The Reproducibility and Stability of *US News and World Report* Primary Care Medical School Rankings

Daniel J. Tancredi, PhD
Department of Pediatrics and Center for Healthcare Policy and Research
UC Davis

Revision of slides originally presented at April 22, 2013 at CHPR Brown Bag Seminar and May 7, 2013 at CHPR Health Services Noon Seminar

Revised on August 20, 2014 to correct “Note” on Slide #11
**Speaker’s Verbal Disclosure Statement:**

Have you (or your spouse/partner) had a personal financial relationship in the last 12 months with the manufacturer of the products or services that will be discussed in this CME activity?

___ Yes

_X_ No

(If yes, please state disclosures and resolutions)
Educational objectives for this seminar:

- To describe how the U.S. News and World Report Primary Care Medical School rankings are computed

- To summarize the distribution of short-term changes in school rankings

- To illustrate that high reliability need not imply that most schools only have small short-term changes in rankings

- To compare rankings based on current methodology with alternative rankings designed to emphasize elements with higher face validity for the quality of primary care training
U.S. News Grad School Rankings

• Stated purposes, according to U.S. News:
  – independent assessment of academic quality
  – help prospective students find best-fitting program

• Medical school rankings began in 1990
  – Targets fully accredited U.S. schools granting MD (n = 126) or Doctor of Osteopathy (n = 23) degrees
  – Rankings limited by available data (n = 114 in latest edition)

• Separate medical school rankings for
  – Research
  – Primary Care (family, general internal & pediatric medicine)
Motivating Questions

• Can the *U.S. News* Primary Care Medical School (PCMS) rankings be reconstructed, using *their* source data and stated methods?

• How reliable are the PCMS rankings? How much does a typical school’s ranking fluctuate?

• How much do the rankings depend on elements less directly related to training quality?
Data Sources

• Four years, 2009 - 12
• Extracted data from *U.S. News* premium website:
  – 2009 ( n = 119 )
  – 2011 ( n = 121 )
  – 2012 ( n = 112 )
• *U.S. News* magazine data from just the 22 schools in “top 20” in 2010
U.S. News Annual Rankings

Methodology

• Survey external experts and collect other measures of inputs, resources and outputs to get $c = 1, \ldots, C$ component scores on $s = 1, \ldots, S$ schools

• Standardize all component scores:

$$Z_{cs} = \frac{X_{cs} - \text{Mean}_c}{\text{SD}_c},$$

where $\text{Mean}_c$ & $\text{SD}_c$ are the sample mean & standard deviation, respectively.

• Form weighted aggregate score for each school:

$$T_s = \frac{\sum w_c Z_{cs}}{\sum |w_c|},$$

using component weights

• Rescale ratings to integers so that maximum is 100 (a step we will forego) and rank in descending order
Primary Care Medical School Components and Weights

• Reputation survey (40% of total weight)
  – Peers (e.g. deans, department heads) 0.25
  – Residency program directors 0.15

• Student selectivity (15% of total weight)
  – Mean composite MCAT score 0.0975
  – Mean undergraduate GPA 0.045
  – Acceptance rate -0.0075

• Resources (faculty : student ratio) 0.15

• Primary care rate (residency placements) 0.30
Reputation survey #1

• **Peer assessment** from school deans, deans of academic affairs, heads of internal medicine or directors of admission

• Respondents rate other schools’ primary care programs on a 5-level Likert item
  1 (marginal), 2 (adequate), 3 (good), 4 (strong), 5 (outstanding)
  Or “don’t know school’s program well enough to rate fairly”

• Mean of numerical ratings is component score

• Low survey response rates, ~ 40 – 50%
Reputation survey #2

• Residency program directors for primary care specialties (internal medicine, pediatrics and family medicine) also rate primary care programs

• Even lower response rates for program director survey, ~ 15 – 20%

Note: Ratings based on a two-year average for this component
Scatter Plot of Reputation Scores, 2012 (r = 0.79)

With 95% Prediction Ellipse

Observations: 115
Correlation: 0.7895
Student Selectivity Measures

• Based on school self-report of data from most recent year’s incoming class

• In case of non-report, imputed z-score of -1 used to compute ratings (personal correspondence with U.S. News)
9.75% of total weight

Box shows 3 schools in Appalachia
Mean Undergrad. Grade Point Average, 2012

Mean Undergraduate GPA, 2012 vs. 2009

With 95% Prediction Ellipse

Observations 105
Correlation 0.8881

4.5% of total weight
Acceptance Rate, 2012

Weight = -0.75%

Acceptance Rates, 2012 vs. 2009
With 95% Prediction Ellipse

Observations 105
Correlation 0.8737

UND
Western (Pomona)
N Texas
UMass-Worch.
OK State
Nova SE
UH-Manoa
Scatter Plot Matrix of Student Selectivity Measures, 2012

- MCAT (r = 0.76)
- UGPA (r = -0.41)
- Acceptance Rate (r = -0.26)
Faculty Resources

• *U.S. News* reports using **Faculty : Student** ratio, computed using full-time science and clinical faculty and full-time students in current year

