NIDUS II Measurement Core: Optimizing Delirium Assessment in Research Proposals

Richard N. Jones, Sc.D. Brown University rich_jones@brown.edu

NIDUS Mentoring Session
October 12, 2022
Zoom
Overview

I. **What** reviewers are looking for in research proposals, with regard to choice in measurement instruments

II. **How** NIDUS resources can help giving reviewers what they want and strengthen the design of your research

III. **Strategies** for choosing an instrument for your proposed research
What reviewers are looking for

Part I
What reviewers want

• **Excellent science**
  - Strong designs that answer well-formed questions (approach, rigor)
  - Questions & answers that advance the field (significance, innovation)
  - Research designs that are ethical and feasible (approach, env., inv.)
  - Clarity and efficiency in presentation
How does this relate to delirium assessment?

• Delirium assessments should
  o Match with the goals of measurement
  o Match with the population being assessed
  o Match with the assessor
  o Have some validity evidence for research context
Match with goals of measurement

• Delirium case identification
• Delirium severity
  ▪ An episode of delirium, or severity of delirium during a stay?
  ▪ Symptom severity (peak of symptom count/sum; sum over all days)
  ▪ Duration of delirium during stay
Match with the population being assessed

- Type of patient
  - Capacity to participate in assessment
Match with the assessor

- Physician?
- Nurse?
- Other caregiver?
- Family?
- Lay interviewer?
Validity evidence for research context

• Has the instrument been used in patients similar to the planned population previously?
• Is there any validity evidence for the use of the chosen instrument in the planned research context?
Please remember reliability and validity statistics are sample-dependent and context-dependent results and do not describe immutable properties of a test.

Figure 2. Coupled forest plot of the 4AT. CI, confidence interval; 4AT, 4 ’A’s Test.

NIDUS resources that might be helpful

Part II
Measurement and Harmonization Core

NIDUS
NIDUS Pilot Awards: Letters of Intent

NIDUS II Pilot Award Letters of Intent (LOI) are due December 7, 2022

NIDUS is a collaborative, multidisciplinary network dedicated to the acceleration of scientific discovery in delirium research, through focused collaboration and creation of sustainable infrastructure to enhance innovative and high-quality research.
Measurement and Harmonization Core

The aims of the NIDUS Measurement and Harmonization Core are to develop a repository of core measures for delirium screening, diagnosis and severity, and to harmonize data across all types of delirium studies. These efforts will enhance the quality and consistency of measurement. It will also enable investigators to compare and share data across studies. Consequently, this allows for accelerated advances in delirium research.

For its harmonization aims, NIDUS will initially focus on harmonizing delirium severity measures. This project includes a systematic review of existing delirium severity measures. In addition, it includes statistical technique for data harmonization across measures. The harmonization method will include a modified Delphi consensus process. If interested in learning more about this effort, please contact us for more information on getting involved.

Learn more about the Measurement and Harmonization Core from Dr. Rich Jones:

Delirium Measurement Resources:

- **Info Cards**: 1-page informational cards summarizing key information and tool characteristics for commonly used delirium screening, diagnostic and severity assessments.
  - Adult Delirium Measurement Info Cards
  - Pediatric Delirium Measurement Info Cards

- **Delirium Severity Measure Crossover Tool**: A score conversion tool for 3 delirium severity instruments - the CAM-S, DR5-R-98, and NIDUS.
Info Cards: 1-page informational cards summarizing key information and test characteristics for commonly used delirium screening, diagnostic and severity assessments.

- Adult Delirium Measurement Info Cards
- Pediatric Delirium Measurement Info Cards

Delirium Severity Measure Crosswalk Tool
A score conversion tool for 3 delirium severity instruments - the CAM-S, DRS-R-98, and MDAS.

Delirium Identification Measures Crosswalk Tool
Linking between the CAM (short and long form), DOSS, DRS-R-98 (severity and total scores), and MDAS instruments.

