Importance of Training and Quality Control of Post-Operative Delirium Assessment:

Hochang Benjamin Lee, M.D.

Associate Professor of Psychiatry
Yale University School of Medicine

Director, Psychological Medicine Service
Yale New Haven Hospital
Disclosure

- **Current funding:**
  - **Relevant:**
    - R01 MH085740 (NIH/NIMH: PI Lee HB)
    - R01AG029656 (NIH/NIA: PI Silverstein J)
    - P30 DK34989 (NIH/NIDDK: Yale Liver Center Pilot Grant: PI: Lee HB)
  - **Unrelated:** RLS Foundation Research Grant (PI: LEE HB)

- **Other financial relationships:** None
Briefly describe a geriatric/CL psychiatrist’s perspective on difficulty in diagnosis of delirium.

Describe three post-surgical delirium prevention trials and compare their training and quality assurance protocols.
- Focus Cognitive Ancillary Study (PI: Gruber-Baldini)
- STRIDE study (PI: Sieber)
- Dexplurium Study (PI: Silverstein)

Describe strengths and limitations of each method while focusing on pitfalls.
Why delirium diagnosis challenging for a psychiatrist as well.

1. Delirium is a longitudinal diagnosis
   - Lack of pre-morbid level of cognition or function
   - “Acute” versus “Gradual” change.

2. Symptoms of delirium commonly overlaps with symptoms of other psychiatric conditions (e.g. dementia and depression).
   - 46% of patients with delirium were misdiagnosed by the referring service personnel (Armstrong 1997)
   - 42% of “depression” referral were delirious (Farrel, 1995)

3. “Hardest diagnosis in psychiatry”:
   - Milder, hypoactive delirium superimposed on dementia.
Change of Diagnostic Criteria in DSM

- DSM III (1980)
- DSM III-R (1987)
- DSM IV (1990)
- DSM V (2013)


- Of 230 geriatric hospital patients, prevalence varied depending on criteria:
  - DSM-IV (24.9% of the subjects) followed by DSM-III-R (19.5%), DSM-III (18.8%) and ICD-10 (10.1%).
Using CAM to make diagnosis of Delirium (Laurila, 2002)

- 81 consecutive elderly patients in geriatric hospital
- Sensitivity rates of the CAM were proved to be only moderate (0.81–0.86) against all DSM criteria of delirium. The specificity rates were lower (0.63–0.84).
- CAM defines delirium in its own way and, ironically and probably provides the most **enduring and generalizable** diagnostic outcome.

Table 5. Sensitivity and specificity rates, positive and negative predictive values, and positive and negative likelihood ratios of CAM compared to DSM-III, DSM-III-R, DSM-IV, and ICD-10 as reference standards

<table>
<thead>
<tr>
<th></th>
<th>DSM-III</th>
<th>DSM-III-R</th>
<th>DSM-IV</th>
<th>ICD-10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Positive CAM score</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>Negative CAM score</td>
<td>3</td>
<td>44</td>
<td>4</td>
<td>43</td>
</tr>
<tr>
<td>Sensitivity rate</td>
<td>0.85</td>
<td>0.81</td>
<td>0.81</td>
<td>0.80</td>
</tr>
<tr>
<td>Specificity rate</td>
<td>0.72</td>
<td>0.72</td>
<td>0.84</td>
<td>0.63</td>
</tr>
<tr>
<td>Positive predictive value</td>
<td>0.50</td>
<td>0.50</td>
<td>0.76</td>
<td>0.24</td>
</tr>
<tr>
<td>Negative predictive value</td>
<td>0.94</td>
<td>0.91</td>
<td>0.87</td>
<td>0.96</td>
</tr>
<tr>
<td>Likelihood ratio for a positive test</td>
<td>3.05</td>
<td>2.86</td>
<td>4.98</td>
<td>2.18</td>
</tr>
<tr>
<td>Likelihood ratio for a negative test</td>
<td>0.21</td>
<td>0.27</td>
<td>0.22</td>
<td>0.32</td>
</tr>
</tbody>
</table>
Use of CAM in PSD clinical trials

- Most of post-surgical delirium prevention trials utilizes CAM
  - Nearly 100%, if secondary outcomes.

