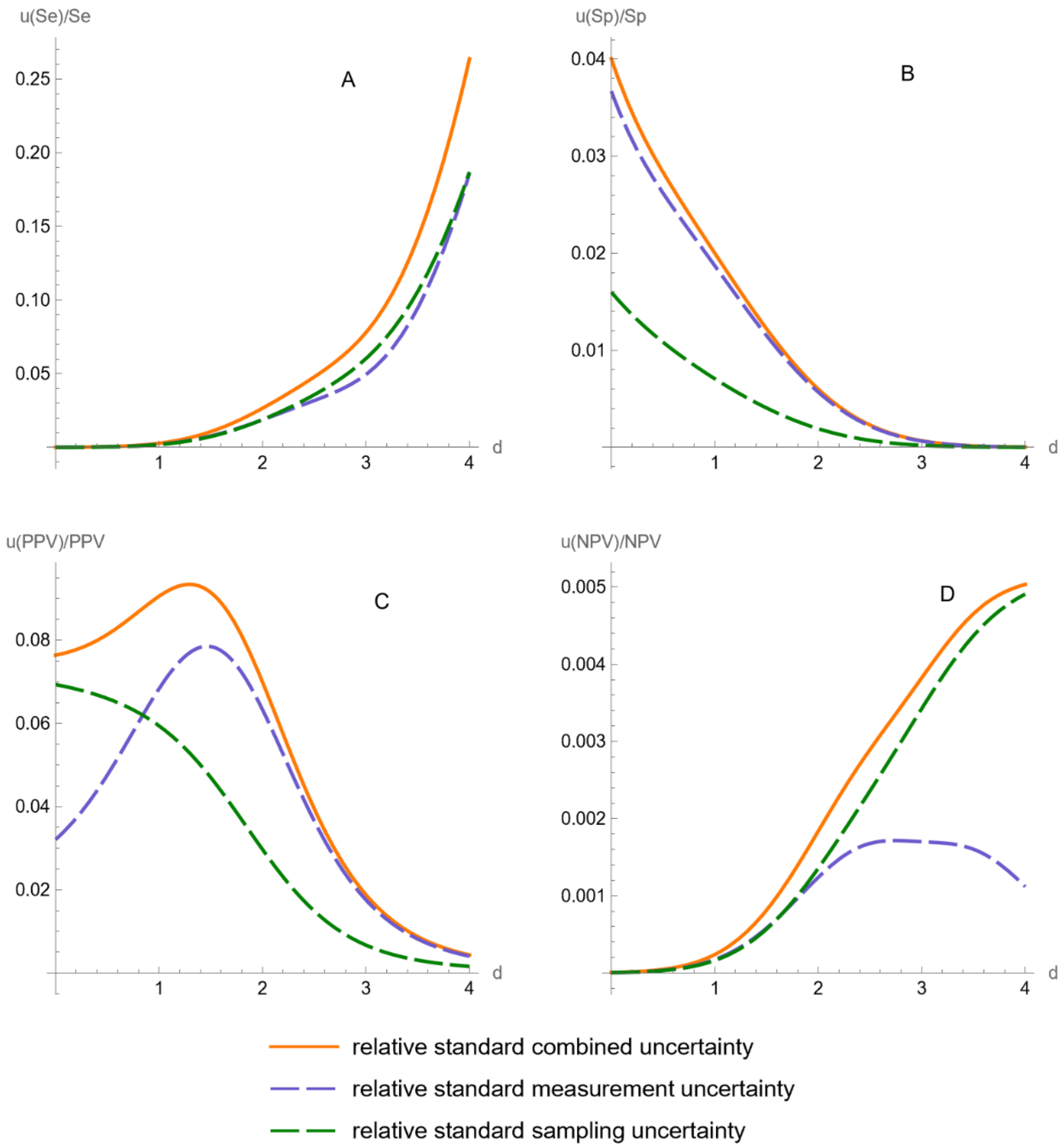


Figure 16. DAM relative standard uncertainties versus diagnostic threshold plots. Plots of relative standard combined, sampling, and measurement uncertainties of (A) sensitivity (*Se*), (B) specificity (*Sp*), (C) positive predictive value (*PPV*), and (D) negative predictive value (*NPV*) versus diagnostic threshold (*d*) curves, with the respective parameters in Table 4**.**

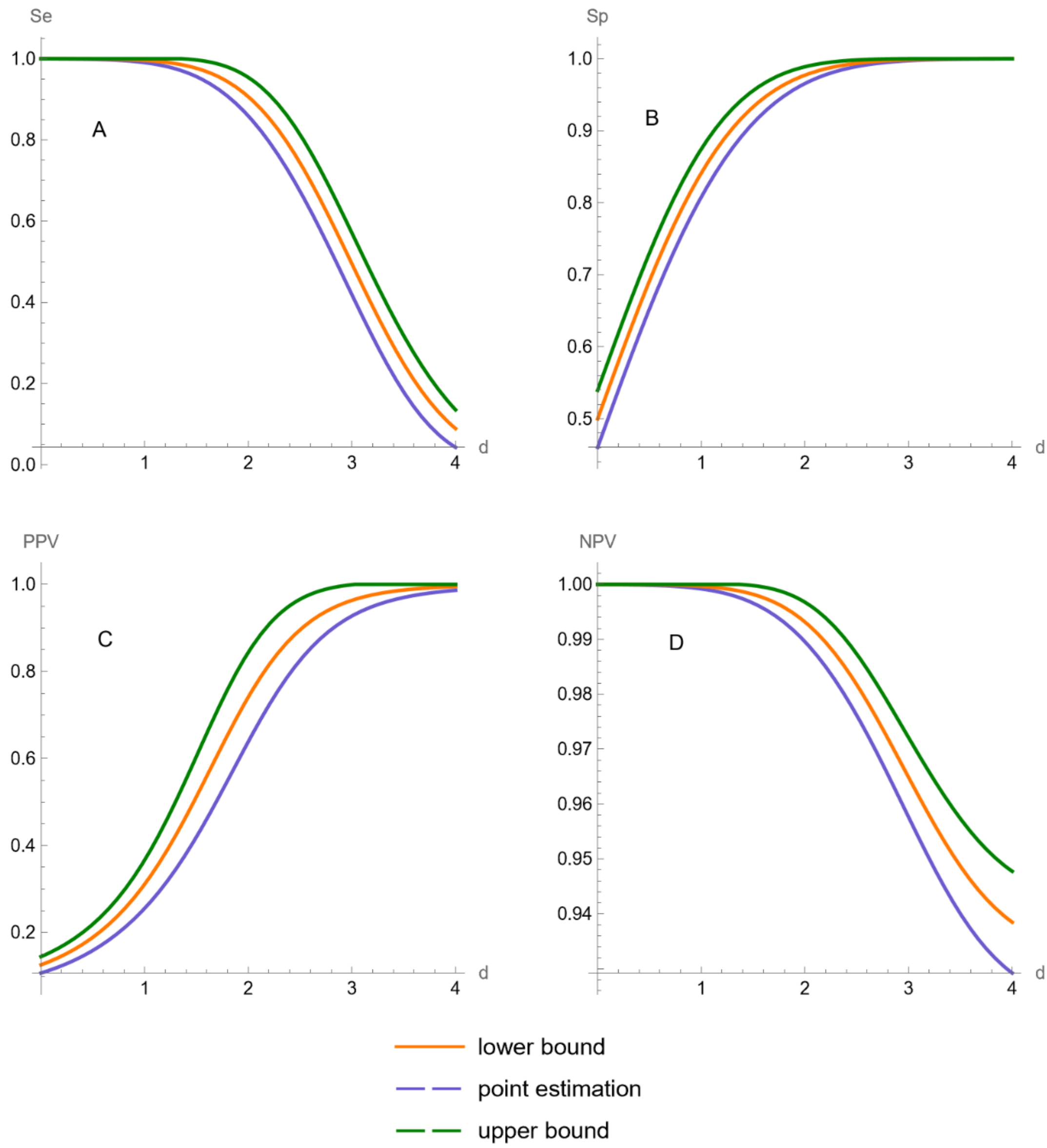


Figure 17. DAM confidence intervals versus diagnostic threshold plots. Plots of confidence intervals of (A) sensitivity (*Se*), (B) specificity (*Sp*), (C) positive predictive value (*PPV*), and (D) negative predictive value (*NPV*) versus diagnostic threshold (*d*) curves with the respective parameters in Table 4**.**

