Diagnostic devices with intermediate/gray zone output

Bipasa Biswas
CDRH, FDA, 10903 New Hampshire Avenue, Silver Spring, MD 20993

Abstract
Medical diagnostic tests often provide a binary response- positive or negative result for a target condition of interest. However, in some cases, the incorporation of a "gray zone" or intermediate zone, leading to more than two results is recommended for clinical management. Using an intermediate/gray zone, to define a 3x2 table is appropriate than ignoring the test scores in these zones. The six-cell matrix (3x2 table), however, would serve limited purpose if clinicians cannot apply the additional information provided by the conditional operating characteristics to effect better patient management decisions. This presentation discusses a decision analytic approach utilizing pre-test and post-test probability to the target condition that can be used for efficient categorization of more than two results for such tests.

Key Words: Binary, Intermediate test results, Decision Analysis.

1. Introduction

A diagnostic device with a dichotomous output has two values (yes or no, present or absent, positive or negative). The test output indicates the presence or absence of the target condition (condition of interest), where a target condition can "refer to a particular disease, a disease stage, health status, or any other identifiable condition within a patient, such as staging a disease already known to be present, or a health condition that should prompt clinical action, such as the initiation, modification or termination of treatment" (STARD, 2003). A qualitative test can provide a dichotomous result indicating the presence or absence or that the test is positive or negative for the target condition. And a quantitative and/or continuous or an ordinal valued test can be dichotomized using a cut-off or a clinical decision point. Diagnostic devices are henceforth referred to as diagnostic tests in this paper.
In general, diagnostic devices with dichotomous output e.g. presence or absence of the condition of interest, is evaluated against a clinical reference standard used to establish the true condition. A test with a binary output is represented by a 2x2confusion matrix:

Table 1: 2x2 table
Study Population

		Clinical Reference Standard	
		D+	D-
Test	T+ (positive)	TP	FP
	T- (negative)	FN	TN
	Total	P=TP+FN	N=FP+TN

The variable D represents the target condition where D+ means the condition is present and D- means the condition is absent and the test is represented by the variable T where T+ means the test is positive and T- means the test is negative. In the above table, "TP" denotes True Positive; "FP" denotes False Positive; "FN" denotes False Negative; "TN" denotes True Negative.

The accuracy of the test is evaluated by either the sensitivity-specificity pair, or the pair of predictive values or the pair of likelihood ratios which are defined as follows.

Sensitivity (TPF) = P(T+|D+) estimated by TP/(TP+FN)
Specificity (1-FPF) =P(T-|D-) estimated by TN/(TN+FP)

Likelihood ratio positive = sensitivity/(1-specificity)
Likelihood ratio negative= (1-sensitivity)/specificity

Positive predictive value (PPV) = P(D+|T+) estimated by TP/(TP+FP)
Negative predictive value (NPV) = P(D-|T-) estimated by TN/(FN+TN)
The performance measures PPV and NPV depend on the prevalence of the true target condition.

1.1 Uninterpretable, Intermediate and Indeterminate

Shinkins et al (2013) discuss diagnostic tests with inconclusive tests results that can be categorized to three main types of "non-positive, non-negative" results as

- Uninterpretable results: those that "do not meet the minimum criteria constituting an adequate test "
- Intermediate test result: those that "confer a likelihood ratio for disease that is more than that conferred by a negative result, but less than that of positive test result"
- Indeterminate test results: those that add no additional diagnostic information to the original probability of disease (a subset of intermediate outcomes and are identified by likelihood ratio of 1)

Begg CB (1986) et al, provides a clear interpretation of uninterpretable test results, distinguishing it from intermediate and or indeterminate test result. If test can be repeated till a result is obtained, then naïve estimates of sensitivity and specificity (i.e. ignoring uninterpretable test results) are unbiased if test un-interpretability is independent of the disease status. However, if they are not, then the recommendation is to repeat the test till a definitive test result is obtained to report sensitivity and specificity. If test cannot be repeated, then un-interpretable test should be reported as a separate result.

1.2 Example for intermediate result

Instantaneous wave-free ratio (iFR) is pressure-derived, hyperemia-free index for the functional assessment of coronary stenoses1. Fractional Flow Reserve (FFR) is used as a reference standard.

