# A software tool for calculating the uncertainty of diagnostic accuracy measures

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

[a] Hellenic Complex Systems Laboratory, Kostis Palamas 21, Drama 66131, Greece

Theodora Chatzimichail, MD

Address: Hellenic Complex Systems Laboratory, Kostis Palamas 21, Drama 66131, Greece

Phone: +30 2521058227

Mobile: +44 7745991039

Email: tc@hcsl.com

Corresponding author:

Aristides T Hatjimihail, MD, PhD

Address: Hellenic Complex Systems Laboratory, Kostis Palamas 21, Drama 66131, Greece

Phone: +30 2521058227

Mobile: +30 6937017876

Email: ath@hcsl.com

# *ABSTRACT*

## BACKGROUND:

Screening and diagnostic tests are used to classify people with and without a disease. 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. The objective for this work is to develop a tool for calculating the uncertainty of diagnostic accuracy measures, as diagnostic accuracy is fundamental to clinical decision-making.

## RESULTS:

For this reason, a freely available interactive program has been developed in Wolfram Language. The program provides six modules with nine submodules, for calculating and plotting the standard and expanded 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 in non-diseased and diseased populations. This is done for differing population sample sizes, mean and standard deviation of the measurand, diagnostic threshold and standard measurement uncertainty of the test.

The application of the program is illustrated with a case study of glucose measurements in diabetic and non-diabetic populations, that demonstrates the calculation of the uncertainty of diagnostic accuracy measures.

## CONCLUSION:

The presented interactive program 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, which are categorized as quantitative and qualitative. The 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 non-diseased populations. To classify the patients with and without a disease using a test based on a measurement, a diagnostic threshold or cutoff point is defined. If the measurement is above the threshold the patient is classified as test-positive, otherwise as test-negative (or vice versa) (Fig. 1). The possible test results are summarized in Table 1.

**Table 1.** A 2x2 contingency table

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

## 1.1 DIAGNOSTIC ACCURACY MEASURES

From the large number of diagnostic accuracy measures (DAM) appearing in literature, only a few are used for evaluating the diagnostic accuracy in clinical research and practice (1). These include:

1. Sensitivity ($Se$), specificity ($Sp$), overall diagnostic accuracy ($ODA$), diagnostic odds ratio ($DOR$), likelihood ratios for positive or negative result ($LR+$ and $LR-$ respectively), that are defined conditionally on the true disease status (2) and are prevalence invariant.
2. Overall diagnostic accuracy ($ODA$), that is defined conditionally on the true disease status and is prevalence dependent.
3. Positive predictive and negative predictive values ($PPV$ and $NPV$), that 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 2:** 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* | $\frac{TN+TP}{TN+FN+TP+FP}$ | $Pr(D)\ Pr(T\|D) + Pr(\overline{D})\ \mathrm{Pr}(\overline{T}\|\overline{D})$ |
| *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* | $\frac{TN\ TP}{(TN+FP)(FN+TP)}$ | $Pr(T\|D)\ Pr(\overline{T}\|\overline{D})$ |

## 1.2 UNCERTAINTY OF DIAGNOSTIC ACCURACY MEASURES

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

### 1.2.1 Measurement uncertainty

As there is inherent variability in any measurement process, there is measurement uncertainty, which 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). Measurement uncertainty is gradually replacing the total analytical error concept (5). Although the estimation of measurement uncertainty is essential for quality assurance in laboratory medicine (6), its effect on clinical decision making and consequently on clinical outcomes is rarely quantified (7). As direct outcome studies are very complex, a feasible first step is exploring the effect of measurement uncertainty on misclassification (8) and subsequently on diagnostic accuracy measures.

### 1.2.2 Sampling uncertainty

Diagnostic accuracy measures are estimated by applying a diagnostic method upon samples of populations. Therefore, sampling variability contributes to their uncertainty (9).

# 2 IMPLEMENTATION

## 2.1 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 non-diseased (10).
2. Either the values of the measurand or their transforms (11, 12) are normally distributed in each of the diseased and non-diseased populations.
3. Measurement uncertainty is normally distributed and homoscedastic in the diagnostic threshold's range.

