Article

A Software Tool for Calculating the Uncertainty of Diagnostic Accuracy Measures

Theodora Chatzimichail [ID] and Aristides T. Hatjimihail *[ID]

Hellenic Complex Systems Laboratory, Kostis Palamas 21, 66131 Drama, Greece; tc@hcsl.com
* Correspondence: ath@hcsl.com; Tel.: +30-2521-058227

Abstract: Screening and diagnostic tests are applied for the classification of people into diseased and non-diseased populations. Although diagnostic accuracy measures are used to evaluate the correctness of classification in clinical research and practice, there has been limited research on their uncertainty. The objective for this work was 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 non-diseased 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 application of the program 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 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; confidence intervals; diagnostic tests; screening tests

1. Introduction

Diagnosis in medicine is the determination of the nature of a disease condition [1]. The term diagnosis is derived from the Greek word "διάγνωσις" meaning "discernment". It is assumed that there is a dichotomy between the populations with and without a disease condition. Diagnostic tests or procedures are applied for the classification of people into the respective disjoint groups. The probability distributions of the measurand of a quantitative diagnostic test in each of the diseased and non-diseased populations are overlapping. The results of a test though can be dichotomized, by assigning a diagnostic threshold or cutoff point (Figure 1) [1]. The possible test results are summarized in Table 1. It is assumed that there is a reference ("gold standard") diagnostic method correctly classifying a subject as diseased or non-diseased [2]. The ratio of the diseased to the total population (diseased and non-diseased) at a single point in time is the prevalence rate (*r*) of the disease.

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 unequivocally [3–7]. There has been extensive research on either diagnostic accuracy or uncertainty, however, extremely limited research has been done on both subjects [8–11].

The program *Diagnostic Uncertainty* 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 [12],
2. Defining the permissible measurement uncertainty is critical to quality and risk management in laboratory medicine [13].
3. Sampling uncertainty is decisive for clinical study design to evaluate a screening or diagnostic test [14].
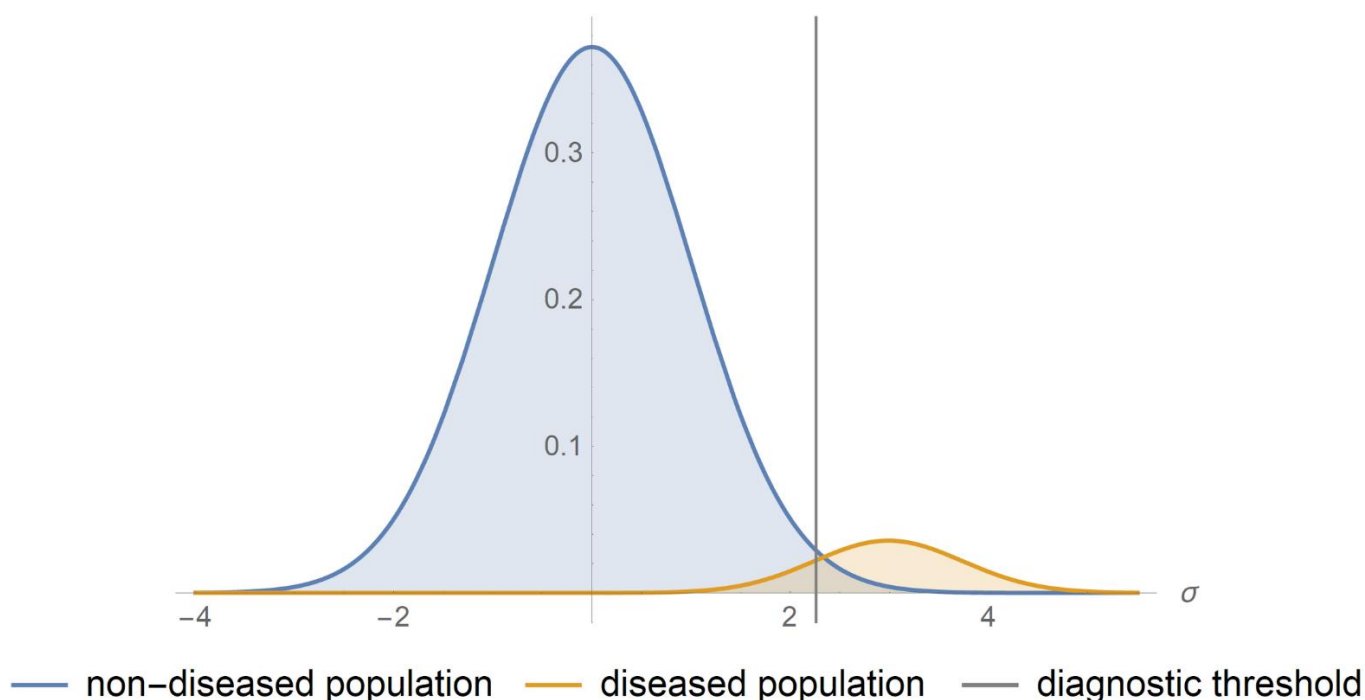

Figure 1. Probability density function plots. The probability density functions plots of a measurand in a non-diseased and diseased population.

Table 1. A 2 × 2 contingency table.

		Populations	
		Non-diseased	**Diseased**
Test Results	**Negative**	true negative (*TN*)	false negative (*FN*)
	Positive	false positive (*FP*)	true positive (*TP*)

1.1. Diagnostic Accuracy Measures

There are diagnostic accuracy measures (DAM) used for evaluating the discriminative ability of a screening or diagnostic test in clinical research and practice [2]. These are [15]:

1. Error-based measures, estimating misclassification rates. These include sensitivity (*Se*), specificity (*Sp*), overall diagnostic accuracy (*ODA*), Youden's index (*J*), Euclidean distance (*ED*) and concordance probability (*CZ*).
2. Information-based measures, assisting the interpretation of each single test result. These include positive predictive value (*PPV*), negative predictive value (*NPV*), likelihood ratio for positive result (*LR+*) and likelihood ratio for negative result (*LR−*).
3. Association-based measures, estimating the strength of the association between the test results and the reference diagnostic method. These include diagnostic odds ratio (*DOR*).

They can be further classified as following:

1. Defined conditionally on

  a. The true disease condition status: sensitivity, specificity, overall diagnostic accuracy, diagnostic odds ratio, likelihood ratio for positive result, likelihood ratio for negative result, Youden's index, Euclidean distance and concordance probability.

  b. The test outcome: positive predictive value and negative predictive value.

2. As prevalence

  a. Invariant: sensitivity, specificity, diagnostic odds ratio, likelihood ratio for positive result, likelihood ratio for negative result, Youden's index, Euclidean distance and concordance probability.

  b. Dependent: positive predictive value, negative predictive value and overall diagnostic accuracy.

  The natural frequency and the probability definitions of the above diagnostic accuracy measures are presented in Table 2. The symbols are explained in Appendix A.

Table 2. Natural frequency and probability definitions of diagnostic accuracy measures.

Measure	Natural Frequency Definition	Probability Definition
Sensitivity (*Se*)	$\frac{TP}{FN+TP}$	$Pr(T\|D)$
Specificity (*Sp*)	$\frac{TN}{TN+FP}$	$Pr(\overline{T}\|\overline{D})$
Positive Predictive Value (*PPV*)	$\frac{TP}{FP+TP}$	$Pr(D\|T)$
Negative Predictive Value (*NPV*)	$\frac{TN}{TN+FN}$	$Pr(\overline{D}\|\overline{T})$
Overall Diagnostic Accuracy (*ODA*)	$\frac{TN+TP}{TN+FN+TP+FP}$	$Pr(D)\,Pr(T\|D)$ $+Pr(\overline{D})\,Pr(\overline{T}\|\overline{D})$
Diagnostic Odds Ratio (*DOR*)	$\frac{TN\ TP}{FN\ FP}$	$\frac{\frac{Pr(T\|D)}{Pr(\overline{T}\|D)}}{\frac{Pr(T\|\overline{D})}{Pr(\overline{T}\|\overline{D})}}$
Likelihood Ratio for a Positive Result (*LR+*)	$\frac{TP(FP+TN)}{FP(FN+TP)}$	$\frac{Pr(T\|D)}{Pr(T\|\overline{D})}$
Likelihood Ratio for a Negative Result (*LR−*)	$\frac{FN(FP+TN)}{TN(FN+TP)}$	$\frac{Pr(\overline{T}\|D)}{Pr(\overline{T}\|\overline{D})}$
Juden's Index (*J*)	$\frac{TN\ TP-FN\ FP}{(TN+FP)(FN+TP)}$	$Pr(T\|D)+Pr(\overline{T}\|\overline{D})-1$
Euclidean Distance (*ED*)	$\sqrt{\left(\frac{FN}{FN+TP}\right)^2+\left(\frac{FP}{TN+FP}\right)^2}$	$\sqrt{Pr(\overline{T}\|D)^2+Pr(T\|\overline{D})^2}$
Concordance Probability (*CZ*)	$\frac{TN\ TP}{(TN+FP)(FN+TP)}$	$Pr(T\|D)\,Pr(\overline{T}\|\overline{D})$

The symbols are explained in Appendix A.

1.2. Uncertainty of Diagnostic Accuracy Measures

  Uncertainty is an expression of imperfect or deficient information. When quantifiable it can be represented with probability [16]. The following components of the combined uncertainty of the diagnostic accuracy measures will be considered:

1.2.1. Measurement Uncertainty

  As measurements are inherently variable, 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" [17]. Measurement uncertainty is replacing the total analytical error concept [18].

1.2.2. Sampling Uncertainty

Diagnostic accuracy measures are estimated by applying a screening or diagnostic test to samples of populations. Sampling heterogeneity contributes to the combined uncertainty of the diagnostic accuracy measures [19]. Even when simple random sampling is applied, there is inherent sample heterogeneity [20]. A sample of size *n* is considered as simple random sample, if all possible samples of the same size are equally probable [21].

2. Materials and Methods

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 [22].
2. Either the values of the measurand or their transforms [23,24] 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.
4. The sampling is simple random.
5. If the measurement is above the threshold the patient is classified as test-positive, otherwise as test-negative.

2.1.1. Calculation of Diagnostic Accuracy Measures

The calculation of the diagnostic accuracy measures is based on their probability definitions (Table 2). The sensitivity and specificity can be defined in terms of the error function and the complementary error function (see Appendix B). The other diagnostic accuracy measures can be expressed in terms of sensitivity, specificity and prevalence rate and calculated as shown in Appendix B.

2.1.2. Calculation of Uncertainty of Diagnostic Accuracy Measures

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 [25].

Measurement Uncertainty

The standard measurement uncertainty u_m of a measurand is estimated from a sample of n_u measurements, as described in "Guide to the expression of uncertainty in measurement"(GUM) and "Expression of measurement uncertainty in laboratory medicine" [17]. Bias may be considered as a component of the standard measurement uncertainty [26].

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:

$$u_s(m_P) = \frac{s_P}{\sqrt{n_P}} \tag{1}$$

$$u_s(s_P) = \frac{s_P}{\sqrt{2(n_P - 1)}} \tag{2}$$

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:

$$u_c(m_P) = \sqrt{\frac{s_P^2}{n_P} + u_m^2} \tag{3}$$

$$u_c(s_P) = \sqrt{\frac{s_P^2}{2(n_P - 1)} + u_m^2} \tag{4}$$

Sampling Uncertainty of Prevalence Rate

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

$$u_s(r) = \sqrt{\frac{(2 + n_{\overline{D}})(2 + n_D)}{(4 + n_{\overline{D}} + n_D)^3}} \tag{5}$$

according to the Agresti–Coull adjustment of the Waldo interval [27].

Combined Uncertainty of Diagnostic Accuracy Measures

The standard combined uncertainty $u_c(x)$ of each diagnostic accuracy measure x is calculated by applying the rules of uncertainty propagation from the input values to the calculated diagnostic accuracy measure (see Appendix B), according to GUM [28,29], with a first-order Taylor series approximation to uncertainty propagation [30].

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

$$u_c(x) = \sqrt{\sum_{i=1}^{l} u_i(x)^2} \tag{6}$$

Expanded Uncertainty of Diagnostic Accuracy Measures

The effective degrees of freedom v_{eff} of the standard combined uncertainty $u_c(x)$ are calculated using the Welch–Satterthwaite formula [31,32]:

$$v_{eff} = \frac{u_c(x)^4}{\sum_{i=1}^{l} \frac{u_i(x)^4}{v_i}} \tag{7}$$

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

$$v_{min} \leq v_{eff} \leq \sum_{i=1}^{l} v_i \tag{8}$$

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

$$U_c(x) = \left(F_v^{-1}\left(\frac{1-p}{2}\right) u_c(x), F_v^{-1}\left(\frac{1+p}{2}\right) u_c(x) \right) \tag{9}$$

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

$$CI_p(x) = \left(x + F_v^{-1}\left(\frac{1-p}{2}\right) u_c(x), x + F_v^{-1}\left(\frac{1+p}{2}\right) u_c(x) \right) \tag{10}$$

2.2. The Program

To calculate the uncertainty of the diagnostic accuracy measures, the interactive program *Diagnostic Uncertainty* was developed in the Wolfram Language [33], using Wolfram Mathematica® Ver. 12.2, Wolfram Research, Inc., Champaign, IL, USA [34]. The program was designed to provide 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 a screening or diagnostic test, applied at a single point in time in non-diseased and diseased population samples. 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 Appendix C).

3. Results

3.1. Flowchart of the Program

The flowchart of the program is presented in Figure 2.

3.2. Interface of the Program

The modules and submodules of the program include panels with controls which allow the interactive manipulation of various parameters, as described in detail in Supplementary File: Diagnostic Uncertainty Interface.pdf. These are the following:

3.2.1. Plots vs. Diagnostic Threshold Module
Diagnostic Accuracy Measures Standard Uncertainty Plots Submodule

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

Diagnostic Accuracy Measures Relative Standard Uncertainty Plots Submodule

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

Confidence Intervals of Diagnostic Accuracy Measures Plots Submodule

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

3.2.2. Plots vs. Measurement Uncertainty Module
Diagnostic Accuracy Measures Standard Uncertainty Plots Submodule

The values of the standard combined, measurement and sampling uncertainties of diagnostic accuracy measures of a screening or diagnostic test are plotted versus the measurement uncertainty of the test (Figure 6).
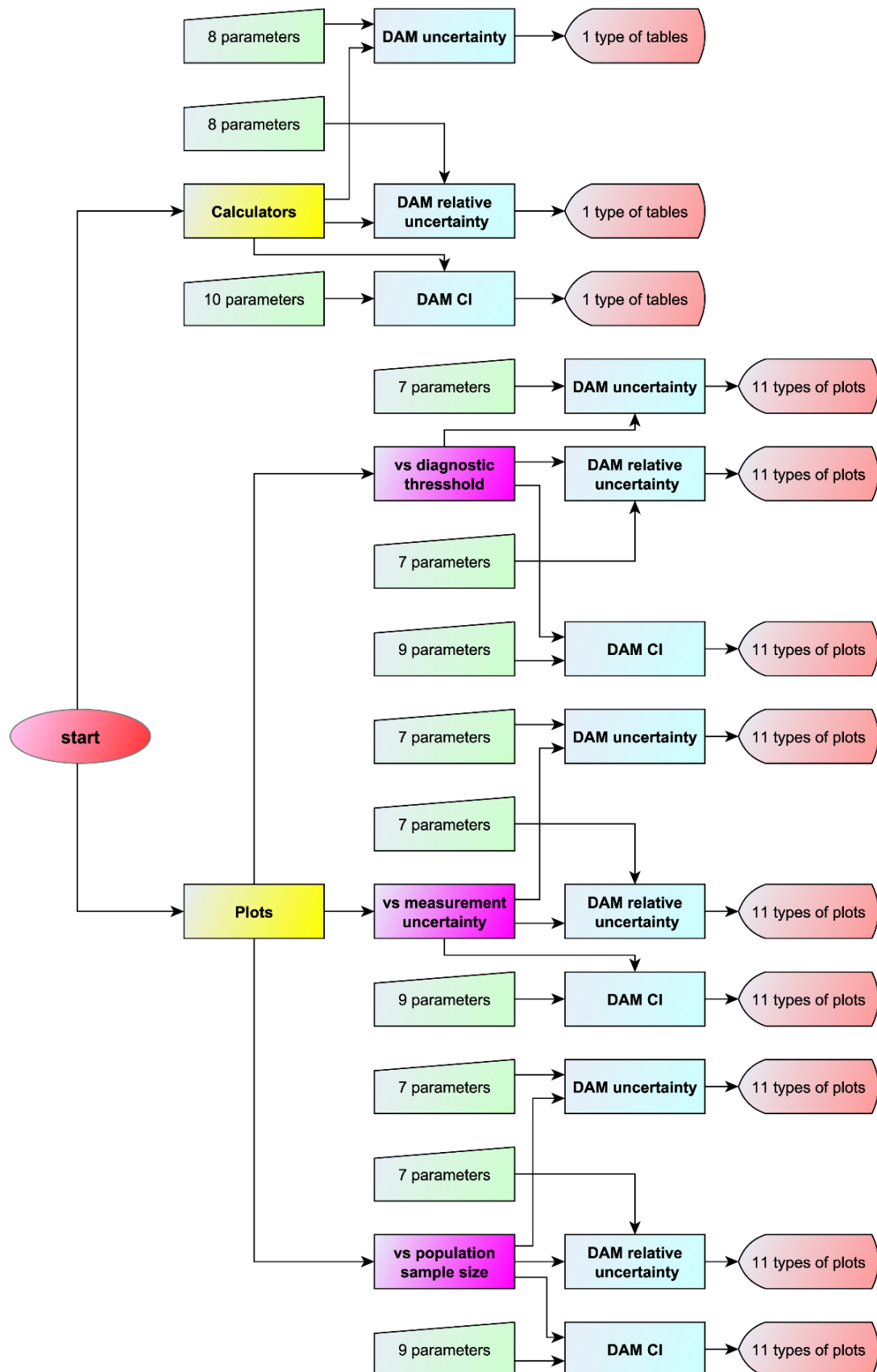
Diagnostic Accuracy Measures Relative Standard Uncertainty Plots Submodule

