knowledge

MDPI

*Article*

# The Uncertainty-Certainty Matrix for Licensing Decision Making, Validation, Reliability, and Differential Monitoring Studies

**Richard Fiene**

**Edna Bennett Pierce Prevention Research Center,**
**The Pennsylvania State University**

**Correspondence: rfiene@rikinstitute.com**

**Abstract: This research article will propose the use of an uncertainty-certainty matrix for licensing decision making in the human services. It will show how the matrix can be used in rule decision making and how it clearly shows when decision making has gone awry. It is also used in making decisions in differential monitoring and in validation and reliability studies.**

**Keywords: Decision Making; Uncertainty; Regulatory Compliance; Licensing, Reliability and Validation Studies.**

## Introduction

This research proposal will take the Confusion Matrix which is a well-known metric in the decision-making research literature and refocus it on regulatory science within the context of the definition of regulatory compliance and licensing measurement. It will also deal with the policy implications of this particular metric. In this paper, it is proposed that the Uncertainty-Certainty Matrix (UCM) is a fundamental building block to licensing decision making. The 2 x 2 matrix has been utilized in regulatory compliance and is the center piece for determining licensing key indicator rules, but it is also a core conceptual framework in licensing measurement and ultimately in program monitoring and reviews.

Academic Editor: Firstname Lastname

Received: date
Revised: date
Accepted: date
Published: date

**Citation:** To be added by editorial staff during production.



The reason for selecting this matrix is the nature of licensing data, it is binary or nominal in measurement. Either a rule/regulation is in compliance or out of compliance. Presently most jurisdictions deal with regulatory compliance measurement in this nominal level or binary level. There is to be no gray area, this is a clear distinction in making a licensing decision about regulatory compliance. The UCM also takes the concept of Inter-Rater Reliability (IRR) a step further in introducing an uncertainty dimension that is very important in licensing decision making which is not as critical when calculating IRR. It is moving from an individual metric to a group metric (See Figures 1 & 2) involving regulatory compliance with rules.

The key pieces to the UCM are the following: the decision (D) regarding regulatory compliance and actual state (S) of regulatory compliance. Plus (+) = In-compliance or Minus (-) = Out of compliance. So, let's build the matrix:

Table 1: Uncertainty-Certainty Matrix (UCM) Logic Model

| **UCM Matrix Logic** | | **Decision (D) Regarding** | **Regulatory Compliance** |
|---|---|---|---|
| | | (+) In Compliance | (-) Not In Compliance |
| **Actual State (S) of** | (+) In Compliance | Agreement | Disagreement |
| **Compliance** | (-) Not In Compliance | Disagreement | Agreement |

The above UCM matrix demonstrates when agreement and disagreement occur which establishes a level of certainty (Agreement Cells) or uncertainty (Disagreement Cells). In a perfect world, there would only be agreements and no disagreements between the decisions made about regulatory compliance and the actual state of regulatory compliance. But from experience, this is not the case based upon reliability testing done in the licensing research field in which a decision is made regarding regulatory compliance with a specific rule or regulation and then that is verified by a second observer who generally is considered the measurement standard.

Disagreements raise concerns in general, but the disagreements are of two types: false positives and false negatives. A false positive is when a decision is made that a rule/regulation is out of compliance when it is in compliance. Not a good thing but its twin disagreement is worse where with false negatives it is decided that a rule/regulation is in compliance when it is out of compliance. False negatives need to be avoided because they place clients at extreme risk, more so than a false positive. False positives should also be avoided but it is more important to deal with the false negatives first before addressing the false positives.

## Methods

Let's look at this from a mathematical point of view in the following matrix (Table 2: UCM Math Model). In order to better understand the above relationships and determine when ameliorative action needs to occur to shore up the differences between the agreements and disagreements, it is easier to do this mathematically than trying to eyeball it.

