Archives of Academic Emergency Medicine. 2025; 13(1): e56

OPEN ACCESS

REVIEW ARTICLE

Calculation of Sensitivity and Specificity from Partial Data for Meta-Analyses: Introducing Some Practical Methods

Reihanesadat Khatami[1•], Mohammadsadegh Faghihi[2•], Hannanesadat Khatami[2], Mahmoud Yousefifard[2*], Seyedhesamoddin Khatami[2†]

1. Technische Universität Berlin, Faculty of Electrical Engineering and Computer Science, Berlin, Germany
2. Physiology Research Center, Iran University of Medical Sciences, Tehran, Iran

Received: March 2025; Accepted: April 2025; Published online: 11 June 2025

Abstract: **Introduction:** Meta-analyses of diagnostic/prognostic studies for calculating the pooled sensitivity and specificity require true positive (TP), true negative (TN), false positive (FP), and false negative (FN) counts. However, few studies report these values directly. This study aimed to consolidate practical methods to reconstruct sensitivity and specificity from minimal data. **Methods:** Our framework addresses three main situations: (1) algebraic rearrangements to compute specificity given partial metrics; (2) digitization of receiver operating characteristic (ROC) curves to obtain threshold-specific sensitivity and specificity; and (3) application of the binormal model when only AUC and prevalence are available. We tested these methods on a dataset related to mortality prediction in myocardial infarction (MI) using machine learning models, assessing how well they reconstructed sensitivity and specificity. **Results:** Algebraic formulas and ROC digitization yielded reliable estimates when partial metrics or graphical curves were sufficiently detailed. However, the binormal model, which assumes equal variances, showed noticeable inaccuracies, especially for sensitivity. Linear regression analyses indicated that higher prevalence and higher AUC reduced estimation errors. **Conclusion:** These methods offer practical alternatives for reconstructing diagnostic accuracy measures when data are incomplete. Relying solely on AUC-based estimations may introduce substantial bias, particularly in low-prevalence contexts. We recommend that primary studies report threshold-specific sensitivity and specificity to support more accurate meta-analytic estimations.

Keywords: sensitivity; specificity; ROC curve; Area Under the Curve; Statistics

Cite this article as: Khatami R, Faghihi M, Khatami H, et al. Calculation of Sensitivity and Specificity from Partial Data for Meta-Analyses: Introducing Some Practical Methods. Arch Acad Emerg Med. 2025; 13(1): e56. https://doi.org/10.22037/aaemj.v13i1.2678.

1. Introduction

In many diagnostic accuracy meta-analyses, investigators combine standard measures such as sensitivity and specificity across multiple studies to assess overall accuracy (1). Ideally, the original articles would report the confusion matrix counts [true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN)] or provide both sensitivity and specificity. However, in practice, such data are often incomplete (2). Some studies may list sensitivity along with another metric but not specificity, forcing reliance on algebraic formulas to recover the missing measure.

Others may present only a plotted ROC curve, requiring digitization to extract sensitivity and 1- specificity points. Still more challenging are articles that report only the Area Under the ROC Curve (AUC) and prevalence, leaving no straightforward way to infer threshold-specific performance (3). Despite its recognized limitations, the binormal model frequently emerges as the only viable solution in these last scenarios (4). By assuming that diseased and non-diseased populations share a common variance, the model can estimate sensitivity and specificity from AUC and prevalence alone, albeit with possible inaccuracies, especially at lower prevalence levels or when variances differ substantially (5).

Because these diverse *missing data* situations regularly arise, we decided to consolidate practical formulas, digitization techniques, and binormal modeling into a single resource. We also evaluated the binormal approach on a real-world dataset, investigating how accurately it recovers sensitivity and specificity when only AUC and prevalence are reported. Our findings underline the utility and the caution needed when applying these methods, particularly in meta-analyses where consistent, threshold-based metrics are paramount.