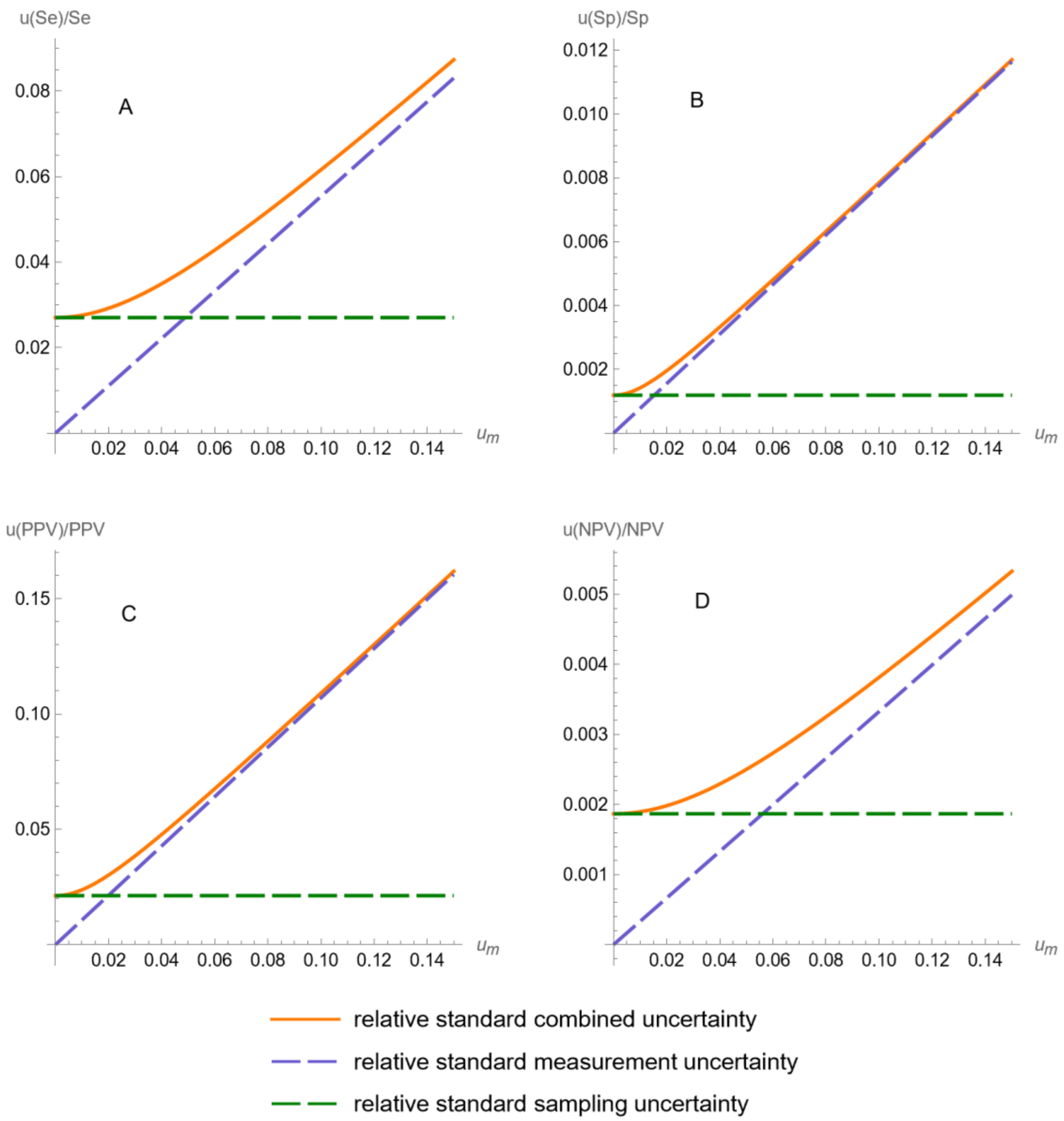


Figure 18. DAM standard uncertainties versus measurement uncertainty plots. Plots of relative standard combined, sampling, and measurement uncertainties of (A) sensitivity (*Se*), (B) specificity (*Sp*), (C) positive predictive value (*PPV*), and (D) negative predictive value (*NPV*) versus standard measurement uncertainty (u_m) curves, with the respective parameters in Table 4**.**