Table 2: Uncertainty-Certainty Matrix (UCM) Math Model

| **UCM Matrix Math Model** | | **Decision (D) Regarding** | **Regulatory Compliance** | **Totals** |
|---|---|---|---|---|
| | | (+) In Compliance | (-) Not In Compliance | |
| **Actual State (S)** | (+) In Compliance | A | B | Y |
| **Of** | (-) Not In | C | D | Z |

| | | | | |
|---|---|---|---|---|
| **Compliance** | Compliance | | | |
| **Totals** | | W | X | |

Formulae based upon above Matrix in Table 2:

Agreements = (A)(D); Disagreements = (B)(C); Randomness = sqrt ((W)(X)(Y)(Z))

UCM Coefficient = ((A)(D)) - ((B)(C)) / sqrt ((W)(X)(Y)(Z)) in which a coefficient closer to 1 indicates agreement (certainty) and a coefficient closer to –1 indicates disagreement (uncertainty). A coefficient closer to 0 indicates randomness. Obviously, we want to see (A)(D) being predominant and very little in (B)(C) which are false positives and negatives where decisions and the actual state of regulatory compliance are not matching. If (WXYZ) is predominant then there is just randomness in the data. Also, not an intended result.

The reason for even suggesting this matrix is the high level of dissatisfaction with the levels of reliability in the results of program monitoring reviews as suggested earlier. If it were not so high, it would not be an issue; but with it being so high the field of licensing needs to take a proactive role in determining the best possible way to deal with increasing inter-rater reliability among licensing inspectors. Hopefully, this organizational schema via the UCM Matrix will help to think through this process related to licensing measurement and monitoring systems.

$$\boldsymbol{UCM} = \ll \boldsymbol{A} \times \boldsymbol{D} \gg - \ll \boldsymbol{B} \times \boldsymbol{C} \gg \div \sqrt{\ll \boldsymbol{W} \times \boldsymbol{X} \times \boldsymbol{Y} \times \boldsymbol{Z} \gg}$$

The above formula provides a means to calculate when action needs to be taken based upon the respective UCM coefficients. A UCM coefficient from +.25 to +1.00 is in the acceptable range; +.24 to -.24 is due to randomness and needs to be addressed with additional inter-rater reliability training; -.25 to –1.00 indicates a severe disagreement problem that needs to be addressed both in reliability training and a full review of the targeted rules/regulations to determine if the specific rule needs additional clarification.

Table 3: Uncertainty-Certainty Matrix (UCM) Licensing Decision Coefficient Ranges

| **UCM Coefficient** | **Licensing Decision** |
|---|---|
| +.25 to +1.00 | Acceptable, No Action Needed, In or Out of Regulatory Compliance Verified through mostly Agreements. (Generally, 90% of cases) |
| +.24 to -.24 | Random, Agreements + Disagreements, Needs Reliability Training. (Generally, 5% of cases) |
| -.25 to –1.00 | Unacceptable, Mostly Disagreements, Needs Training & Rule/Regulation Revision. (Generally, 5% of cases) |

Figures 1 and 2 provide the formulae for Inter-Rater Reliability (IRR) in figure 1 and the formulae used for calculating the Uncertainty-Certainty Coefficient (UCM):

**Figure 1: Kappa Coefficient**

Observed agreement

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

Expected agreement if random judgment

**Figure 2: Uncertainty-Certainty Coefficient**

$$\phi = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

$$\phi = \sqrt{\frac{\chi^2}{n}}$$

Let's provide an example of how this could work. A standard/rule/regulation that is common is the following:

Do all caregivers/teachers and children wash their hands often, especially before eating and after using the bathroom or changing diapers?

This is obviously an observation item where the licensing staff would observe in a sample of classrooms in a child care center for a set period of time. During their observations, there were several opportunities where the necessary behavior was required, and the staff complied with the rule and washed their hands. So, on the surface this specific rule was in compliance and there would appear to be full compliance with this rule based upon the observation.