[*] **Corresponding Author:** Mahmoud Yousefifard; Physiology Research Center, Iran University of Medical Sciences, Tehran, Iran. Phone: +98 939 157 3164, Email: yousefifard20@gmail.com, ORCID: https://orcid.org/0000-0001-5181-4985.

[†] **Corresponding Author:** Seyedhesamoddin Khatami; Physiology Research Center, Iran University of Medical Sciences, Tehran, Iran. Phone: +989129279205, Email: khatamihesam@gmail.com, ORCID: https://orcid.org/0009-0006-4810-1662.

[•] **R.K. and M.F. have contributed equally to this work.**

2. Definitions

In diagnostic test evaluation, several widely used metrics helps quantify performance based on the confusion matrix's four counts: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). Below are the principal measures (6):

1. **Sensitivity** = $\frac{TP}{TP+FN}$

This measures how effectively a test detects true positives (i.e., the proportion of actual positives correctly identified) (7).

2. **Specificity** = $\frac{TN}{TN+FP}$

This measures how effectively a test avoids false positives (i.e., the proportion of actual negatives correctly identified) (7).

3. **Accuracy** = $\frac{TP+TN}{TP+TN+FP+FN}$

Accuracy reflects the proportion of correct classifications (positives and negatives) (7).

4. **Precision** = $\frac{TP}{TP+FP}$

Also called the positive predictive value, precision is the reliability of positive predictions.

5. **F1-Score** = $2 \times \frac{Precision \times Sensitivity}{Precision + Sensitivity}$

The F1-score is the harmonic mean of precision and Sensitivity and is particularly informative in datasets with class imbalance.

3. Practical Methods for calculating the sensitivity and specificity

3.1. Algebraic formulas to derive specificity

Studies sometimes report only partial metrics, making it challenging to recover specificity. To address these scenarios, the following rearrangements help derive specificity when sensitivity and certain other metrics (e.g., accuracy, precision, F1) are available, along with knowledge of total sample size N and event (positive) count P. The algebraic methods presented assume that prevalence (or event count) and sample size are reported, as these are essential for calculating the confusion matrix counts required in diagnostic meta-analyses. Studies lacking these data may not be suitable for inclusion.

Specificity from sensitivity and accuracy

Given sample size N, event count P, Sensitivity (Se), and accuracy (Acc):

$$\textbf{Specificity} = \frac{(N \times Acc) - (P \times Se)}{N - P}$$

Specificity from sensitivity and precision

If only Sensitivity (Se) and precision (Prec) are reported:

$$\textbf{Specificity} = 1 - [\ \frac{\frac{P \times Se}{Prec - (P \times Se)}}{N - P}\ ]$$

Specificity from Sensitivity and F1-Score

In cases reporting the F1-score ($F1$) and Sensitivity (Se)

$$\textbf{Specificity} = 1 - [\ \frac{P \times Se \times (1 - \frac{F1 \times Se}{2\ Se - F1})}{(N - P) \times \frac{F1 \times Se}{2\ Se - F1}}\ ]$$

Although more complex in appearance, this rearrangement still allows the computation of specificity with minimal original information.

3.2. Digitizing ROC Curves Using Three-Point Calibration

In some studies, ROC curves are presented as graphical plots without numerical values, making it challenging to extract key diagnostic accuracy metrics (8). In such cases, digitization software (e.g., *PlotDigitizer* (9)) provides a systematic approach to approximate Sensitivity and specificity values:

a. **Three-Point Calibration:** To accurately map pixel positions to real sensitivity and specificity values, the process begins by selecting three reference points on the graph. Commonly, (0, 0) and (1, 1) serve as fixed calibration points, while a third precisely known coordinate ensures accurate scaling.

b. **Curve Tracing:** Once calibrated, the ROC curve is manually traced by selecting multiple points along its path. This process generates a series of coordinate pairs (x_i, y_i), where x_i= 1 - Specificity and y_i = Sensitivity. These extracted points reconstruct the curve numerically.

c. **Estimating Sensitivity and Specificity:** Unlike conventional threshold selection methods, this approach focuses on identifying the optimal point on the ROC curve. After digitizing the curve, the optimal threshold is determined by selecting the point closest to (0, 1)—the upper-left corner—representing the highest sensitivity with reasonable specificity (7). This point is often chosen as it minimizes the Euclidean distance to perfect classification.

d. **Extracting Diagnostic Accuracy Metrics:** Once the optimal point is identified, its sensitivity and specificity values are extracted and used for further meta-analysis. This method allows for consistent estimation across studies, even when numerical tables are unavailable (10).

