

4 STATISTICAL TERMINOLOGY

Authors of reports and studies on risk assessment tools adopt a range of descriptions of statistical outcomes. For the sake of consistency in the tables below, the RMA has adopted the following descriptors drawing from the most prevalently used terminology in the literature.

There is a useful explanation that uses an analogy of the bull's-eye on a dartboard: actually hitting the bull's-eye represents accuracy; landing shots together indicates good reliability. Considering this, hitting the bull's-eye and landing all the shots together would convey both accuracy and precision (<u>Viera and Garrett, 2005</u>).

Statistical Analyses

| | Definition | Interpretation |
|-------------------------|------------------------------|------------------------|
| Brier Scores | The Brier score is | The Brier score |
| | defined as the average | ranges from 0 to 1. |
| | quadratic difference | The best possible |
| | between the predicted | score is 0, indicating |
| | probability and the | total accuracy. The |
| | binary outcome. Its | lowest possible |
| | purpose is to measure | score is 1, meaning |
| | the 'calibration' of a set | that the forecast |
| | of probabilistic | was wholly |
| | predictions. | inadequate. |
| Intra Class Correlation | An ICC score | ICC values range |
| Coefficient (ICC) – | represents the | from 0 to 1 and are |
| descriptors; | estimation of the | typically reported |
| poor/moderate/excellent | correlation between two | with two decimal |
| | scores. It measures the | points, e.g75. |
| | magnitude of agreement | Cicchetti (1994) |
| | for inter-rater reliability. | recommends the |
| | | following thresholds: |
| | | • <.40 = 'poor' |
| | | • .40 to .75 = |
| | | 'moderate' |



| | | • .75 to 1.0 = |
|-------------------------|----------------------------|-----------------------|
| | | 'excellent' |
| Kappa (.) Coefficient – | The Kappa Coefficient | A Kappa is always |
| descriptors; | (.) measures the | less than or equal to |
| poor/average/excellent | agreement between two | a value of '1.' A |
| | individuals. | value of '1' implies |
| | | perfect agreement |
| | | and a value of '-1' |
| | | implies perfect |
| | | disagreement. |
| | | Recommended |
| | | thresholds are: |
| | | • < 0 Less than |
| | | chance |
| | | agreement |
| | | • 0.01– 0.20 poor |
| | | agreement |
| | | • 0.21– 0.60 |
| | | average |
| | | agreement |
| | | • 0.61– 0.99 |
| | | excellent |
| | | agreement |
| Effect sizes | The effect size | Effect sizes can be |
| | quantifies the size of the | interpreted in terms |
| | difference between two | of the percentiles or |
| | groups. It is calculated | ranks at which two |
| | as the standardised | distributions |
| | mean difference | overlap. |
| | between the two groups: | Interpretations of |
| | mean of Group A minus | effect sizes are |
| | mean of Group B; the | dependent on the |
| | total of which is divided | assumptions that |
| | | the two groups are |



| | by the standard | normally distributed |
|-----------------------------|----------------------------|-------------------------|
| | deviation. | and have the same |
| | | standard deviations. |
| Pearson Correlation | Pearson correlation | The values of r can |
| Coefficient– descriptors; | coefficient measures the | range from '-1' to '1,' |
| small/moderate/large | association between a | with '0' indicating |
| omaii/modorato/largo | predictor variable and | that there is no |
| | the outcome. | relationship |
| | the outcome. | between the |
| | | |
| | | predictor variable |
| | | and the outcome. |
| | | Positive values |
| | | indicate that the |
| | | high scores are |
| | | associated with |
| | | increased |
| | | recidivism; whereas |
| | | negative values |
| | | indicate that high |
| | | scores are |
| | | associated with |
| | | decreased |
| | | recidivism. |
| | | |
| | | According to Cohen |
| | | (1988), r values may |
| | | be interpreted as |
| | | follows: |
| | | .10 are small, |
| | | .25 are moderate, |
| | | .40 are large. |
| Sensitivity and Specificity | Sensitivity is the ability | The higher the value |
| Conditivity and Openholty | of a test to correctly | of sensitivity, the |
| | | _ |
| | classify an individual as | greater the ability of |
| | possessing a particular | the measure being |



| | characteristic (e.g. | tested to correctly |
|------------------------------|----------------------------|-----------------------------|
| | offending). Specificity is | identify individuals. |
| | the ability of a test to | For instance, a |
| | classify an individual as | sensitivity of 62% on |
| | not possessing a | a risk assessment |
| | particular characteristic | tool indicates that it |
| | (e.g. not offending). | has the ability to |
| | Sensitivity is inversely | correctly classify |
| | proportional with | just under two-thirds |
| | specificity, meaning that | of individuals who |
| | as the sensitivity | will reoffend. For |
| | increases, the specificity | specificity, higher |
| | decreases. | values indicate the |
| | | ability of a measure |
| | | to correctly identify |
| | | who will <i>not</i> possess |
| | | certain |
| | | characteristics (e.g. |
| | | who will not go on to |
| | | reoffend). |
| Z+ - descriptors; | Measures the | According to Cohen |
| small/moderate/large | association between the | (1988), d values |
| | predictor variable and | may be interpreted |
| | the outcome. These two | as follows: |
| | groups usually comprise | .20 is considered |
| | of the (1) recidivists and | 'small'; |
| | (2) the non-recidivists, | .50 is considered |
| | separated by the | 'moderate'; |
| | difference in their scores | .80 is considered |
| | obtained on risk | 'large'. |
| | assessments. | |
| Receiver Operating | In the context of risk | Several different |
| Characteristic (ROC) Curve – | assessment, the ROC | indicators can be |
| descriptors; | curve is a plot that | calculated from |
| low/moderate/high | shows the probability | ROC curves. The |



| | that a measure will | most commonly |
|----------------------------|-----------------------------|------------------------|
| | correctly identify | used indicator is the |
| | persons as recidivists or | 'area under the |
| | non recidivists | curve' (AUC) (see |
| | (Mossman, 1994; Rice | below). |
| | and Harris, 1995). | |
| | It is a plot of the 'hits' | |
| | (the proportion of | |
| | recidivists correctly | |
| | identified as recidivists) | |
| | against the 'false | |
| | alarms' (the proportion | |
| | of non-recidivists | |
| | identified as recidivists). | |
| Area Under the Curve (AUC) | The area under the | AUC values can |
| - descriptors; | curve (AUC) for the | range from '0' to '1'. |
| low/moderate/high | ROC curve is a useful | They can be |
| | summary statistic for the | interpreted as the |
| | extent to which a | probability that a |
| | measure discriminates | randomly-selected |
| | between recidivists and | recidivist has a |
| | non recidivists. | worse score than a |
| | | randomly-selected |
| | | non-recidivist. |
| | | Values between |
| | | '.51' and '1.0' |
| | | indicate positive |
| | | associations with |
| | | recidivism; whilst |
| | | values between '0' |
| | | and '.50' indicate |
| | | that predictions are |
| | | no better than |
| | | chance. |



| | Using <u>Cohen's</u> |
|-----|----------------------|
| | (1988) d values as a |
| | guide, AUC values |
| | may be interpreted |
| | as follows: |
| | .556 is considered |
| | 'low'; |
| | .639 is considered |
| | 'moderate'; |
| | .714 is considered |
| | 'high.' |
| - 1 | 1 |