A second scenario is where the observation is made, and the licensing staff observes the child care staff not washing their hands on several occasions. Then this specific rule would be out of compliance, and it would be duly noted by the licensing staff. These two scenarios establish a certain level of certainty during this observation session. However, there are other outcomes, for example, possibly one of the classrooms that was not observed had the opposite finding than what was observed in these particular classrooms. If data were being aggregated and a specific percentage was to be used the final decision about this rule could be different. Now we are getting into the uncertainty cells of the matrix where a false positive or negative could be the result. The licensing staff records the rule as being in compliance when in reality it is not = false negative or the rule is recorded as being out of compliance when in reality it is in compliance = false positive.

Another example which involves either Random Clinical Trials (RCT) or the use of abbreviated inspections (AI) and the results from these two interventions. The decision making in both RCT and AI is basically the same. We want to make sure that the results match reality. Every time an abbreviated review is done the following four regulatory compliance results should occur based upon the UCM matrix: 1) no additional random non-compliance is found; 2) there are no false negatives (abbreviated review finds no non-compliance but in reality there is); 3) when there is non-compliance found in abbreviated inspections, other related non-compliance is found; and 4) lastly the level of false positives (abbreviated review finds non-compliance but in reality there are no other related non-compliances) is kept to a minimum. This last result based upon copious research is that it is difficult to obtain but as the regulatory science moves forward hopefully this will become more manageable.

Hopefully these above examples provided some context for how the Uncertainty-Certainty Matrix (UCM) can be used in making specific licensing decisions based upon the regulatory compliance results which we will turn our attention to now.

# Results

## Uncertainty-Certainty Matrix for Validation and Reliability Studies

The purpose of this part of this research proposal is to explore the possibility of utilizing the Uncertainty-Certainty Matrix (UCM) in validation and reliability studies in licensing decision making. The UCM has been proposed for use in licensing decision making but this would be an extension of this thinking to studies that involve validating licensing decisions such as when key indicators are used in comparison with comprehensive reviews of rules, and in reliability studies to determine individual inspector bias in regulatory compliance.

The basic premise of the UCM is that individual decision-making matches reality. When it comes to regulatory compliance decision making a 2 x 2 matrix can be drawn with the possible outcomes as is indicated in the following table (Table 4).

**Table 4: Uncertainty-Certainty Matrix (UCM) Logic Model**

| **UCM Matrix Logic** | | **Decision Regarding** | **Regulatory Compliance** |
|---|---|---|---|
| | | (+) In Compliance | (-) Not In Compliance |
| **Actual State of** | (+) In Compliance | *Agreement (++)* | *Disagreement (+-)* |
| **Compliance** | (-) Not In Compliance | *Disagreement (-+)* | *Agreement (--)* |

In using this table, the hope is that the decision regarding regulatory compliance matches the actual state of compliance where the coefficient is as close to +1.0 as possible, in other words, perfect agreement. So, the agreement cells are heavily weighted. We do not want to see all the cells, both agreement and disagreement cells, equally weighted. That would indicate a random response rate and a coefficient close to 0.0.

But there is another possibility which involves bias on the part of the licensing inspector in which they have certain biases or tendencies when it comes to making regulatory compliance decisions about individual rules. So, it is possible that decisions made regarding regulatory compliance could be either overall (+) positive In-Compliance or (-) negative Not-In-Compliance when in reality the actual state of compliance is more random.

When this occurs, the coefficient falls off the range category and is not between 0 and +/-1.0 because there is no variance detected in the data. It is always biased either positively or negatively.

The UCM can be used for both reliability and validity testing as suggested in the above. Just look for different results. For validity, false positives and negatives should either be eliminated or reduced as well as possible and the remaining results should show the typical diagonal pattern as indicated by the agreement cells.

For reliability, the same pattern should be observed as in the validity testing above but there is an additional test in which bias is tested for. Bias will be ascertained if the patterns in the results indicate a horizontal or vertical pattern in the data with little or no diagonal indication. Bias can be found at the individual inspector level as well as at the standard level or the actual state of compliance.