By employing this approach, researchers can systematically extract diagnostic performance metrics from graphical ROC representations, ensuring that essential data are not lost

when authors report only visualized results. However, potential errors may arise if there is visual misjudgment in selecting the optimal point or if the graph resolution is low, leading to inaccuracies in extracted sensitivity and specificity values.

3.3. Binormal Model: Estimating Sensitivity and Specificity from AUC and Prevalence

In such minimal-information scenarios (AUC + prevalence, without threshold-specific metrics), no purely nonparametric or alternative parametric method can fully recover how Sensitivity and specificity trade off across different thresholds (11-13). Nonparametric approaches require partial ROC data or at least a known pair (Se, Sp), while more flexible parametric methods demand additional distributional parameters.

Under these circumstances, a parametric binormal model is commonly employed. This approach— sometimes referred to simply as the "binormal ROC model" or the "equal-variance normal model"—assumes that the test's continuous output (or underlying latent variable) follows a normal distribution in non-diseased individuals and another normal distribution of the same variance but shifted mean in diseased individuals. Despite its simplifications, this model often becomes the only practical means of estimating Sensitivity and specificity from AUC and prevalence alone (7, 14). Below is a step-by-step explanation of how it works, why it is used, and what each symbol in the formulas means (15).

A Receiver Operating Characteristic (ROC) curve is created by plotting the test's true positive rate (TPR, also called Sensitivity) against its false positive rate (FPR, which is 1- specificity) across every possible decision threshold (7). The Area Under the ROC Curve (AUC) represents the probability that, when comparing one diseased individual and one non-diseased individual, the test score for the diseased individual will be higher (16). An AUC of 1.0 means perfect discrimination; an AUC of 0.5 means no discriminative ability beyond chance. Typically, a point on the ROC curve (a threshold t) is chosen in practice to balance Sensitivity and specificity according to clinical or practical needs.

However, if we only have the AUC (a single statistic) rather than the full curve, we lose all other threshold-specific details.

4. Rationale for the Binormal Model

1) *Why do We Need an Assumption About Distribution?*

Simply knowing the single AUC value does not tell us exactly where the test's operating threshold should lie or how Sensitivity and specificity should trade off. To bridge this gap, we need to assume a shape for how test results are distributed in each group (diseased and non-diseased). One of the most widespread assumptions is:

Non-diseased distribution: X ~ $N(\mu_0, \sigma^2)$, Diseased distribution: X ~ $N(\mu_1, \sigma^2)$ (16)

Here, $\mu_1 > \mu_0$ (so that diseased scores are, on average, higher than non-diseased scores), and both groups share the same variance, σ^2.

i.e., $b = \frac{\sigma_1}{\sigma_0} = 1$ (b: ratio of standard deviations between the test result distributions of the diseased (event) group and the non-diseased (non-event) group).

This is the simplest version of a binormal model. As we will see, the binormal model links the AUC directly to the standardized separation of these two normal distributions.

2) *Why is it often the Only Practical Option?*

When all we have is AUC, Prevalence (p), and Sample size (N), we lack the data needed to fit more flexible models. Specifically:

• *Single Summary Statistic (AUC):* Different ROC shapes can yield the same AUC, so a single numeric value does not specify threshold-dependent trade-offs.

• *Lack of Threshold Information:* Without at least one known (Se, Sp) pair or partial ROC points, nonparametric methods cannot reconstruct the full curve.