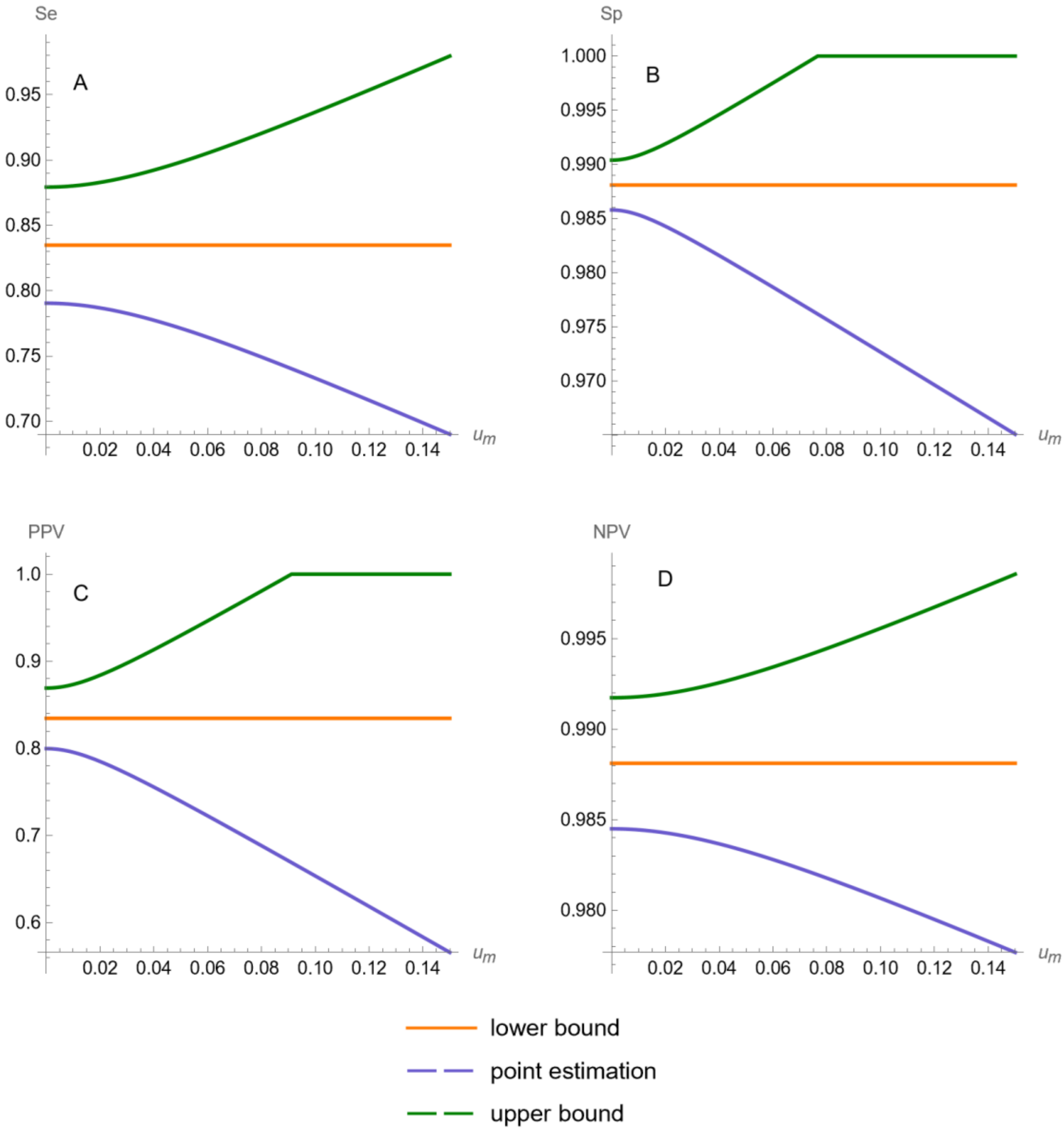


Figure 19. DAM confidence intervals versus measurement uncertainty plots. Plots of confidence intervals of (A) sensitivity (*Se*), (B) specificity (*Sp*), (C) positive predictive value (*PPV*), and (D) negative predictive value (*NPV*) versus standard measurement uncertainty (u_m) curves, with the respective parameters in Table 4**.**

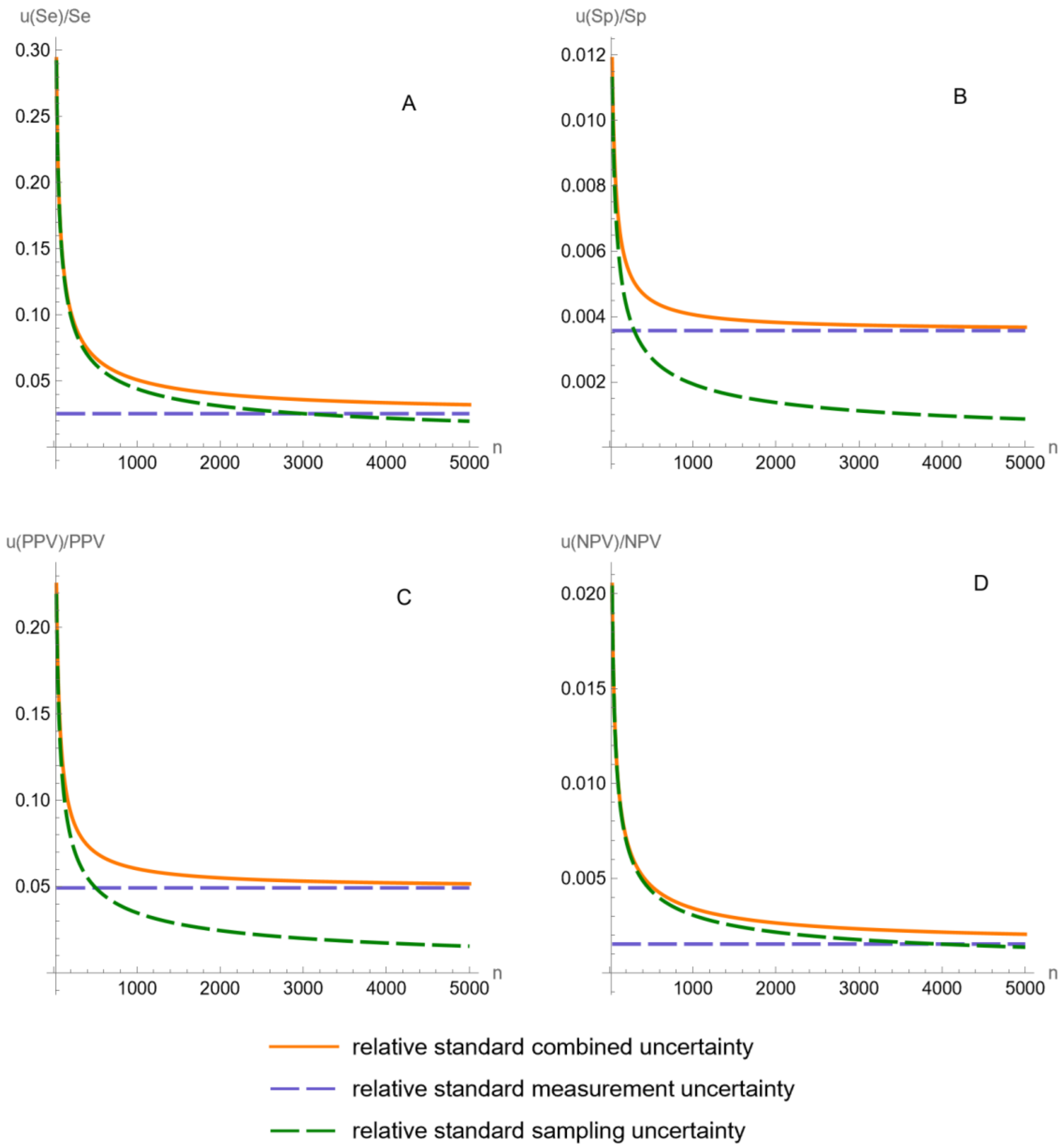


Figure 20. DAM standard uncertainties versus population sample size plots. Plots of relative standard combined, sampling, and measurement uncertainties of (A) sensitivity (*Se*), (B) specificity (*Sp*), (C) positive predictive value (*PPV*), and (D) negative predictive value (*NPV*) versus total population sample size (*n*) curves, with the respective parameters in Table 4**.**