In both reliability and validity testing, random results in which each of the cells are equally filled is not a desirable result either.

The following tables 5-10 depict the above relationships with results

highlighted in red:

Table 5: Valid and Reliable Results

| Valid & Reliable Results | (+) In Compliance | (-) Not In Compliance |
|---|---|---|
| (+) In Compliance | *Agreement (++)* | *Disagreement (+-)* |
| (-) Not In Compliance | *Disagreement (-+)* | *Agreement (--)* |

Table 6:  Random Results

| Random Results | (+) In Compliance | (-) Not In Compliance |
|---|---|---|
| (+) In Compliance | *Agreement (++)* | *Disagreement (+-)* |
| (-) Not In Compliance | *Disagreement (-+)* | *Agreement (--)* |

Table 7: Positive Bias Results Individual Assessor

| Positive Bias Results Individual | (+) In Compliance | (-) Not In Compliance |
|---|---|---|
| (+) In Compliance | *Agreement (++)* | *Disagreement (+-)* |
| (-) Not In Compliance | *Disagreement (-+)* | *Agreement (--)* |

Table 8: Negative Bias Results Individual Assessor

| Negative Bias Results Individual | (+) In Compliance | (-) Not In Compliance |
|---|---|---|
| (+) In Compliance | *Agreement (++)* | *Disagreement (+-)* |
| (-) Not In Compliance | *Disagreement (-+)* | *Agreement (--)* |

Table 9: Positive Bias Results Standard

| Positive Bias Results Standard | (+) In Compliance | (-) Not In Compliance |
|---|---|---|
| (+) In Compliance | *Agreement (++)* | *Disagreement (+-)* |
| (-) Not In Compliance | *Disagreement (-+)* | *Agreement (--)* |

Table 10: Negative Bias Results Standard

| Negative Bias Results Standard | (+) In Compliance | (-) Not In Compliance |
|---|---|---|
| (+) In Compliance | *Agreement (++)* | *Disagreement (+-)* |
| (-) Not In Compliance | *Disagreement (-+)* | *Agreement (--)* |

Tables 5 – 10 demonstrate the different results based upon individual response rates when making regulatory compliance decisions about rules. Table 5 is what needs to be attained and tables 6 – 10 need to be avoided. Only in table 5 are false negatives and positives eliminated or avoided. In tables 6 – 10, false negatives and/or false positives are introduced which is not desirable when making validity or reliability decisions.

Table 6 results clearly indicate that a great deal of randomness has been introduced in the regulatory compliance decision making in which the individual licensing inspector decisions do not match reality. Tables 7 and 8, demonstrate bias in the decision-making process either positively (inspector always indicates in compliance) or negatively (inspector always indicates out of compliance). It is also possible that the standard being used has bias built into it, this is less likely but is still a possibility. The results in Tables 9 and 10 demonstrate where this could happen.

All these scenarios need to be avoided and should be monitored by agency staff to determine if there are patterns in how facilities are being monitored.

## Uncertainty-Certainty Matrix for Differential Monitoring Studies

The purpose of this part of the research proposal is to explore the possibility of utilizing the Uncertainty-Certainty Matrix (UCM) not only in validation and reliability studies in licensing decision making but also with differential monitoring studies. The UCM has been proposed for use in licensing decision making but this would be an extension of this thinking to studies that involve validating licensing decisions such as when key indicators are used in comparison with comprehensive reviews of rules, and in the development of risk rules as part of the risk assessment methodology. This new Differential Monitoring 2x2 Matrix can also be used to depict the relationship between full and substantial regulatory compliance and the nature of rulemaking.

The basic premise of the DMM: Differential Monitoring Matrix is similar to the original thinking with the UCM but there are some changes in the formatting of the various cells in the matrix (see Table 11). When it comes to regulatory compliance decision making a 2 x 2 matrix can be drawn with the possible outcomes as is indicated in Table 11 where each individual rule is either in (+) or out (-) of compliance. Also, there is the introduction of a high regulatory compliant group (+) and a low regulatory compliant group (-) which is different from the original UCM.