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

### 2.1.1 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, therefore of the error function and the complementary error function.

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

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

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

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

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

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

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

### 2.1.2 Uncertainty

The uncertainty of an input parameter or a diagnostic accuracy measure $x$, can be expressed in the forms of standard and expanded uncertainty. The former, denoted as $u(x)$, equals the standard deviation of $x$. The later, denoted as $U(x)$, is defined as an interval around $x$ including $x$ with probability $p$ (13).

### 2.1.3 Measurement uncertainty of means and standard deviations

The standard measurement uncertainty is estimated as described in "Guide to the expression of uncertainty in measurement"(GUM) (4) and "Expression of measurement uncertainty in laboratory medicine" (6). Bias may be considered as a component of the standard measurement uncertainty (14).

### 2.1.4 Sampling uncertainty of means and standard deviations

If $m_P$ and $s_P$ 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:

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

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

### 2.1.5 Combined uncertainty of means and standard deviations

If $u_m$ the standard measurement uncertainty of a screening or diagnostic test measuring a measurand, and $m_P$ and $s_P$ the mean and standard deviation of the measurand in a population sample of size $n_P$, then the standard combined uncertainties of the mean $m_p$ and standard deviation $s_p$ are:

$$u_c(m_P) = \sqrt{\frac{s_p^2}{n_p} + u_m^2} \quad 7$$

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

#### 2.1.5.1 *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:

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

according to the Agresti-Coull adjustment of Waldo interval (15).

### 2.1.6 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, according to GUM, with a first order Taylor series approximation to uncertainty propagation (16).

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

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

### 2.1.7 Expanded uncertainty of diagnostic accuracy measures

The effective degrees of freedom $v_{eff}$ of the combined uncertainty $u_c$ are calculated using the Welch–Satterthwaite formula (17, 18):

$$v_{eff} = \frac{u_c^4}{\sum_{i=1}^{l} \frac{u_i^4}{v_i}} \quad 11$$

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

$$v_{min} \leq v_{eff} \leq \sum_{i=1}^{l} v_i \quad 12$$

If $F_v(z)$ 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:

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

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

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

## THE PROGRAM

To calculate the uncertainty of the diagnostic accuracy measures, an interactive program written in Wolfram Language (19) was developed using Wolfram Mathematica® (20). 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 non-diseased 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 Uncertainty.nb). It can be run on Wolfram Player® or Wolfram Mathematica® (see Section 6).

# 3 RESULTS

## 3.1 INTERFACE OF THE PROGRAM

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

### 3.1.1 Plots

#### 3.1.1.1 *Plots vs diagnostic threshold module*

##### 3.1.1.1.1 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 of the test (Fig. 3).

#### 3.1.1.1.2 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 of the test (Fig. 4).

#### 3.1.1.1.3 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).

### 3.1.1.2 ***Plots vs measurement uncertainty module***

#### 3.1.1.2.1 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).

#### 3.1.1.2.2 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).

#### 3.1.1.2.3 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).

#### 3.1.1.3 *Plots vs population sample size module*

##### 3.1.1.3.1 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).

##### 3.1.1.3.2 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).

##### 3.1.1.3.3 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 total population sample size (Fig. 11).

### 3.1.2 Calculators

#### 3.1.2.1 *Diagnostic accuracy measures standard uncertainties calculator module*

The values of 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).

#### 3.1.2.2 *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).

#### 3.1.2.3 *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.** 8).

## 3.2 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) (21). Briefly, after the ingestion of 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 non-diseased 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 non-diseased population. The normalized log-transformed diagnostic threshold 2.26 corresponds to the American Diabetes Association (ADA) diagnostic threshold for diabetes of the 2-h postprandial glucose during OGTT, that is equal to 11.1 mmol/l (22). 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 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.

The combined uncertainty and the resultant confidence interval increase with measurement uncertainty (Fig. 6-8, 18, and 19) and decrease with 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, on likelihood ratio for a positive or negative result, and Euclidean distance,

in accordance with previous findings (23-25).