• *No distributional parameters:* Without means (μ_0, μ_1) or variances (σ_0^2, σ_1^2), we cannot estimate unequal variances ($b \neq 1$) or non-normal distributions.

• *Identifiability:* Without assuming b=1, the model is underdetermined—infinite combinations of μ_0, μ_1, σ_0, σ_1 can produce the same AUC. Equal variances resolve this ambiguity, ensuring unique solutions for Sensitivity and specificity.

• *Empirical Robustness:* Studies show that equal-variance assumptions yield reasonable Sensitivity/specificity estimates even when variances differ slightly (e.g., b = 0.8 – 1.2) (4, 5, 17).

For example, a true b = 1.5 introduces only modest bias (3–5% error in Se/Sp for AUC = 0.8).

• *Clinical Interpretability:* Assuming b = 1 implies symmetry in misclassification costs (Se Sp at the optimal threshold). This aligns with clinical guidelines where false positives and negatives are equally weighted.

Consequently, the equal-variance binormal model is not just a default—it is often the only feasible choice.

5. Step-by-Step Binormal Model Method

I. *Converting AUC to d*

First, we define a useful parameter often called the standardized separation, denoted d'. Conceptually, d' measures how far apart the diseased and non-diseased distributions are in standardized units. Under the binormal model, there is a well-known relationship between AUC and d':

$d' = \sqrt{1 + b^2}\,\phi^{-1}(\text{AUC}) = \sqrt{2}\,\phi^{-1}(\text{AUC})$ (b = 1)

Where $\phi^{-1}(\cdot)$ is the inverse of the standard normal cumulative distribution function (CDF). For a given AUC, ϕ^{-1}(AUC) tells us the z-score corresponding to that cumulative probability.

We multiply by $\sqrt{2}$ due to a standard result in binormal ROC theory.

II. *Determining the Threshold t^**

A widely cited formula gives the threshold t^* that maximizes overall classification accuracy in a population whose disease prevalence is p. Under the binormal assumption, that threshold is:

$$t^* = \frac{d'^2 + 2\ln(\frac{1-p}{p})}{2\,d'}$$

Here:
• $\ln(\frac{1-p}{p})$ is the natural logarithm of the odds ratio for being non-diseased vs. diseased (when p is the probability of disease).
• The numerator combines the separation d'^2 with a term that depends on the prevalence.
• Dividing by 2 d' yields the *best* threshold in standard-normal units if our goal is purely to maximize accuracy (i.e., the fraction of correctly classified individuals).

III. *Computing Sensitivity and Specificity*

Once t^* is found, the Sensitivity and Specificity implied by that threshold follow from normal probability theory:
Sensitivity = $1 - \phi\,(t^* - d')$
Specificity = $\phi\,(t^*)$
• In these formulas, ϕ (.) is the standard normal CDF, i.e., ϕ(z) gives the probability that a standard normal variable $Z \sim N$ (0, 1) is less than or equal to z.
• Notice that Specificity is simply the Area under the standard normal curve up to t^*. Meanwhile, Sensitivity involves a shift by d', reflecting the distribution of diseased scores.
Why 1 $\phi(t^* - d')$?
A diseased individual's test score is modeled as $Z + d'$, where $Z \sim N$ (0, 1). The test is "positive" if that score exceeds t^*. Hence:
Pr (test positive | diseased) = Pr $(Z + d' \geq t^*)$ = Pr $(Z \geq t^* - d')$ = $1 - \phi\,(t^* - d')$
That is precisely the Sensitivity under the binormal assumption.

a. **Practical Usage, Limitations, and Cautions**

Applicability. When an article reports only AUC, prevalence (p), and possibly sample size (N), the binormal model offers a way to infer Sensitivity and specificity at a presumed *optimal* threshold (i.e., the threshold maximizing accuracy).
Limitations. Real data may violate the equal-variance assumption, leading to inaccuracies. Moreover, a real-world study's chosen threshold could reflect clinical priorities (e.g., very high Sensitivity) rather than theoretical maximum accuracy.
Approximation. Different shapes of the ROC curve can yield the same AUC, and forcing $b = 1$ can introduce bias if the true variance ratio differs substantially (5).
Educational Value. Despite its shortcomings, the binormal approach illustrates how AUC, prevalence, and threshold selection interact, highlighting that a single number (AUC) rarely suffices to depict real-world performance fully (7).