FFR $\leq$ 0.80 indicates stenosis severity
FFR > 0.80

Adoption of fractional flow reserve (FFR) remains low (6-8%) (Petraca et al 2013), partly because of the time, cost and potential inconvenience associated with vasodilator

administration. The instantaneous wave-Free Ratio (iFR) is a pressure-only index of stenosis severity calculated without vasodilator drugs.
Since the introduction of iFR, a hybrid iFR-FFR diagnostic strategy has been proposed (Escaned et al 2015), where upper and lower iFR cut-offs are used to restrict decisions on the basis of iFR to those regions in which its post-test probability is very high, and FFR use is limited to the intermediate iFR range of values called the "adenosine zone". Thus,

- iFR $\leq$ 0.85 Treat
- 0.85 < iFR <0.94 "Adenosine zone"
- iFR $\geq$ 0.94 Defer

was proposed as a diagnostic strategy to manage patients where FFR use is limited to the "adenosine zone".

2. Literature review of different methods for intermediate/gray zone

There are three methods discussed in literature (Grenier et al 1995, Coste et al 2003 and 2006, Landsheer 2018) for defining and determining intermediate test results- 1) TG-ROC,2) Gray Zone and 3) Uncertain Interval. The three methods use different definitions and differ in clinical relevance.

2.1 TG-ROC

The two-graph Receiver Operating Characteristic (TG ROC) (Grenier et al 1995) establishes an intermediate zone based on generating two cut-points (thresholds), defined by a chosen high sensitivity and specificity and then using the percentiles of the distribution of the test results from subjects with and without target condition (e.g. disease) to select the two cut-points.
The TG-ROC based intermediate zone can be based using either non-parametric method or parametric method.
The advantages of TG-ROC method for determining intermediate zone is that it is based on target sensitivity and specificity of test which is independent of prevalence and is applicable to distribution free assumptions. The disadvantage is for tests with significant overlap between distribution of results from subjects with and without target condition, which can lead to exclusion of substantial number of patients from a decision.

2.2 Gray Zone

The gray-zone method (Coste et al 2003, 2005) establishes an intermediate zone based on prevalence of the target condition or pre-test probability of target condition and the clinical requirements for positive and negative post-test probability. In order to define the intermediate zone, one needs to analyze the clinical context, estimate pre-test probability and then using Bayes theorem, minimum required positive likelihood ratio and negative likelihood ratio (and/or sensitivity and specificity) are estimated. The gray zone is determined using similar approach as in TG-ROC in the final step.
The advantages of gray-zone method are that it is based on target sensitivity and specificity of test which is independent of prevalence and is applicable for distribution free assumptions (similar as TG-ROC). The disadvantages are that it requires the information about pre-test probability (unlike the TG-ROC) and like the TG-ROC approach, test results with significant overlap between distribution of results from subjects with and without target condition, can lead to exclusion of substantial number of patients from a decision.

2.3 Uncertain interval

The uncertain interval method (Landsheer 2018) establishes an intermediate zone around the intersection between distribution of test results of subjects with and without the target condition. It is defined as an interval where the probability of a correct diagnosis is almost equal as an incorrect diagnosis.

Defining intermediate zone would be based on find the point of intersection of the distribution of test results for subjects with and without the target condition and then the area around the intersection where the area around true negatives is balanced by the area of the false positives (likewise an area around the intersection where the area with true positives is balanced by the area of the false negatives).

The advantages of Uncertain Interval method are that it focuses on test results with insufficient validity and reliability and generally excludes fewer patients in clinical decision than the other two methods. Additionally, it is independent of prevalence

The disadvantage lies with its portability across settings (amount of permissible uncertainty can vary from setting to setting).

3. Challenges in presenting and/or analyzing

Simel et al (1987) report that there is no consensus analytic method for reporting the results of diagnostic tests for non-positive, non-negative results as patients with such outcomes may be excluded from the study and subsequently from analysis or that patients' non-positive, non-negative results are often forced inappropriately into a binary outcome in the standard 2x2 (four-cell) confusion matrix.

