BASIL Reference Standard (Clinical Assessor)

Doug Tommet and Rich Jones for Anna MacKay-Brandt, Tamara G Fong, and the BASIL-II team

2025-05-30

## Aims of manuscript:

1. Describe clinical assessor inter-rater reliability for *delirium severity* ratings:
   1. Overall agreement among clinical assessors
      1. Before consensus meeting *(for generalizability - this is the result that is [more or less] generalizable to other research studies using a single clinical rater for delirium)*
      2. After consensus meeting *(this is the result that describes the reliability of the clinical assessor rating of delirium in the BASIL-II study)*.
   2. Variability in agreement as a function delirium severity (Bland-Altman plots)
2. Explore if there are *delirium features* that are associated with delirium severity inter-rater agreement (e.g., features present include Inattention, Disorganized Thinking, Altered Level of Consciousness, Disorientation, Memory Impairment, Perceptual Disturbances, Psychomotor Agitation, Psychomotor Retardation, Altered Sleep-Wake, Inappropriate Behavior, Delusions).
3. Explore if there are *patient characteristics* that may influence delirium severity inter-rater agreement (e.g., pre-acute cognitive status (NC, MCI, Dementia), three distinct study sites: (1) acute medical inpatients; (2) scheduled surgical inpatients, and (3) long-term and post-acute care residents; informant (family member/friend, nurse, CNA), pre-acute depression)

Note: Aims of manuscript copied from Statistical Analysis Plan ([GDoc](https://docs.google.com/document/d/1GPIzetmxAJxbsQnpXG5_sj9buN0inlTDNljtyg4V1vE/edit?usp=sharing))

## This analysis report (30 May 2025)

* This report includes heatmaps and Bland-Altman figures for rater agreement.
* **ACTION**: Need to create IPTW weights for selection into the sample
* **ACTION**: Need to generate VDSs for ADL and IADL characteristics
* **ACTION**: Replace REDCap PHQ, MoCA, and AD-8 summary measures with VDS versions
* **ACTION** Aim 2 additionally requires individual CAM-S delirium features (inattention, disorganized thinking, altered level of consciousness, disorientation, memory impairment, perceptual disturbances, psychomotor agitation, psychomotor retardation, altered sleep-wake, inappropriate behavior, delusions), need to modify vds for CAM-S to generate and save individual indicators for the CAM features (0/1)