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.

**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 | - |
| $\mu_D$ | 2.99 | 2.99 | 2.99 | 2.99 | 2.99 | 2.99 | 2.99 |
| $\sigma_D$ | 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.*

# 4 DISCUSSION

There is a persistent need of estimating the uncertainty of diagnostic accuracy measures, especially regarding screening and diagnostic tests of life-threatening diseases. The current pandemic of novel corona virus disease 2019 (COVID-19) has exposed this convincingly (26-31).

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

1. Diagnostic accuracy is fundamental to clinical decision-making (32),
2. Defining the permissible measurement uncertainty is critical to quality and risk management in laboratory medicine (33)
3. Sampling uncertainty is decisive for clinical study design to evaluate a screening or diagnostic test (34).

There has been extensive research on either diagnostic accuracy or uncertainty, however, very limited research has been done on both subjects (7, 25, 35, 36).

This program explores the measurement, sampling and 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 on 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 the calculation of the components of uncertainty 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), in this instance uncertainty has relatively little effect on specificity, overall diagnostic accuracy, and negative predictive value. It affects more sensitivity, positive predictive value, Youden's index, and concordance probability, while it has a considerable impact on 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.

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

1. The assumptions used for the calculations:
    1.1. The existence of a “gold standard” diagnostic method. If a “gold standard” does not exist, there are alternative approaches for the estimation of diagnostic accuracy measures (37).
    1.2. The normality of either the measurements or their applicable transforms (11, 12, 38, 39), however, this is usually valid. There is related literature on the distribution of measurements of diagnostic tests, in the context of reference intervals and diagnostic thresholds or clinical decision limits (40-44).
    1.3. The bimodality of the measurands, that is generally accepted, although unimodal distributions could be considered (45, 46).
    1.4. The measurement uncertainty homoscedasticity in the diagnostic thresholds range. Nevertheless, if measurement uncertainty is heteroscedastic, thus skewing the measurements distribution, appropriate transformations may restore homoscedasticity (47).

    If the above assumptions are not valid, there are other components of uncertainty which are not calculated by this program.

2. The first order Taylor series approximations for the uncertainty propagation calculations (4) (16). Higher order approximations may improve the accuracy.
3. The uncertainty of prevalence rate approximation by the Agresti-Coull adjusted Waldo interval (15), although there are more exact methods (48).

However, addressing these limitations, would increase exponentially the computational complexity.

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 different types of plots and 3 different 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 above-mentioned programs 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 and can be used as a flexible, user-friendly, interactive educational or research tool in medical decision-making.

# 6 AVAILABILITY AND REQUIREMENTS

**Project name:** Uncertainty

**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® Pentium™ Dual-Core or equivalent CPU and 2GB of RAM

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

**Any restrictions to use by non-academics:** No other restrictions

# 7 LIST OF ABBREVIATIONS

DAM: diagnostic accuracy measure

GUM: Guide to the expression of uncertainty in measurement (4)

OGTT: oral glucose tolerance test

ADA: American Diabetes Association

COVID-19: novel corona virus disease 2019

# 8 DECLARATIONS

## 8.1 ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

## 8.2 CONSENT FOR PUBLICATION

Not applicable.

## 8.3 AVAILABILITY OF DATA AND MATERIALS

All data of the case study are included in this published article (Subsection Sample in Research Design and Methods, and Table 3):

Lim TO, Bakri R, Morad Z, Hamid MA. Bimodality in blood glucose distribution: is it universal? Diabetes Care. 2002;25(12):2212-7. Epub 2002/11/28. doi: 10.2337/diacare.25.12.2212. PubMed PMID: 12453963.

## 8.4 COMPETING INTERESTS

The authors declare that they have no competing interests.

## 8.5 FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## 8.6 AUTHORS' CONTRIBUTIONS

TC substantially contributed to the conception and design of the research project, analyzed and interpreted the case data, substantially contributed to the creation of the software and drafted the article; she has approved the submitted version; she has agreed both to be personally accountable for her own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which she was not personally involved, are appropriately investigated, resolved and the resolution documented in the literature.