- Why CAM in PSD studies?
  - Generalizable.
  - Validity is well-established
  - “Simplicity, ” however, rigorous training is essential.
    - Minimally trained bedside nurses – 23.8% and 66.7% sensitivity based on two scoring methods for CAM (Lemiengre et, JAGS 2006)
    - Partially trained research nurses – 13% detection (Rolfson, IJP 1999)
“Myth of Simplicity”

- Wong CL, et al. JAMA 2010; 304:779-786
  - CAM has the best available supportive data as a bedside delirium instrument (summary-positive LR, 9.6; 95% CI, 5.8-16.0; summary-negative LR, 0.16; 95% CI, 0.09-0.29).”
  - Conclusion: “The choice of instrument may be dictated by the amount of time available and the discipline of the examiner; however, the best evidence supports use of the CAM, which takes 5 minutes to administer.”
  - “…But how long does it take to get to CAM?”
But, how long does it take to GET TO CAM?

- Gathering information for each component of CAM takes time and clinical judgment
  - Acute cognitive change – testing, review of records,
  - Attention – testing cognition
  - Disorganized thought – interview with the patient
  - Level of consciousness – observation of the patient

- Delirium Diagnosis Methodology Used by Reference Raters in Research: A Survey-Based Study (Neufeld KJ, et al, under review)
  - 33 of 39 studies from 3 recent systematic reviews of delirium detection instruments.
  - Tremendous variability in diagnostic methods and rater backgrounds
Tremendous variability in incidence of acute post-hip fracture delirium with CAM: 5 – 40% (Bruce 2006)

<table>
<thead>
<tr>
<th>Authors</th>
<th>Setting</th>
<th>Excluded Cognitive Impairment</th>
<th>Age (Years)</th>
<th>N</th>
<th>Rating Scale</th>
<th>Postop Days Tested</th>
<th>Incidence 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Johansson et al. 2002</td>
<td>Örebro County Hospital, Sweden</td>
<td>No</td>
<td>Mean = 80.4</td>
<td>47</td>
<td>NEECHAM</td>
<td>Day 7</td>
<td>4%</td>
</tr>
<tr>
<td>Brauer et al. 2000</td>
<td>Four New York hospitals, New York, U.S.A.</td>
<td>No</td>
<td>Median = 85</td>
<td>546</td>
<td>CAM</td>
<td>5 days/week</td>
<td>5.30%</td>
</tr>
<tr>
<td>Edlund et al. 1999</td>
<td>Pite River Valley Hospital, Sweden</td>
<td>No</td>
<td>Mean = 77.1</td>
<td>44</td>
<td>DSM-III-R/OBS</td>
<td>Daily</td>
<td>11.40%</td>
</tr>
<tr>
<td>Morrison et al. 2003</td>
<td>Four New York hospitals, New York, U.S.A.</td>
<td>Not stated</td>
<td>525</td>
<td></td>
<td>CAM</td>
<td>5 days/week</td>
<td>14.00%</td>
</tr>
<tr>
<td>Schuurmans et al. 2003</td>
<td>General hospital, Netherlands, Netherlands</td>
<td>No</td>
<td>Mean = 82.7</td>
<td>92</td>
<td>DSM-IV</td>
<td>Daily for 6 days</td>
<td>19.60%</td>
</tr>
<tr>
<td>Edlund et al. 2001</td>
<td>University Hospital, Umeå, Sweden</td>
<td>No</td>
<td>Mean = 79.5</td>
<td>71</td>
<td>DSM-IV/OBS</td>
<td>Daily</td>
<td>26.80%</td>
</tr>
<tr>
<td>Formiga et al. 2003</td>
<td>Two university hospitals, Barcelona</td>
<td>No</td>
<td>Mean = 92.4</td>
<td>89</td>
<td>CAM</td>
<td>Day 1 or 2, and at discharge</td>
<td>28.10%</td>
</tr>
<tr>
<td>Gustafson et al. 1988</td>
<td>University Hospital, Umeå, Sweden</td>
<td>No</td>
<td>Mean = 79.3</td>
<td>74</td>
<td>DSM-III/OBS</td>
<td>Daily</td>
<td>41.90%</td>
</tr>
<tr>
<td>Zakriya et al. 2004</td>
<td>Medical Centre, U.S.A.</td>
<td>No – all had dementia</td>
<td>Mean = 78–79</td>
<td>10</td>
<td>CAM</td>
<td>Daily</td>
<td>50%</td>
</tr>
<tr>
<td>Thakur et al. 2002</td>
<td>Community teaching hospital, U.S.A.</td>
<td>No</td>
<td>Mean = 66–98</td>
<td>30</td>
<td>CAM</td>
<td>Not given</td>
<td>53.30%</td>
</tr>
<tr>
<td>Kagansky et al. 2004</td>
<td>Community teaching hospital, Israel</td>
<td>Excluded severe dementia</td>
<td>Mean = 82</td>
<td>96</td>
<td>CAM/DRS</td>
<td>Day 7</td>
<td>6.30%</td>
</tr>
<tr>
<td>Andersson et al. 2001</td>
<td>Lund County City Hospital, Sweden</td>
<td>Confusional states excluded</td>
<td>Range = 65–96</td>
<td>208</td>
<td>DSM-IV/OBS</td>
<td>Daily</td>
<td>20.20%</td>
</tr>
<tr>
<td>Duppils et al. 2000</td>
<td>County hospital, Sweden</td>
<td>Dementia and MMSE ≤ 10</td>
<td>65+</td>
<td>149</td>
<td>DSM-IV</td>
<td>At least twice daily</td>
<td>24.30%</td>
</tr>
<tr>
<td>Zakriya et al. 2002</td>
<td>Medical Centre, U.S.A.</td>
<td>Dementia excluded</td>
<td>Mean = 77.6</td>
<td>168</td>
<td>CAM</td>
<td>Day 2 until discharge</td>
<td>28.00%</td>
</tr>
<tr>
<td>Galanakis et al. 2001</td>
<td>Munich hospital, Germany</td>
<td>Severe dementia excluded</td>
<td>Mean = 74.9</td>
<td>37</td>
<td>CAM</td>
<td>Days 1–7</td>
<td>40.50%</td>
</tr>
<tr>
<td>Bowman 1997</td>
<td>Manitoba teaching hospital, Canada</td>
<td>Dementia and MMSE ≤ 23</td>
<td>Mean = 80</td>
<td>17</td>
<td>DSM-III</td>
<td>Days 1–5, twice daily</td>
<td>47.10%</td>
</tr>
</tbody>
</table>
Importance of Case Ascertainment methods in Delirium Prevention or Treatment Trials