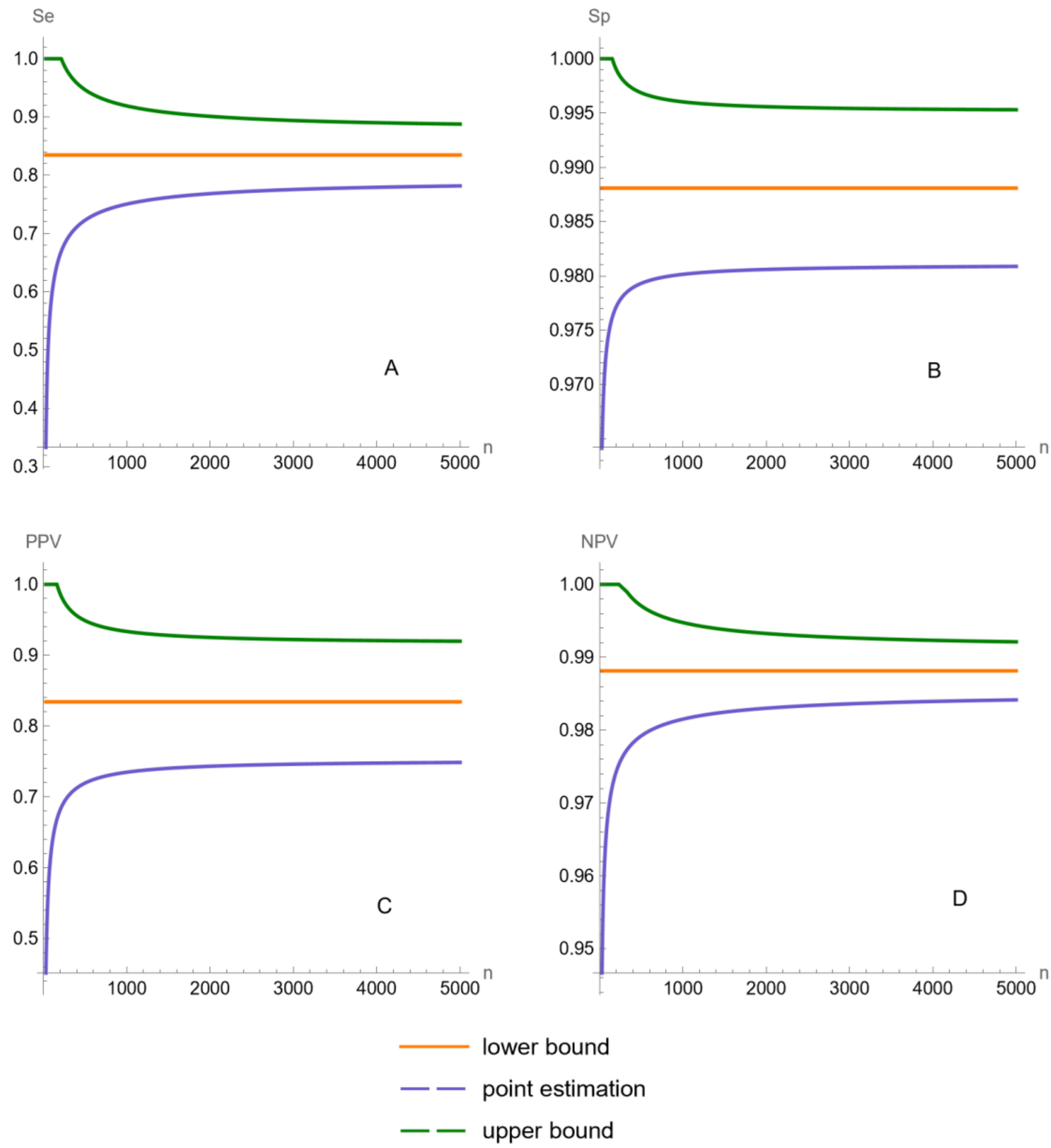


Figure 21. DAM confidence intervals versus population sample size plots. Plots of confidence intervals of (A) sensitivity (*Se*), (B) specificity (*Sp*), (C) positive predictive value (*PPV*), and (D) negative predictive value (*NPV*) versus total population sample size (*n*) curves, with the respective parameters in Table 4**.**

The combined uncertainty and the resultant confidence interval increase with measurement uncertainty (Fig. 6-8, 18, and 19) and decrease with the total population sample size (Fig. 9-11, 20, and 21).

In the illustrative case study, combined uncertainty has (see Fig. 13 and Fig. 22):

1. Little effect ($(u_c(x)/x) < 0.5\%$) on specificity, overall diagnostic accuracy, and negative predictive value,
2. Intermediate effect ($3.5\% < (u_c(x)/x) < 5.5\%$) on sensitivity, positive predictive value, Youden's index, and concordance probability,
3. Greater effect ($18\% < u_c(x)/x < 39\%$) on diagnostic odds ratio, likelihood ratio for a positive or negative result, and Euclidean distance, following previous findings (21-23).

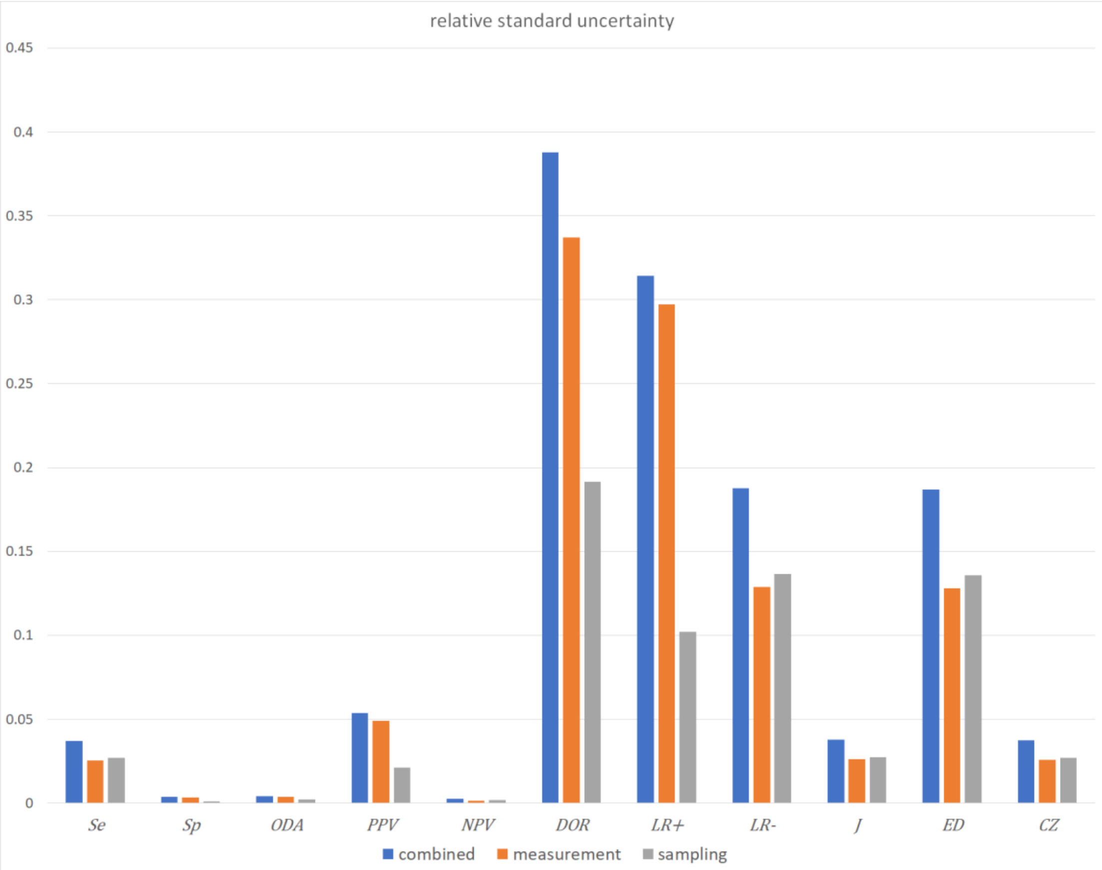


Figure 22. DAM relative standard uncertainties of diagnostic accuracy measures. Histogram of standard combined, sampling, and measurement uncertainties of diagnostic accuracy measures, with the respective parameters in Table 4**.**

In addition, measurement uncertainty is the main component of the combined uncertainty of specificity, overall diagnostic accuracy, positive predictive value, diagnostic odds ratio, and likelihood ratio for a positive result.

4. Discussion

There is a persistent need to estimate the uncertainty of diagnostic accuracy measures, especially regarding screening and diagnostic tests for life-threatening diseases. The current pandemic of novel coronavirus disease 2019 (COVID-19) has exposed this convincingly (24-29).

This program has been developed to explore the combined measurement and sampling uncertainty of diagnostic accuracy measures because:

1. Diagnostic accuracy is fundamental to clinical decision-making (30),
2. Defining the permissible measurement uncertainty is critical to quality and risk management in laboratory medicine (31)
3. Sampling uncertainty is decisive for clinical study design to evaluate a screening or diagnostic test (32).