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

Figure 2. Program flowchart. The flowchart of the program with the number of the input parameters and of the output types for each module or submodule.
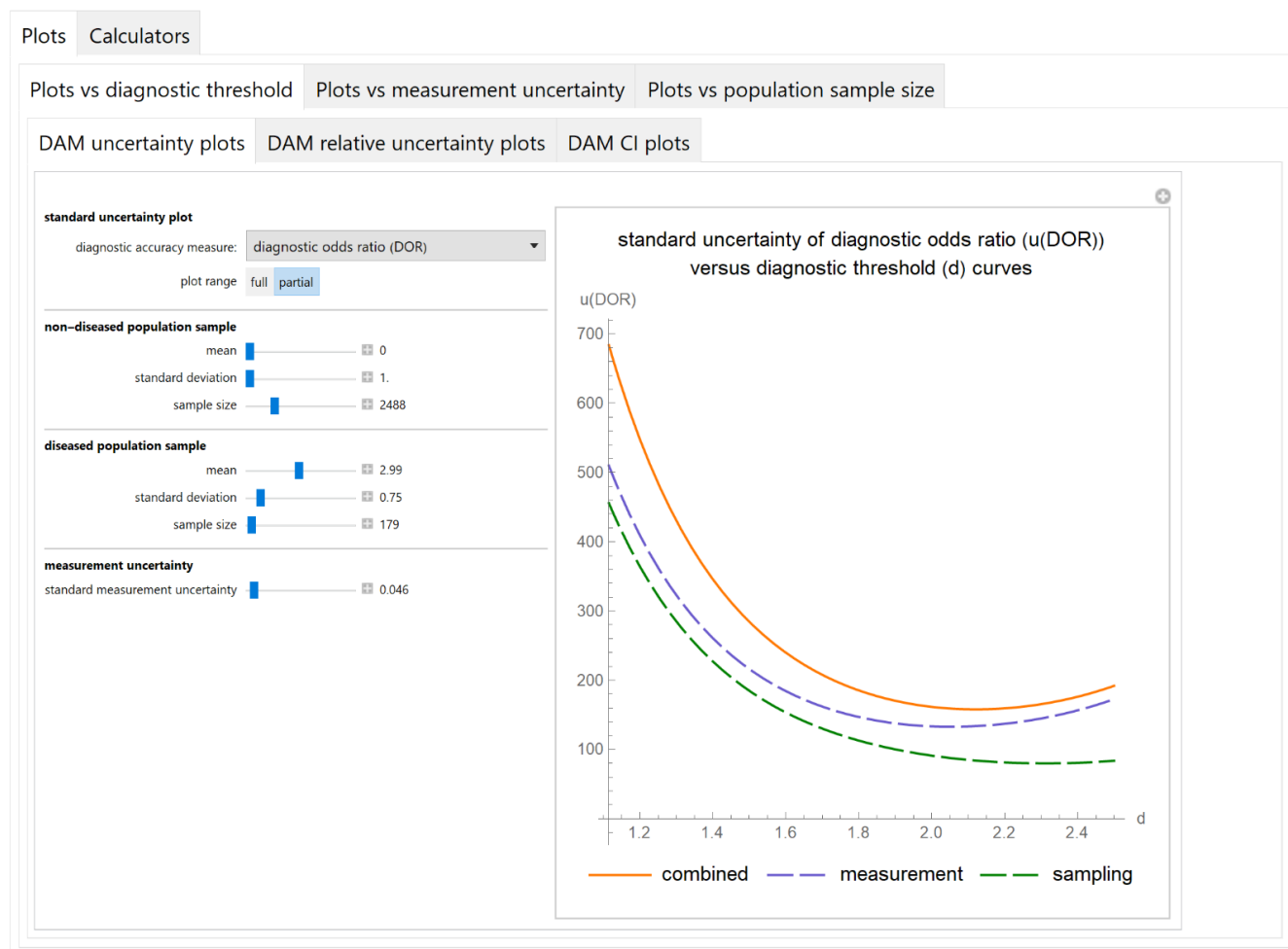
Figure 3. Plots vs. diagnostic threshold module, DAM uncertainty plots submodule screenshot. Standard combined, measurement and sampling uncertainty of diagnostic odds ratio (*u(DOR)*) versus diagnostic threshold (*d*) curve plot, with the settings shown on the left. The respective parameter settings are also shown in Table 3.

Table 3. The parameter settings of Figures 3–14.

Settings	Figure 3	Figure 4	Figure 5	Figures 6 and 7	Figure 8	Figures 9 and 10	Figure 11	Figures 12 and 13	Figure 14
p	-	-	0.95	-	0.95	-	0.95	-	0.95
d	1.1–2.5	0–4.0	2.26	2.26	2.26	2.26	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	2.99	2.99
σ_D	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75
n_D	179	179	179	179	179	-	-	179	179
$\mu_{\overline{D}}$	0.0	0.0	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	1.0	1.0
$n_{\overline{D}}$	2488	2488	2488	2488	2488	-	-	2488	2488
n	-	-	-	-	-	30–5000	30–5000	-	-
u_m	0.046	0.046	0.046	0–0.15	0–0.15	0.046	0.046	0.046	0.046
n_u	-	-	80	-	80	-	80	-	80

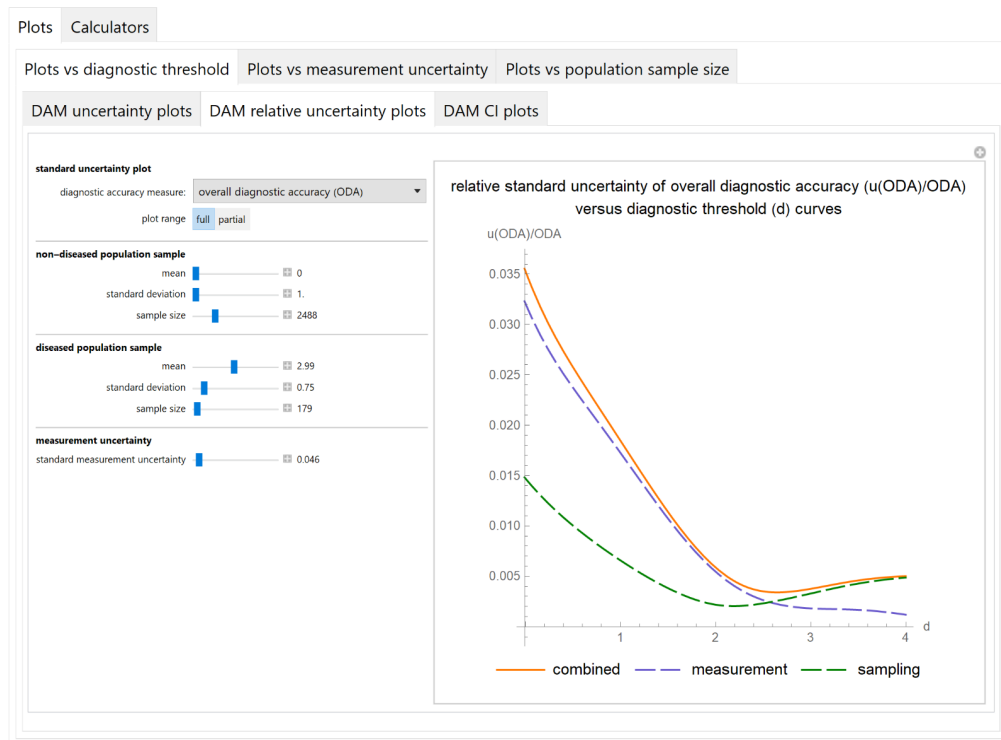The symbols are explained in Appendix A.

Figure 4. Plots vs. diagnostic threshold module, DAM relative uncertainty plots submodule screenshot. Relative standard combined, measurement and sampling uncertainty of overall diagnostic accuracy ($u(ODA)/ODA$) versus diagnostic threshold (d) curve plot, with the settings shown on the left. The respective parameter settings are also shown in Table 3.
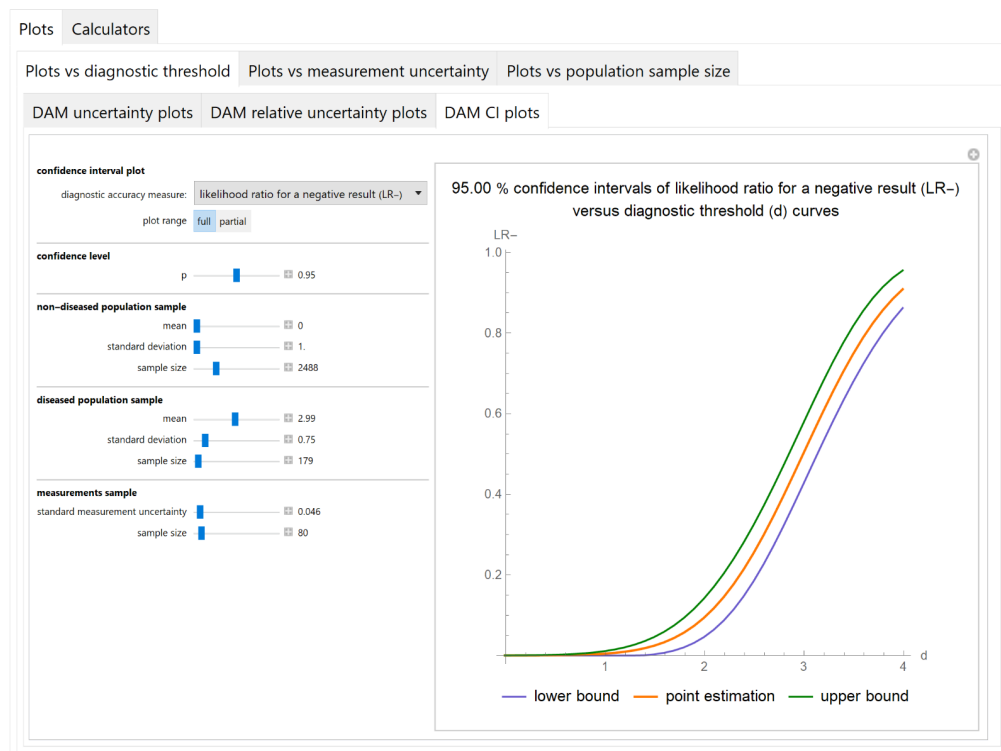

Figure 5. Plots vs. diagnostic threshold module, DAM CI plots submodule screenshot. Confidence intervals of likelihood ratio for a negative test result ($LR-$) versus diagnostic threshold (d) curves plot, with the settings shown on the left. The respective parameter settings are also shown in Table 3.
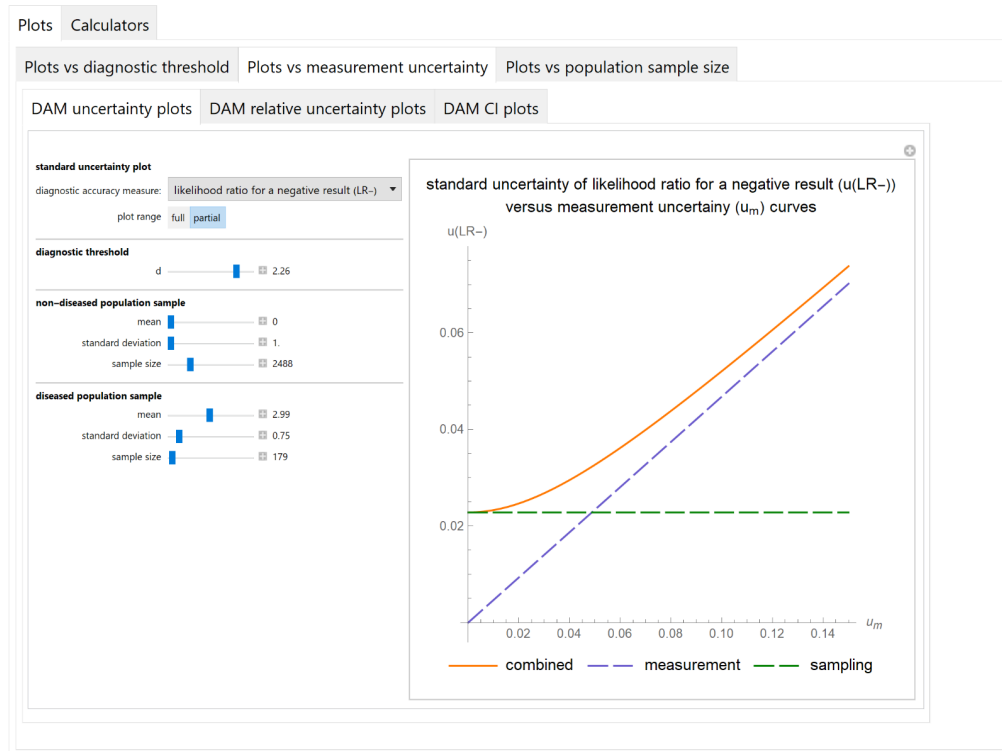
Figure 6. Plots vs. measurement uncertainty module, DAM uncertainty plots submodule screenshot. Standard combined, measurement and sampling uncertainty of likelihood ratio for a negative test result ($u(LR-)$) versus standard measurement uncertainty (u_m) curve plot, with the settings shown on the left. The respective parameter settings are also shown in Table 3.
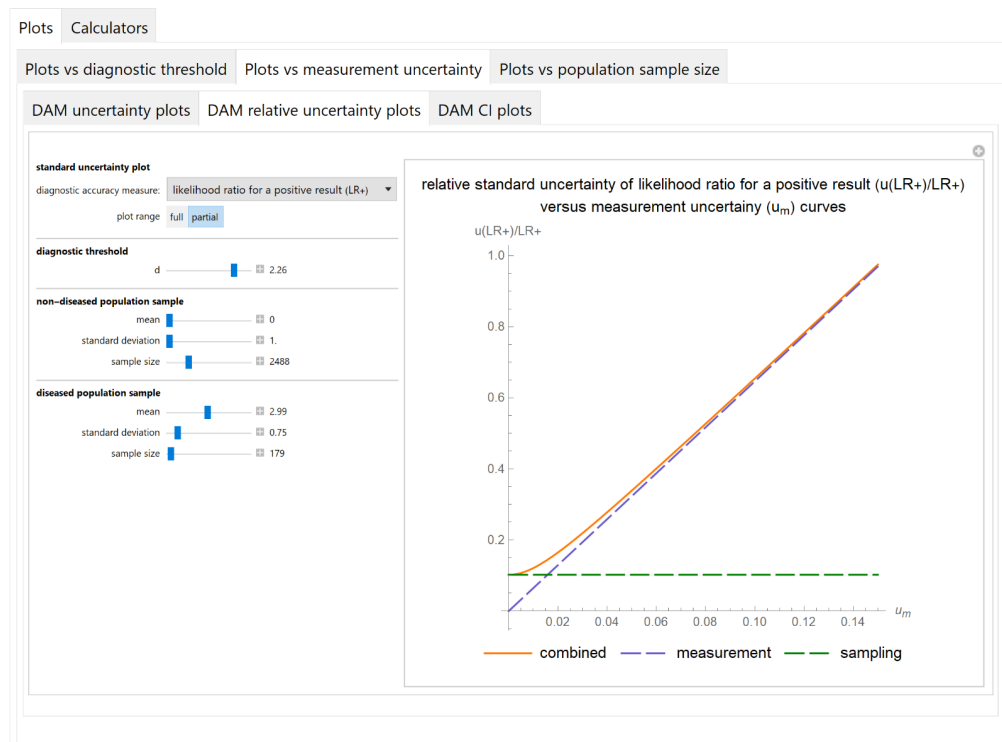

Figure 7. Plots vs. measurement uncertainty module, DAM relative uncertainty plots submodule screenshot. Relative standard combined, measurement and sampling uncertainty of likelihood ratio for a positive test result ($u(LR+)/LR+$) versus measurement uncertainty (u_m) curves plot, with the settings shown on the left. The respective parameter settings are also shown in Table 3.