Delirium Item Bank and Harmonization Tool
Statistical harmonization code (Stata version 16.1) to create the crosswalks and the Delirium Item Bank.

Delirium Severity Measurement Systematic Review Overview (PDF)
A brief overview of Jones et al. (2019) systematic review of delirium severity instruments and additional resources for delirium severity measurement.

Delirium Severity Measure Summary Table
Summary information on 14 delirium severity measures. These include number of items, approximate time to administer, certification or training required, and notes on background for development of the scale.

Delirium Severity Measure Symptom Coverage Table
Compares 17 commonly-used delirium severity measures by symptoms included in each measure.

In addition to these resources, the Measurement and Harmonization Core has several additional projects and publications in process. In time, these will be made available on this site.
Adult Delirium Measurement Info Cards

There is considerable variability in delirium measurement tools, with a wide variety of instruments for screening, diagnosis and severity available for clinical and research use. As a result, it can be difficult for researchers and clinicians to choose a tool that is most appropriate for their purpose. To assist with this issue, we created “info cards” that provide a standardized summary of commonly used measurement tools for delirium in adults that were identified through a systematic review and an online survey that we conducted in 2017. The instruments below include assessments for delirium identification as well as delirium severity.

Adult Delirium Measurement Tool Info Cards (PDFs)

- 3-Minute Diagnostic Confusion Assessment Method (3D-CAM)
- 4AT Rapid Clinical Test for Delirium (4AT)
- Bedside Confusion Scale (BCS)
- Brief Confusion Assessment Method (bCAM)
- Chart-Based Delirium Identification Instrument (CHART-DEL)
- Clinical Assessment of Confusion – A (CAC-A)
- Clinical Assessment of Confusion – B (CAC-B)
- Confusion Assessment Method (CAM)
### 3-Minute Diagnostic Confusion Assessment Method (3-D CAM)

**Purpose:** The tool is designed for use in clinical settings to assist healthcare providers in identifying and assessing confusion in hospitalized patients. It is based on the CAM (Confusion Assessment Method) and is intended to be used within the first 3 minutes of a patient encounter.

#### Instrument:
- **Name:** 3-D CAM
- **Sources:**
  - Dave L. Delirium Network: [deliriumnetwork.org/wp-content/uploads/2022/05/3D-CAM.pdf](https://deliriumnetwork.org/wp-content/uploads/2022/05/3D-CAM.pdf)

#### Features:
- A 6-item screening tool designed to be administered within the first 3 minutes of patient encounter.
- The tool assesses the presence of specific symptoms and changes in mental status.

#### Scoring:
- **Total:** 24 points
- **Minimum Score:** 3

#### Administration:
- **Primary use:** Screening
- **Number of questions:** 6
- **Description:** A short, structured interview to assess recent onset or fluctuating course of delirium.

#### Content:
- **Cognitive Testing:**
  - Orientation: patient's name, date, hospital, current symptoms.
  - Attention: ability to follow commands.
  - Language: ability to respond to questions.
  - Construction: ability to draw a simple object.