There has been extensive research on either diagnostic accuracy or uncertainty. However, minimal research has been done on both subjects (23, 33-35).

This program explores the combined uncertainty of diagnostic accuracy measures of a screening or diagnostic test (Fig. 3, 4, 6, 7, 9, 10, and 11-13) and the resultant confidence limits (Fig. 5, 8, 11 and 14). Combined uncertainty and the resultant confidence limits depend on the diagnostic threshold (Fig. 3-5, and 15-17), on measurement uncertainty (Fig. 6-8, 18 and 19), and population parameters, including total population sample size (Fig. 9-11, 20 and 21).

In antithesis to the complexity of the calculations, the program simplifies its exploration with a user-friendly interface. Furthermore, it provides calculators for calculating uncertainty components on the diagnostic accuracy measures and the resultant confidence intervals (Fig. 12-14).

As demonstrated by the illustrative case study described above (Section 3.2), the uncertainty has little effect on specificity, overall diagnostic accuracy, and negative predictive value. However, it affects sensitivity, positive predictive value, Youden's index, and concordance probability. At the same time, it has a considerable impact on the diagnostic odds ratio, likelihood ratio for a positive or negative result, and Euclidean distance (Fig. 22). However, further research is needed to explore the uncertainty of diagnostic accuracy measures with different clinically and laboratory relevant parameter settings.

Shortcomings of this program that could be improved by further research are the following:

1. *Assumptions Underlying the Calculations*
    - 1.1. *Existence of a "Gold Standard" Diagnostic Method*: Calculations assume the presence of a "gold standard" diagnostic method. If such a standard is unavailable, alternative methods exist for estimating diagnostic accuracy measures (36).
    - 1.2. *Normality of Measurements or Their Transformations*: The calculations assume the normality of either the raw measurements or suitable transformations of these measurements), a condition typically met in practice. Related studies discuss the distribution of diagnostic test measurements in the contexts of reference intervals, diagnostic thresholds, and clinical decision limits (39-43).
    - 1.3. *Bimodality of Measurands*: The underlying measurands are generally assumed to be bimodal; however, unimodal distributions may also be considered under certain conditions (44, 45).
    - 1.4. *Homoscedasticity of Measurement Uncertainty Within the Diagnostic Threshold Range*: The calculations assume homoscedasticity of measurement uncertainty around diagnostic thresholds. If measurement uncertainty is heteroscedastic and distorts the distribution, suitable transformations may be applied to achieve homoscedasticity (46).
2. *First-Order Taylor Series Expansion for Uncertainty Propagation*
   The uncertainty propagation calculations are based on a first-order Taylor series expansion approximation (14).
3. Prevalence Rate Uncertainty Approximation Using the Agresti-Coull Adjusted Wald Interval: The uncertainty in prevalence rates is approximated using the Agresti-Coull adjusted Wald interval (13), though more precise methods are available (47).

However, employing these more accurate methods increases computational complexity significantly.

The program complements the previously published tool for exploring the relation between diagnostic accuracy and measurement uncertainty (23).

All major general or medical statistical software packages (Matlab®, NCSS®, R, SAS®, SPSS®, Stata® and MedCalc®) include routines for the calculation and plotting of various diagnostic accuracy measures and their confidence intervals. The program presented in this work provides 99 types of plots and three types of comprehensive tables of the uncertainty of diagnostic accuracy measures and the resultant confidence intervals (Fig. 2), many of which are novel. To the best of our knowledge, no one of the programs mentioned above or any other software provides this range of plots and tables without advanced statistical programming.

5. Conclusion

This program calculates the uncertainty of diagnostic accuracy measures, their components, and the resultant confidence intervals. It can also be a flexible, user-friendly, interactive educational or research tool in medical decision-making.

6. Availability and requirements

Project name: *Diagnostic Uncertainty*

Version: 2.0.0.1

Project home page: https://www.hcsl.com/Tools/Uncertainty/

Operating systems: Microsoft Windows, Linux, Apple iOS

Programming language: Wolfram Language

Other software requirements: Wolfram Player®, freely available at: https://www.wolfram.com/player/ or Wolfram Mathematica®

System requirements: Intel Core i9 or equivalent CPU and 32GB+ of RAM

License: Attribution - NonCommercial - ShareAlike 4.0 International Creative Commons License

7. Supplemental Material

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

a) Supplemental File I:
Uncertainty.nb: The program *Diagnostic Uncertainty* as a Wolfram Notebook.

b) Supplemental File II:
DiagnosticUncertaintyInterface.pdf: A brief interface documentation of the program.

c) Supplemental File III:
DiagnosticUncertaintyCalculations.nb: The calculations for estimating diagnostic accuracy measures and their standard uncertainty in a Wolfram Notebook

8. References

1. Šimundić A-M. Measures of Diagnostic Accuracy: Basic Definitions. EJIFCC. 2009;19(4):203-11.
2. Shiu S-Y, Gatsonis C. The predictive receiver operating characteristic curve for the joint assessment of the positive and negative predictive values. Philos Trans A Math Phys Eng Sci. 2008;366(1874):2313-33.
3. Ayyub BM, Klir GJ. Uncertainty Modeling and Analysis in Engineering and the Sciences: Chapman and Hall/CRC; 2006 2006/1/1.
4. Metrology JCfGi. Evaluation of measurement data — Guide to the expression of uncertainty in measurement. Joint Committee for Guides in Metrology; 2008.
5. Oosterhuis WP, Theodorsson E. Total error vs. measurement uncertainty: revolution or evolution? Clin Chem Lab Med. 2016;54(2):235-9.
6. Rostron MHRSLREP. Measurement uncertainty arising from sampling - A guide to methods and approaches. Guideline. EURACHEM/CITAC; 2019 2019.
7. Bloch DA. Comparing Two Diagnostic Tests against the Same "Gold Standard" in the Same Sample. Biometrics. 1997;53(1):73-85.
8. Gillard J. A generalised Box–Cox transformation for the parametric estimation of clinical reference intervals. J Appl Stat. 2012;39(10):2231-45.
9. Atkinson AB. The box-cox transformation: review and extensions. Stat Sci. 2020.
10. Hund E, Massart DL, Smeyers-Verbeke J. Operational definitions of uncertainty. Trends Analyt Chem. 2001;20(8):394-406.
11. Kallner A, Boyd JC, Duewer DL, Giroud C, Hatjimihail AT, Klee GG, et al. Expression of Measurement Uncertainty in Laboratory Medicine; Approved Guideline: Clinical and Laboratory Standards Institute; 2012.
12. White GH. Basics of estimating measurement uncertainty. Clin Biochem Rev. 2008;29 Suppl 1:S53-60.
13. Agresti A, Coull BA. Approximate is Better than "Exact" for Interval Estimation of Binomial Proportions. Am Stat. 1998;52(2):119-26.

