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

McManus,B.N; Carr,N.J; Anderson,F.H.; Holloway,J.A.; Field,J.H; & Rushworth,S.M.

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

Background and Purpose

We wished to model and pilot a novel use of the confidence interval (CI) and standard error of the measurement (SEM) with the borderline regression method, in line with recommendations by PMETB/GMC,[1,2] and in place of simple examiner global judgements.

Methodology

Students must satisfy two criteria to pass the BM finals OSCE: aggregate score and minimum number of stations passed. The SEM has been equated with CI[1] and applied to aggregate score[3-5]. We wished to introduce it into our examination, and also proposed a novel strategy to calculate the CI in the cut score for a single station. Using the standard error of the intercept and gradient we calculated the CI for these values, and used them in the regression equation to interpolate a new value of y when x is constant. We modelled these techniques to maximise the sensitivity and specificity of both criteria.

Results

In a cohort of 242 students, 6 failed >3 stations on global judgement. For 2 of them the mean grade was also below the threshold but none failed this criterion alone. Introducing borderline regression without adjustment, 23 students failed >3 stations but none on aggregate score. Recalculating the aggregate pass mark as mean cut score plus 1.96xSEM (upper 95%CI) considerably improved the sensitivity of the aggregate score criterion, which 6 students now failed.

For individual stations, using the gradient and intercept minus 1.96xSEM (lower 95%CI) provided an adjusted cut score for each and considerably improved the specificity of this criterion. Students failed if their actual scores were below the cut score for >3 stations. 7 failed on this criterion.

Considering both criteria 8 failed the OSCE, 5 of whom failed both criteria. Observed agreement with global assessments rose from 92.1% to 98.35% (Kappa 0.32 to 0.71).

Discussion and Conclusions

The adjusted cut scores showed improved sensitivity and specificity for both criteria and improved agreement with global judgements. It was perceived to be fair to students, affording them the benefit of the doubt when considering individual stations, but protecting patient safety when decisions could be reliably based on 16 assessments. Since most students who failed did so on both criteria, the method was perceived to be more robust. The authors plan to remodel this on another cohort of students before considering incorporating into the exam regulations.

References

1. Postgraduate Medical Education and Training Board. Developing and maintaining an assessment system - a PMETB guide to good practice. PMETB 2007.

2. General Medical Council. Assessment in undergraduate medical education - Advice supplementary to Tomorrow's Doctors (2009). GMC 2010.

3. Dauphinee, W.D., Blackmore, D.E., et al. . Using the Judgments of Physician Examiners in setting the Standards for a National Multi-center High Stakes OSCE. *Advances in Health Sciences Education* 1997; **2:** 201–211.

4. Smee, S.M., Blackmore D.E. Setting standards for an objective structured clinical examination: the borderline group method gains ground on Angoff . *Med Educ* 2001; **35:** 1009-1010.

5. Kilminster, S., Roberts, T. Standard Setting for OSCEs: Trial of Borderline Approach. *Advances in Health Sciences Education* 2004*;* **9:** 201–2097.