Modelling the use of confidence intervals with the borderline regression method for final year undergraduate OSCE at the University of Southampton.

McInnes,B.; Curtin,J.; Anderton,P.; Holloway,A.; Field,J.; & Rushworth,S.M.

Abstract:
Background and Purpose
Several studies have noted Objective Structured Clinical Examinations (OSCEs) are used in the BM exams, and in a high stakes examination, improvements in its reliability and validity of these exams is desirable. We aim to model and pilot a novel use of the borderline regression method (BRM) to standard set BM exams at Southampton. The exam was designed with a test–retest reliability of 0.71.

Methodology
Students who pass more than 11 of the 16 BM OSCE stations were certified as passing, with a view to adoption as the formal standard setting procedure at Southampton.

Results
A total of 219 BM finals failed OSCEs. 22 failed the OSCE based on passing the OSCE to pass the BM finals. 15 failed based on the pass rate of >3 stations. 20 failed based on both of these criteria. Students who failed the OSCE were remodeled using option 2. 20 failed based on both of these criteria. Students who failed the OSCE were remodeled using option 2.

Conclusion
The problem here is that while relatively simple to administer this system is highly reliant on single examiners making very high stakes decisions on a single exam. We have proposed and evaluated a well described alternative method in the borderline regression method (BRM) to standard set these exams. A new adjusted lower CI cut score can be interpolated from this new equation (see Figure 1 blue line).

Results demonstrated in Table 2

<table>
<thead>
<tr>
<th>Station</th>
<th>Total</th>
<th>% Chance</th>
<th>No. who fail</th>
<th>% Fail</th>
<th>Borderline System, pass rate 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>242</td>
<td>97.62%</td>
<td>9</td>
<td>0.00%</td>
<td>[0.00%, 0.00%]</td>
</tr>
<tr>
<td>2</td>
<td>196</td>
<td>92.14%</td>
<td>14</td>
<td>0.00%</td>
<td>[0.00%, 0.00%]</td>
</tr>
<tr>
<td>3</td>
<td>205</td>
<td>99.36%</td>
<td>1</td>
<td>0.00%</td>
<td>[0.00%, 0.00%]</td>
</tr>
<tr>
<td>4</td>
<td>158</td>
<td>99.36%</td>
<td>1</td>
<td>0.00%</td>
<td>[0.00%, 0.00%]</td>
</tr>
<tr>
<td>5</td>
<td>219</td>
<td>89.30%</td>
<td>24</td>
<td>12.31%</td>
<td>[11.30%, 13.32%]</td>
</tr>
<tr>
<td>6</td>
<td>216</td>
<td>96.69%</td>
<td>13</td>
<td>6.13%</td>
<td>[5.12%, 7.14%]</td>
</tr>
</tbody>
</table>

4) Discussion & Conclusions:

Table 2

<table>
<thead>
<tr>
<th>Station</th>
<th>No. who fail</th>
<th>% Fail</th>
<th>Borderline System, pass rate 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9</td>
<td>0.00%</td>
<td>[0.00%, 0.00%]</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>0.00%</td>
<td>[0.00%, 0.00%]</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0.00%</td>
<td>[0.00%, 0.00%]</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0.00%</td>
<td>[0.00%, 0.00%]</td>
</tr>
<tr>
<td>5</td>
<td>24</td>
<td>12.31%</td>
<td>[11.30%, 13.32%]</td>
</tr>
<tr>
<td>6</td>
<td>13</td>
<td>6.13%</td>
<td>[5.12%, 7.14%]</td>
</tr>
</tbody>
</table>

5) Method: Standard Error of Measurement

- Use of the Standard Error of Measurement was calculated to assess the overall stability of the OSCE.
- The standard error of measurement (SEM) is defined as the square root of the variance of the difference between repeated measurements.
- SEM = \( \sqrt{\text{Var}(\text{errors})} \)

6) Method: Confidence intervals for the cut score on a single station

- Use of the standard error of measurement was calculated to assess the overall stability of the OSCE.
- The standard error of measurement (SEM) is defined as the square root of the variance of the difference between repeated measurements.
- SEM = \( \sqrt{\text{Var}(\text{errors})} \)

7) Results of re-modelling:

<table>
<thead>
<tr>
<th>Overall OSCE</th>
<th>Pass</th>
<th>Fail</th>
<th>Total</th>
<th>Borderline System, pass rate 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current</td>
<td>219</td>
<td>24</td>
<td>242</td>
<td>[0.00%, 0.00%]</td>
</tr>
<tr>
<td>Option 2</td>
<td>219</td>
<td>24</td>
<td>242</td>
<td>[0.00%, 0.00%]</td>
</tr>
</tbody>
</table>

Table 1

8) Possible Solutions

- For criterion 1:
  - Suggest criterion lacks sensitivity as a screening tool for minimal competence.
- For criterion 2:
  - Suggest criterion lacks sensitivity as a screening tool for minimal competence.

9) Problems

- How to choose a screening criterion between current and SEM criteria.
- How to choose an appropriate compensation method (criterion 2)
- How to handle multiple criteria (criterion 1 & 2)
- How to handle multiple criteria in the borderline regression method (BRM) to standard set these exams.

References:

3. McManus, B.N, BM finals OSCE coordinator, Faculty of Medicine, University of Southampton, Southampton General Hospital MP801, Tremona Road, Southampton. 
5. McManus, B.N, BM finals OSCE coordinator, Faculty of Medicine, University of Southampton, Southampton General Hospital MP801, Tremona Road, Southampton.