- **Treatment Trials**
  - Recruitment of delirious study subjects
    - Under-detection: Cannot run the trials
    - Over-detection: Weakened signal of intervention by recruiting wrong subjects

- **Prevention Trials:**
  - Primary outcomes: Delirium
    - Under-detection: Need a large sample size
    - Over-detection: Results in erroneously negative or positive findings

- Must balance practicality and science based on available personnel, setting, and budget.
Lessons from Three NIA-sponsored Post-Surgical Delirium Prevention Trials.

- **Focus Cognitive Ancillary Study (PI: Gruber-Baldini)**
  - Completed – my role: peripheral involvement.

- **STRIDE study (PI: Sieber)**
  - On-going - designed the study outcomes/ training protocol/ quality assurance while at Hopkins

- **Dexlirium Study (PI: Silverstein)**
  - Ongoing – primary delirium “expert”- responsible for training and quality assurance.
## Summary

<table>
<thead>
<tr>
<th></th>
<th>Focus Cognition</th>
<th>Dexlirium</th>
<th>STRIDE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
<td>139</td>
<td>708 (planned)</td>
<td>200</td>
</tr>
<tr>
<td>Site #</td>
<td>17</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Intervention</td>
<td>Transfusion</td>
<td>Dexmedotimidine</td>
<td>Sedation Level</td>
</tr>
<tr>
<td>Instruments</td>
<td>DIS, MDAS, CAM</td>
<td>DIS, MDAS, CAM</td>
<td>DRS-98, CAM, DI</td>
</tr>
<tr>
<td>Training</td>
<td>Web-based</td>
<td>Web-based/supplemented</td>
<td>Fully in-person</td>
</tr>
<tr>
<td></td>
<td>certification</td>
<td>by in-person at each</td>
<td></td>
</tr>
<tr>
<td></td>
<td>and limited in-person training in the beginning.</td>
<td>site by coordinating center</td>
<td></td>
</tr>
<tr>
<td>Quality Assurance</td>
<td>Weekly Teleconference</td>
<td>Teleconference and monthly case presentation and data review</td>
<td>Consensus Panel Case presentation</td>
</tr>
</tbody>
</table>
FOCUS Cognitive Ancillary Study
(PI: Ann Gruber-Baldini)

- **Goal**: To examine the impact of the hemoglobin interventions on delirium in a subsample of 200 subjects (100 per randomization group).
- **Reality**: 17 sites and short duration of intense recruitment.
  - Must weigh the issue of fidelity and practicality of outcome measure (delirium: case versus severity)
  - Need for multiple raters in multiple sites
    - Cannot utilize clinical psychiatrists for all sites
    - Alternative: Train available research staff (including non-clinical Research Assistants)
      - Delirium Symptom Interview (structured)
      - Memorial Delirium Assessment Scale (severity)
      - Confusion Assessment Method – primary outcome (case)
FOCUS Cognitive Study: Training and Quality Assurance

● Training:
  ● Investigator kick-off meeting
  ● Introductory lectures and training
  ● Certification: Web-based training – certification process -3 video cases
  ● Individual ratings of DSI, MDAS and CAM submitted to the coordinating site for Case #3
  ● Case #3 answers are compared to the master answers with individual feedback.