• Actually based on **Student : Faculty** ratio!
Component Z-scores would be > 5 for Harvard and Mayo.
Weight = - 0.75%

Z-score is 3.32 in 2012 for the schools with a S:F ratio of 10
Primary Care Rate

• Percentage of graduates entering family medicine, internal medicine and pediatrics residencies, averaged over the most recent 3 years

• But these specialties really aren’t alike…
  – ~95% of family medicine residents will become primary care providers
  – but only ~25-50% of physicians going into internal and pediatric residencies stick with primary care...the rest pursue subspecialties
Scatter Matrix of high-weight components, 2012

Peer Score: $r = 0.41$

MCAT: $r = -0.61$

Stud : Fac: $r = -0.49$

PC rate: $r = -0.02$

Stud : Fac: $r = -0.48$

PC rate: $r = 0.55$
Reconstructed PCMS Scores, 2012

Count

SD = 0.52
Reconstructed scores highly correlated with published scores

<table>
<thead>
<tr>
<th>Year</th>
<th>N</th>
<th>Correlation (n)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>119</td>
<td>99.9% (61)</td>
</tr>
<tr>
<td>2010</td>
<td>22</td>
<td>99.5% (22)</td>
</tr>
<tr>
<td>2011</td>
<td>121</td>
<td>99.6% (98)</td>
</tr>
<tr>
<td>2012</td>
<td>112</td>
<td>99.7% (89)</td>
</tr>
</tbody>
</table>

* U.S. News does not publish lower scores, so correlation is for subsample with published scores
Reliability Estimates for US News PCMS Ratings and Rankings

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Standard Error of Measurement</th>
<th>Within-school correlation coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reconstructed PCMS rating</td>
<td>0.13</td>
<td>93%</td>
</tr>
<tr>
<td>Reconstructed PCMS ranking</td>
<td>9.4</td>
<td>91%</td>
</tr>
</tbody>
</table>

REML estimates from mixed-effects model, n=107 schools w/ 2009, 11 & 12 data
Bland-Altman type plot of ranking ranges versus mean rank for 107 schools with at least three reconstructed scores in 2009-12

Half of the schools have 14 or more positions between their best and worst ranking in just 4 years
**Top-20 type schools are more stable: median range = 4**

Median range in the remaining schools is 17 (i.e. half of them moved at least 17 positions in the rankings!)
Is it plausible that rankings could be so unstable when they are based on ratings with 90 to 95% reliability?

A brief simulation study
Boxplot of within-school ranges in annual rankings
Simulated data, 100 schools rated over 4 years
95% Reliability

Within-school range (Max-Min)
Let’s call the median range, the spread statistic.
Boxplot of within-school ranges in annual rankings
Simulated analyses, 100 schools rated over 4 years
95% Reliability

Within-school range (Max-Min)

Analysis: All 100 Schools, 'Bottom 80', 'Top 20'
Boxplots of within-school ranges in annual rankings
Simulated analyses of 100 schools rated over 4 years
Varying Rating Reliability

Reliability of Annual Rating

Analysis • All 100 Schools □ 'Bottom 80' □ 'Top 20'
Empirical sampling distributions of Spread Statistic

Each boxplot describes 13,000 simulated analyses
Spreads can be relatively large, despite high reliability

Spread Statistic

Reliability of Annual Rating

Analysis  All 100 Schools  'Bottom 80'  'Top 20'
Alternative Rankings

1. Would rankings based only on the higher face validity components differ?
   
   *Reputation Scores and Primary Care Rate*
   
   *Student Selectivity and Faculty Resources*
Alternative Rankings

1. Would rankings based only on the higher face validity components differ?

   Reputation Scores and Primary Care Rate
   Student Selectivity and Faculty Resources
Alternative Rankings

1. Would rankings based only on the higher face validity components differ?

   *Reputation Scores and Primary Care Rate*
   *Student Selectivity and Faculty Resources*

2. What if we also changed the PC Rate computation to account for differential propensity of residents to pursue primary care careers?
   - 50% weight to Internal Medicine and Pediatrics
   - 95% weight to Family Medicine
Osteopathic schools would move up in the standings.

Some choosier and wealthier schools would move down.

High validity components are Rep Scores and PC Rate.
25+ position changers are labelled.
Hotbeds for future subspecialists would move down.

Schools producing more general practice docs would move up.

$r = 0.81$

Discounts PC Rate to adjust for subspecialists. 25+ position changers are labelled.
Conclusions

• US News PCMS Scores can be reconstructed, but require undocumented adjustments
• Ratings have good reliability coefficients
• Even with good rating reliability, most schools will have differences between best and worst ranking > 10% of the number of schools ranked
• Alternative scoring methods could yield big shifts
• Treat rankings with due regard for high uncertainty!
Reference and acknowledgments

Tancredi DJ, Bertakis KD, Jerant A. Short Term Stability and Spread of the U.S. News & World Report Primary Care Medical School Rankings. Academic Medicine. in-press

Nicholas Clark

UC Davis CHPR

Bob Morse (U.S. News & World Report)