b. **Challenging the Binormal Model in a Real-World Dataset**

We compiled 183 observations from our systematic review and meta-analysis of studies examining mortality prediction in Myocardial Infarction using various machine learning models. Each observation provided an AUC, a disease prevalence (p), and the model's reported Sensitivity and specificity. Our goal was to see how the binormal model's estimates compared with the reported (actual) values.
For each observation, we used the AUC and prevalence to derive d', then calculated the theoretical threshold t^* and the implied Sensitivity and specificity. These values were computed using the previously mentioned functions in Microsoft Excel. The following results were obtained by comparing the estimated sensitivity and specificity with the reported values in the studies using STATA 18.

6. Results

Absolute Errors
On average, the absolute error for Sensitivity was 0.58 (SD = 0.18), whereas that for specificity was 0.21 (SD = 0.10). A paired t-test confirmed that Sensitivity estimates showed significantly larger deviations than specificity estimates ($p < 0.0001$).
Correlation with Reported Values
Pearson correlations (0.26 for Sensitivity, $p < 0.001$; 0.17 for specificity, $p < 0.05$) demonstrated only modest agreement.
Regression on AUC and Prevalence
Linear models revealed both AUC and prevalence as strong predictors of estimation error ($p < 0.0001$), with negative coefficients indicating that higher AUC and prevalence lead to reduced error.
Prevalence Thresholds
Solving for the prevalence needed to keep absolute error under 0.1 showed a mean threshold of ~0.46 (SD = 0.06) for Sensitivity and ~0.38 (SD = 0.15) for specificity. In low-prevalence or lower-AUC scenarios, the error for Sensitivity could be substantial. In these 183 cases, AUC values ranged from 0.618 to 0.982, while prevalence varied between 0.013 and 0.625.
These findings underscore that the binormal approach tends to overestimate or underestimate Sensitivity more than specificity, particularly in lower-prevalence scenarios or when the AUC is not very high. Specificity estimates, however, often remain within more acceptable bounds, especially when prevalence and AUC are sufficiently large. Researchers should thus interpret binormal-derived Sensitivity and specificity with caution and ideally cross-check against additional data if available. Despite the theoretical rationale

for using the binormal approach, these empirical results indicate practical limitations, particularly for the estimation of Sensitivity.

Table 1 summarizes the six most common missing-data scenarios and the corresponding formulas or procedures we propose for reconstructing sensitivity, specificity, and the four confusion-matrix counts.

7. Discussion

The range of methodologies presented here reflects the common dilemmas researchers face when constructing a meta-analysis of diagnostic accuracy and prognostic test performance. Ideally, authors should report both sensitivity and specificity if they do not provide the confusion matrix counts for each study. In practice, such comprehensive reporting is rare, and investigators are often left to reconstruct missing metrics using algebraic formulas or by digitizing ROC curves. These approaches, while not without limitations, can yield reasonably accurate estimates, particularly if the required input data (e.g., Sensitivity and accuracy or a well-calibrated ROC plot) are reliable and complete. In contrast, the binormal model, which provides Sensitivity and specificity estimates from AUC and prevalence alone, should be approached with caution. Although it offers a solution when no threshold-based performance data are available, our findings confirm that low-prevalence conditions (such as mortality in our dataset) pose a serious challenge. The binormal assumption of equal variances can lead to large estimation errors in Sensitivity, as demonstrated in our analysis. This issue becomes even more pronounced when prevalence is low and the AUC is not sufficiently high, making sensitivity estimates less reliable in such conditions. Consequently, for meta-analyses of diagnostic tests in low-prevalence conditions, excluding studies that report only AUC (and no threshold-specific measures) may be more prudent. Inaccurate estimates of sensitivity and specificity can have significant clinical consequences, particularly in low-prevalence settings. For example, overestimating sensitivity may lead to missed diagnoses, while underestimating specificity could result in unnecessary interventions. Our findings highlight the risks of relying solely on AUC-based estimates, especially when prevalence is low. Such exclusions help ensure that pooled estimates do not introduce unwarranted bias from an overly simplistic or potentially misapplied model. Taken together, these observations underscore the importance of transparent reporting of threshold-based performance metrics in diagnostic research. In circumstances where ROC curves are given without numeric points, digitization can be a helpful fallback, so long as the curve is adequately labeled and the software calibration is carefully performed. Once sensitivity and specificity are reconstructed using the methods described, researchers should employ appropriate meta-analytic techniques, such as the bivariate model or HSROC, to account for variability in diagnostic thresholds across studies.

