

ACGME Summary Questions

Xilin Chen

11/18/2020

Q1: Differences between excluded milestone cohort and cohort included in analysis

Milestones rating for each sub-competency is different between the 3001 (=3521-520) and 520 residents in terms of distribution. Were the latter residents rated lower or higher than the former group?

Rating	Excluded, N = 2,978 ¹	Included, N = 520 ¹	p-value ²
Overall	8.75 (8.12, 9.00)	8.66 (8.06, 9.00)	0.052
Overall ≥ 8	2,429 (82%)	420 (81%)	0.7
Professional	9.00 (8.00, 9.00)	8.67 (8.00, 9.00)	0.3
Professional ≥ 8	2,424 (81%)	421 (81%)	0.8
Operative	8.67 (8.00, 9.00)	8.67 (8.00, 9.00)	0.4
Operative ≥ 8	2,515 (85%)	425 (82%)	0.12
Leadership	9.00 (8.25, 9.00)	8.75 (8.25, 9.00)	0.085
Leadership ≥ 8	2,577 (87%)	449 (86%)	>0.9

¹Statistics presented: Median (IQR); n (%)

²Statistical tests performed: Wilcoxon rank-sum test; chi-square test of independence

≥ 8 : dichotomizing the mean ratings as 1: ≥ 8 ; 0: < 8 .

The reason the excluded from original cohort had 2,978 trainees instead of 3,001 was that 23 trainees didn't have "Year-end" ratings when they are at PGY5. So they were excluded.

We eventually decided to use the binary milestone ratings for our model analysis. I think from the table comparison above, there were no significant differences between the original group and the cohort we used.

Q2: How did we define the ratings at the time of their graduation?

We examined the most proximal Milestone ratings for each trainee to the time of their graduation. Does this mean that there were missing Milestones ratings for some trainees at the time of graduation?

A small number of trainees didn't have last year graduation ratings and were excluded. The way we defined the time of graduation was *resident year = 5* and *eval period = "Year-End Milestone Evaluations"*. Also, about 2% trainees in the original cohort had 2 academic years associated with PGY5, we kept the most recent academic year as their final graduation year. In our final cohort, 3/520 had 2 PGY5 academic year in the raw data. Overall, the data quality was pretty good.

Q3: How did we define the clustering structure in our data?

Are patients completely nested within surgeons, who in turn are completely nested within hospitals? Could surgeons join multiple hospitals? Could patients be seen by multiple physicians and those data be included in the dataset?

Patients were not considered completely nested within surgeons, or surgeons within hospitals. We simply hypothesized that the observations within the same hospital or the same physician are not independent. So, no observations were excluded based on their clustering structure.

In the mixed model setting, it doesn't matter if the data are strictly hierarchical, which is a special case of clustering, or crossed. The mixed model package lme4 and some other tools doesn't treat "nested" or "crossed" data any differently. The same computational methods are used whether the patients nested or not nested within physicians and hospitals.