# Table 1

| **Characteristic** | **Overall** N = 46*1* | **BIDMC** N = 10*1* | **UF** N = 14*1* | **TNJH** N = 22*1* |
| --- | --- | --- | --- | --- |
| Age | 80.0 (6.2) | 80.1 (7.2) | 77.4 (3.9) | 81.5 (6.7) |
| Female |  |  |  |  |
| Male | 19 (41%) | 4 (40%) | 7 (50%) | 8 (36%) |
| Female | 27 (59%) | 6 (60%) | 7 (50%) | 14 (64%) |
| Non-white | 15 (33%) | 2 (20%) | 1 (7.1%) | 12 (55%) |
| Hispanic ethnicity |  |  |  |  |
| Not Hispanic | 46 (100%) | 10 (100%) | 14 (100%) | 22 (100%) |
| Hispanic | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) |
| English is second language | 1 (2.2%) | 0 (0%) | 0 (0%) | 1 (4.5%) |
| Education | 14.3 (3.2) | 14.5 (2.7) | 14.7 (2.6) | 14.0 (3.7) |
| Unknown | 1 | 0 | 0 | 1 |
| Lives alone | 9 (20%) | 5 (50%) | 4 (29%) | 0 (0%) |
| Married | 15 (33%) | 2 (20%) | 10 (71%) | 3 (14%) |
| Lives in NH | 22 (48%) | 0 (0%) | 0 (0%) | 22 (100%) |
| Vision impairment | 16 (39%) | 2 (20%) | 3 (21%) | 11 (65%) |
| Unknown | 5 | 0 | 0 | 5 |
| Hearing impairment | 14 (30%) | 2 (20%) | 4 (29%) | 8 (36%) |
| Charlson Comorbidity Index |  |  |  |  |
| 0 | 11 (24%) | 2 (20%) | 4 (31%) | 5 (23%) |
| 1 | 3 (6.7%) | 1 (10%) | 2 (15%) | 0 (0%) |
| 2 | 10 (22%) | 2 (20%) | 3 (23%) | 5 (23%) |
| 3 | 6 (13%) | 2 (20%) | 2 (15%) | 2 (9.1%) |
| 4 | 4 (8.9%) | 0 (0%) | 0 (0%) | 4 (18%) |
| 5 | 6 (13%) | 2 (20%) | 1 (7.7%) | 3 (14%) |
| 8 | 5 (11%) | 1 (10%) | 1 (7.7%) | 3 (14%) |
| Unknown | 1 | 0 | 1 | 0 |
| Major neurocognitive disorder (GS) | 27 (59%) | 3 (30%) | 3 (21%) | 21 (95%) |
| Minor neurocognitive disorder (GS) | 9 (20%) | 5 (50%) | 3 (21%) | 1 (4.5%) |
| Delirium diagnosis (GS) | 21 (46%) | 6 (60%) | 6 (43%) | 9 (41%) |
| Delirium severity (GS) |  |  |  |  |
| 0 | 21 (46%) | 3 (30%) | 8 (57%) | 10 (45%) |
| 1 | 4 (8.7%) | 1 (10%) | 0 (0%) | 3 (14%) |
| 2 | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) |
| 3 | 2 (4.3%) | 1 (10%) | 1 (7.1%) | 0 (0%) |
| 4 | 7 (15%) | 3 (30%) | 3 (21%) | 1 (4.5%) |
| 5 | 4 (8.7%) | 1 (10%) | 0 (0%) | 3 (14%) |
| 6 | 5 (11%) | 1 (10%) | 1 (7.1%) | 3 (14%) |
| 7 | 2 (4.3%) | 0 (0%) | 1 (7.1%) | 1 (4.5%) |
| 8 | 1 (2.2%) | 0 (0%) | 0 (0%) | 1 (4.5%) |
| 9 | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) |
| 10 | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) |
| Dementia severity (GS) |  |  |  |  |
| No dementia | 10 (22%) | 2 (20%) | 8 (57%) | 0 (0%) |
| Very Mild | 9 (20%) | 5 (50%) | 3 (21%) | 1 (4.5%) |
| Mild | 17 (37%) | 3 (30%) | 2 (14%) | 12 (55%) |
| Moderate | 10 (22%) | 0 (0%) | 1 (7.1%) | 9 (41%) |
| Severe | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) |
| vdcamsrs | 3.22 (2.60) | 4.20 (2.82) | 0.71 (0.73) | 4.36 (2.17) |
| PHQ-9 score | 5.5 (3.8) | 5.6 (2.7) | 4.9 (2.7) | 5.9 (4.7) |
| MOCA score | 17 (6) | 20 (5) | 18 (6) | 14 (6) |
| AD8 total score |  |  |  |  |
| 0 | 25 (54%) | 4 (40%) | 8 (57%) | 13 (59%) |
| 1 | 4 (8.7%) | 1 (10%) | 2 (14%) | 1 (4.5%) |
| 2 | 8 (17%) | 2 (20%) | 2 (14%) | 4 (18%) |
| 3 | 2 (4.3%) | 0 (0%) | 1 (7.1%) | 1 (4.5%) |
| 4 | 3 (6.5%) | 1 (10%) | 0 (0%) | 2 (9.1%) |
| 5 | 1 (2.2%) | 0 (0%) | 0 (0%) | 1 (4.5%) |
| 7 | 2 (4.3%) | 2 (20%) | 0 (0%) | 0 (0%) |
| 8 | 1 (2.2%) | 0 (0%) | 1 (7.1%) | 0 (0%) |
| *1*Mean (SD); n (%) | | | | |

**?(caption)**

# Rater agreement

## Delirium severity

|  | Delirium Severity (Rater A) | | | | | | | | | |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Total |
| Delirium Severity (Rater B) |  |  |  |  |  |  |  |  |  |  |  |
| 0 | 218 | 57 | 40 | 18 | 2 | 3 | 0 | 0 | 0 | 0 | 338 |
| 1 | 60 | 31 | 19 | 17 | 7 | 7 | 0 | 0 | 0 | 0 | 141 |
| 2 | 7 | 7 | 5 | 5 | 1 | 0 | 0 | 0 | 0 | 0 | 25 |
| 3 | 14 | 6 | 7 | 2 | 7 | 8 | 3 | 1 | 0 | 0 | 48 |
| 4 | 14 | 12 | 5 | 12 | 24 | 18 | 4 | 7 | 0 | 0 | 96 |
| 5 | 0 | 1 | 0 | 4 | 22 | 36 | 30 | 17 | 3 | 5 | 118 |
| 6 | 0 | 1 | 0 | 1 | 14 | 27 | 21 | 22 | 6 | 6 | 98 |
| 7 | 0 | 0 | 0 | 0 | 3 | 16 | 19 | 13 | 2 | 5 | 58 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 7 | 9 | 19 |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 5 | 4 | 13 |
| Total | 313 | 115 | 76 | 59 | 80 | 115 | 82 | 62 | 23 | 29 | 954 |

**?(caption)**

|  |
| --- |
| Figure 1: Heatmap showing agreement of paired initial ratings of delirium severity |

|  | Delirium Severity (Rater A) | | | | | | | | | |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Total |
| Delirium Severity (Rater B) |  |  |  |  |  |  |  |  |  |  |  |
| 0 | 212 | 54 | 42 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 308 |
| 1 | 51 | 55 | 28 | 16 | 0 | 0 | 0 | 0 | 0 | 0 | 150 |
| 2 | 14 | 17 | 26 | 6 | 9 | 0 | 0 | 0 | 0 | 0 | 72 |
| 3 | 0 | 10 | 4 | 11 | 7 | 0 | 0 | 0 | 0 | 0 | 32 |
| 4 | 0 | 0 | 7 | 5 | 31 | 16 | 5 | 0 | 0 | 0 | 64 |
| 5 | 0 | 0 | 0 | 1 | 18 | 45 | 54 | 8 | 0 | 0 | 126 |
| 6 | 0 | 0 | 0 | 0 | 5 | 34 | 51 | 16 | 0 | 0 | 106 |
| 7 | 0 | 0 | 0 | 0 | 0 | 18 | 24 | 12 | 5 | 3 | 62 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 9 | 10 | 21 |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 7 | 4 | 13 |
| Total | 277 | 136 | 107 | 39 | 70 | 113 | 134 | 40 | 21 | 17 | 954 |