8. Conclusions

This article consolidates various methods for reconstructing sensitivity and specificity when they are not reported, as these metrics are essential for deriving TP, TN, FP, and FN in diagnostic and prognostic meta-analyses. First, we provide algebraic formulas for deriving specificity from sensitivity and another metric. Second, we discuss the digitization of ROC curves for estimating sensitivity and specificity. Lastly, we examined the binormal model for estimating these metrics when only AUC and prevalence are available. We explained this approach and challenged it using a real-world dataset. Ultimately, we demonstrated that this method lacks sufficient accuracy, particularly in low prevalence and AUC. Then, researchers should apply it with caution or exclude such studies from meta-analyses.

9. Declarations

9.1. Acknowledgments

None.

9.2. Authors Contributions

All authors contributed to the conceptualization, data collection, mathematical and statistical validation, writing, and revision of the manuscript, and approved the final version.

9.3. Using Artificial Intelligence Chatbots

The editing of this manuscript was assisted by ChatGPT for language and clarity improvement.

9.4. Conflict of Interest

The authors declare that they have no conflicts of interest relevant to this manuscript. Mahmoud Yousefifard, one of the corresponding authors of this manuscript, is also an editor of the Archives of Academic Emergency Medicine. However, he was not involved in the peer-review or decision-making process for this article.

9.5. Funding

None.

References

1. Cleophas TJ, Zwinderman AHJCc, medicine l. Meta-analyses of diagnostic studies. 2009;47(11):1351-4.
2. Ter Riet G, Bachmann LM, Kessels AG, Khan KSJBE-BM. Individual patient data meta-analysis of diagnostic studies: opportunities and challenges. 2013;18(5):165-9.
3. Walter S, Sinuff TJJoce. Studies reporting ROC curves of diagnostic and prediction data can be incorporated into meta-analyses using corresponding odds ratios. 2007;60(5):530-4.
4. Hanley JAJMdm. The robustness of the" binormal" assumptions used in fitting ROC curves. 1988;8(3):197-203.

5. Hillis SLJAr. Simulation of unequal-variance binormal multireader ROC decision data: an extension of the Roe and Metz simulation model. 2012;19(12):1518-28.
6. Stapor K, editor Evaluation of classifiers: current methods and future research directions. FedCSIS (Position Papers); 2017.
7. Zhou X-H, Obuchowski NA, McClish DK. Statistical methods in diagnostic medicine: John Wiley & Sons; 2014.
8. Carter JV, Pan J, Rai SN, Galandiuk SJS. ROC-ing along: Evaluation and interpretation of receiver operating characteristic curves. 2016;159(6):1638-45.
9. Huwaldt JA, Steinhorst SJUhpsn. Plot digitizer. 2013.
10. Kumar R, Indrayan AJIp. Receiver operating characteristic (ROC) curve for medical researchers. 2011;48:277-87.
11. Wald NJ, Bestwick JPJJoMS. The area under the ROC curve: is it a valid measure of screening performance? : SAGE Publications Sage UK: London, England; 2014. p. 220-.
12. Hajian-Tilaki KO, Hanley JA, Joseph L, Collet J-PJMDM. A comparison of parametric and nonparametric approaches to ROC analysis of quantitative diagnostic tests. 1997;17(1):94-102.
13. Jiménez-Valverde AJB, Conservation. Threshold-dependence as a desirable attribute for discrimination assessment: implications for the evaluation of species distribution models. 2014;23:369-85.
14. Bandos AI, Guo B, Gur DJAR. Estimating the area under ROC curve when the fitted binormal curves demonstrate improper shape. 2017;24(2):209-19.
15. Hanley JAJSim. The use of the 'binormal'model for parametric ROC analysis of quantitative diagnostic tests. 1996;15(14):1575-85.
16. Metz CE, editor Basic principles of ROC analysis. Seminars in nuclear medicine; 1978: Elsevier.
17. WALSH SJJSim. Limitations to the robustness of binormal ROC curves: effects of model misspecification and location of decision thresholds on bias, precision, size and power. 1997;16(6):669-79.