Confidence Intervals of Diagnostic Accuracy Measures Plots Submodule

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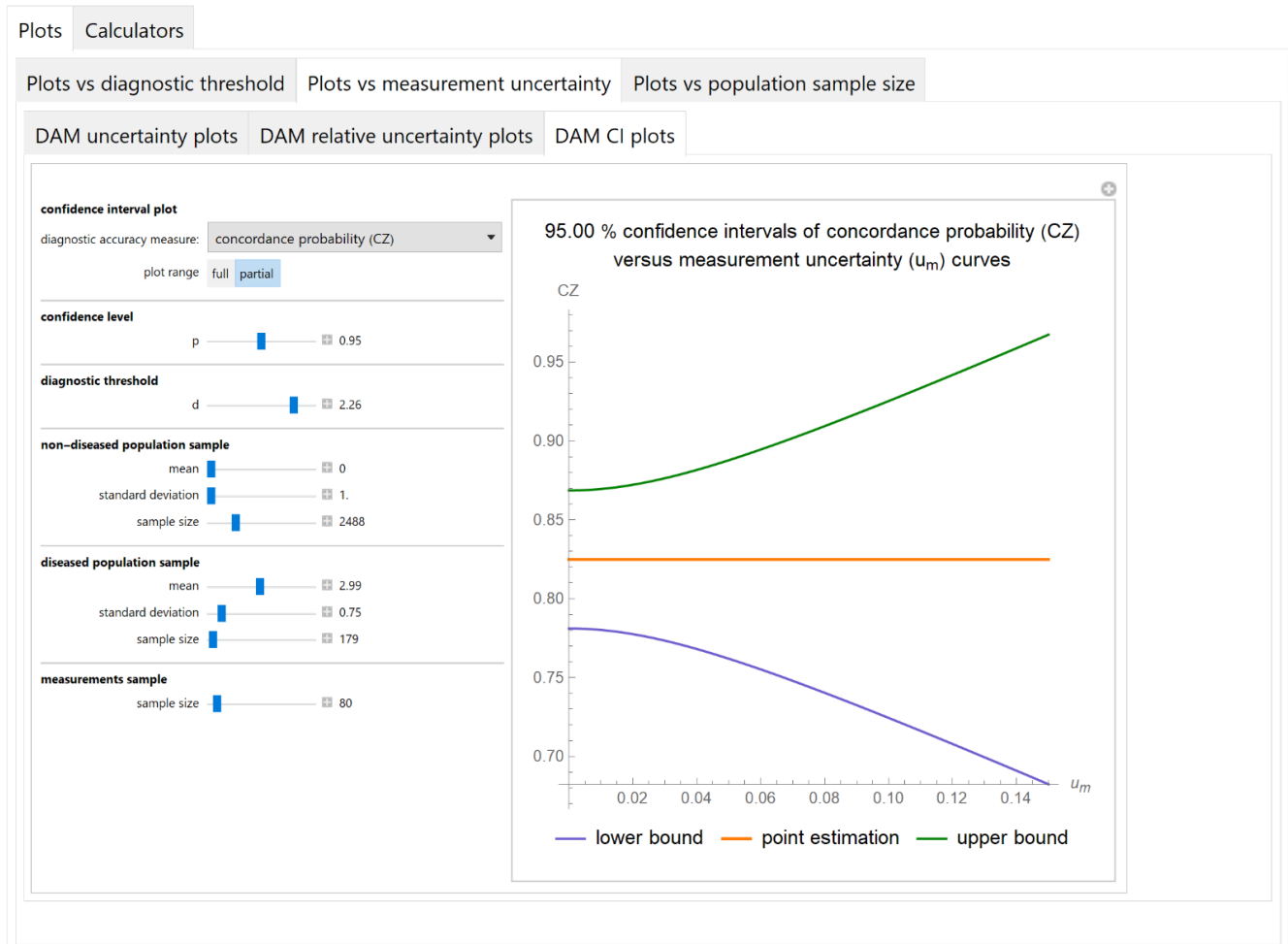

Figure 8. Plots vs. measurement uncertainty module, DAM CI plots submodule screenshot. Confidence intervals of concordance probability (*CZ*) versus standard measurement uncertainty (*u_m*) curves plot, with the settings shown on the left. The respective parameter settings are also shown in Table 3.

3.2.3. Plots vs. Population Sample Size Module

Diagnostic Accuracy Measures Standard Uncertainty Plots Submodule

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

Diagnostic Accuracy Measures Relative Standard Uncertainty Plots Submodule

The values of the relative standard combined, measurement and sampling uncertainties of diagnostic accuracy measures of a screening or diagnostic test are plotted versus the total population sample size (Figure 10).
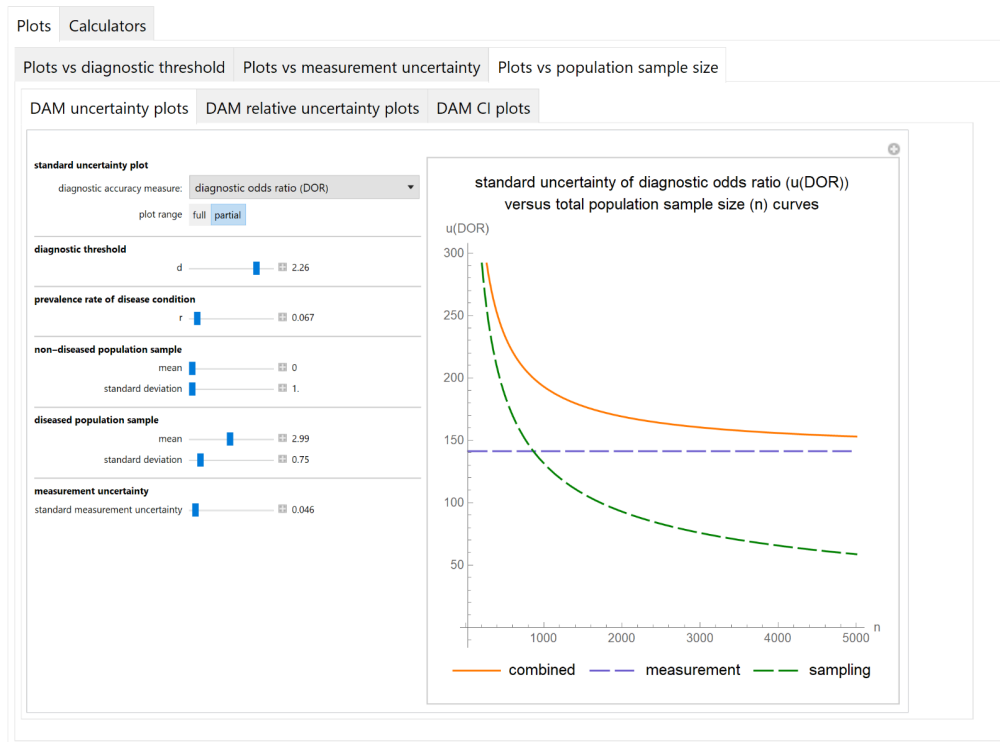
Figure 9. Plots vs. population sample size module, DAM uncertainty plots submodule screenshot. Standard combined, measurement and sampling uncertainty of diagnostic odds ratio (*u(DOR)*) versus total population sample size (*n*) curves plot, with the settings shown on the left. The respective parameter settings are also shown in Table 3.
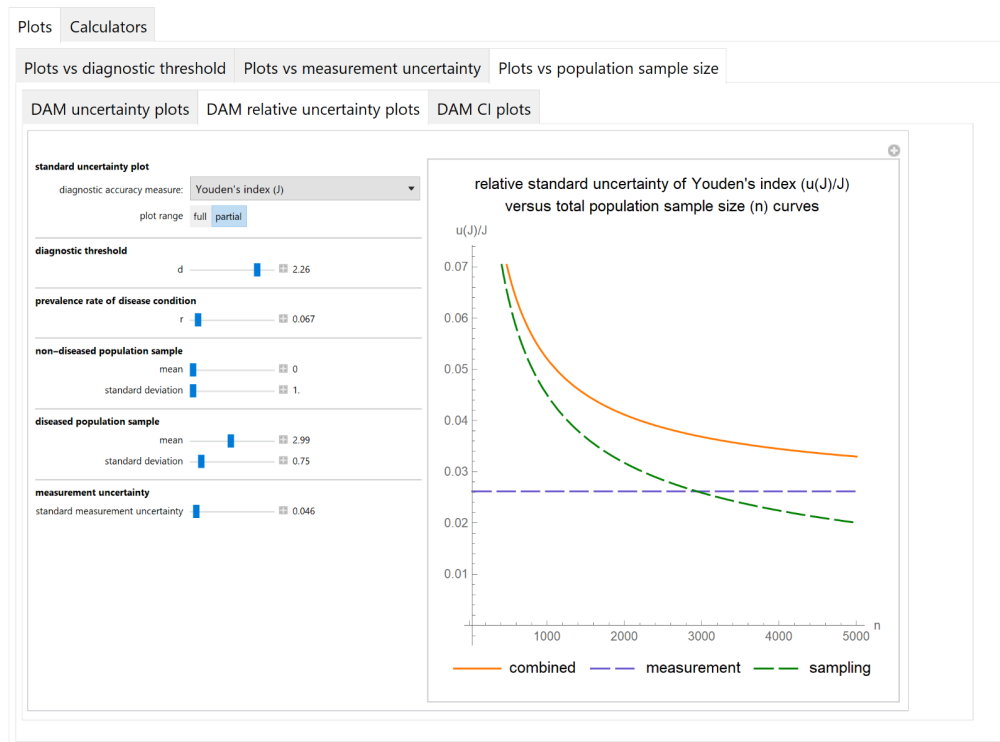

Figure 10. Plots vs. population sample size module, DAM relative uncertainty plots submodule screenshot. Relative standard combined, measurement and sampling uncertainty of Youden's index (*u(J)/J*) versus total population sample size (*n*) curves plot, with the settings shown on the left. The respective parameter settings are also shown in Table 3.

Confidence Intervals of Diagnostic Accuracy Measures Plots Submodule

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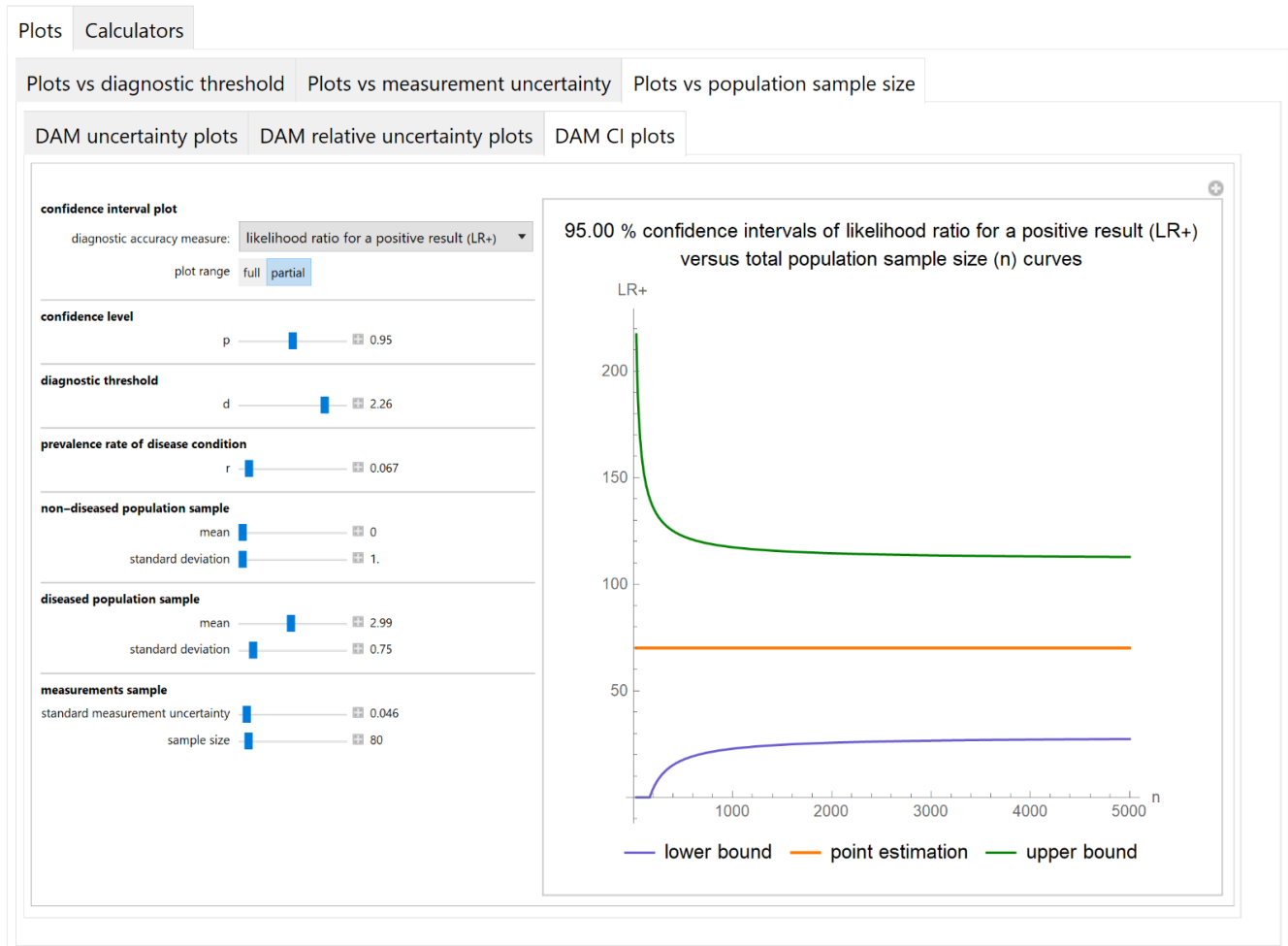

Figure 11. Plots vs. population sample size module, DAM CI plots submodule screenshot. Confidence intervals of likelihood ratio for a positive test result (*LR*+) versus total population sample size (*n*) curves plot, with the settings shown at the left. The respective parameter settings are also shown in Table 3.