**?(caption)**

|  |
| --- |
| Figure 2: Heatmap showing agreement of paired ratings of delirium severity after consensus process |

The correlation between rater A and rater B delirium severity ratings (0-9) prior to the consensus process was 0.84. After the consensus process, the correlation was 0.93. The kappa statistic went from 0.84 (0.82, 0.86) to 0.93 (0.92, 0.94).

## Delirium presence

|  | Delirium Presence (Rater A) | |  |
| --- | --- | --- | --- |
|  | No | Yes | Total |
| Delirium Presence (Rater B) |  |  |  |
| No | 480 | 47 | 527 |
| Yes | 60 | 367 | 427 |
| Total | 540 | 414 | 954 |

**?(caption)**

|  | Delirium Presence (Rater A) | |  |
| --- | --- | --- | --- |
|  | No | Yes | Total |
| Delirium Presence (Rater B) |  |  |  |
| No | 474 | 41 | 515 |
| Yes | 48 | 391 | 439 |
| Total | 522 | 432 | 954 |

**?(caption)**

The correlation between rater A and rater B delirium presence ratings (Yes/No) prior to the consensus process was 0.77. After the consensus process, the correlation was 0.81. The kappa statistic went from 0.77 (0.73, 0.81) to 0.81 (0.77, 0.85).

## Dementia severity

|  | Dementia Severity (Rater A) | | | | |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | No dementia | Very mild | Mild | Moderate | Severe | Total |
| Dementia Severity (Rater B) |  |  |  |  |  |  |
| No dementia | 170 | 57 | 10 | 0 | 0 | 237 |
| Very mild | 59 | 111 | 88 | 12 | 0 | 270 |
| Mild | 5 | 41 | 140 | 72 | 0 | 258 |
| Moderate | 0 | 1 | 35 | 110 | 32 | 178 |
| Severe | 0 | 0 | 0 | 6 | 5 | 11 |
| Total | 234 | 210 | 273 | 200 | 37 | 954 |

**?(caption)**

|  |
| --- |
| Figure 3: Heatmap showing agreement of paired initial ratings of dementia severity |

|  | Dementia Severity (Rater A) | | | | |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | No dementia | Very mild | Mild | Moderate | Severe | Total |
| Dementia Severity (Rater B) |  |  |  |  |  |  |
| No dementia | 180 | 41 | 8 | 0 | 0 | 229 |
| Very mild | 45 | 124 | 77 | 3 | 0 | 249 |
| Mild | 2 | 36 | 204 | 47 | 0 | 289 |
| Moderate | 0 | 1 | 21 | 117 | 31 | 170 |
| Severe | 0 | 0 | 0 | 5 | 12 | 17 |
| Total | 227 | 202 | 310 | 172 | 43 | 954 |

**?(caption)**

|  |
| --- |
| Figure 4: Heatmap showing agreement of paired ratings of dementia severity after consensus process |

The correlation between rater A and rater B dementia severity ratings (No dementia/Very mild/Mild/Moderate/Severe) prior to the consensus process was 0.81. After the consensus process, the correlation was 0.86. The kappa statistic went from 0.8 (0.77, 0.82) to 0.85 (0.83, 0.87).

## NCD presence

|  | NCD Presence (Rater A) | | |  |
| --- | --- | --- | --- | --- |
|  | No NCD | Minor NCD | Major NCD | Total |
| NCD Presence (Rater B) |  |  |  |  |
| No NCD | 174 | 59 | 7 | 240 |
| Minor NCD | 59 | 116 | 100 | 275 |
| Major NCD | 4 | 47 | 388 | 439 |
| Total | 237 | 222 | 495 | 954 |

**?(caption)**

|  | NCD Presence (Rater A) | | |  |
| --- | --- | --- | --- | --- |
|  | No NCD | Minor NCD | Major NCD | Total |
| NCD Presence (Rater B) |  |  |  |  |
| No NCD | 180 | 41 | 8 | 229 |
| Minor NCD | 45 | 126 | 81 | 252 |
| Major NCD | 2 | 38 | 433 | 473 |
| Total | 227 | 205 | 522 | 954 |

**?(caption)**

The correlation between rater A and rater B NCD presence ratings (No/Minor/Major) prior to the consensus process was 0.77. After the consensus process, the correlation was 0.81. The kappa statistic went from 0.76 (0.73, 0.79) to 0.81 (0.78, 0.84).

# Bland Altman plots

## Delirium severity

|  |
| --- |
| Figure 5: Bland-Altman plot of initial ratings of delirium severity |

|  |
| --- |
| Figure 6: Bland-Altman plot of consensus ratings of delirium severity |