#### References:
<table>
<thead>
<tr>
<th>Instrument</th>
<th>3-Minute Diagnostic Confusion Assessment Method</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Acronym</strong></td>
<td>3D-CAM</td>
</tr>
<tr>
<td><strong>Primary use</strong></td>
<td>Screening</td>
</tr>
<tr>
<td><strong>Area assessed (Number of questions)</strong></td>
<td>Addresses 4 core features: Acute onset or fluctuating course (feature 1); Inattention (feature 2); Disorganized thinking (feature 3); Altered level of consciousness (feature 4) 10 interview questions, 10 observational items, 2 supplementary questions</td>
</tr>
<tr>
<td><strong>Description</strong></td>
<td>A short interview and rating scale that uses verbal responses and observations by the rater to rate the Confusion Assessment Method (CAM) diagnostic algorithm. The clinical version includes skip patterns that can shorten the instrument, while the research version is designed for systematic case-finding for delirium in a research setting and does not include skip patterns.</td>
</tr>
<tr>
<td>Version</td>
<td>2 (for clinical or research use)</td>
</tr>
<tr>
<td>---------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>Scoring Information</td>
<td>Considered positive if 3 out of 4 features are present (features 1 and 2, and either 3 or 4), according to the original CAM diagnostic algorithm. Each of the 20 items pertains to a specific feature and is coded either yes/no or correct/incorrect.</td>
</tr>
<tr>
<td>Cognitive testing</td>
<td>Cognitive testing is embedded within the 3D-CAM interview.</td>
</tr>
<tr>
<td>Estimated time to rate</td>
<td>3 minutes</td>
</tr>
<tr>
<td>Require trained rater</td>
<td>Yes – can be trained lay raters or clinicians</td>
</tr>
<tr>
<td>Administer to</td>
<td>Patient, in-person</td>
</tr>
<tr>
<td>How to obtain</td>
<td>Detailed free instructions (registration required) at <a href="http://hospitalelderlifeprogram.org">http://hospitalelderlifeprogram.org</a></td>
</tr>
<tr>
<td>Licensing Fee*</td>
<td>No charge for nonprofit or educational use</td>
</tr>
<tr>
<td>Languages available</td>
<td>English, Danish, Italian (clinical version only)</td>
</tr>
<tr>
<td>Highest COSMIN** rating</td>
<td>4/6†</td>
</tr>
<tr>
<td>Languages available</td>
<td>English, Danish, Italian (clinical version only)</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Highest COSMIN** rating</td>
<td>4/6†</td>
</tr>
<tr>
<td>Test Performance Characteristics</td>
<td>Marcantonio 2014</td>
</tr>
<tr>
<td></td>
<td><strong>Reference standard:</strong> diagnosis by clinical psychologists and practice nurses based on face-to-face interview, medical record review, input from nurse and family members, Montreal Cognitive Assessment (MoCA), Geriatric Depression Scale (GPS) and adjudicated by study panel using DSM-IV criteria</td>
</tr>
<tr>
<td></td>
<td>• Reliability (inter-rater agreement 95%)</td>
</tr>
<tr>
<td></td>
<td>• Sensitivity (compared to reference standard, 95% [95% CI of 84-99%])</td>
</tr>
<tr>
<td></td>
<td>• Specificity (compared to reference standard, 94% [95% CI of 90-97%])</td>
</tr>
</tbody>
</table>

* Fees and licensing information is effective as of 2018, but is subject to change over time

** COSMIN is used to rate a study's evaluation of a survey or test's measurement properties. COSMIN does NOT rate the instrument itself, but helps readers understand if they can have confidence in the results of studies evaluating measurement properties of surveys and tests. For example, a rigorous study evaluating a test with poor measurement properties will receive a “good” COSMIN rating, while a poorly-conducted study evaluating a test with good measurement properties will receive a “poor” COSMIN rating. Small sample size can impact all COSMIN ratings. You must consider both the COSMIN rating and the results of studies provided when forming your opinion about that test. COSMIN ratings shown are based solely on the instrument's original validation study.

† COSMIN breakdown: content validity: GOOD, effect indicators: GOOD, internal consistency: NONE, inter-rater reliability: GOOD, construct validity: NONE, external validity: GOOD

**Reference:**
COSMIN

Consensus-based Standards for the selection of health Measurement Instruments


The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes.