3.2.4. Diagnostic Accuracy Measures Standard Uncertainty Calculator Module

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

3.2.5. Diagnostic Accuracy Measures Relative Standard Uncertainty Calculator Module

The values of the relative standard combined, measurement and sampling uncertainties of diagnostic accuracy measures of a screening or diagnostic test, at a selected diagnostic threshold, are calculated and presented in a table (Figure 13).
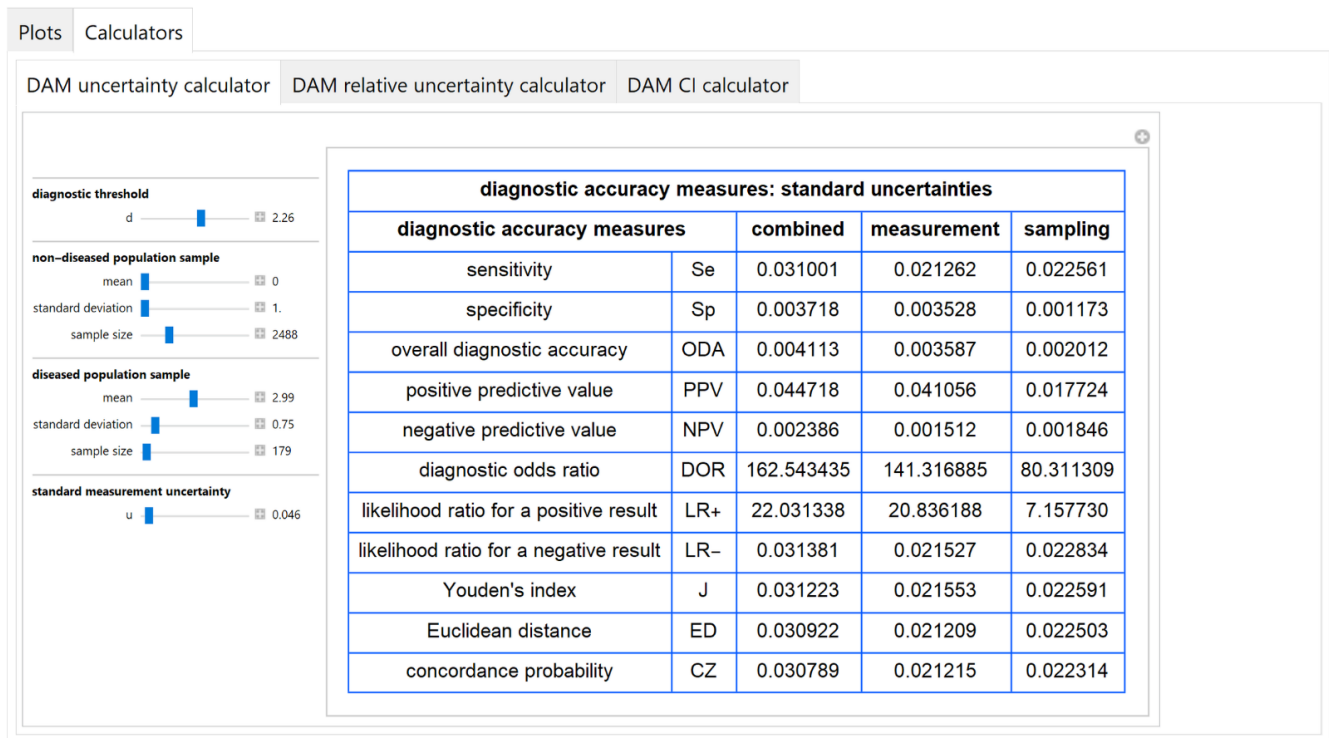
Figure 12. DAM uncertainty calculator module screenshot. Calculated standard combined, measurement and sampling uncertainties of diagnostic accuracy measures, with the settings shown on the left. The respective parameter settings are also shown in Table 3.
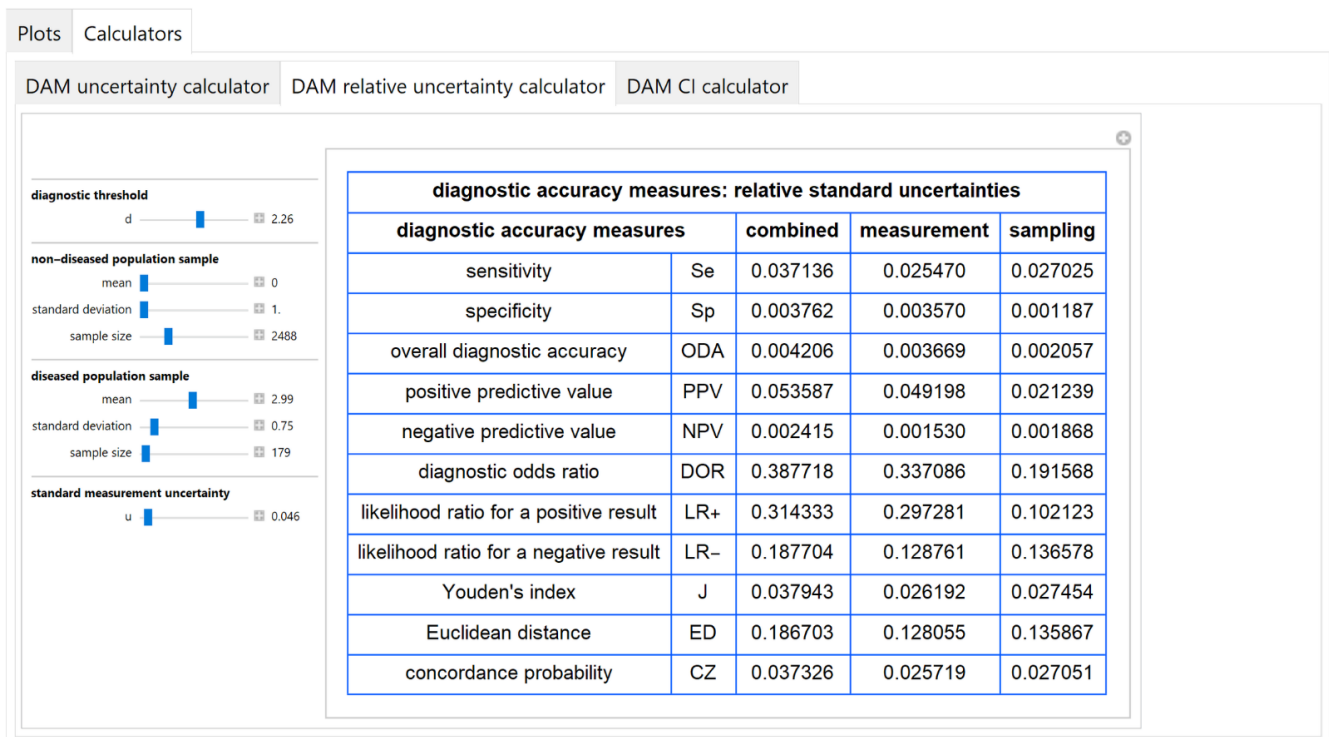

Figure 13. DAM relative uncertainty calculator submodule screenshot. Calculated relative standard combined, measurement and sampling uncertainty of diagnostic accuracy measures, with the settings shown on the left. The respective parameter settings are also shown in Table 3.

3.2.6. Diagnostic Accuracy Measures Confidence Intervals Calculator Module

The point estimations and the lower and upper bounds of the confidence intervals of diagnostic accuracy measures of a screening or diagnostic, at a selected confidence level and diagnostic threshold, are calculated and presented in a table (Figure 14).
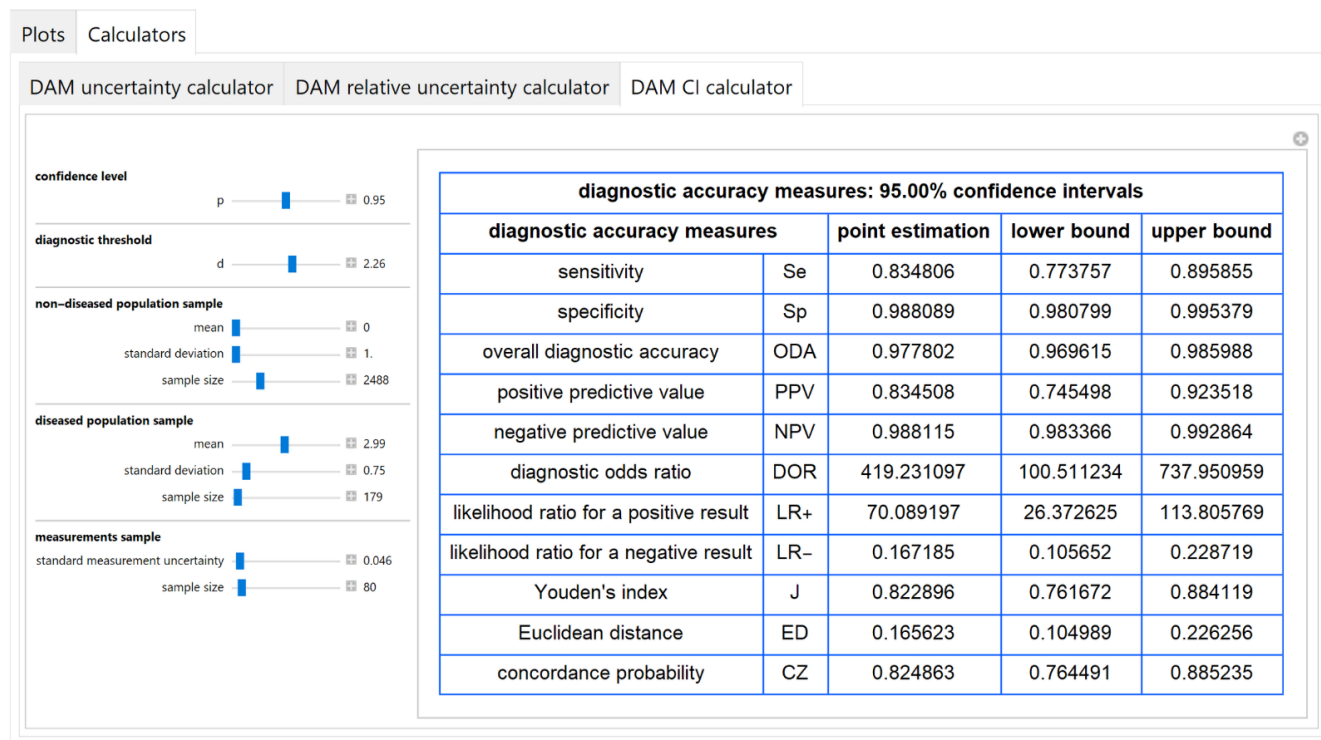

Figure 14. DAM CI calculator module screenshot. Calculated point estimations and confidence intervals of diagnostic accuracy measures, with the settings shown on the left. The respective parameter settings are also shown in Table 3.

3.3. Illustrative Example

The program was applied to a bimodal distribution of log-transformed blood glucose measurements in samples of non-diabetic and diabetic populations. The data were derived from a national health survey conducted in Malaysia in 1996 [35]. A glucose tolerance test (OGTT) was performed on 2667 Malay adults, aged 40–49 years. The respective sizes of the samples of the diseased and non-diseased populations were 179 and 2488. Glucose was measured with reflectance photometry, after the ingestion of 75 g glucose monohydrate. 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 Figure 1. The standardized log-transformed measurand means and standard deviations of the samples of 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 of the sample of the non-diseased population. The standardized log-transformed standard measurement uncertainty 0.046 of the test corresponds to coefficient of variation equal to 2%. The standardized log-transformed American Diabetes Association (ADA) diagnostic threshold for diabetes of the 2-h postprandial glucose during OGTT is equal to 2.26 [36].

The results of the illustrative example are presented:

1. In the plots of Figures 3–11 and 15–21.
2. In the chart of Figure 22
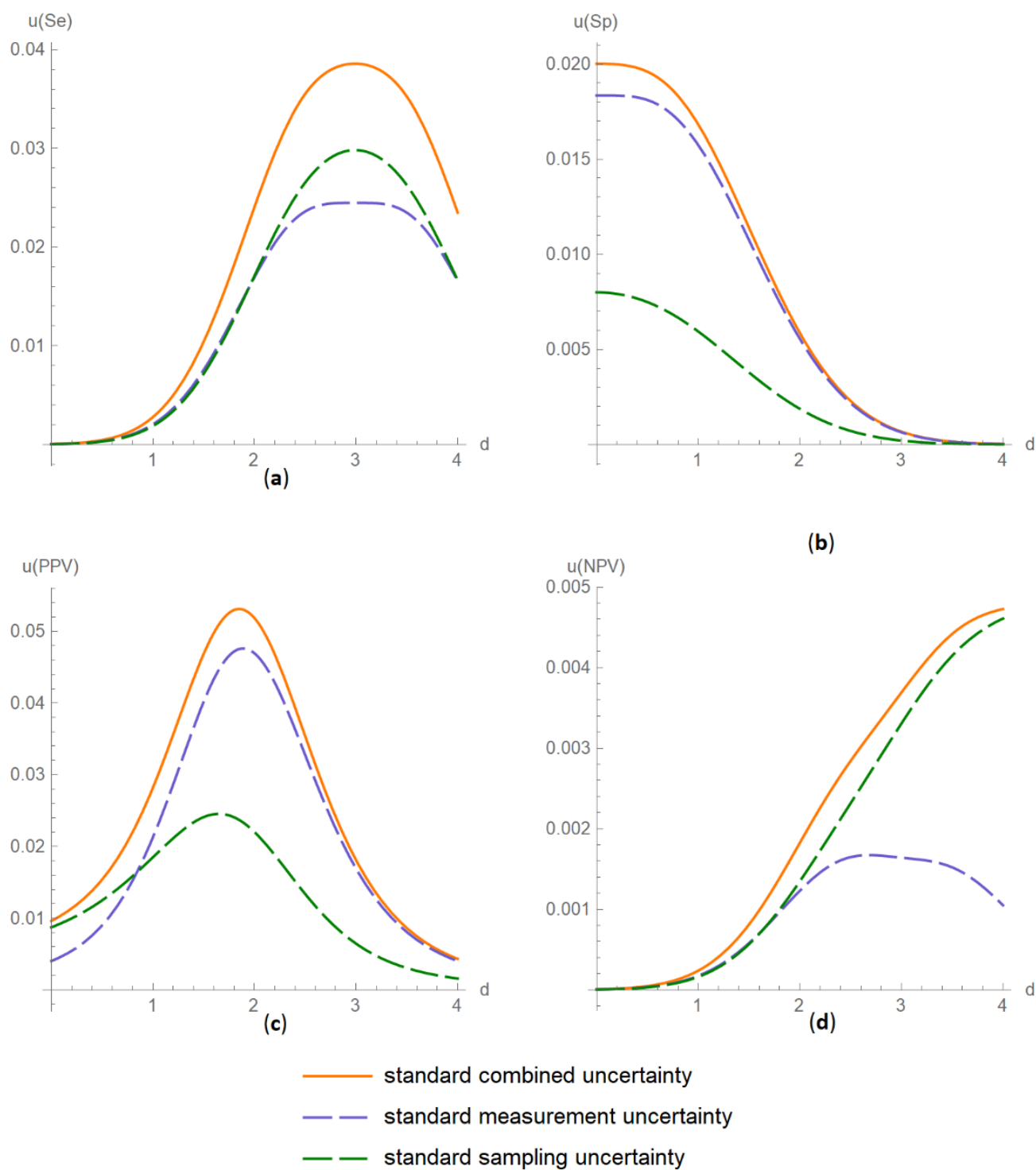3. In the tables of Figures 12–14.

Figure 15. Histogram of standard combined, measurement and sampling uncertainties of diagnostic accuracy measures, with the respective parameter settings in Table 4.

Table 4. The parameter settings of Figures 15–22.

Settings	Figures 15 and 16	Figure 17	Figure 18	Figure 19	Figure 20	Figure 21	Figure 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 are explained in Appendix A.

The parameter settings of Figures 3–14 are presented in Table 3 and of Figures 15–22 in Table 4. Figures 15–21 present the standard combined, measurement and sampling uncertainty and the resultant confidence intervals of sensitivity, specificity, positive and negative predictive value versus diagnostic threshold, measurement uncertainty and total population sample size.

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

In the illustrative example, combined uncertainty $u_c(x)$ has (see Figures 13 and 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 [37,38].