ATH supervised the research project, substantially contributed to the creation of the software and substantially revised the draft of the article; he has approved the submitted version; he has agreed both to be personally accountable for his own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which he was not personally involved, are appropriately investigated, resolved and the resolution documented in the literature.

# 9 ACKNOWLEDGEMENTS

Not applicable.

# 10 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. 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 2012.
7. 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.
8. Ceriotti F, Fernandez-Calle P, Klee GG, Nordin G, Sandberg S, Streichert T, et al. Criteria for assigning laboratory measurands to models for analytical performance specifications defined in the 1st EFLM Strategic Conference. Clin Chem Lab Med. 2017;55(2):189-94.

9. Rostron MHRSLREP. Measurement uncertainty arising from sampling - A guide to methods and approaches. Guideline. EURACHEM/CITAC; 2019 2019.
10. Bloch DA. Comparing Two Diagnostic Tests against the Same "Gold Standard" in the Same Sample. Biometrics. 1997;53(1):73-85.
11. Gillard J. A generalised Box–Cox transformation for the parametric estimation of clinical reference intervals. J Appl Stat. 2012;39(10):2231-45.
12. Atkinson AB. The box-cox transformation: review and extensions. Stat Sci. 2020.
13. Hund E, Massart DL, Smeyers-Verbeke J. Operational definitions of uncertainty. Trends Analyt Chem. 2001;20(8):394-406.
14. White GH. Basics of estimating measurement uncertainty. Clin Biochem Rev. 2008;29 Suppl 1:S53-60.
15. Agresti A, Coull BA. Approximate is Better than “Exact” for Interval Estimation of Binomial Proportions. Am Stat. 1998;52(2):119-26.
16. 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.
17. Welch BL. The Generalization of `Student's' Problem when Several Different Population Variances are Involved. Biometrika. 1947;34(1/2):28-35.
18. Satterthwaite FE. An approximate distribution of estimates of variance components. Biometrics. 1946;2(6):110-4.
19. Wolfram S. An Elementary Introduction to the Wolfram Language. Second ed2017. 340 p.
20. Wolfram Research I. Mathematica, Version 12.0. Champaign, IL2019.
21. Lim TO, Bakri R, Morad Z, Hamid MA. Bimodality in blood glucose distribution: is it universal? Diabetes Care. 2002;25(12):2212-7.
22. American Diabetes A. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2019. Diabetes Care. 2019;42(Suppl 1):S13-S28.
23. 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.
24. 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 2011.
25. Chatzimichail T, Hatjimihail AT. A software tool for exploring the relation between diagnostic accuracy and measurement uncertainty. In Review. 2020.
26. 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.
27. Tang Y-W, Schmitz JE, Persing DH, Stratton CW. The Laboratory Diagnosis of COVID-19 Infection: Current Issues and Challenges. J Clin Microbiol. 2020.
28. 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.
29. 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.

30. Mahase E. Covid-19: "Unacceptable" that antibody test claims cannot be scrutinised, say experts. BMJ. 2020;369:m2000.
31. 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).
32. 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.
33. 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.
34. Hajian-Tilaki K. Sample size estimation in diagnostic test studies of biomedical informatics. J Biomed Inform. 2014;48:193-204.
35. Theodorsson E. Uncertainty in Measurement and Total Error: Tools for Coping with Diagnostic Uncertainty. Clin Lab Med. 2017;37(1):15-34.
36. 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.
37. Collins J, Albert PS. Estimating diagnostic accuracy without a gold standard: A continued controversy. J Biopharm Stat. 2016;26(6):1078-82.
38. Sakia RM. The Box-Cox Transformation Technique: A Review. Journal of the Royal Statistical Society Series D (The Statistician). 1992;41(2):169-78.
39. Box GEP, Cox DR. An Analysis of Transformations. J R Stat Soc Series B Stat Methodol. 1964;26(2):211-43.
40. 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.
41. Pavlov IY, Wilson AR, Delgado JC. Reference interval computation: which method (not) to choose? Clin Chim Acta. 2012;413(13-14):1107-14.
42. 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.
43. 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.
44. 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.
45. Wilson JMG, Jungner G. Principles and practice of screening for disease. Geneva: World Health Organization; 1968 1968. 163 p.
46. 2.3 Clinical test evaluation. Unimodal and bimodal approaches. Scand J Clin Lab Invest. 1992;52(sup208):51-7.
47. Analytical Methods Committee AN. Why do we need the uncertainty factor? Anal Methods. 2019;11(15):2105-7.
48. Brown LD, Cai TT, DasGupta A. Interval Estimation for a Binomial Proportion. Stat Sci. 2001;16(2):101-17.