Lidwine B. Mokkink, Caroline B. Terwee, Donald L. Patrick, Jordi Alonso, Paul W. Stratford, Dirk L. Knol, Lex M. Bouter, Henrica C.W. de Vet

*Department of Epidemiology and Biostatistics, The EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, The Netherlands*

1Department of Health Services, University of Washington, Seattle, WA, USA

2Health Services Research Unit, Institut Municipal d’Investigacio Medic (IMIM-Hospital del Mar), Barcelona, Spain

3Ciber de Epidemiologia y Salud Publica (CIBERESP), Barcelona, Spain

4Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Canada

5School of Rehabilitation Science, McMaster University, Hamilton, Canada

6Executive Board of VU University Amsterdam, Amsterdam, The Netherlands

Accepted 5 February 2010
Reliability
Validity
Responsiveness†

† Interpretability was thought to be important but not a measurement property, *per se*

Fig. 2. COSMIN taxonomy of relationships of measurement properties. *Abbreviations:* COSMIN, COnsensus-based Standards for the selection of health Measurement INstruments; HR-PRO, health related-patient reported outcome.

Reliability  The degree to which the measurement is free from measurement error

Validity  The degree to which [the] instrument measures the construct(s) it purports to measure

Responsiveness  The ability of [the] instrument to detect change over time in the construct to be measured

Interpretability  The degree to which one can assign qualitative meaning -- that is, clinical or commonly understood connotations -- to an instrument’s quantitative scores or change in scores.
### Validity

**Content validity**
The degree to which the content of [the] instrument is an adequate reflection of the construct to be measured

**Hypothesis testing**
The degree to which the scores of [the] instrument are consistent with hypotheses (for instance with regard to internal relationships, relationships to scores of other instruments, or differences between relevant groups) based on the assumption that the instrument validly measures the construct to be measured

### Reliability

**Internal consistency**
The degree of the interrelatedness among the items
Four step procedure for completing the checklist
When completing the COSMIN checklist, four steps should be taken (Figure 2), which will be further explained in the next paragraphs.

Figure 2. Four-step procedure for completing the COSMIN checklist.

Box B. Reliability: relative measures (including test-retest reliability, inter-rater reliability and intra-rater reliability)

<table>
<thead>
<tr>
<th>Design r</th>
<th>yes</th>
<th>no</th>
<th>?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 War:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 War:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 War:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Wer:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Wer:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 War:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 Wer:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 Wer:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 Wer:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 Wer:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 for r</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 for r</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13 for ordinal scores: Was a weighted kappa calculated?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 for ordinal scores: Was the weighting scheme described? e.g. linear, quadratic</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Effect indicators
A "COSMIN-guided" review of measurement properties

Does the scale consist ENTIRELY of effect indicators?
Effect indicators are CAUSED by delirium.
Effect indicators are appropriate for use in a reflective measurement model.
Cause or formative indicators are factors that might be risk factors for, or otherwise determine levels of, delirium or delirium severity.
Acknowledging that the pathophysiology of delirium is imperfectly understood, please use your best judgement.
Content validity
A “COSMIN-guided” review of measurement properties

Do the authors describe procedures for ensuring that all items refer to relevant aspects of delirium or delirium severity?

Content validity describes the extent to which the items included in a scale sample from the potential universe of possible questions that could be used to assess the target construct.

In the COSMIN framework (Mokkink et al 2010 BMC Med Res Meth https://doi.org/10.1186/1471-2288-10-22) the assessment of content validity is a judgement of the (a) relevance and (b) comprehensiveness of the items. Relevance refers to the match of questions to the target population. Comprehensiveness refers to the extent that the items included address the breadth of the domain or construct being assessed, and the clarity with which those constructs and or domains are defined. Both aspects of content validity are to be defended and adjudged with content expertise.
Internal consistency
A "COSMIN-guided" review of measurement properties

If internal consistency reliability was reported (e.g., Cronbach's alpha) what was the estimate?
If an internal consistency statistic was reported, please put check the other box and type of statistic in the other bullet. (e.g., "0.91 Cronbach's alpha").

What was the sample size used for internal consistency estimation?
Inter-rater reliability

A “COSMIN-guided” review of measurement properties

Inter-rater reliability refers to assessments of the agreement of two or more raters when making ratings on a single patient or research participant.

Was inter-rater reliability assessed?