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.
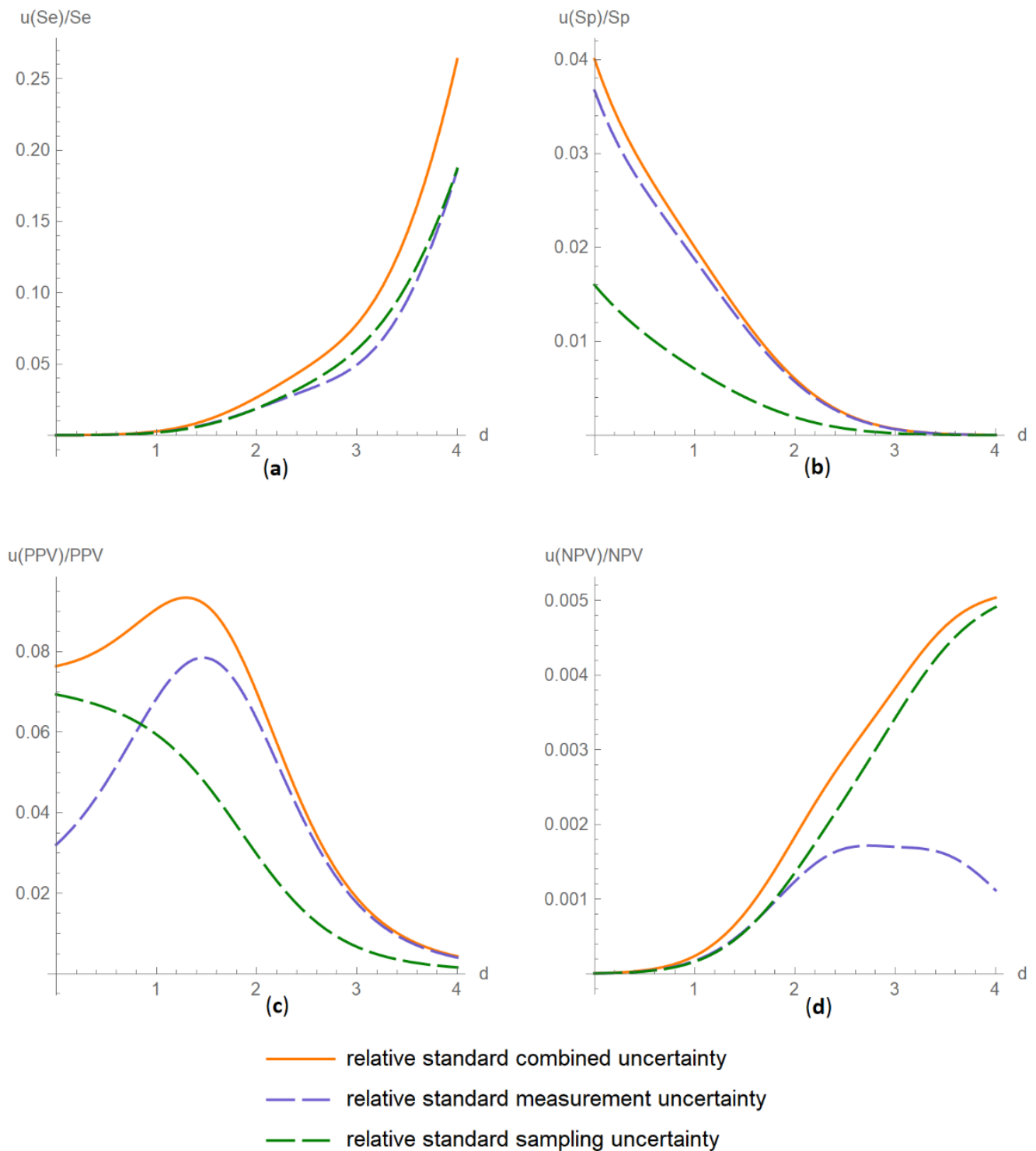
Figure 16. DAM standard uncertainties versus diagnostic threshold plots. Plots of standard combined, measurement and sampling uncertainties of (**a**) sensitivity ($u(Se)$), (**b**) specificity ($u(Sp)$), (**c**) positive predictive value ($u(PPV)$) and (**d**) negative predictive value ($u(NPV)$) versus diagnostic threshold (d) curves, with the respective parameter settings in Table 4.
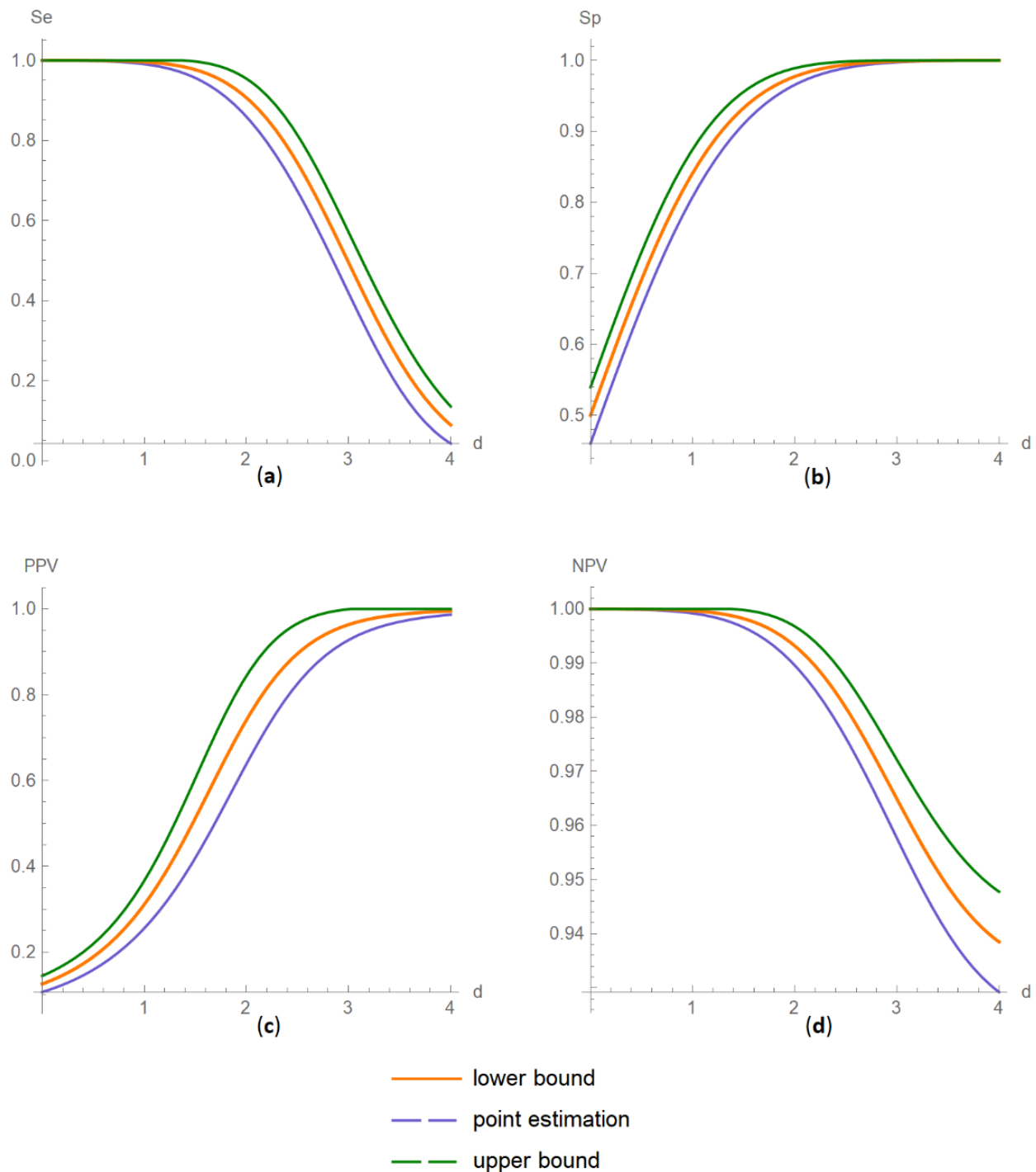
Figure 17. DAM relative standard uncertainties versus diagnostic threshold plots. Plots of relative standard combined, measurement and sampling uncertainties of (**a**) sensitivity ($u(Se)/Se$), (**b**) specificity ($u(Sp)/Sp$), (**c**) positive predictive value ($u(PPV)/PPV$) and (**d**) negative predictive value ($u(NPV)/NPV$) versus diagnostic threshold (d) curves, with the respective parameter settings in Table 4.
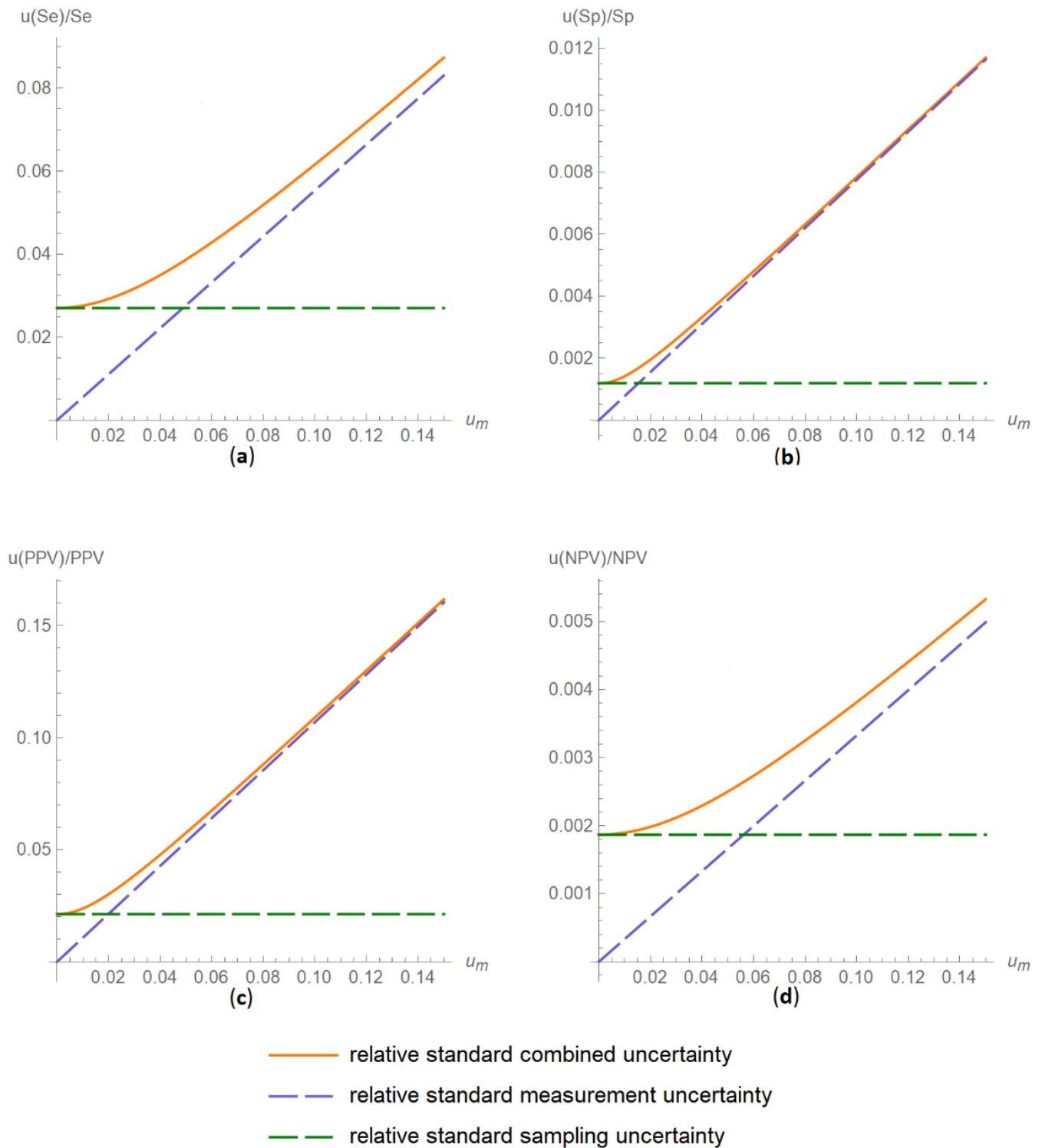
Figure 18. DAM confidence intervals versus diagnostic threshold plots. Plots of confidence intervals of (**a**) sensitivity (*Se*), (**b**) specificity (*Sp*), (**c**) positive predictive value (*PPV*) and (**d**) negative predictive value (*NPV*) versus diagnostic threshold (*d*) curves, with the respective parameter settings in Table 4.
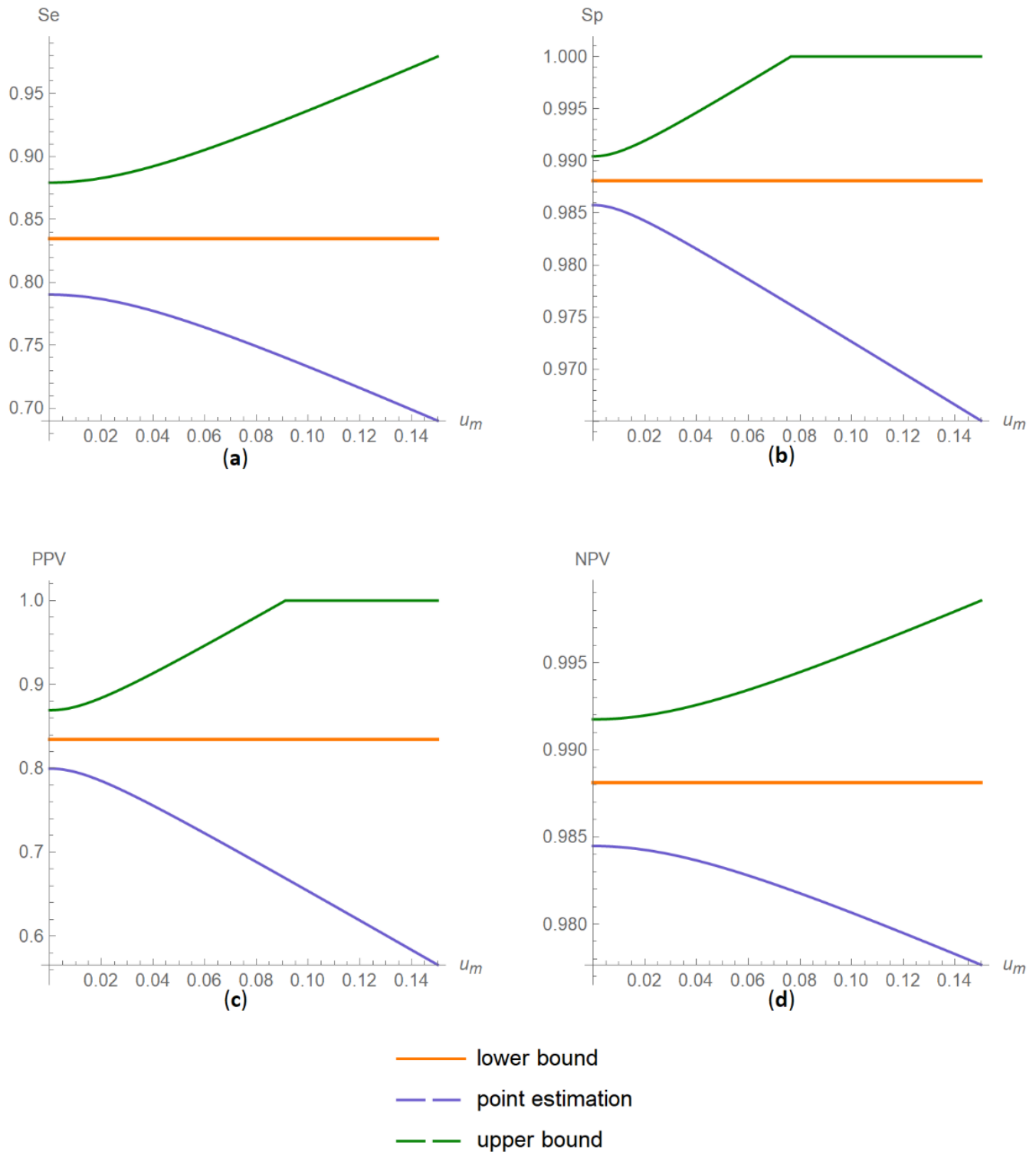
Figure 19. DAM relative standard uncertainties versus measurement uncertainty plots. Plots of relative standard combined, measurement and sampling uncertainties of (**a**) sensitivity ($u(Se)/Se$), (**b**) specificity ($u(Sp)/Sp$), (**c**) positive predictive value ($u(PPV)/PPV$) and (**d**) negative predictive value ($u(NPV)/NPV$) versus standard measurement uncertainty (u_m) curves, with the respective parameter settings in Table 4.