**Table 11: DMM - Differential Monitoring Matrix**

| **DMM Matrix** | High Group (+) | Low Group (-) |
|---|---|---|
| (+) Rule is In Compliance | (++) | (+-) |
| (-) Rule is Not In Compliance | (-+) | (--) |

By utilizing the format of Table 11, several key components of differential monitoring can be highlighted, such as key indicators and risk assessment rules, as well as the relationship between full and substantial regulatory compliance.

Regulatory compliance is grouped into a high group (+), generally this means that there is either full or substantial regulatory compliance with all rules. The low group (-) usually has 10 or more regulatory compliance violations. Individual rules being in (+) or out (-) of regulatory compliance is self-explanatory.

Tables 12-18 below will demonstrate the following relationships:

Table 12 depicts the key indicator relationship between individual rules and the high/low groups as indicated in red. In this table, the individual rule is in compliance with the high group and is out of compliance with the low group. This result occurs on a very general basis and should have a .50 coefficient or higher with a p value of less than .0001.

Table 13 depicts what most rules look like in the 2x2 DMM. Most rules are always in full compliance since they are standards for basic health and safety for individuals. This is especially the case with rules that have been weighted as high-risk rules. Generally, one never sees non-compliance with these rules. There will be a substantial number of false positives (+-) found with high-risk rules but that is a good thing.

Table 14 depicts what happens when full compliance is used as the only criterion for the high group. Notice that the cell right below (++) is eliminated (-+). This is highly recommended since it eliminates false negatives (-+) from occurring in the high group. As will be seen in Table 15, when substantial compliance is used as part of the high group sorting, false negatives are re-introduced. If possible, this should be avoided, however in some cases because of the regulatory compliance data distribution it is not always possible where not enough full compliant programs are present.

Table 15 depicts what occurs when substantial compliance is used as part of determining the high group. False negatives can be reintroduced into the matrix which needs to be either eliminated or reduced as best as possible. If substantial compliance needs to be used in determining the high group, then there is a mathematical adjustment that can be made which will impact the equation and essentially eliminate false negatives mathematically (see the research note at the end of this research abstract).

Table 16 depicts what happens if the individual rule is particularly difficult to comply with. Both the high performers as well as the low performers are out of compliance with the rule.

Table 17 depicts a situation where the programs are predominantly in a low group with few at full or substantial regulatory compliance which is indicative of poor performing programs. Very honestly, this is generally not seen in the research literature, but it is a possibility and one to be in tune with.

Table 18 depicts a terrible individual rule which predicts just the opposite of what we are trying to do with programs. Obviously, this rule would need to be rewritten so that it fits with the essence of regulatory compliance in helping to protect individuals.

The following tables 12-18 will depict the above relationships with results

highlighted in red:

Table 12: Key Indicators

| **Key Indicators** | High Group (+) | Low Group (-) |
|---|---|---|
| (+) Rule is In Compliance | (++) | (+-) |
| (-) Rule is Not In Compliance | (-+) | (--) |

Table 13: Risk Rules

| **Risk Rules** | High Group (+) | Low Group (-) |
|---|---|---|
| (+) Rule is In Compliance | (++) | (+-) |
| (-) Rule is Not In Compliance | (-+) | (--) |

Table 14: Full Compliance with All Rules

| **Full Compliance** | High Group (+) | Low Group (-) |
|---|---|---|
| (+) Rule is In Compliance | (++) | (+-) |
| (-) Rule is Not In Compliance | | (--) |

Table 15: Substantial Compliance with All Rules

| **Substantial Compliance** | High Group (+) | Low Group (-) |
|---|---|---|
| (+) Rule is In Compliance | (++) | (+-) |
| (-) Rule is Not In Compliance | (-+) | (--) |