# 11 ADDITIONAL FILES

**File name:** Uncertainty.nb (It can be run on Wolfram Player®, available at:

https://www.wolfram.com/player/)

**Description of the file:** It is the software tool presented in the article.

# 12 NOTATION

## 12.1 POPULATIONS

$\overline{\mathrm{D}}$: nondiseased population

$D$: diseased population

## 12.2 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

## 12.3 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

## 12.4 PARAMETERS

$\mu_P$ : mean of the measurand of a test in the population *P*

$\sigma_P$ : 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 sambling 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 $x$

## 12.5 FUNCTIONS AND RELATIONS

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

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

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

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

# 13 APPENDIX

## 13.1 ABOUT THE PROGRAM CONTROLS

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

## 13.2 RANGE OF PARAMETERS

### 13.2.1 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

$\mu_D$: 0.1 – 6

$\sigma_D$: 0.1 – 6

$n_D$ : 2 – 10000

$r$: 0.001 – 0.999

$u_m$ : 0 – 6

$n_u$ : 20 – 1000

## 13.3 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 in the distribution are not considered.

### 13.3.1 Abscissas

#### 13.3.1.1 *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 $\{\left[\left[\frac{2}{r}\right]\right], \left[\left[\frac{2}{1-r}\right]\right], 20\}$ – max $\{10 \left[\left[\frac{2}{r}\right]\right], 10 \left[\left[\frac{2}{1-r}\right]\right], 10000\}$)

#### 13.3.1.2 *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 $\{\left[\left[\frac{2}{r}\right]\right], \left[\left[\frac{2}{1-r}\right]\right], 20\}$ – max $\{5 \left[\left[\frac{2}{r}\right]\right], 5 \left[\left[\frac{2}{1-r}\right]\right], 5000\}$)

## 13.4 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 non-diseased 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.

### 13.4.1 Plots

#### 13.4.1.1 *Plots vs diagnostic threshold*

##### 13.4.1.1.1 Diagnostic accuracy measures uncertainty plots

###### *13.4.1.1.1.1 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.

4) The range of the coordinate points to be plotted:

    a) Full,

    b) Partial.

*13.4.1.1.1.2 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 ($d$).

13.4.1.1.2 Diagnostic accuracy measures relative uncertainty plots

*13.4.1.1.2.1 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.

4) The range of the coordinate points to be plotted:

a) Full,

b) Partial.

*13.4.1.1.2.2 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 ($d$).

13.4.1.1.3 Diagnostic accuracy measures confidence intervals plots

*13.4.1.1.3.1 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.

6) The range of the coordinate points to be plotted:

   a) Full,

   b) Partial.

*13.4.1.1.3.2 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.

#### 13.4.1.2 ***Plots vs measurement uncertainty***

##### 13.4.1.2.1 Diagnostic accuracy measures uncertainty plots submodule

###### *13.4.1.2.1.1 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 ($CZ$).
4) The range of the coordinate points to be plotted:
    a) Full,
    b) Partial.

*13.4.1.2.1.2 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 ($u_m$).

13.4.1.2.2 Diagnostic accuracy measures relative uncertainty plots

*13.4.1.2.2.1 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.

4) The range of the coordinate points to be plotted:

a) Full,

b) Partial.

*13.4.1.2.2.2 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 ($u_m$).