What was the sample size for the assessment of inter-rater reliability?
Convergent validity
A "COSMIN-guided" review of measurement properties

Convergent validity describes the extent to which measure of a construct correlates with other measures of the same construct.

Sometimes this is called "Construct validity". However, nowadays we take all aspects of validity as evidence of construct validity.

What was the sample size for the assessment of convergent validity?
Criterion validity, Predictive validity, or Responsiveness

A “COSMIN-guided” review of measurement properties

This section contains items grouped under the heading "Hypothesis testing" in the COSMIN framework. Evidence for criterion validity would be relationship of a test score to a reference standard (e.g., a Psychiatrist's diagnosis). Evidence for Predictive validity would come from the prediction of a clinically relevant outcome (e.g., death, length of stay, costs). Evidence for Responsiveness would be something like the measure is sensitive to change due to treatments or risk factors for the target condition.

Does the manuscript contain a description of the instrument's ability to associate predictably with external criteria, an outcome, or be influenced by a treatment or group with known difference on delirium. This includes a priori hypothesized mean differences and correlations with external variables.

What was the sample size used for describing criterion validity, predictive validity, or responsiveness. If more than one sample size would be appropriate, report the largest.
### CAM

<table>
<thead>
<tr>
<th></th>
<th>Number evaluated</th>
<th>Prevalence of delirium</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Del+ Test+</td>
<td>205</td>
<td>0.25</td>
<td>0.82</td>
<td>0.99</td>
</tr>
<tr>
<td>Del- Test-</td>
<td>45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>250</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 4AT

<table>
<thead>
<tr>
<th></th>
<th>Number evaluated</th>
<th>Prevalence of delirium</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Del+ Test+</td>
<td>220</td>
<td>0.25</td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td>Del- Test-</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>250</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 3D-CAM

<table>
<thead>
<tr>
<th></th>
<th>Number evaluated</th>
<th>Prevalence of delirium</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Del+ Test+</td>
<td>238</td>
<td>0.25</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td>Del- Test-</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>250</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Positive predictive value: CAM 0.97, 4AT 0.71, 3D-CAM 0.84.
Negative predictive value: CAM 0.94, 4AT 0.96, 3D-CAM 0.98.
A' (AUC 1-point ROC): CAM 0.95, 4AT 0.93, 3D-CAM 0.97.
Level of test (prop screen +): CAM 0.21, 4AT 0.31, 3D-CAM 0.28.


Scoring
A “COSMIN-guided” review of measurement properties

Assign 1 point if each of (1) CONTENT VALIDITY, (2) ALL EFFECT INDICATORS, (3) INTERNAL CONSISTENCY, (4) any aspect of RELIABILITY (other than internal consistency, e.g., inter-rater), (5) CONVERGENT VALIDITY or construct validity, and (6) CRITERION validity or predictive validity or external validity were assessed.

Subtract 1/2 point if INTERNAL CONSISTENCY was based on fewer than 50 observations,
Subtract 1/2 point if RELIABILITY was based on less than 50 observations
Subtract 1/2 point if CONVERGENT validity was based on less than 50 persons,
Subtract 1/2 point if CRITERION validity was based on less than 50 persons.
NIDUS Measurement core COSMIN rating

• Is a very high-level summary of the original publication describing the instrument

• Does not reflect any validation research subsequent to the original publication

• Only partially represents the full COSMIN framework

• Might be unfairly applied to instruments described before the circa 2010 COSMIN framework was described
Strategies for choosing an instrument

Part III
Feasibility

• What instrument(s) is/are used in your lab/hospital/city by mentors/collaborators?
• Do you have access to training or other resources to make effective use of the instruments?
Reliability & Validity

• Are the instruments suitable for the target population?
• Do you have the right assessors?
• Has the instrument be used in your target population previously?
• With success?
• Do instruments maximize sensitivity and specificity in a way most beneficial to your question?
Final thought

• If you would like to know which of two or more instruments is the “best” for your target population (sensitivity, specificity, predictive value, reliability)