Figure 20. 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 parameter settings in Table 4.
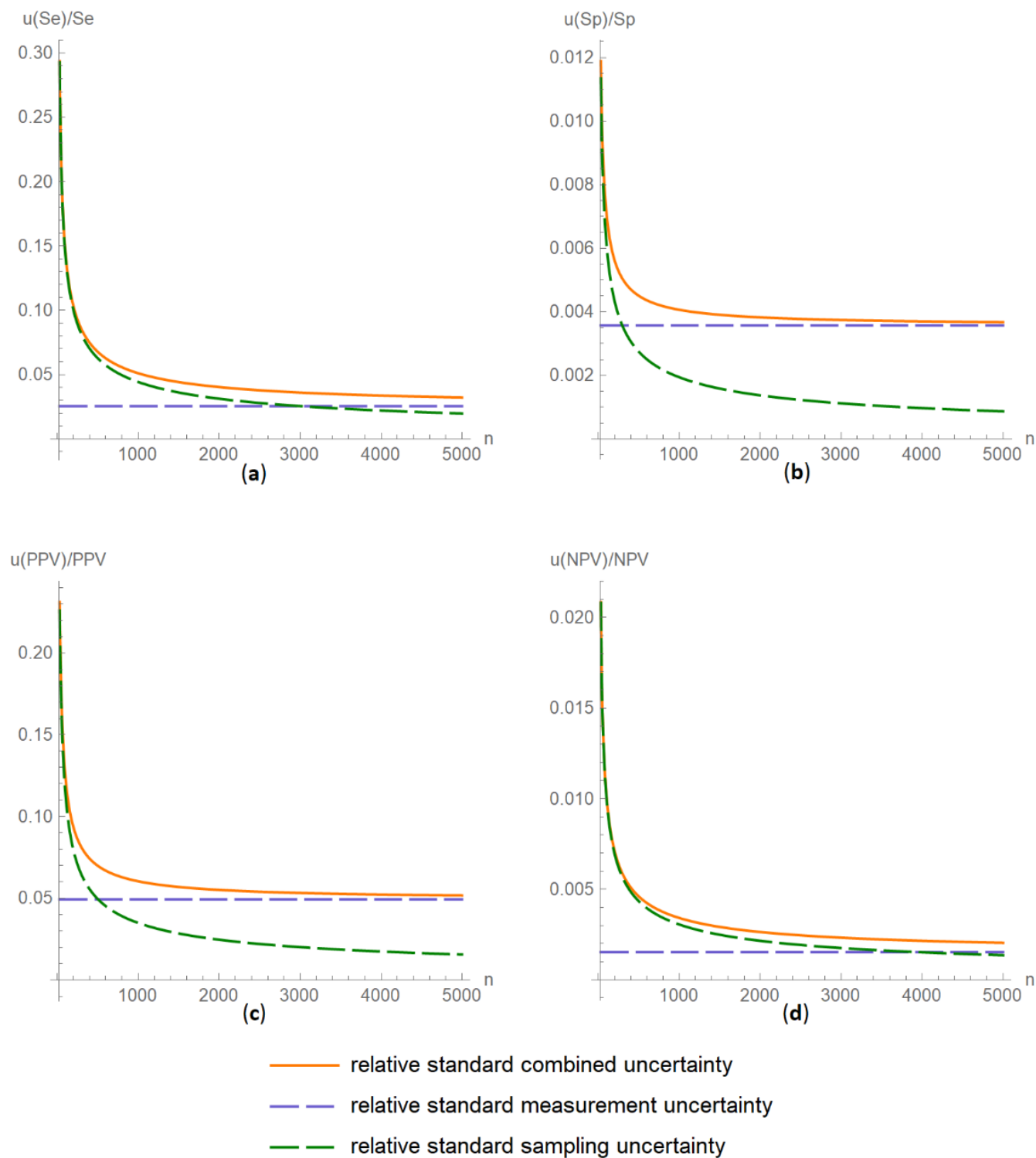
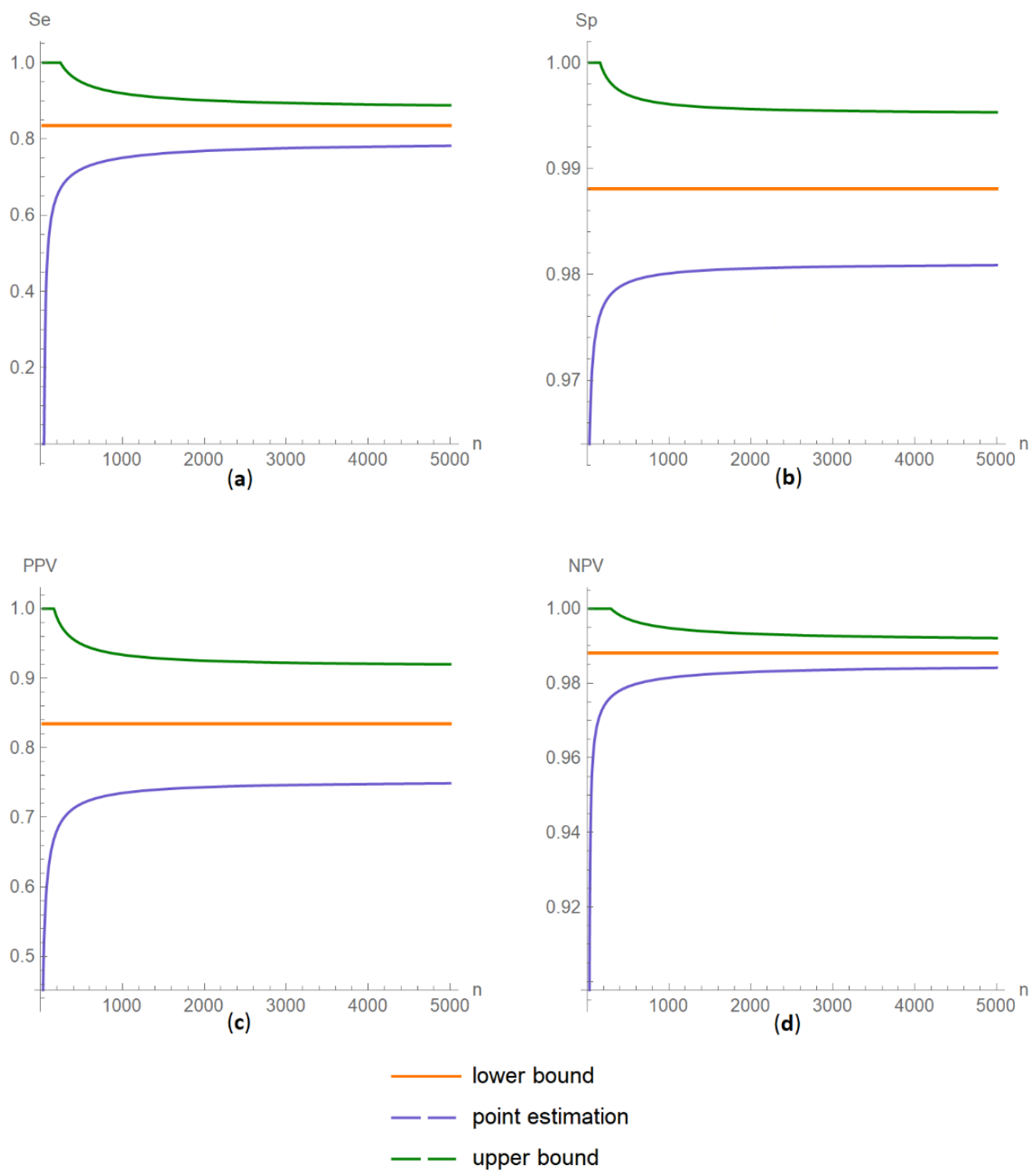**Figure 21.** DAM relative standard uncertainties versus population sample size plots. Plots of relative standard combined, measurement and sampling uncertainties of (**a**) sensitivity ($u(Se)/Se$), (**b**) specificity ($u(Sp)/Sp$), (**c**) positive predictive value ($u(PPV)/PPV$) and (**d**) negative predictive value ($u(NPV)/NPV$) versus total population sample size (n) curves, with the respective parameter settings in Table 4.
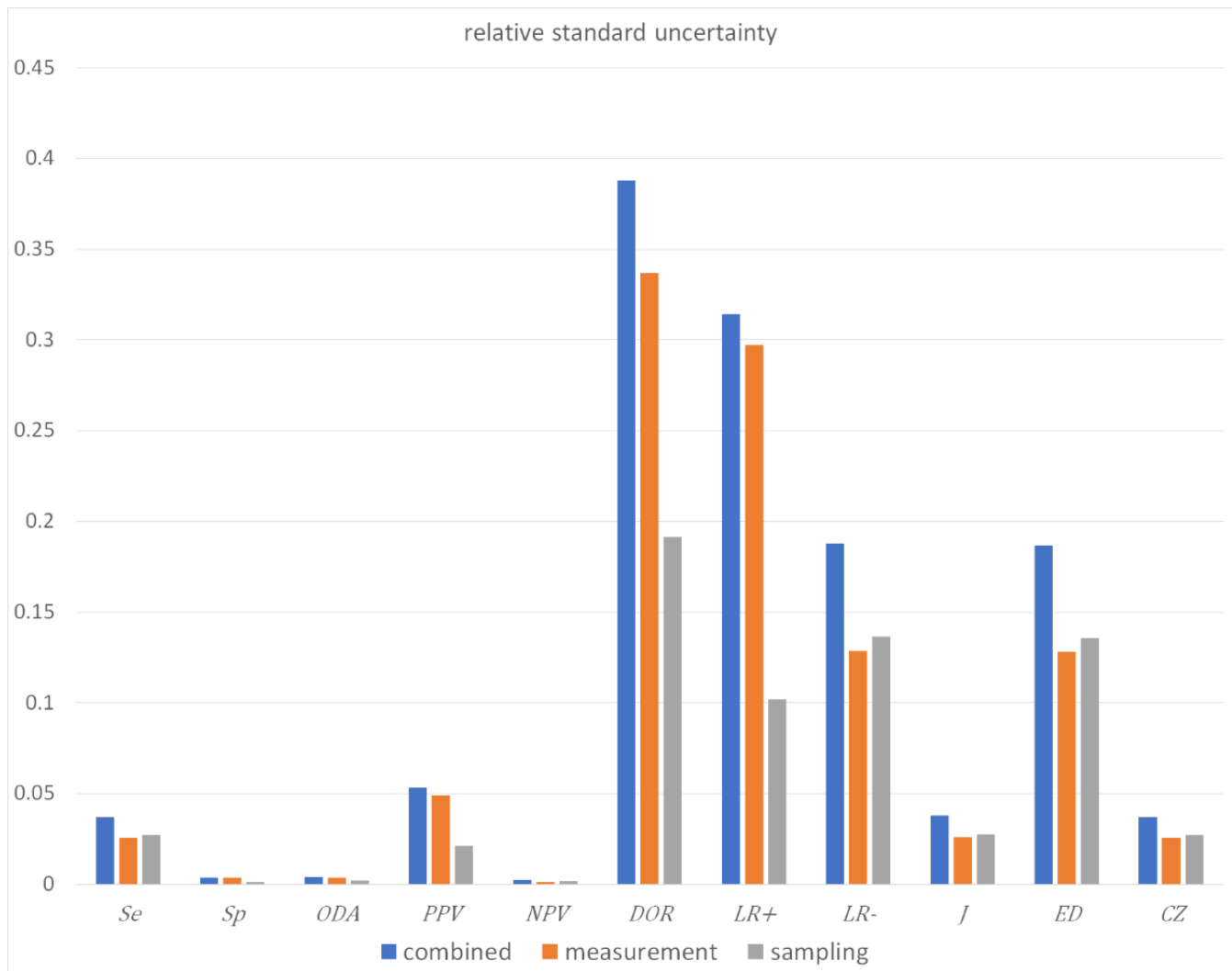
Figure 22. DAM confidence intervals versus population sample size plots. Plots of confidence intervals of (**a**) sensitivity (*Se*), (**b**) specificity (*Sp*), (**c**) positive predictive value (*PPV*) and (**d**) negative predictive value (*NPV*) versus total population sample size (*n*) curves, with the respective parameter settings in Table 4.

4. Discussion

The program *Diagnostic Uncertainty* explores the combined, measurement and sampling uncertainty of diagnostic accuracy measures of a screening or diagnostic test (Figures 3, 4, 6, 7 and 9–13) and the resultant confidence intervals (Figures 5, 8, 11 and 14). Combined uncertainty and the resultant confidence intervals depend on the diagnostic threshold (Figures 3–5 and 15–17), on measurement uncertainty (Figures 6–8, 18 and 19) and on population parameters, including the total population sample size (Figures 9–11, 20 and 21).

The complexity of the calculations of the confidence intervals of the diagnostic accuracy measures is considerable. In antithesis of 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 of diagnostic accuracy measures and the resultant confidence intervals (Figures 12–14).

As demonstrated by the illustrative example described above, 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 (Figure 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:

a. 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 [39].

b. The normality of either the measurements or their applicable transforms [23,24,40,41], 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 [42–46].

c. The bimodality of the measurands, that is generally accepted, although unimodal distributions could be considered [47,48].

d. The simple random sampling.

e. 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 [49].

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 [28,30]. Higher order approximations may improve the accuracy.

(3) The uncertainty of prevalence rate approximation by the Agresti–Coull adjusted Waldo interval [25], although there are more exact methods [50].

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

The program presented in this work complements our previously published software [11], which explores the effects of measurement uncertainty on diagnostic accuracy measures applied to populations. This program calculates the standard and expanded combined, measurement and sampling uncertainty and the resultant confidence intervals of diagnostic accuracy measures of diagnostic tests, applied to samples of populations, providing 99 different types of plots and three different types of comprehensive tables (Figure 2), many of which are novel. To the best of our knowledge, no software, including all major general or medical statistical and uncertainty related software packages (Matlab®, NCSS®, R, SAS®, SPSS®, Stata®, MedCalc®, NIST Uncertainty Machine, UQLab, metRology), provides this range of plots and tables without advanced programming.

5. Conclusions

The presented program *Diagnostic Uncertainty* calculates the combined, measurement and sampling uncertainty of diagnostic accuracy measures and the resultant confidence intervals and can be used as a flexible, user-friendly, interactive educational or research tool in medical decision-making.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in [35].

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

Appendix A

Notation

1. Populations

    $\overline{D}$: Non-diseased population
    D: Diseased population

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

3. Parameters

    m_P: Mean of the measurand of a test in a sample of population P
    s_P: Standard deviation of the measurand of a test in a sample of population P
    n_P: Size of a sample of population P
    n: Size of a sample of total population
    n_u: Size of a measurements sample
    r: Prevalence rate of the disease
    d: Diagnostic threshold of a test
    u_m : Standard measurement uncertainty of a test
    p: Confidence level
    v: Degrees of freedom
    v_{eff} : Effective degrees of freedom

4. Diagnostic accuracy measures: Abbreviations

    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 receiver operating characteristic curve point from the point (0,1)
    CZ: Concordance probability

5. Diagnostic accuracy measures: Functions

    $se(\ldots)$: Sensitivity of a test
    $sp(\ldots)$: Specificity of a test
    $oda(\ldots)$: Overall diagnostic accuracy of a test
    $ppv(\ldots)$: Positive predictive value of a test
    $npv(\ldots)$: Negative predictive value of a test
    $plr(\ldots)$: Likelihood ratio for a positive test result
    $nlr(\ldots)$: Likelihood ratio for a negative test result
    $dor(\ldots)$: Diagnostic odds ratio of a test
    $ed(\ldots)$: Euclidean distance of a test

$j(\ldots)$: Youden's index of a test

$cz(\ldots)$: Concordance probability of a test

6. Other functions and relations

$u(x)$: Standard uncertainty of x

$u_s(x)$: 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 ith component of the standard combined uncertainty of x

$u_c(dam(\ldots),\ldots)$: Standard combined uncertainty of the diagnostic accuracy measure $dam(\ldots)$

$\Phi(x)$: Cumulative distribution function of the standard normal distribution evaluated at x

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

$F_v(x)$: Cumulative distribution function of the Student's t-distribution with v degrees of freedom, 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

Appendix B

Appendix B.1. Uncertainty Propagation Rules

$$u(x+y) = \sqrt{u(x)^2 + u(y)^2}$$

$$u(x-y) = \sqrt{u(x)^2 + u(y)^2}$$

$$u(\alpha + x) = u(x)$$

$$u(\alpha x) = \alpha u(x)$$

$$u(x^\alpha) = |\alpha| x^{\alpha-1} u(x)$$

$$u(x\,y) = x\,y \sqrt{\left(\frac{u(x)}{x}\right)^2 + \left(\frac{u(y)}{y}\right)^2}$$

$$u\left(\frac{x}{y}\right) = \frac{x}{y} \sqrt{\left(\frac{u(x)}{x}\right)^2 + \left(\frac{u(y)}{y}\right)^2}$$

$$H(x) = \int h(x)dt \Rightarrow u\left(\int_\alpha^\beta h(t)dt\right) = u(H(\beta) - H(\alpha)) = \sqrt{u(H(\beta))^2 + u(H(\alpha))^2}$$

Appendix B.2. Definitions and Calculations

Appendix B.2.1. Error Function

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

Appendix B.2.2. Complementary Error Function

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

Appendix B.2.3. Standard Normal Distribution Cumulative Density Function

$$\Phi(x) = \frac{1}{2\sqrt{\pi}} \int_{-\infty}^x e^{-t^2} dt = \frac{1}{2} \left(1 + erf\left(\frac{x}{\sqrt{2}}\right) \right) = 1 - \frac{1}{2} erfc\left(\frac{x}{\sqrt{2}}\right)$$

Appendix B.2.4. Normal Distribution Cumulative Density Function

$$\Psi(x, m_P, s_P) = \Phi\left(\frac{x - m_P}{s_P}\right) = \tfrac{1}{2}\left(1 + erf\left(\frac{x - m_P}{\sqrt{2}s_P}\right)\right)$$
$$= 1 - \tfrac{1}{2} erfc\left(\frac{x - m_P}{\sqrt{2}s_P}\right)$$

Appendix B.2.5. Prevalence Rate (*r*)

$$r = \frac{n_D}{n_{\overline{D}} + n_D}$$

Appendix B.2.6. Sensitivity (*Se*)

2.6.1. Measure

$$se(d, m_D, s_D) = 1 - \Psi(x, m_D, s_D) = \frac{1}{2} erfc\left(\frac{m_D - d}{\sqrt{2}s_D}\right)$$

2.6.2. Standard Combined Uncertainty

$$u_c(se(d|m_D, s_D)|u, n_D) = \sqrt{\frac{e^{-\frac{(m_D - d)^2}{s_D^2}}\left((m_D - d)^2\left(\frac{s_D^2}{2(n_D-1)} + u^2\right) + s_D^2\left(\frac{s_D^2}{n_D} + u^2\right)\right)}{2\pi \, s_D^4}}$$

Appendix B.2.7. Specificity (*Sp*)

2.7.1. Measure

$$sp(d, m_{\overline{D}}, s_{\overline{D}}) = \Psi(x, m_{\overline{D}}, s_{\overline{D}}) = 1 - \frac{1}{2} erfc\left(\frac{m_{\overline{D}} - d}{\sqrt{2}s_{\overline{D}}}\right)$$

2.7.2. Standard Combined Uncertainty

$$u_c\left(sp(d, m_{\overline{D}}, s_{\overline{D}}), u, n_{\overline{D}}\right) = \sqrt{\frac{e^{-\frac{(m_{\overline{D}} - d)^2}{s_{\overline{D}}^2}}\left((m_{\overline{D}} - d)^2\left(\frac{s_{\overline{D}}^2}{2(n_{\overline{D}}-1)} + u^2\right) + s_{\overline{D}}^2\left(\frac{s_{\overline{D}}^2}{n_{\overline{D}}} + u^2\right)\right)}{2\pi \, s_{\overline{D}}^4}}$$

Appendix B.2.8. Overall Diagnostic Accuracy (*ODA*)

2.8.1. Measure

$$oda\left(d, m_D, m_{\overline{D}}, s_D, s_{\overline{D}}, n_D, n_{\overline{D}}\right) = se(v, m_D, s_D)\, r + sp(v, m_{\overline{D}}, s_{\overline{D}})\,(1 - r) =$$
$$\frac{n_{\overline{D}} erfc\left(\frac{m_{\overline{D}} - d}{\sqrt{2}s_{\overline{D}}}\right) + n_D erf\left(\frac{m_D - d}{\sqrt{2}s_D}\right) + n_D}{2\left(n_{\overline{D}} + n_D\right)}$$

2.8.2. Standard Combined Uncertainty

$$u_c\big(oda\big(d, m_D, m_{\overline{D}}, s_D, s_{\overline{D}}, n_D, n_{\overline{D}}\big), u\big) = \sqrt{\frac{A_1 + A_2 + A_3 + A_4 + A_5}{\big(n_{\overline{D}} + n_D\big)^2}}$$

where:

$$A_1 = \frac{(n_D + 2)\big(n_{\overline{D}} + 2\big)\big(n_{\overline{D}} + n_D\big)\left(erfc\left(\frac{m_{\overline{D}} - d}{\sqrt{2}s_{\overline{D}}}\right) + erfc\left(\frac{m_D - d}{\sqrt{2}s_D}\right) - 2\right)^2}{\big(n_{\overline{D}} + n_D + 4\big)^2}$$

$$A_2 = \frac{2n_{\overline{D}}\, e^{-\frac{(m_{\overline{D}} - d)^2}{s_{\overline{D}}^2}}\left(u^2 n_{\overline{D}} + s_{\overline{D}}^2\right)}{\pi s_{\overline{D}}^2}$$

$$A_3 = \frac{2n_{\overline{D}}^2\big(m_{\overline{D}} - d\big)^2 e^{-\frac{(m_{\overline{D}} - t)^2}{s_{\overline{D}}^2}}\left(\frac{s_{\overline{D}}^2}{2\big(n_{\overline{D}} - 1\big)} + u^2\right)}{\pi s_{\overline{D}}^4}$$

$$A_4 = \frac{2n_D\, e^{-\frac{(m_D - d)^2}{s_D^2}}\left(u^2 n_D + s_D^2\right)}{\pi s_D^2}$$

$$A_5 = \frac{2n_D^2(m_D - d)^2 e^{-\frac{(m_D - d)^2}{s_D^2}}\left(\frac{s_D^2}{2(n_D - 1)} + u^2\right)}{\pi s_D^4}$$