Table 1: Summary of Methods for Reconstructing Sensitivity, Specificity, and Confusion Matrix Counts Under Incomplete Reporting Scenarios

Scenario	Reported Data/ Metrics	Additional Info Needed	Key Formula / Method	Steps to Obtain TP, TN, FP, FN	Notes
1	Se & Sp both reported	- Total sample size (N) - Number of positives (P) or prevalence (p)	No extra formula is needed; Se and Sp are directly taken from the study.	Use: - TP = Se × P - FN = (1 - Se) × P - TN = Sp × (N - P) - FP = (1 - Sp) × (N - P)	*Ideal scenario.* Directly convert reported Se, Sp into counts.
2	Se + Accuracy	- N, P (or p) - Se - Accuracy (Acc)	The formula for Sp: Specificity = [(N × Acc) - (P × Se)] / (N - P)	1. Calculate Sp from the formula above. 2. Then compute TP, FN, TN, and FP as in Scenario 1.	Must ensure Acc and Se are accurately reported.
3	Se + Precision (PPV)	- N, P - Se - Precision (Prec)	Formula for Sp: Specificity = 1 - [(P × Se) / (Prec - (P × Se))] / (N - P)	1. Compute Sp from the above expression. 2. Then compute TP, FN, TN, and FP as usual.	Precision is also called PPV = TP / (TP + FP).
4	Se + F1-Score	- N, P - Se - F1-Score	Formula for Sp: Specificity = 1 - { [P × Se × (1 - (F1 × Se) / (2Se - F1))] / [(N - P) × ((F1 × Se) / (2Se - F1))] }	1. Use the formula above to get Sp. 2. Then obtain TP, FN, TN, and FP similarly.	Must have consistent F1, Se values.
5	Only a ROC Curve (graph) without a numeric table	- N, P (for final counts) - Digitization software (e.g., PlotDigitizer)	3-step ROC digitization: 1. Calibrate the axes with 3 known points. 2. Trace the ROC curve to obtain (1Sp, Se) coordinates. 3. Pick the "optimal" point to get the final Se, Sp.	1. Once Se and Sp are extracted, use them as in Scenario 1. 2. TP = Se × P, etc.	Watch out for resolution/graph quality; digitization errors can occur.
6	Only AUC & Prevalence (p)	- AUC - p (or P) and N (to get absolute counts)	Binormal Model (equal variances): 1 $\sqrt{2}\times \Phi^{-1}$(AUC) 2. t* = [d'² + 2 ln((1p)/p)] / (2 d') 3. Sensitivity = 1 - Φ ($t^* - d'$) and Specificity = Φ (t^*)	1. Get Se, Sp from the model's formulas. 2. Then compute TP, FN, TN, and FP from Se, Sp, and N, P.	Potentially large estimation error for Se, especially in low prevalence.