Table 16: Very Difficult Rules

| **Very Difficult Rule** | High Group (+) | Low Group (-) |
|---|---|---|
| (+) Rule is In Compliance | (++) | (+-) |
| (-) Rule is Not In Compliance | (-+) | (--) |

Table 17: Poor Performing Programs

| **Poor Performing Programs** | High Group (+) | Low Group (-) |
|---|---|---|
| (+) Rule is In Compliance | (++) | (+-) |
| (-) Rule is Not In Compliance | (-+) | (--) |

Table 18: Terrible Rule

| **Terrible Rule** | High Group (+) | Low Group (-) |
|---|---|---|
| (+) Rule is In Compliance | (++) | (+-) |
| (-) Rule is Not In Compliance | (-+) | (--) |

Tables 12 – 18 demonstrate the different results based on the relationship between individual regulatory compliance and if a program is either a high performer or a low performer. These tables are provided as guidance for understanding the essence of differential monitoring and regulatory compliance which has various nuances when it comes to data distributions. This research abstract hopefully can be used as a guide in determining from a data utilization point of view how to make important regulatory compliance policy decisions, such as: which rules are excellent key indicator rules, which are performing as high risk rules, importance of full compliance, what to do when substantial compliance needs to be employed, are there difficult rules to comply with, how well are our programs performing, and do we have less than optimal rules that are in need of revision.

## Discussion

Over the past decade in doing research on the Regulatory Compliance Key Indicator Metric (RCKIm) it has become very clear that false negatives needed to be controlled for because of their potential to increase morbidity and mortality. When dealing with regulatory compliance and full compliance as the threshold for the high grouping variable in the 2 x 2 Regulatory Compliance Key Indicator Matrix (RCKIm) (see matrix below in Table 19), false negatives could be either eliminated or reduced to the point of no concern.

However, if substantial compliance rather than full compliance is used as the threshold for the high grouping variable in the 2 x 2 Regulatory Compliance Key Indicator Matrix (RCKIm) this becomes a problem again. There is the need to introduce a weighting factor. In utilizing the RCKIm, the following equation/algorithm is used to produce the UCM Coefficient:

$$\mathbf{UCM = ((A)(D)) – ((B)(C)) / sqrt (WXYZ)}$$

This RCKIm needs to be revised/updated to the following to consider the need to again eliminate false negatives being generated by the results of the equation/algorithm; this can be accomplished by cubing B:

**UCM* = ((A)(D)) – ((B^3)(C)) / sqrt (WXYZ)**

By this simple adjustment to cube (B = False Negatives) it will basically eliminate the use of any results in which a false negative occurs when substantial compliance is determined. The table below (Table 19) displays the variables of the Regulatory Compliance Key Indicator Matrix (RCKIm).

| Table 19: RCKIm | High RC Group | RC Low Group | |
|---|---|---|---|
| KI In Compliance | A | B^3 | Y |
| KI Violations | C | D | Z |
| Totals | W | X | |

In the above examples, UCM can be used when the High RC Group is at full regulatory compliance, but UCM* needs to be used when the High RC Group is including substantial as well as full regulatory compliance. By using both equations/algorithms, it better deals with the results of the Regulatory Compliance Theory of Diminishing Returns.

The results should clearly show that only positive (+) coefficients will become Regulatory Compliance Key Indicators versus those rules that do not show any relationship to overall regulatory compliance (0), but now the negative (-) coefficients will more clearly show when any false negatives appear and clearly not include them as Regulatory Compliance Key Indicators. This is a major improvement in the Regulatory Compliance Key Indicator methodology which clearly demonstrates the differences in the results. It provides a gateway in regulatory compliance data distributions where substantial regulatory compliance is heavily present while full regulatory compliance is not. This could become a problem as the regulatory science field moves forward with the use of the Regulatory Compliance Theory of Diminishing Returns.

## Conclusion

The Uncertainty-Certainty Matrix provides a useful tool for assessing the effectiveness of licensing decision making in the human services via validation and reliability studies within differential monitoring systems. It is hoped that licensing researchers and regulatory scientists will experiment with it and test it out in different arenas. It appears to have broad applicability across regulatory disciplines. The matrix has helped to identify the need to address false positives and negatives in the human services licensing decision making process which undermines this effort.