14. Wilson BM, Smith BL. Taylor-series and Monte-Carlo-method uncertainty estimation of the width of a probability distribution based on varying bias and random error. Meas Sci Technol. 2013;24(3):035301.
15. Welch BL. The Generalization of `Student's' Problem when Several Different Population Variances are Involved. Biometrika. 1947;34(1/2):28-35.
16. Satterthwaite FE. An approximate distribution of estimates of variance components. Biometrics. 1946;2(6):110-4.
17. Wolfram S. An Elementary Introduction to the Wolfram Language. Second ed2017. 340 p.
18. Wolfram Research I. Mathematica, Version 12.1. Champaign, IL 2020.
19. Lim TO, Bakri R, Morad Z, Hamid MA. Bimodality in blood glucose distribution: is it universal? Diabetes Care. 2002;25(12):2212-7.
20. American Diabetes A. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2019. Diabetes Care. 2019;42(Suppl 1):S13-S28.
21. Kupchak P, Wu AHB, Ghani F, Newby LK, Ohman EM, Christenson RH. Influence of imprecision on ROC curve analysis for cardiac markers. Clin Chem. 2006;52(4):752-3.
22. Kroll MH, Biswas B, Budd JR, Durham P, Gorman RT, Gwise TE, et al. Assessment of the Diagnostic Accuracy of Laboratory Tests Using Receiver Operating Characteristic Curves; Approved Guideline—Second Edition: Clinical and Laboratory Standards Institute; 2011.
23. Chatzimichail T, Hatjimihail AT. A software tool for exploring the relation between diagnostic accuracy and measurement uncertainty. In Review. 2020.
24. Lippi G, Simundic A-M, Plebani M. Potential preanalytical and analytical vulnerabilities in the laboratory diagnosis of coronavirus disease 2019 (COVID-19). Clin Chem Lab Med. 2020.
25. Tang Y-W, Schmitz JE, Persing DH, Stratton CW. The Laboratory Diagnosis of COVID-19 Infection: Current Issues and Challenges. J Clin Microbiol. 2020.
26. Deeks JJ, Dinnes J, Takwoingi Y, Davenport C, Leeflang MMG, Spijker R, et al. Diagnosis of SARS-CoV-2 infection and COVID-19: accuracy of signs and symptoms; molecular, antigen, and antibody tests; and routine laboratory markers. Cochrane Database Syst Rev. 2020;26:1896.
27. Infantino M, Grossi V, Lari B, Bambi R, Perri A, Manneschi M, et al. Diagnostic accuracy of an automated chemiluminescent immunoassay for anti-SARS-CoV-2 IgM and IgG antibodies: an Italian experience. J Med Virol. 2020.
28. Mahase E. Covid-19: "Unacceptable" that antibody test claims cannot be scrutinised, say experts. BMJ. 2020;369:m2000.
29. Kontou PI, Braliou GG, Dimou NL, Nikolopoulos G, Bagos PG. Antibody Tests in Detecting SARS-CoV-2 Infection: A Meta-Analysis. Diagnostics (Basel). 2020;10(5).
30. Owen RK, Cooper NJ, Quinn TJ, Lees R, Sutton AJ. Network meta-analysis of diagnostic test accuracy studies identifies and ranks the optimal diagnostic tests and thresholds for health care policy and decision-making. J Clin Epidemiol. 2018;99:64-74.
31. Haeckel R, Wosniok W, Gurr E, Peil B, on behalf of Arbeitsgruppe R. Supplements to a recent proposal for permissible uncertainty of measurements in laboratory medicine. 2016.
32. Hajian-Tilaki K. Sample size estimation in diagnostic test studies of biomedical informatics. J Biomed Inform. 2014;48:193-204.
33. Smith AF, Shinkins B, Hall PS, Hulme CT, Messenger MP. 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. Clin Chem. 2019;65(11):1363-74.
34. Theodorsson E. Uncertainty in Measurement and Total Error: Tools for Coping with Diagnostic Uncertainty. Clin Lab Med. 2017;37(1):15-34.
35. Padoan A, Sciacovelli L, Aita A, Antonelli G, Plebani M. Measurement uncertainty in laboratory reports: A tool for improving the interpretation of test results. Clin Biochem. 2018;57:41-7.
36. Collins J, Albert PS. Estimating diagnostic accuracy without a gold standard: A continued controversy. J Biopharm Stat. 2016;26(6):1078-82.
37. Sakia RM. The Box-Cox Transformation Technique: A Review. Journal of the Royal Statistical Society Series D (The Statistician). 1992;41(2):169-78.
38. Box GEP, Cox DR. An Analysis of Transformations. J R Stat Soc Series B Stat Methodol. 1964;26(2):211-43.
39. Solberg HE. Approved recommendation (1987) on the theory of reference values. Part 5. Statistical treatment of collected reference values. Determination of reference limits. Clin Chim Acta. 1987;170(2):S13-S32.

40. Pavlov IY, Wilson AR, Delgado JC. Reference interval computation: which method (not) to choose? Clin Chim Acta. 2012;413(13-14):1107-14.
41. Sikaris K. Application of the stockholm hierarchy to defining the quality of reference intervals and clinical decision limits. Clin Biochem Rev. 2012;33(4):141-8.
42. Daly CH, Liu X, Grey VL, Hamid JS. A systematic review of statistical methods used in constructing pediatric reference intervals. Clin Biochem. 2013;46(13-14):1220-7.
43. Ozarda Y, Sikaris K, Streichert T, Macri J, intervals ICoR, Decision L. Distinguishing reference intervals and clinical decision limits - A review by the IFCC Committee on Reference Intervals and Decision Limits. Crit Rev Clin Lab Sci. 2018;55(6):420-31.
44. Wilson JMG, Jungner G. Principles and practice of screening for disease. Geneva: World Health Organization; 1968. 163 p.
45. 2.3 Clinical test evaluation. Unimodal and bimodal approaches. Scand J Clin Lab Invest. 1992;52(sup208):51-7.
46. Analytical Methods Committee AN. Why do we need the uncertainty factor? Anal Methods. 2019;11(15):2105-7.
47. Brown LD, Cai TT, DasGupta A. Interval Estimation for a Binomial Proportion. Stat Sci. 2001;16(2):101-17.

9. List of abbreviations

DAM: diagnostic accuracy measure

OGTT: oral glucose tolerance test

ADA: American Diabetes Association

COVID-19: novel corona virus disease 2019

10. Notation

Populations

$\overline{D}$: nondiseased population

D: diseased population

Test outcomes

$\overline{T}$: negative test result

T: positive test result

TN: true negative test result

TP: true positive test result

FN: false negative test result

FP: false positive test result

Diagnostic accuracy measures

Se: sensitivity

Sp : specificity

PPV: positive predictive value

NPV: negative predictive value

ODA : overall diagnostic accuracy

DOR : diagnostic odds ratio

$LR+$: likelihood ratio for a positive test result

$LR-$: likelihood ratio for a negative test result

J: Youden's index

ED : Euclidean distance of a *ROC* curve point from the point (0,1)

CZ : concordance probability

Parameters

μ_P: mean of the measurand of a test in the population *P*

σ_P: standard deviation of the measurand of a test in the population *P*

n_P: size of a sample of the population *P*

r : prevalence of the disease

d : diagnostic threshold of a test

p : confidence level

$u_s(x)$: standard sampling uncertainty of x

$u_m(x)$: standard measurement uncertainty of x

$u_c(x)$: standard combined uncertainty of x

$u_i(x)$: the i^{th} component of the standard combined uncertainty of x

Functions and relations

$se\,(d,\dots)$: sensitivity function of a single test versus its diagnostic threshold d

$sp\,(d,\dots)$: specificity function of a single test versus its diagnostic threshold d

$\Psi(x,\mu,\sigma)$: cumulative distribution function of a normal distribution with mean μ and standard deviation σ, evaluated at x

$erf(x)$: error function, evaluated at x

$erfc(x)$: complementary error function, evaluated at x

$Pr\,(a)$: probability of an event a

$Pr(a\,|b)$: probability of an event a given the event b

$CI_p(x)$: confidence interval of x at confidence level p

$F^{-1}(\dots)$: the inverse function F

11. Appendix

About the program controls

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

Dragging with the mouse rotates the three-dimensional plots while dragging with the mouse while pressing the *ctrl*, *alt*, or *opt* keys zooms in or out.