Appendix B.2.9. Positive Predictive Value (*PPV*)

2.9.1. Measure

$$dor\big(d, m_D, m_{\overline{D}}, s_D, s_{\overline{D}}\big) = \frac{\frac{se(d, m_D, s_D)}{1 - se(d, m_D, s_D)}}{\frac{1 - sp\big(d, m_{\overline{D}}, s_{\overline{D}}\big)}{sp\big(d, m_{\overline{D}}, s_{\overline{D}}\big)}} =$$

$$\frac{\left(erfc\left(\frac{m_D - d}{\sqrt{2}s_D}\right) - 2\right)erfc\left(\frac{m_{\overline{D}} - d}{\sqrt{2}s_{\overline{D}}}\right)}{erfc\left(\frac{m_D - d}{\sqrt{2}s_D}\right)\left(erfc\left(\frac{m_{\overline{D}} - d}{\sqrt{2}s_{\overline{D}}}\right) - 2\right)}$$

2.9.2. Standard Combined Uncertainty

$$u_c\big(ppv\big(d, m_D, m_{\overline{D}}, s_D, s_{\overline{D}}, n_D, n_{\overline{D}}\big), u\big) = \sqrt{\frac{B_1 + B_2 + B_3 + B_4 + B_5}{\pi\, B_6}}$$

where:

$$B_1 = (n_D - 1)n_D^2 s_D^4 n_{\overline{D}}^2\big(n_{\overline{D}} + n_D + 4\big)^2\big(m_{\overline{D}} - d\big)^2 e^{-\frac{(m_{\overline{D}} - d)^2}{s_{\overline{D}}^2}}\left(2u^2\big(n_{\overline{D}} - 1\big) + s_{\overline{D}}^2\right)\left(erfc\left(\frac{m_D - d}{\sqrt{2}s_D}\right) - 2\right)^2$$

$$B_2 = 2(n_D - 1)n_D^2 s_D^4\big(n_{\overline{D}} - 1\big)n_{\overline{D}}\big(n_{\overline{D}} + n_D + 4\big)^2 s_{\overline{D}}^2 e^{-\frac{(m_{\overline{D}} - d)^2}{s_{\overline{D}}^2}}\left(u^2 n_{\overline{D}} + s_{\overline{D}}^2\right)\left(erfc\left(\frac{m_D - d}{\sqrt{2}s_D}\right) - 2\right)^2$$

$$B_3 = 2(n_D - 1)n_D s_D^2\big(n_{\overline{D}} - 1\big)n_{\overline{D}}^2\big(n_{\overline{D}} + n_D + 4\big)^2 s_{\overline{D}}^4 e^{-\frac{(m_D - d)^2}{s_D^2}}\left(u^2 n_D + s_D^2\right)\left(erfc\left(\frac{m_{\overline{D}} - d}{\sqrt{2}s_{\overline{D}}}\right) - 2\right)^2$$

$$B_4 = n_D^2(n_{\overline{D}}-1)n_{\overline{D}}^2(n_{\overline{D}}+n_D+4)^2 s_{\overline{D}}^4 (m_D-d)^2 e^{-\frac{(m_D-d)^2}{s_D^2}}\left(2u^2(n_D-1)+s_D^2\right)\left(erfc\left(\frac{m_{\overline{D}}-d}{\sqrt{2}s_{\overline{D}}}\right)-2\right)^2$$

$$B_5 = \pi(n_D-1)(n_D+2)s_D^4(n_{\overline{D}}-1)(n_{\overline{D}}+2)(n_{\overline{D}}+n_D)^3 s_{\overline{D}}^4\left(erfc\left(\frac{m_D-d}{\sqrt{2}s_D}\right)-2\right)^2\left(erfc\left(\frac{m_{\overline{D}}-d}{\sqrt{2}s_{\overline{D}}}\right)-2\right)^2$$

$$B_6 = (n_D-1)s_D^4(n_{\overline{D}}-1)(n_{\overline{D}}+n_D+4)^2 s_{\overline{D}}^4\left(n_{\overline{D}}erfc\left(\frac{m_{\overline{D}}-d}{\sqrt{2}s_{\overline{D}}}\right)-2(n_{\overline{D}}+n_D)+n_D erfc\left(\frac{m_D-d}{\sqrt{2}s_D}\right)\right)^4$$

Appendix B.2.10. Negative Predictive Value (*NPV*)

2.10.1. Measure

$$npv\left(d,m_D,m_{\overline{D}},s_D,s_{\overline{D}},n_D,n_{\overline{D}}\right) \quad = \frac{sp\left(d,m_{\overline{D}},s_{\overline{D}}\right)\,(1-r)}{sp\left(d,m_{\overline{D}},s_{\overline{D}}\right)\,(1-r)+(1-se(d,m_D,s_D))r} \quad = \frac{1}{\frac{n_D erfc\left(\frac{m_D-d}{\sqrt{2}s_D}\right)}{n_{\overline{D}}erfc\left(\frac{m_{\overline{D}}-d}{\sqrt{2}s_{\overline{D}}}\right)}+1}$$

2.10.2. Standard Combined Uncertainty

$$u_c\left(npv\left(d,m_D,m_{\overline{D}},s_D,s_{\overline{D}},n_D,n_{\overline{D}}\right),u\right) = \sqrt{\frac{C_1+C_2+C_3+C_4+C_5}{\left(n_{\overline{D}}erfc\left(\frac{m_{\overline{D}}-d}{\sqrt{2}s_{\overline{D}}}\right)+n_D erfc\left(\frac{m_D-d}{\sqrt{2}s_D}\right)\right)^4}}$$

where:

$$C_1 = \frac{2n_D n_{\overline{D}}^2 e^{-\frac{(m_D-d)^2}{s_D^2}}\left(u^2 n_D+s_D^2\right)erfc\left(\frac{m_{\overline{D}}-d}{\sqrt{2}s_{\overline{D}}}\right)^2}{\pi s_D^2}$$

$$C_2 = \frac{n_D^2 n_{\overline{D}}^2(m_D-d)^2 e^{-\frac{(m_D-d)^2}{s_D^2}}\left(2u^2(n_D-1)+s_D^2\right)erfc\left(\frac{m_{\overline{D}}-d}{\sqrt{2}s_{\overline{D}}}\right)^2}{\pi(n_D-1)s_D^4}$$

$$C_3 = \frac{2n_D^2 n_{\overline{D}} e^{-\frac{(m_{\overline{D}}-d)^2}{s_{\overline{D}}^2}}\left(u^2 n_{\overline{D}}+s_{\overline{D}}^2\right)erfc\left(\frac{m_D-d}{\sqrt{2}s_D}\right)^2}{\pi s_{\overline{D}}^2}$$

$$C_4 = \frac{n_D^2 n_{\overline{D}}^2(m_{\overline{D}}-d)^2 e^{-\frac{(m_{\overline{D}}-d)^2}{s_{\overline{D}}^2}}\left(2u^2(n_{\overline{D}}-1)+s_{\overline{D}}^2\right)erfc\left(\frac{m_D-d}{\sqrt{2}s_D}\right)^2}{\pi(n_{\overline{D}}-1)s_{\overline{D}}^4}$$

$$C_5 = \frac{(n_D+2)(n_{\overline{D}}+2)(n_{\overline{D}}+n_D)^3 erfc\left(\frac{m_D-d}{\sqrt{2}s_D}\right)^2 erfc\left(\frac{m_{\overline{D}}-d}{\sqrt{2}s_{\overline{D}}}\right)^2}{(n_{\overline{D}}+n_D+4)^2}$$

Appendix B.2.11. Diagnostic Odds Ratio (*DOR*)

2.11.1. Measure

$$dor\big(d, m_D, m_{\overline{D}}, s_D, s_{\overline{D}}\big) = \frac{\frac{se(d,m_D,s_D)}{1-se(d,m_D,s_D)}}{\frac{1-sp(d,m_{\overline{D}},s_{\overline{D}})}{sp(d,m_{\overline{D}},s_{\overline{D}})}}$$

$$= \frac{\left(erfc\left(\frac{m_D-d}{\sqrt{2}s_D}\right)-2\right)erfc\left(\frac{m_{\overline{D}}-d}{\sqrt{2}s_{\overline{D}}}\right)}{erfc\left(\frac{m_D-d}{\sqrt{2}s_D}\right)\left(erfc\left(\frac{m_{\overline{D}}-d}{\sqrt{2}s_{\overline{D}}}\right)-2\right)}$$

2.11.2. Standard Combined Uncertainty

$$u_c\big(dor\big(d, m_D, m_{\overline{D}}, s_D, s_{\overline{D}}\big), n_D, n_{\overline{D}}, u\big) = 2\sqrt{\frac{D_1 + D_2 + D_3 + D_4}{\pi\,D_5}}$$

where:

$$D_1 = 2(n_D-1)n_D s_D^4(n_{\overline{D}}-1)s_{\overline{D}}^2 e^{-\frac{(m_{\overline{D}}-d)^2}{s_{\overline{D}}^2}}\left(u^2 n_{\overline{D}} + s_{\overline{D}}^2\right)\left(erfc\left(\frac{m_D-d}{\sqrt{2}s_D}\right)-2\right)^2 erfc\left(\frac{m_D-d}{\sqrt{2}s_D}\right)^2$$

$$D_2 = (n_D-1)n_D s_D^4 n_{\overline{D}}(m_{\overline{D}}-d)^2 e^{-\frac{(m_{\overline{D}}-d)^2}{s_{\overline{D}}^2}}\left(2u^2(n_{\overline{D}}-1) + s_{\overline{D}}^2\right)\left(erfc\left(\frac{m_D-d}{\sqrt{2}s_D}\right)-2\right)^2 erfc\left(\frac{m_D-d}{\sqrt{2}s_D}\right)^2$$

$$D_3 = 2(n_D-1)s_D^2(n_{\overline{D}}-1)n_{\overline{D}}s_{\overline{D}}^4 e^{-\frac{(m_D-d)^2}{s_D^2}}\left(u^2 n_D + s_D^2\right)\left(erfc\left(\frac{m_{\overline{D}}-d}{\sqrt{2}s_{\overline{D}}}\right)-2\right)^2 erfc\left(\frac{m_{\overline{D}}-d}{\sqrt{2}s_{\overline{D}}}\right)^2$$

$$D_4 = n_D(n_{\overline{D}}-1)n_{\overline{D}}s_D^4(m_D-d)^2 e^{-\frac{(m_D-d)^2}{s_D^2}}\left(2u^2(n_D-1) + s_D^2\right)\left(erfc\left(\frac{m_{\overline{D}}-d}{\sqrt{2}s_{\overline{D}}}\right)-2\right)^2 erfc\left(\frac{m_{\overline{D}}-d}{\sqrt{2}s_{\overline{D}}}\right)^2$$

$$D_5 = (n_D-1)n_D s_D^4(n_{\overline{D}}-1)n_{\overline{D}}s_{\overline{D}}^4 erfc\left(\frac{m_D-d}{\sqrt{2}s_D}\right)^4\left(erfc\left(\frac{m_{\overline{D}}-d}{\sqrt{2}s_{\overline{D}}}\right)-2\right)^4$$

Appendix B.2.12. Likelihood Ratio for a Positive Result (*LR+*)

2.12.1. Measure

$$plr\big(d, m_D, m_{\overline{D}}, s_D, s_{\overline{D}}\big) = \frac{se(d,m_D,s_D)}{1-sp(d,m_{\overline{D}},s_{\overline{D}})}$$

$$= \frac{erf\left(\frac{m_D-d}{\sqrt{2}s_D}\right)+1}{erf\left(\frac{m_{\overline{D}}-d}{\sqrt{2}s_{\overline{D}}}\right)+1}$$

2.12.2. Standard Combined Uncertainty

$$u_c\big(plr\big(d, m_D, m_{\overline{D}}, s_D, s_{\overline{D}}\big), u, n_D, n_{\overline{D}}\big) = \sqrt{\frac{2(E_1 + E_2 + E_3 + E_4)}{\pi\left(erf\left(\frac{m_{\overline{D}}-d}{\sqrt{2}s_{\overline{D}}}\right)+1\right)^4}}$$

where:

$$E_1 = \frac{e^{-\frac{(m_D-d)^2}{s_D^2}}\left(\frac{s_D^2}{n_D}+u^2\right)\left(erf\left(\frac{m_{\overline{D}}-d}{\sqrt{2}s_{\overline{D}}}\right)+1\right)^2}{s_D^2}$$

$$E_2 = \frac{(m_D-d)^2 e^{-\frac{(m_D-d)^2}{s_D^2}}\left(\frac{s_D^2}{2(n_D-1)}+u^2\right)\left(erf\left(\frac{m_{\overline{D}}-d}{\sqrt{2}s_{\overline{D}}}\right)+1\right)^2}{s_D^4}$$

$$E_3 = \frac{e^{-\frac{(m_{\overline{D}}-d)^2}{s_{\overline{D}}^2}}\left(\frac{s_{\overline{D}}^2}{n_{\overline{D}}}+u^2\right)\left(erf\left(\frac{m_D-d}{\sqrt{2}s_D}\right)+1\right)^2}{s_{\overline{D}}^2}$$

$$E_4 = \frac{(m_{\overline{D}}-t)^2 e^{-\frac{(m_{\overline{D}}-d)^2}{s_{\overline{D}}^2}}\left(\frac{s_{\overline{D}}^2}{2(n_{\overline{D}}-1)}+u^2\right)\left(erf\left(\frac{m_D-d}{\sqrt{2}s_D}\right)+1\right)^2}{s_{\overline{D}}^4}$$

Appendix B.2.13. Likelihood Ratio for a Negative Result (*LR*−)

2.13.1. Measure

$$nlr\left(d, m_D, m_{\overline{D}}, s_D, s_{\overline{D}}\right) = \frac{1-se(d,m_D,s_D)}{sp\left(d,m_{\overline{D}},s_{\overline{D}}\right)}$$

$$= \frac{erfc\left(\frac{m_D-d}{\sqrt{2}s_D}\right)}{erfc\left(\frac{m_{\overline{D}}-d}{\sqrt{2}s_{\overline{D}}}\right)}$$

2.13.2. Standard Combined Uncertainty

$$u_c\left(nlr\left(d, m_D, m_{\overline{D}}, s_D, s_{\overline{D}}\right), u, n_D, n_{\overline{D}}\right) = \sqrt{\frac{2(F_1 + F_2 + F_3 + F_4)}{\pi \, erfc\left(\frac{m_{\overline{D}}-d}{\sqrt{2}s_{\overline{D}}}\right)^4}}$$

where:

$$F_1 = \frac{e^{-\frac{(m_D-d)^2}{s_D^2}}\left(\frac{s_D^2}{n_D}+u^2\right)erfc\left(\frac{m_{\overline{D}}-d}{\sqrt{2}s_{\overline{D}}}\right)^2}{s_D^2}$$

$$F_2 = \frac{(m_D-d)^2 e^{-\frac{(m_D-d)^2}{s_D^2}}\left(\frac{s_D^2}{2(n_D-1)}+u^2\right)erfc\left(\frac{m_{\overline{D}}-d}{\sqrt{2}s_{\overline{D}}}\right)^2}{s_D^4}$$

$$F_3 = \frac{e^{-\frac{(m_{\overline{D}}-d)^2}{s_{\overline{D}}^2}}\left(\frac{s_{\overline{D}}^2}{n_{\overline{D}}}+u^2\right)erfc\left(\frac{m_D-d}{\sqrt{2}s_D}\right)^2}{s_{\overline{D}}^2}$$

$$F_4 = \frac{(m_{\overline{D}}-d)^2 e^{-\frac{(m_{\overline{D}}-d)^2}{s_{\overline{D}}^2}}\left(\frac{s_{\overline{D}}^2}{2(n_{\overline{D}}-1)}+u^2\right)erfc\left(\frac{m_D-d}{\sqrt{2}s_D}\right)^2}{s_{\overline{D}}^4}$$

Appendix B.2.14. Yuden's Index (*J*)

2.14.1. Measure

$$j\left(d, m_D, m_{\overline{D}}, s_D, s_{\overline{D}}\right) = se(d, m_D, s_D) + sp\left(d, m_{\overline{D}}, s_{\overline{D}}\right) - 1$$

$$= \frac{1}{2}\left(erf\left(\frac{m_D-d}{\sqrt{2}s_D}\right) - erf\left(\frac{m_{\overline{D}}-d}{\sqrt{2}s_{\overline{D}}}\right)\right)$$