## References

Fiene, R. (2019). A Treatise on Regulatory Compliance. *Journal of Regulatory Science*, *7*, 1-3. https://doi.org/10.21423/JRS-V07FIENE

Fiene, R. (2025). Finding the Rules that Work: An emerging paradigm promised to close the gap between regulatory compliance scores and the quality of childcare services, *American Scientist, Volume 113, January-February 2025, 16-21*.

Miranda, B., Ekyalongo, Y., Franchett, A., & Maxwell, K. (2022). *Monitoring Practices Used in Child Care and Early Education Licensing*. OPRE Report #2022-137. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Service.

Trivedi, P. A. (2015). *Innovation in monitoring in early care and education: Options for states*. Washington, DC: Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of health and Human Services.

| | | Predicted condition | | Sources: [12][13][14][15][16][17][18][19] | |
|---|---|---|---|---|---|
| | Total population $= \mathrm{P} + \mathrm{N}$ | **Predicted positive** | **Predicted negative** | Informedness, bookmaker informedness (BM) $= \mathrm{TPR} + \mathrm{TNR} - 1$ | Prevalence threshold (PT) $= \frac{\sqrt{\mathrm{TPR} \times \mathrm{FPR}} - \mathrm{FPR}}{\mathrm{TPR} - \mathrm{FPR}}$ |
| **Actual condition** | **Positive (P)** [a] | **True positive** (TP), hit[b] | **False negative** (FN), miss, underestimation | True positive rate (TPR), recall, sensitivity (SEN), probability of detection, hit rate, power $= \frac{\mathrm{TP}}{\mathrm{P}} = 1 - \mathrm{FNR}$ | False negative rate (FNR), miss rate type II error [c] $= \frac{\mathrm{FN}}{\mathrm{P}} = 1 - \mathrm{TPR}$ |
| **Actual condition** | **Negative (N)**[d] | **False positive** (FP), false alarm, overestimation | **True negative** (TN), correct rejection[e] | False positive rate (FPR), probability of false alarm, fall-out type I error [f] $= \frac{\mathrm{FP}}{\mathrm{N}} = 1 - \mathrm{TNR}$ | True negative rate (TNR), specificity (SPC), selectivity $= \frac{\mathrm{TN}}{\mathrm{N}} = 1 - \mathrm{FPR}$ |
| | Prevalence $= \frac{\mathrm{P}}{\mathrm{P} + \mathrm{N}}$ | Positive predictive value (PPV), precision $= \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} = 1 - \mathrm{FDR}$ | False omission rate (FOR) $= \frac{\mathrm{FN}}{\mathrm{TN} + \mathrm{FN}} = 1 - \mathrm{NPV}$ | Positive likelihood ratio (LR+) $= \frac{\mathrm{TPR}}{\mathrm{FPR}}$ | Negative likelihood ratio (LR−) $= \frac{\mathrm{FNR}}{\mathrm{TNR}}$ |
| | Accuracy (ACC) $= \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{P} + \mathrm{N}}$ | False discovery rate (FDR) $= \frac{\mathrm{FP}}{\mathrm{TP} + \mathrm{FP}} = 1 - \mathrm{PPV}$ | Negative predictive value (NPV) $= \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FN}} = 1 - \mathrm{FOR}$ | Markedness (MK), deltaP (Δp) $= \mathrm{PPV} + \mathrm{NPV} - 1$ | Diagnostic odds ratio (DOR) $= \frac{\mathrm{LR+}}{\mathrm{LR-}}$ |
| | Balanced accuracy (BA) $= \frac{\mathrm{TPR} + \mathrm{TNR}}{2}$ | $F_1$ score $= \frac{2\,\mathrm{PPV} \times \mathrm{TPR}}{\mathrm{PPV} + \mathrm{TPR}} = \frac{2\,\mathrm{TP}}{2\,\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}$ | Fowlkes–Mallows index (FM) $= \sqrt{\mathrm{PPV} \times \mathrm{TPR}}$ | Matthews correlation coefficient (MCC) $= \sqrt{\mathrm{TPR} \times \mathrm{TNR} \times \mathrm{PPV} \times \mathrm{NPV}} - \sqrt{\mathrm{FNR} \times \mathrm{FPR} \times \mathrm{FOR} \times \mathrm{FDR}}$ | Threat score (TS), critical success index (CSI), Jaccard index $= \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN} + \mathrm{FP}}$ |