13.4.1.2.3 Diagnostic accuracy measures confidence intervals plots submodule

*13.4.1.2.3.1 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.

5) The range of the coordinate points to be plotted:

a) Full,

b) Partial.

*13.4.1.2.3.2 Output*

Plots of the values of the:

1) Lower bound,

2) Point estimation,

3) Upper bound

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

### 13.4.1.3 ***Plots vs population size***

#### 13.4.1.3.1 Diagnostic accuracy measures uncertainty plots

*13.4.1.3.1.1 Input*

The user defines:

5) The diseased and nondiseased population samples parameters:
    a) The prevalence of the disease,
    b) The measurand means,
    c) The measurand standard deviations,
6) 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 ($CZ$),
7) The standard measurement uncertainty of the test.
8) The range of the coordinate points to be plotted:
    a) Full,
    b) Partial.

*13.4.1.3.1.2 Output*

Plots of the values of the:

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

of the measure versus total population size ($n$).

#### 13.4.1.3.2 Diagnostic accuracy measures relative uncertainty plots submodule

##### *13.4.1.3.2.1 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.
4) The range of the coordinate points to be plotted:
    a) Full,
    b) Partial.

*13.4.1.3.2.2 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 total population size ($n$).

13.4.1.3.3 Diagnostic accuracy measures confidence intervals plots

*13.4.1.3.3.1 Input*

The user defines:

1) The confidence level ($p$),
2) The diseased and nondiseased population samples parameters:
    a) The prevalence of the disease,
    b) The measurand means,
    c) The measurand standard deviations,
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.

6) The range of the coordinate points to be plotted:

a) Full,

b) Partial.

*13.4.1.3.3.2 Output*

Plots of the values of the:

1) Lower bound,

2) Point estimation,

3) Upper bound

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

## 13.4.2 Calculators

### 13.4.2.1 ***Diagnostic accuracy measures uncertainty calculator***

#### 13.4.2.1.1 Input

1) The diagnostic threshold ($d$)

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 standard measurement uncertainty of the test.

#### 13.4.2.1.2 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 ($CZ$).

### 13.4.2.2 ***Diagnostic accuracy measures relative uncertainty calculator***

#### 13.4.2.2.1 Input

1) The diagnostic threshold ($d$),
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 standard measurement uncertainty of the test.

#### 13.4.2.2.2 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 ($CZ$).

### 13.4.2.3 ***Diagnostic accuracy measures confidence intervals calculator***

#### 13.4.2.3.1 Input

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.

13.4.2.3.2 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 ($CZ$).

# 14 FIGURE LEGENDS

---

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

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

**Figure 3. Plots vs diagnostic threshold module, DAM uncertainty plots 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.

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

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

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

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

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

**Figure 9. Plots vs total population sample size module, DAM uncertainty plots 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.

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

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

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

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

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

**Figure 15. DAM standard uncertainties versus diagnostic threshold plots.** Plots of standard combined, sampling, and measurement uncertainties of (A) sensitivity ($u(Se)$), (B) specificity ($u(Sp)$), (C) positive predictive value ($u(PPV)$), and (D) negative predictive value ($u(NPV)$) 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 of (A) sensitivity ($u(Se)/Se$), (B) specificity ($u(Sp)/Sp$), (C) positive predictive value ($u(PPV)/PPV$), and (D) negative predictive value ($u(NPV)/NPV$) 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 relative standard uncertainties versus measurement uncertainty plots.** Plots of relative standard combined, sampling, and measurement uncertainties of (A) sensitivity ($u(Se)/Se$), (B) specificity ($u(Sp)/Sp$), (C) positive predictive value ($u(PPV)/PPV$), and (D) negative predictive value ($u(NPV)/NPV$) 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 relative standard uncertainties versus population sample size plots.** Plots of relative standard combined, sampling, and measurement uncertainties of (A) sensitivity ($u(Se)/Se$), (B) specificity ($u(Sp)/Sp$), (C) positive predictive value ($u(PPV)/PPV$), and (D) negative predictive value ($u(NPV)/NPV$) 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**.**

**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**.**