Range of parameters

Input parameters

p: 0.90 – 0.999

d: $\max\{\mu_D - 4, \mu_{\overline{D}} - 4\sigma_{\overline{D}}\} - \min\{\mu_D + 4\sigma_D, \mu_{\overline{D}} + 4\sigma_{\overline{D}}\}$

$\mu_{\overline{D}}$: 0 – 6

$\sigma_{\overline{D}}$: 1 – 6

$n_{\overline{D}}$: 2 – 10000

μ_D: 0.1 – 6

σ_D: 0.1 – 6

n_D : 2 – 10000

r: 0.001 – 0.999

u_m : 0 – 6

n_u : 20 – 1000

Range of the coordinates of the plots

There are two options for the range of coordinates to be included in each plot:

1) Full: All the calculated coordinate points are plotted.
2) Partial: The distribution of coordinate values is found, and any points sufficiently far out are not considered.

Abscissas

Full range

d: $(\max\{\mu_D - 4\sigma_{\overline{D}}, \mu_{\overline{D}} - 4\sigma_{\overline{D}}\} - \min\{\mu_D + 4\sigma_D, \mu_{\overline{D}} + 4\sigma_{\overline{D}}\})$

u_m : $(0 - 0.5\,\sigma_{\overline{D}})$

n : $(\max\{[\![\frac{2}{r}]\!], [\![\frac{2}{1-r}]\!], 20\} - \max\{10\,[\![\frac{2}{r}]\!], 10\,[\![\frac{2}{1-r}]\!], 10000\})$

Partial range

d: $(\max\{\mu_D - 2.5\sigma_{\overline{D}}, \mu_{\overline{D}} - 2.5\sigma_{\overline{D}}\} - \min\{\mu_D + 2.5\sigma_D, \mu_{\overline{D}} + 2.5\sigma_{\overline{D}}\})$

u_m : $(0 - 0.15\,\sigma_{\overline{D}})$

n : $(\max\{[\![\frac{2}{r}]\!], [\![\frac{2}{1-r}]\!], 20\} - \max\{5\,[\![\frac{2}{r}]\!], 5\,[\![\frac{2}{1-r}]\!], 5000\})$

Input and output

The program provides in six modules and nine submodules plots and tables of the uncertainty and the resultant confidence intervals of diagnostic accuracy measures of a screening or diagnostic test for a measurand, applied at a single point in time in samples of diseased and a nondiseased population.

Singularity points are excluded from the plots.

Indeterminate results of the calculation modules represent numerical quantities whose magnitudes cannot be determined because they are either too small or too large.

Plots

Plots vs diagnostic threshold module

Diagnostic accuracy measures uncertainty plots submodule

Input

The user defines:

1) The diseased and nondiseased population samples parameters:
    a) The measurand means,
    b) The measurand standard deviations,
    c) The sizes of the samples,
2) The diagnostic accuracy measure:
    a) Sensitivity (Se),
    b) Specificity (Sp),
    c) Overall diagnostic accuracy (ODA),
    d) Positive predictive value (PPV),
    e) Negative predictive value (NPV),
    f) Diagnostic odds ratio (DOR),
    g) Likelihood ratio for a positive test result (LR+),
    h) Likelihood ratio for a negative test result (LR-),
    i) Youden's index (J),
    j) Euclidean distance (ED),
    k) Concordance probability (CZ),
3) The standard measurement uncertainty of the test.

Output

Plots of the values of the:

1) Standard combined uncertainty,
2) Standard measurement uncertainty,
3) Standard sampling uncertainty

of the measure versus diagnostic threshold (d).

Diagnostic accuracy measures relative uncertainty plots submodule

Input

The user defines:

1) The diseased and nondiseased population samples parameters:
    a) The measurand means,
    b) The measurand standard deviations,
    c) The sizes of the samples,
2) The diagnostic accuracy measure:
    a) Sensitivity (Se),
    b) Specificity (Sp),
    c) Overall diagnostic accuracy (ODA),
    d) Positive predictive value (PPV),

e) Negative predictive value (NPV),
f) Diagnostic odds ratio (DOR),
g) Likelihood ratio for a positive test result ($LR+$),
h) Likelihood ratio for a negative test result ($LR-$),
i) Youden's index (J),
j) Euclidean distance (ED),
k) Concordance probability (CZ),

3) The standard measurement uncertainty of the test.

Output

Plots of the values of the:

1) Relative standard combined uncertainty,
2) Relative standard measurement uncertainty,
3) Relative standard sampling uncertainty

of the measure versus diagnostic threshold (d).

Diagnostic accuracy measures confidence intervals plots submodule

Input

The user defines:

1) The confidence level (p),
2) The diseased and nondiseased population samples parameters:
   a) The measurand means,
   b) The measurand standard deviations,
   c) The sizes of the samples,
3) The diagnostic accuracy measure:
   a) Sensitivity (Se),
   b) Specificity (Sp),
   c) Overall diagnostic accuracy (ODA),
   d) Positive predictive value (PPV),
   e) Negative predictive value (NPV),
   f) Diagnostic odds ratio (DOR),
   g) Likelihood ratio for a positive test result ($LR+$),
   h) Likelihood ratio for a negative test result ($LR-$),
   i) Youden's index (J),
   j) Euclidean distance (ED),
   k) Concordance probability (CZ),
4) The standard measurement uncertainty of the test,
5) The size of the sample of the measurements for the estimation of the measurement uncertainty.

Output

Plots of the values of the:

1) Lower bound,
2) Point estimation,
3) Upper bound

of the measure versus diagnostic threshold (d) at the selected confidence level.

Plots vs. measurement uncertainty module

Diagnostic accuracy measures uncertainty plots submodule

Input

The user defines:

1) The diseased and nondiseased population samples parameters:
2) The diseased and nondiseased population samples parameters:
    a) The measurand means,
    b) The measurand standard deviations,
    c) The sizes of the samples,
3) The diagnostic accuracy measure:
    a) Sensitivity (Se),
    b) Specificity (Sp),
    c) Overall diagnostic accuracy (ODA),
    d) Positive predictive value (PPV),
    e) Negative predictive value (NPV),
    f) Diagnostic odds ratio (DOR),
    g) Likelihood ratio for a positive test result (LR+),
    h) Likelihood ratio for a negative test result (LR-),
    i) Youden's index (J),
    j) Euclidean distance (ED),
    k) Concordance probability (CZ).

Output
Plots of the values of the:

1) Standard combined uncertainty,
2) Standard measurement uncertainty,
3) Standard sampling uncertainty

of the measure versus standard measurement uncertainty (u_m).

Diagnostic accuracy measures relative uncertainty plots submodule

Input
The user defines:

1) The diseased and nondiseased population samples parameters:
    a) The prevalence of the disease,
    b) The measurand means,
    c) The measurand standard deviations,
2) The diagnostic accuracy measure:
    a) Sensitivity (Se),
    b) Specificity (Sp),
    c) Overall diagnostic accuracy (ODA),
    d) Positive predictive value (PPV),
    e) Negative predictive value (NPV),
    f) Diagnostic odds ratio (DOR),
    g) Likelihood ratio for a positive test result (LR+),
    h) Likelihood ratio for a negative test result (LR-),
    i) Youden's index (J),
    j) Euclidean distance (ED),
    k) Concordance probability (CZ),
3) The standard measurement uncertainty of the test.

Output
Plots of the values of the:

1) Relative standard combined uncertainty,
2) Relative standard measurement uncertainty,
3) Relative standard sampling uncertainty

of the measure versus standard measurement uncertainty (u_m).