a. the number of real positive cases in the data
b. A test result that correctly indicates the presence of a condition or characteristic
c. Type II error: A test result which wrongly indicates that a particular condition or attribute is absent
d. the number of real negative cases in the data
e. A test result that correctly indicates the absence of a condition or characteristic
f. Type I error: A test result which wrongly indicates that a particular condition or attribute is present

# Confusion matrices with more than two categories

Confusion matrix is not limited to binary classification and can be used in multi-class classifiers as well. The confusion matrices discussed above have only two conditions: positive and negative. For example, the table below summarizes communication of a whistled language between two speakers, with zero values omitted for clarity.[20]

**Classification Matrix & Sensitivity Analysis for Validating Licensing Key indicator Systems (Fiene, 2017)**

| | 1 | 2 | 3 | 5 | 7 | 8 | 10 | Comments |
|---|---|---|---|---|---|---|---|---|
| A | **1** | **1** | **1** | **0** | **0** | **1** | **1** | **Perfect** |
| B | **.52** | **.52** | **.52** | **.48** | **.48** | **.52** | **.04** | **Random** |
| C | **.71** | **.96** | **.94** | **.04** | **.29** | **.84** | **.70** | **False (-)** |
| D | **.94** | **.78** | **.71** | **.22** | **.06** | **.81** | **.70** | **False (+)** |
| E | **---** | **0** | **0** | **1** | **---** | **0** | **---** | **False +100%** |
| F | **0** | **0** | **0** | **1** | **1** | **0** | **-1** | **False+-100** |
| H | **.45** | **.46** | **.40** | **.54** | **.55** | **.46** | **-.08** | **Random** |

***Measures:***

***1 = Sensitivity*** ***TPR = TP / (TP + FN)***
***2 = Specificity*** ***SPC = TN / (FP + TN)***
***3 = Precision*** ***PPV = TP / (TP + FP)***
***5 = False Positive*** ***FPR = FP / (FP + TN)***
***7 = False Negative*** ***FNR = FN / (FN + TP)***
***8 = Accuracy*** ***ACC = (TP +TN) (P + N)***
***10 = Correlation*** ***((TP)(TN)) - ((FP)(FN)) / SQRT((TP + FP)(TP + FN)(TN + FP)(TN + FN))***

***PP = Predicted Positive = CI+***
***PN = Predicted Negaive = CI-***
***TP= True Positive = KI+***
***TN = True Negative =KI-***

| | ***TRUE POSITIVE (TP)(KI+)*** | ***TRUE NEGATIVE (TN)(KI-)*** |
|---|---|---|
| ***PREDICTED POSITIVE (PP)(CI+)*** | ***++*** | ***+-*** |
| ***PREDICTED NEGATIVE (PN)(CI-)*** | ***-+*** | ***--*** |

***CI+/CI-/KI+/KI-***
***A = 25/0/0/25 – Perfect match between CI and KI.***
***B = 13/12/12/13 – Random matching between CI and KI.***
***C = 17/7/1/25 – KI+ x CI- (False-)***
***D = 17/1/7/25 – KI- x CI+ (False+)***
***E = 0/0/50/0 – KI- x CI+ unlikely***
***F = 0/25/25/0 - False + & - 100% unlikely***
***H = 20/24/30/26 – Random matching between CI and KI.***