Table 2: 3x2 table

Study Population

		Clinical Reference Standard	
		D+	D-
Test	Positive	A1	B1
	Intermediate	A2	B2
	Negative	A3	B3
	Total	A=A1+A2+A3	B=B1+B2+B3

Usual but not recommended forms of reporting are, either ignoring intermediate or use conservative or anti-conservative approaches. By ignoring intermediate results, the report eliminates clinically useful information and stretches the performance metrics (sensitivity =A1/A1+A3, specificity= B3/B1+B3). The conservative reporting (also known as "worst case" scenario) treats non-positive, non-negative result counted as negative when patients have disease (conservative estimate of sensitivity (=A1/A)) and treats non-positive, non-negative result counted as positive when patients do not have disease (conservative estimate of specificity (=B3/B)). While the anti-conservative/optimistic reporting (also known as "best case" scenario) treats non-positive, non-negative result counted as positive when patients have disease (optimistic estimate of sensitivity (=((A1+A2)/A)) and treats non-positive, non-negative result counted as negative when patients do not have disease (optimistic estimate of specificity (=((B2+B3)/B)).

None of these methods of reporting provides the appropriate information to evaluate the test performance where the test provides an intermediate result.

4. Decision Analytic approach to six cell approach

Based on the six-cell (3x2 table) in Table 2, the counts (A1-A3, B1-B3) populate the cells and there are the following relations.

Relation between the sensitivities:

$A1/(A1+A2+A3) \leq A1/(A1+A3) \leq (A1+A2)/(A1+A2+A3)$

"Worst" case Sensitivity — Ignoring intermediate — "Best" case Sensitivity

Relation between specificities:

$B3/(B1+B2+B3) \leq B3/(B1+B3) \leq (B2+B3)/(B1+B2+B3)$

"Worst" case Specificity — Ignoring intermediate — "Best" case Specificity

4.1 Decision Analytic approach

Simel et al (1987) discuss how the possibility of obtaining "non-positive, non-negative" results factor in the decision to order a test. They define a new numerical expression, the test's yield, which provides the proportion of obtaining a "non-positive, non-negative" test result. Th e overall test yield describes the probability of obtaining either a positive or a negative test result, without regard whether the test result is a false positive or a false negative.

Thus, numerically define the yield of test (YD) as

$YD= (A1+A3+B1+B3)/(A1+A2+A3+B1+B2+B3)$

The yield for test positive is numerically defined as

$YD+= (A1+A3)/(A1+A2+A3)$

The yield for test negative is numerically defined as

$YD-=(B1+B3)/(B1+B2+B3)$

The following definitions are expressed as functions of test yield

Conditional sensitivity (Csens)$= A1/(A1+A3)$
$=(A1/(A1+A2+A3))/((A1+A3)/(A1+A2+A3))$
$=$ "Worst" case sensitivity/YD+

Conditional specificity (Cspec) $=B3/(B1+B3)$
$=(B1/(B1+B2+B3))/((B1+B3)/(B1+B2+B3))$
="Worst" case specificity/YD-

The likelihood for intermediate zone is defined as a function of the test yield as $LR_I = (1-YD+)/(1-YD-)$.

Additional notations are used to further illustrate the decision to test, treat or neither:

R_x += Treat the test positive

R_x I= Treat test positive and intermediate

NOR_x = No treatment
R_t =Risk of test
R_{rx} =Risk of treatment in a patient without disease
B_{rx} =Benefit of treating a patient with disease

Using the above notations one can derive the disease probability (prevalence) for threshold for no treatment versus treating test positives, treating test positives versus treating test positives and intermediates and treating test positives and intermediates versus treating all:

$$NOR_X - R_x + = \frac{(YD-)((1-Cspec)R_{rx}+R_t}{(YD-)((1-Cspec)R_{rx}+(YD+)(Csens)B_{rx}}$$

$$R_x + - R_x I = \frac{1}{1+(B_{rx}/R_{rx})(1-YD+/1-YD-)}$$

$$R_x I - R_x = \frac{(YD-)((Cspec)R_{rx}-R_t}{(YD-)((Cspec)R_{rx}+(YD+)(1-Csens)B_{rx}}$$

Thus, the threshold probabilities for disease can be expressed as functions of the test yield. The test yield can be easily be estimated from a six-cell matrix. And for same test yield the approach to treat none or treat test positive or test positive and intermediate or treat all depends on the disease prevalence (probabilities of disease).