• The only **trustworthy** data to inform this decision would be
  - Head-to-head comparison in same sample (e.g., randomized design)
  - Individual participant data meta-analysis (mega-analysis)

• Individual (but separately conducted) studies and meta-analyses are not directly comparable (selection of patients, other design and analysis choices), publication bias, etc.
Measurement and Harmonization Core:
On the web

https://deliriumnetwork.org/measurement/

Questions
Rich Jones
Rich_jones@brown.edu

NIDUS
nidus@hsl.harvard.edu
### CAM

<table>
<thead>
<tr>
<th>Del+</th>
<th>Del-</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test+</td>
<td>198</td>
<td>15</td>
</tr>
<tr>
<td>Test-</td>
<td>52</td>
<td>735</td>
</tr>
<tr>
<td>Total</td>
<td>250</td>
<td>750</td>
</tr>
</tbody>
</table>

- Positive predictive value: 0.93
- Negative predictive value: 0.93
- A' (AUC 1-point ROC): 0.94
- Level of test (prop screen +): 0.21
- Hit rate (correct decision): 0.93
- Efficiency: 0.27
- Kraemer's k(1,0): 0.73
- Kraemer's k(0,0): 0.90
- Kraemer's k(5,0): 0.81

### 4AT

<table>
<thead>
<tr>
<th>Del+</th>
<th>Del-</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test+</td>
<td>220</td>
<td>90</td>
</tr>
<tr>
<td>Test-</td>
<td>30</td>
<td>660</td>
</tr>
<tr>
<td>Total</td>
<td>250</td>
<td>750</td>
</tr>
</tbody>
</table>

- Positive predictive value: 0.71
- Negative predictive value: 0.96
- A' (AUC 1-point ROC): 0.93
- Level of test (prop screen +): 0.31
- Hit rate (correct decision): 0.88
- Efficiency: 0.33
- Kraemer's k(1,0): 0.83
- Kraemer's k(0,0): 0.61
- Kraemer's k(5,0): 0.70

### 3D-CAM

<table>
<thead>
<tr>
<th>Del+</th>
<th>Del-</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test+</td>
<td>238</td>
<td>45</td>
</tr>
<tr>
<td>Test-</td>
<td>12</td>
<td>705</td>
</tr>
<tr>
<td>Total</td>
<td>250</td>
<td>750</td>
</tr>
</tbody>
</table>

- Positive predictive value: 0.84
- Negative predictive value: 0.98
- A' (AUC 1-point ROC): 0.97
- Level of test (prop screen +): 0.28
- Hit rate (correct decision): 0.94
- Efficiency: 0.34
- Kraemer's k(1,0): 0.93
- Kraemer's k(0,0): 0.79
- Kraemer's k(5,0): 0.85

### Random Forest using EHR

- Number evaluated: 1000
- Sensitivity: 0.95
- Specificity: 0.94

<table>
<thead>
<tr>
<th>Del+</th>
<th>Del-</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test+</td>
<td>195</td>
<td>90</td>
</tr>
<tr>
<td>Test-</td>
<td>55</td>
<td>660</td>
</tr>
<tr>
<td>Total</td>
<td>250</td>
<td>750</td>
</tr>
</tbody>
</table>

- Positive predictive value: 0.68
- Negative predictive value: 0.92
- A' (AUC 1-point ROC): 0.90
- Level of test (prop screen +): 0.29
- Hit rate (correct decision): 0.86
- Efficiency: 0.30
- Kraemer's k(1,0): 0.69
- Kraemer's k(0,0): 0.59
- Kraemer's k(5,0): 0.63


Kraemer’s k(1,0) is chance-corrected sensitivity (or chance-corrected positive predictive value); Kraemer’s k(0,0) is chance-corrected specificity (or chance-corrected negative predictive value); Kraemer’s k(5,0) is the weighted average of the k(1,0) and k(0,0) and is also known as Cohen’s kappa when sensitivity and specificity are equally weighted.