Diagnostic accuracy measures confidence intervals plots submodule

Input
The user defines:

1) The confidence level (p),
2) The diseased and nondiseased population samples parameters:
   a) The measurand means,
   b) The measurand standard deviations,
   c) The sizes of the samples,
3) The diagnostic accuracy measure:
   a) Sensitivity (Se),
   b) Specificity (Sp),
   c) Overall diagnostic accuracy (ODA),
   d) Positive predictive value (PPV),
   e) Negative predictive value (NPV),
   f) Diagnostic odds ratio (DOR),
   g) Likelihood ratio for a positive test result (LR+),
   h) Likelihood ratio for a negative test result (LR-),
   i) Youden's index (J),
   j) Euclidean distance (ED),
   k) Concordance probability (CZ),
4) The size of the sample of the measurements for the estimation of the measurement uncertainty.

Output
Plots of the values of the:

4) Lower bound,
5) Point estimation,
6) Upper bound

of the measure versus standard measurement uncertainty (u_m) at the selected confidence level.

Plots vs population size module

Diagnostic accuracy measures uncertainty plots submodule

Input
The user defines:

4) The diseased and nondiseased population samples parameters:
   a) The prevalence of the disease,
   b) The measurand means,
   c) The measurand standard deviations,
5) The diagnostic accuracy measure:
   a) Sensitivity (Se),
   b) Specificity (Sp),
   c) Overall diagnostic accuracy (ODA),
   d) Positive predictive value (PPV),
   e) Negative predictive value (NPV),
   f) Diagnostic odds ratio (DOR),
   g) Likelihood ratio for a positive test result (LR+),
   h) Likelihood ratio for a negative test result (LR-),
   i) Youden's index (J),
   j) Euclidean distance (ED),
   k) Concordance probability (CZ),
6) The standard measurement uncertainty of the test.

Output
Plots of the values of the:

4) Standard combined uncertainty,
5) Standard measurement uncertainty,
6) Standard sampling uncertainty

of the measure versus total population size (n).

Diagnostic accuracy measures relative uncertainty plots submodule

Input
The user defines:

4) The diseased and nondiseased population samples parameters:
    a) The prevalence of the disease,
    b) The measurand means,
    c) The measurand standard deviations,
5) The diagnostic accuracy measure:
    a) Sensitivity (Se),
    b) Specificity (Sp),
    c) Overall diagnostic accuracy (ODA),
    d) Positive predictive value (PPV),
    e) Negative predictive value (NPV),
    f) Diagnostic odds ratio (DOR),
    g) Likelihood ratio for a positive test result (LR+),
    h) Likelihood ratio for a negative test result (LR-),
    i) Youden's index (J),
    j) Euclidean distance (ED),
    k) Concordance probability (CZ),
6) The standard measurement uncertainty of the test.

Output
Plots of the values of the:

4) Relative standard combined uncertainty,
5) Relative standard measurement uncertainty,
6) Relative standard sampling uncertainty

of the measure versus total population size (n).

Diagnostic accuracy measures confidence intervals plots submodule

Input
The user defines:

5) The confidence level (p),
6) The diseased and nondiseased population samples parameters:
    a) The prevalence of the disease,
    b) The measurand means,
    c) The measurand standard deviations,
7) The diagnostic accuracy measure:
    a) Sensitivity (Se),
    b) Specificity (Sp),
    c) Overall diagnostic accuracy (ODA),
    d) Positive predictive value (PPV),
    e) Negative predictive value (NPV),
    f) Diagnostic odds ratio (DOR),
    g) Likelihood ratio for a positive test result (LR+),
    h) Likelihood ratio for a negative test result (LR-),

i) Youden's index (J),
j) Euclidean distance (ED),
k) Concordance probability (CZ),

8) The standard measurement uncertainty of the test,
9) The size of the sample of the measurements for the estimation of the measurement uncertainty.

Output

Plots of the values of the:

7) Lower bound,
8) Point estimation,
9) Upper bound

of the measure versus total population size (n) at the selected confidence level.

Calculators

Diagnostic accuracy measures uncertainty calculator

Input

1) The number of decimal digits of the results
2) The diagnostic threshold (d)
3) The diseased and nondiseased population samples parameters:
   a) The measurand means,
   b) The measurand standard deviations,
   c) The sizes of the samples,
4) The standard measurement uncertainty of the test.

Output

A table of the values of:

1) The standard combined uncertainty,
2) The standard measurement uncertainty,
3) The standard sampling uncertainty,

of the following diagnostic accuracy measures at the selected diagnostic threshold:

1) Sensitivity (Se),
2) Specificity (Sp),
3) Overall diagnostic accuracy (ODA),
4) Positive predictive value (PPV),
5) Negative predictive value (NPV),
6) Diagnostic odds ratio (DOR),
7) Likelihood ratio for a positive test result (LR+),
8) Likelihood ratio for a negative test result (LR-),
9) Youden's index (J),
10) Euclidean distance (ED),
11) Concordance probability (CZ).

Diagnostic accuracy measures relative uncertainty calculator

Input

1) The number of decimal digits of the results
2) The diagnostic threshold (d),
3) The diseased and nondiseased population samples parameters:
   a) The measurand means,
   b) The measurand standard deviations,
   c) The sizes of the samples,

4) The standard measurement uncertainty of the test.

Output

A table of the values of:

1) The relative standard combined uncertainty,
2) The relative standard measurement uncertainty and
3) The relative standard sampling uncertainty,

of the following diagnostic accuracy measures at the selected diagnostic threshold:

1) Sensitivity (Se),
2) Specificity (Sp),
3) Overall diagnostic accuracy (ODA),
4) Positive predictive value (PPV),
5) Negative predictive value (NPV),
6) Diagnostic odds ratio (DOR),
7) Likelihood ratio for a positive test result (LR+),
8) Likelihood ratio for a negative test result (LR-),
9) Youden's index (J),
10) Euclidean distance (ED),
11) Concordance probability (CZ).

Diagnostic accuracy measures confidence intervals calculator

Input

5) The number of decimal digits of the results
1) The confidence level (p),
2) The diagnostic threshold (d),
3) The diseased and nondiseased population samples parameters:
   a) The measurand means,
   b) The measurand standard deviations.
   c) The sizes of the samples,
4) The standard measurement uncertainty of the test,
5) The size of the sample of the measurements for the estimation of the measurement uncertainty.

Output

A table of the values and the confidence intervals of the following diagnostic accuracy measures at the selected confidence level and diagnostic threshold:

1) Sensitivity (Se),
2) Specificity (Sp),
3) Overall diagnostic accuracy (ODA),
4) Positive predictive value (PPV),
5) Negative predictive value (NPV),
6) Diagnostic odds ratio (DOR),
7) Likelihood ratio for a positive test result (LR+),
8) Likelihood ratio for a negative test result (LR-),
9) Youden's index (J),
10) Euclidean distance (ED),
11) Concordance probability (CZ).

12. Permanent Citation

Chatzimichail T, Hatjimihail AT. *A Software Tool for Calculating the Uncertainty of Diagnostic Accuracy Measures.* Technical Report XXVIII. Hellenic Complex Systems Laboratory; 2024. https://www.hcsl.com/TR/hcsltr28/hcsltr28.pdf

13. License

First Published in *Diagnostics* on February 21, 2021:

Chatzimichail T, Hatjimihail AT. A Software Tool for Calculating the Uncertainty of Diagnostic Accuracy Measures. *Diagnostics* 2021;11(3):406. Available at: https://www.mdpi.com/2075-4418/11/3/406

Revised: November 8, 2024