4.2 Reporting

In reporting, one cannot ignore intermediate zone as discarding these from analysis gives biased estimates of sensitivity and specificity. In addition, the terms sensitivity and specificity are not appropriate measures for evaluating the performance of the test. A six-cell matrix (3x2) provides the transparency in performance and the following report provides performance evaluation of a test:

Table 3: Report

	Results	Clinical Reference Standard		Total		Percentage of the total	Risk of Disease	Likelihood Ratios
		D+	D-					
Test	Positive	A1	B1	A1+B1		100%x(A1+B1)/N	100%x(A1/(A1+B1))	(A1/A)/(B1/B)
	Intermediate	A2	B2	A2+B2		100%x(A2+B2)/N	100%x(A2/(A2+B2))	(A2/A)/(B2/B)
	Negative	A3	B3	A3+B3		100%x(A3+B3)/N	100%x(A3/(A3+B3))	(A3/A)/(B3/B)
	Total	A=A1+A2+A3	B=B1+B2+B3			Pre-test risk of disease= 100% x (A/N)		

5. Conclusions

There are some quantitative tests with two cutoffs/thresholds instead of a single threshold based on the clinical context as in the example for a cardiovascular test where the need was to have a high rule-in and rule-out claim. However, these thresholds need to be fixed prior to clinical validation studies to evaluate performance of the test in a clinical study. It should be noted that intermediate test result is different from uninterpretable or invalid test results. Each of the three methods discussed for determining an intermediate zone can lead to different intermediate zones. While evaluating the performance of test with intermediate tests result, adopting a 3x2 cell matrix for reporting is recommended. The intermediate test results are not to be ignored and should be reported in the test performance evaluation. Reporting likelihood ratios along with the corresponding percentage of each outcome and the post-test risk of the disease provides transparency of the performance.

References

1. Bossuyt, P.M., Reitsma, J.B., Bruns, D.E., Gatsonis, C.A., Glasziou, P.P., Irwig, L.M., Moher, D., Rennie, D., deVet, H.C.W., & Lijmer, J.G. (2003). The STARD statement for reporting studies of diagnostic accuracy: Explanation and elaboration. Clinical Chemistry, 49(1), 7–18.
2. Biswas B. AdvaMed (2019) Cardiovascular tests with intermediate/gray zone.
3. Begg CB, Greenes RA, Iglewicz B. The influence of unitepretability on the assessment of diagnostic tests. J Chron Dis Vol 39(8): 575-584, 1986.
4. Shinkins B Thompson M, Mallet S, Perera R. Diagnostic accuracy studies: how to report and analyse inconclusive test results. BMJ 2013; 346: f2778.
5. Shinkins B. Diagnostic uncertainty: dichotomies are not the answer. B J of Gen Pract 2013; 122-123.
6. Escaned J, Echavrria-pinto M, Garcia-Garcia HM, et al. Prospective Assessment of the diagnostic accuracy of instantaneous wave-free ratio to assess coronary stenosis relevance. J Am Coll Cardiol intv 2015; 8:824-833.
7. Petraco R, Park JJ, Sen S, et al. Hybrid iFR-FFR decision-making strategy: implications for enhancing universal adoption of physiology-guided coronary revascularisation. EuroIntervention 2013; 8:1157– 65.
8. Simel DL, Feussner JR, Delong ER, Matchar DB. Intermediate, indeterminate and uninterpretable diagnostic test results. Med Decis Making 1987; 7:107-114.
9. Grenier M, Sohr D, Gobel P. A modified ROC analysis for the selection of cut-off values and the definition of intermediate results of serodiagnostic tests. J of Immun Meth 1995; 185: 123-132.
10. Coste J, Jourdain P, Pouchot J. A gray zone assigned to inconclusive results of quantitative diagnostic tests: application to the use of brain natriuretic peptide for diagnosis of heart failure in acute dyspenic patients. Clin Chem 2006; 52: 2229-2235.
11. Coste J, Pouchot J. A gray zone for quantitative diagnostic and screening tests. Int J Epidemiol 2003; 32: 304-313.
12. Landsheer JA. The clinical relevance of methods for handling inconclusive medical test results: Quantification of uncertainty in medical-decision making and screening. Diagnostics 2018; 8, 32.
13. Petrides VH et al. Assessing performance for assays reporting Equivocal Results. AACC Annual Meeting Atlanta GA 2015.
14. Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. J Clin Epidemiol 1991; 44(8):763-770.