● Quality Assurance
  ● Site visits and teleconference
FOCUS Cognitive Study: Strengths and limitations

- Balancing fidelity and practicality
  - Wide range of raters (physicians, nurses, non-clinical RAs)
    - What is the “gold standard”? 
  - Web-based training has its strengths and weaknesses
    - Covers multiple sites distributed widely in geography
    - Cannot provide close oversight over the training
    - RA turnover is difficult to overcome
- However, for multi-site clinical trial of short duration and limited budget, probably no other choice.
A Strategy to Reduce the Incidence of Post-Operative Delirium in Elderly patients: The STRIDE Study (PI: Frederick Sieber)

- Sponsor: NIA
- Design: Single-site randomized double-blinded clinical trial.
- Aim: to determine whether limiting the level of sedation in elderly patients during spinal anesthesia for surgical repair of a hip fracture will lead to a lower rate of post-operative delirium.
- Intervention: To give one group of elderly traumatic hip fracture patients standard spinal anesthesia, with light-to-moderate sedation, and the other group standard spinal anesthesia with deeper sedation.
STRIDE Study: Training and QA

- Single site study with experienced research nurses as the rater.
- CL Psychiatrist trainer is on-site and available at all times.
- Introductory Group Seminar – 6+ hours to go over the manuals for CAM and DRS-98.
- In person training – three practice cases prior to data collection.
- Bi-weekly case presentation by the Research RN to the consensus panel.
  - Multi-disciplinary consensus panel consists of psychiatrist, geriatrician, anesthesiologist, and surgeon.
STRIDE Study: Strengths and limitations

- Single-site design allows more rigorous training protocol and quality assurance.
- High personnel cost for rater/trainer/consensus panel.
  - Consensus Panel blind to the group assignment affords more “gold-standard”-like comparison.
  - Took a long lead-in time.
- Consensus panel methods “adapted” from other dementia prevention studies.
Dexlirium Study (PI: Jeff Silverstein)

- Sponsor: NIA
- randomized double blinded, parallel group, placebo-controlled study of the effects of perioperative dexmedetomidine on the incidence of postoperative delirium and postoperative cognitive dysfunction
- Sites: 8 sites
- Duration: “5 years”
- Target sample: 706 elderly patients undergoing elective “major” general surgery under general anesthesia
- dexmedetomidine vs. placebo
Dexlirium Study: before QA

- Limitation imposed by the multi-site study design and limited personnel and budget.
  - Also, rapid turnover of raters
- Similar to Focus Cognitive Study – Wide range of clinical background among the raters: “non-clinical” RA to nurses and MDs.
- Formal training and QA protocol was implemented in the middle of the study
  - Concern about low delirium incidence
Dextririum Study: Training and QA

- RAs asked to read the protocol manuals and go through the web-based certification process first.
- 3 video cases from the FOCUS cognition study
- Site visits by PI and Delirium Trainer
- Monthly teleconference with delirium assessment case presentation from each site.
- Data review of every delirium assessment by the neuropsychiatrist for detection of data inconsistency and data reconciliation.
Detection of incident delirium before and after QA program

- Unpublished data
Delirium Incidence before and after QA implemented

- Unpublished data
Dexlirium Study: Strengths and Limitations

- Balancing fidelity and feasibility with limited budget and personnel
  - Who is the gold standard/ reference rater in each site?
- Widely varied background of RAs
  - From post-doc fellow/ junior faculty, MDs, RNs, and RAs with no clinical background who just graduated from college.
    - Individual attention is absolutely necessary
    - Clinical background – not necessarily an advantage.
- Rapid turn-over rate of RAs in some sites.
  - High training burden.
Lessons learned

- Training and quality assurance for delirium assessment is an arduous, but absolutely necessary task.
- Rigorous training protocol and continuous quality assurance effort is necessary
  - A clinical trial is as good as the fidelity of its clinical outcome
- Need for more standardized assessment strategy before applying diagnostic instrument like CAM.
  - Especially for the non-clinical raters.

- WE ARE IN THE EARLY STAGE OF DEVELOPING THE FIELD – NEED TO LEARN FROM EACH OTHER
Gratitude

Without their generous guidance and help, this presentation would not have been possible.

- Focus Cognitive Ancillary Study
  - Anne Gruber-Baldini
  - Ed Marcantonio

- STRIDE study
  - Frederick Sieber

- Dexlirium Study
  - Jeff Silverstein

- NIA for sponsoring the studies above