2.14.2. Standard Combined Uncertainty

$$u_c\left(j\left(d, m_D, m_{\overline{D}}, s_D, s_{\overline{D}}\right), u, n_D, n_{\overline{D}}\right) = \frac{\sqrt{G_1 + G_2 + G_3 + G_4}}{\sqrt{2\pi}}$$

where:

$$G_1 = \frac{e^{-\frac{(m_{\overline{D}}-d)^2}{s_{\overline{D}}^2}}\left(\frac{s_{\overline{D}}^2}{n_{\overline{D}}}+u^2\right)}{s_{\overline{D}}^2}$$

$$G_2 = \frac{(m_{\overline{D}}-d)^2 e^{-\frac{(m_{\overline{D}}-d)^2}{s_{\overline{D}}^2}}\left(\frac{s_{\overline{D}}^2}{2(n_{\overline{D}}-1)}+u^2\right)}{s_{\overline{D}}^4}$$

$$G_3 = \frac{e^{-\frac{(m_D-d)^2}{s_D^2}}\left(\frac{s_D^2}{n_D}+u^2\right)}{s_D^2}$$

$$G_4 = \frac{(m_D-d)^2 e^{-\frac{(m_D-d)^2}{s_D^2}}\left(\frac{s_D^2}{2(n_D-1)}+u^2\right)}{s_D^4}$$

Appendix B.2.15. Euclidean Distance (*ED*)

2.15.1. Measure

$$mced\big(d, m_D, m_{\overline{D}}, s_D, s_{\overline{D}}\big) = \sqrt{(1-se(d, m_D, s_D))^2 + \big(1-sp(d, m_{\overline{D}}, s_{\overline{D}})\big)^2}$$

$$= \tfrac{1}{2}\sqrt{\left(erf\left(\frac{m_{\overline{D}}-d}{\sqrt{2}s_{\overline{D}}}\right)+1\right)^2 + erfc\left(\frac{m_D-d}{\sqrt{2}s_D}\right)^2}$$

2.15.2. Standard Combined Uncertainty

$$u_c\big(ed\big(d, m_D, m_{\overline{D}}, s_D, s_{\overline{D}}\big), u, n_D, n_{\overline{D}}\big) = \sqrt{\frac{H_1 + H_2 + H_3 + H_4}{2\pi\,H_5}}$$

where:

$$H_1 = \frac{e^{-\frac{(m_{\overline{D}}-d)^2}{s_{\overline{D}}^2}}\left(\frac{s_{\overline{D}}^2}{n_{\overline{D}}}+u^2\right)\left(erf\left(\frac{m_{\overline{D}}-d}{\sqrt{2}s_{\overline{D}}}\right)+1\right)^2}{s_{\overline{D}}^2}$$

$$H_2 = \frac{(m_{\overline{D}}-d)^2 e^{-\frac{(m_{\overline{D}}-d)^2}{s_{\overline{D}}^2}}\left(\frac{s_{\overline{D}}^2}{2(n_{\overline{D}}-1)}+u^2\right)\left(erf\left(\frac{m_{\overline{D}}-d}{\sqrt{2}s_{\overline{D}}}\right)+1\right)^2}{s_{\overline{D}}^4}$$

$$H_3 = \frac{e^{-\frac{(m_D-d)^2}{s_D^2}}\left(\frac{s_D^2}{n_D}+u^2\right)erfc\left(\frac{m_D-d}{\sqrt{2}s_D}\right)^2}{s_D^2}$$

$$H_4 = \frac{(m_D-d)^2 e^{-\frac{(m_D-d)^2}{s_D^2}}\left(\frac{s_D^2}{2(n_D-1)}+u^2\right)erfc\left(\frac{m_D-d}{\sqrt{2}s_D}\right)^2}{s_D^4}$$

$$H_5 = \left(erf\left(\frac{m_{\overline{D}}-d}{\sqrt{2}s_{\overline{D}}}\right)+1\right)^2 + erfc\left(\frac{m_D-d}{\sqrt{2}s_D}\right)^2$$

Appendix B.2.16. Concordance Probability (*CZ*)

2.16.1. Measure

$$cz\left(d, m_D, m_{\overline{D}}, s_D, s_{\overline{D}}\right) = se\left(d, m_D, s_D\right) sp\left(d, m_{\overline{D}}, s_{\overline{D}}\right) =$$
$$-\frac{1}{4}\left(erfc\left(\frac{m_D - d}{\sqrt{2}s_D}\right) - 2\right)erfc\left(\frac{m_{\overline{D}} - d}{\sqrt{2}s_{\overline{D}}}\right)$$

2.16.2. Standard Combined Uncertainty

$$u_c\left(cz\left(d, m_D, m_{\overline{D}}, s_D, s_{\overline{D}}\right), u, n_D, n_{\overline{D}}\right) = \frac{1}{2}\sqrt{\frac{I_1 + I_2 + I_3 + I_4}{2\pi}}$$

where:

$$I_1 = \frac{e^{-\frac{(m_D - d)^2}{s_D^2}}\left(\frac{s_D^2}{n_D} + u^2\right)erfc\left(\frac{m_{\overline{D}} - d}{\sqrt{2}s_{\overline{D}}}\right)^2}{s_D^2}$$

$$I_2 = \frac{(m_D - d)^2 e^{-\frac{(m_D - d)^2}{s_D^2}}\left(\frac{s_D^2}{2(n_D - 1)} + u^2\right)erfc\left(\frac{m_{\overline{D}} - d}{\sqrt{2}s_{\overline{D}}}\right)^2}{s_D^4}$$

$$I_3 = \frac{e^{-\frac{(m_{\overline{D}} - d)^2}{s_{\overline{D}}^2}}\left(\frac{s_{\overline{D}}^2}{n_{\overline{D}}} + u^2\right)\left(erfc\left(\frac{m_D - d}{\sqrt{2}s_D}\right) - 2\right)^2}{s_{\overline{D}}^2}$$

$$I_4 = \frac{(m_{\overline{D}} - v)^2 e^{-\frac{(m_{\overline{D}} - d)^2}{s_{\overline{D}}^2}}\left(\frac{s_{\overline{D}}^2}{2(n_{\overline{D}} - 1)} + u^2\right)\left(erfc\left(\frac{m_D - d}{\sqrt{2}s_D}\right) - 2\right)^2}{s_{\overline{D}}^4}$$

Appendix C

Software Availability and Requirements

Program name: *Diagnostic Uncertainty*
Project home page: https://www.hcsl.com/Tools/Uncertainty/ (accessed 24 February 2021)
Operating systems: Microsoft Windows, Linux, Apple iOS
Programming language: Wolfram Language
Other software requirements: Wolfram Player®, freely available at: https://www.wolfram.com/player/ (accessed 12 February 2021) or Wolfram Mathematica®
System requirements: Intel® i7™ or equivalent CPU and 16 GB of RAM
License: Attribution—Noncommercial—ShareAlike 4.0 International Creative Commons License

References

1. Zou, K.H.; O'Malley, A.J.; Mauri, L. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation* **2007**, *115*, 654–657. [CrossRef] [PubMed]
2. Šimundić, A.-M. Measures of Diagnostic Accuracy: Basic Definitions. *EJIFCC* **2009**, *19*, 203–211.
3. 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**. [CrossRef]
4. Tang, Y.-W.; Schmitz, J.E.; Persing, D.H.; Stratton, C.W. The Laboratory Diagnosis of COVID-19 Infection: Current Issues and Challenges. *J. Clin. Microbiol.* **2020**. [CrossRef] [PubMed]
5. Deeks, J.J.; Dinnes, J.; Takwoingi, Y.; Davenport, C.; Leeflang, M.M.G.; Spijker, R.; Hooft, L.; Van den Bruel, A.; Emperador, D.; Dittrich, S. 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. [CrossRef]
6. Axell-House, D.B.; Lavingia, R.; Rafferty, M.; Clark, E.; Amirian, E.S.; Chiao, E.Y. The estimation of diagnostic accuracy of tests for COVID-19: A scoping review. *J. Infect.* **2020**, *81*, 681–697. [CrossRef] [PubMed]

7. Lisboa Bastos, M.; Tavaziva, G.; Abidi, S.K.; Campbell, J.R.; Haraoui, L.-P.; Johnston, J.C.; Lan, Z.; Law, S.; MacLean, E.; Trajman, A.; et al. Diagnostic accuracy of serological tests for covid-19: Systematic review and meta-analysis. *BMJ* **2020**, *370*, m2516. [CrossRef]

8. Smith, A.F.; Shinkins, B.; Hall, P.S.; Hulme, C.T.; Messenger, M.P. 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*, 1363–1374. [CrossRef] [PubMed]

9. Theodorsson, E. Uncertainty in Measurement and Total Error: Tools for Coping with Diagnostic Uncertainty. *Clin. Lab. Med.* **2017**, *37*, 15–34. [CrossRef]

10. 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*, 41–47. [CrossRef] [PubMed]

11. Chatzimichail, T.; Hatjimihail, A.T. A Software Tool for Exploring the Relation between Diagnostic Accuracy and Measurement Uncertainty. *Diagnostics* **2020**, *10*, 610. [CrossRef] [PubMed]

12. Owen, R.K.; Cooper, N.J.; Quinn, T.J.; Lees, R.; Sutton, A.J. 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. [CrossRef] [PubMed]

13. Haeckel, R.; Wosniok, W.; Gurr, E.; Peil, B. Supplements to a recent proposal for permissible uncertainty of measurements in laboratory medicine. *J. Lab. Med.* **2016**. [CrossRef]

14. Hajian-Tilaki, K. Sample size estimation in diagnostic test studies of biomedical informatics. *J. Biomed. Inform.* **2014**, *48*, 193–204. [CrossRef]

15. Bossuyt, P.M.M. Interpreting diagnostic test accuracy studies. *Semin. Hematol.* **2008**, *45*, 189–195. [CrossRef] [PubMed]

16. Ayyub, B.M.; Klir, G.J. *Uncertainty Modeling and Analysis in Engineering and the Sciences*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2006. [CrossRef]

17. Kallner, A.; Boyd, J.C.; Duewer, D.L.; Giroud, C.; Hatjimihail, A.T.; Klee, G.G.; Lo, S.F.; Pennello, G.; Sogin, D.; Tholen, D.W.; et al. *Expression of Measurement Uncertainty in Laboratory Medicine, Approved Guideline*; Clinical and Laboratory Standards Institute: Annapolis Junction, MD, USA, 2012.

18. Oosterhuis, W.P.; Theodorsson, E. Total error vs. measurement uncertainty: Revolution or evolution? *Clin. Chem. Lab. Med.* **2016**, *54*, 235–239. [CrossRef] [PubMed]

19. Ramsey, M.H.; Ellison, S.L.R.; Rostron, P. (Eds.) Eurachem/EUROLAB/ CITAC/Nordtest/AMC Guide. In *Measurement Uncertainty Arising from Sampling: A Guide to Methods and Approaches*, 2nd ed.; Eurachem: Vilnius, Lithuania, 2019; 109p.

20. Ellison, S.L.R.; Williams, A. Eurachem CITAC Guide. In *Quantifying Uncertainty in Analytical Measurement*; Eurachem: Vilnius, Lithuania, 2019; 133p.

21. McLeod, A.I. Simple Random Sampling. In *Encyclopedia of Statistical Sciences*, 2nd ed.; Kotz, S., Ed.; John Wiley & Sons Inc.: Hoboken, NJ, USA, 2006; Volume 12, pp. 7740–7742.

22. Bloch, D.A. Comparing Two Diagnostic Tests against the Same "Gold Standard" in the Same Sample. *Biometrics* **1997**, *53*, 73–85. [CrossRef] [PubMed]

23. Gillard, J. A generalised Box–Cox transformation for the parametric estimation of clinical reference intervals. *J. Appl. Stat.* **2012**, *39*, 2231–2245. [CrossRef]

24. Atkinson, A.B. The box-cox transformation: Review and extensions. *Stat. Sci.* **2020**, in press.

25. Hund, E.; Massart, D.L.; Smeyers-Verbeke, J. Operational definitions of uncertainty. *Trends Analyt. Chem.* **2001**, *20*, 394–406. [CrossRef]

26. White, G.H. Basics of estimating measurement uncertainty. *Clin. Biochem. Rev.* **2008**, *29*, S53–S60. [PubMed]

27. Agresti, A.; Coull, B.A. Approximate is Better than "Exact" for Interval Estimation of Binomial Proportions. *Am. Stat.* **1998**, *52*, 119–126. [CrossRef]

28. Joint Committee for Guides in Metrology. *Evaluation of Measurement Data—Guide to the Expression of Uncertainty in Measurement*; BIPM: Sèvres, France, 2008; 120p.

29. Farrance, I.; Frenkel, R. Uncertainty of Measurement: A Review of the Rules for Calculating Uncertainty Components through Functional Relationships. *Clin. Biochem. Rev.* **2012**, *33*, 49–75. [PubMed]

30. Wilson, B.M.; Smith, B.L. 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*, 035301. [CrossRef]

31. Welch, B.L. The Generalization of 'Student's' Problem when Several Different Population Variances are Involved. *Biometrika* **1947**, *34*, 28–35. [CrossRef] [PubMed]

32. Satterthwaite, F.E. An approximate distribution of estimates of variance components. *Biometrics* **1946**, *2*, 110–114. [CrossRef]

33. Wolfram Research Inc. *An Elementary Introduction to the Wolfram Language*, 2nd ed.; Wolfram Research Inc.: Champaign, IL, USA, 2017; 340p.

34. Wolfram Research Inc. *Mathematica, Ver. 12.2*; Wolfram Research Inc.: Champaign, IL, USA, 2020.

35. Lim, T.-O.; Bakri, R.; Morad, Z.; Hamid, M.A. Bimodality in blood glucose distribution: Is it universal? *Diabetes Care* **2002**, *25*, 2212–2217. [CrossRef]

36. Cefalu, W.T.; Berg, E.G.; Saraco, M.; Petersen, M.P.; Uelmen, S.; Robinson, S. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2019. *Diabetes Care* **2019**, *42*, S13–S28. [CrossRef]

37. Kupchak, P.; Wu, A.H.B.; Ghani, F.; Newby, L.K.; Ohman, E.M.; Christenson, R.H. Influence of imprecision on ROC curve analysis for cardiac markers. *Clin. Chem.* **2006**, *52*, 752–753. [CrossRef] [PubMed]
38. Kroll, M.H.; Biswas, B.; Budd, J.R.; Durham, P.; Gorman, R.T.; Gwise, T.E.; Halim, A.-B.; Hatjimihail, A.T.; Hilden, J.; Song, K. *Assessment of the Diagnostic Accuracy of Laboratory Tests Using Receiver Operating Characteristic Curves, Approved Guideline*, 2nd ed.; Clinical and Laboratory Standards Institute: Annapolis Junction, MD, USA, 2011.
39. Collins, J.; Albert, P.S. Estimating diagnostic accuracy without a gold standard: A continued controversy. *J. Biopharm. Stat.* **2016**, *26*, 1078–1082. [CrossRef] [PubMed]
40. Sakia, R.M. The Box-Cox Transformation Technique: A Review. *J. R. Stat. Soc. Ser. D Stat.* **1992**, *41*, 169–178. [CrossRef]
41. Box, G.E.P.; Cox, D.R. An Analysis of Transformations. *J. R. Stat. Soc. Series B Stat. Methodol.* **1964**, *26*, 211–243. [CrossRef]
42. Solberg, H.E. 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*, S13–S32. [CrossRef]
43. Pavlov, I.Y.; Wilson, A.R.; Delgado, J.C. Reference interval computation: Which method (not) to choose? *Clin. Chim. Acta* **2012**, *413*, 1107–1114. [CrossRef] [PubMed]
44. Sikaris, K. Application of the stockholm hierarchy to defining the quality of reference intervals and clinical decision limits. *Clin. Biochem. Rev.* **2012**, *33*, 141–148. [PubMed]
45. Daly, C.H.; Liu, X.; Grey, V.L.; Hamid, J.S. A systematic review of statistical methods used in constructing pediatric reference intervals. *Clin. Biochem.* **2013**, *46*, 1220–1227. [CrossRef]
46. Ozarda, Y.; Sikaris, K.; Streichert, T.; Macri, J.; IFCC Committee on Reference intervals and Decision Limits (C-RIDL). Distinguishing reference intervals and clinical decision limits—A review by the IFCC Committee on Reference Intervals and Decision Limits. *Crit. Rev. Clin. Lab. Sci.* **2018**, *55*, 420–431. [CrossRef]
47. Wilson, J.M.G.; Jungner, G. *Principles and Practice of Screening for Disease*; World Health Organization: Geneva, Switzerland, 1968; Volume 34, 163p.
48. Petersen, P.H.; Horder, M. 2.3 Clinical test evaluation. Unimodal and bimodal approaches. *Scand. J. Clin. Lab. Invest.* **1992**, *52*, 51–57. [CrossRef]
49. Analytical Methods Committee, A.N. Why do we need the uncertainty factor? *Anal. Methods* **2019**, *11*, 2105–2107. [CrossRef]
50. Brown, L.D.; Cai, T.T.; DasGupta, A. Interval Estimation for a Binomial Proportion. *Stat. Sci.* **2001**, *16*, 101–117.