Se: sensitivity; Sp: Specificity;

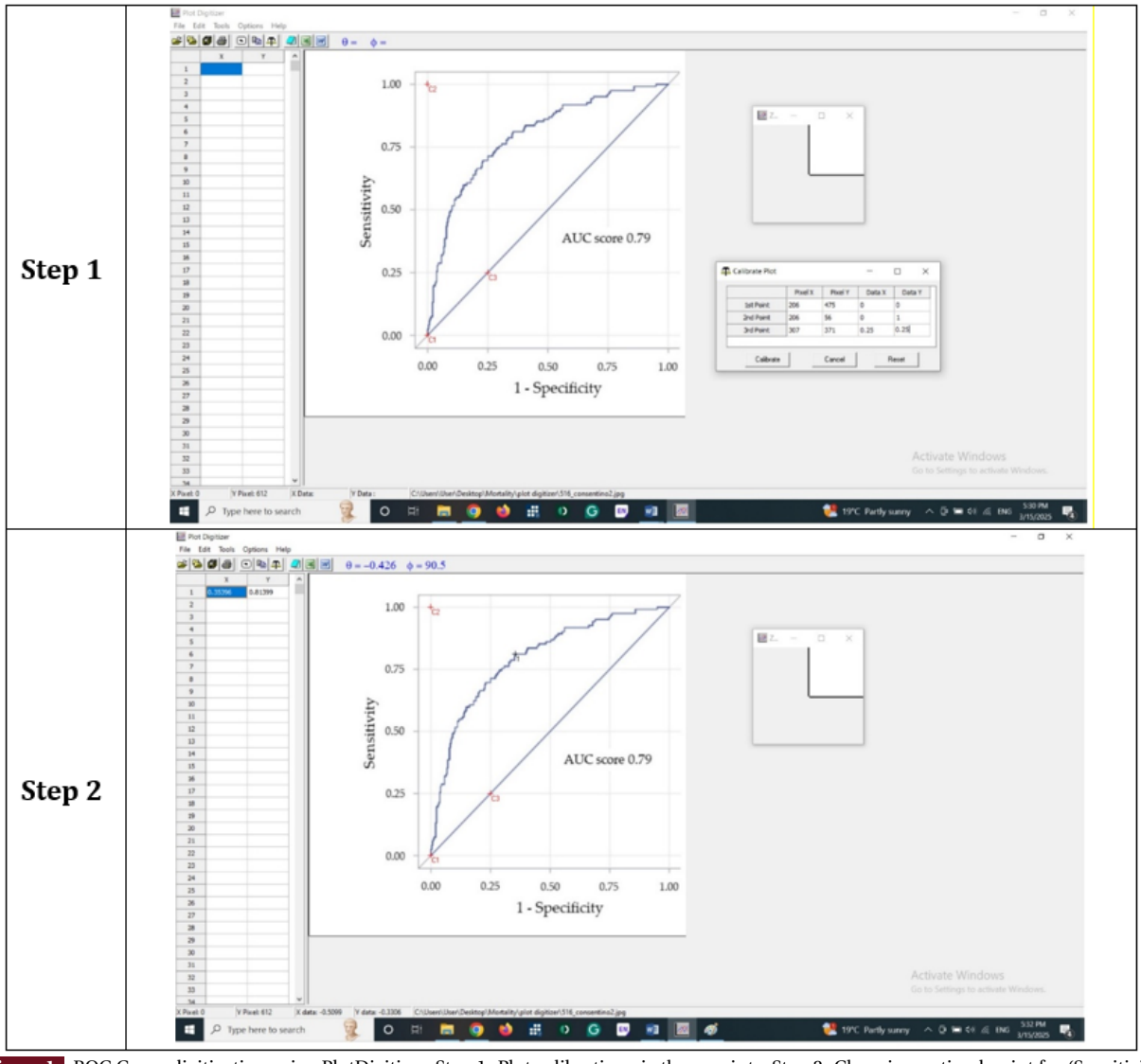

Figure 1: ROC Curve digitization using PlotDigitizer. Step 1: Plot calibration via three points; Step 2: Choosing optimal point for (Sensitivity, Specificity) pair.