Understanding and Interpreting the NRC’s “Data-Based Assessment of Research-Doctorate Programs in the United States (2010)"

Jaxk Reeves, SCC Director
Kim Love-Myers, SCC Associate Director

Presented at UGA Graduate School Workshop
10/19/2010

About the SCC

- The Statistical Consulting Center (SCC) is part of the Department of Statistics at UGA
- We offer statistical advice to on- and off-campus clients
- Web page: www.stat.uga.edu/consulting
Topical Outline

- Brief Overview
- NRC Methodology
- NRC Excel File Explanation and Demonstration
- Interpreting Individual Program Results

Overview

- The NRC study covers 5,004 doctoral programs in 62 fields at 212 universities (50 programs at UGA)
- Provides ranking information in 59 fields (48 ranked programs at UGA)
- To be ranked, a field must have produced at least 500 Ph.D.s in the five academic years from 1999-2004, and a program must have produced at least 5 Ph.D.s in the five academic years from 2000-2005
Overview

- NRC surveys conducted about once every 15 years (1983, 1995, 2010)
- One goal is to evaluate (rank) PhD-granting programs in various fields
- Current survey has more quantifiable aspects to counter previous criticisms of being a ‘popularity contest’ that unduly favored established programs

Overview

- Rankings provided as a range of possible rankings for each program overall (two methods, R and S) and in each of three dimensions (Research Activity, Student Support and Outcomes, and Diversity of the Academic Environment)
- According to the NRC, “Users are encouraged to look at groups of programs that are in the same range as their own programs, as well as programs whose ranges are above or below, in trying to answer the question ‘Where do we stand?’”
NRC Methodology

• This is a summary of the NRC’s methodology
• For a complete SCC presentation and handout on the NRC’s methodology prepared in 2009, visit http://www.grad.uga.edu/nrc/index.html
• The NRC’s published methodology guide can be found at http://www.nap.edu/rdp/, along with the complete Assessment; the methodology guide is the basis for this presentation

NRC Methodology

• The NRC developed two methods of assigning overall ranks to programs: a regression (“R”) ranking, and a survey (“S”) ranking
  • The survey ranking has also been modified to create the three dimensional rankings mentioned
• Program data collected from universities and opinions of tenure-track faculty members contributed to both ranking methods
NRC Methodology

- Program data (used in both ranking methods) included information about students, faculty, and characteristics of the program
- Data used in the final rankings were obtained on 20 “key variables” from three groups: Faculty Characteristics, Student Characteristics, and Program Characteristics

NRC Methodology

Faculty Characteristic Variables

- Publications per faculty member per year [2000-2006] R
- Avg. citations per publication [2000-2006] R
- % Faculty with Grant [2005-06 AY] R
- % Interdisciplinary Faculty [2005-06 AY]
- % non-Asian Minority Faculty [2005-06 AY] D
- % Female Faculty [2005-06 AY] D
- Awards/Faculty Member [2001-2006] R
### NRC Methodology

#### Student Characteristic Variables
- Median GRE V or Q Score of Entering Students [2004-2006]
- % First-year Receiving Full Support [Fall 2005] S
- % First-year With External Support [Fall 2005]
- Average # of student pubs/presentations [not collected]
- % non-Asian Minority Students [Fall 2005] D
- % Female Students [Fall 2005] D
- % International Students [Fall 2005] D

#### Program Characteristic Variables
- Average # completing PhD/year [2004-2006] S
- % PhDs graduating within 6 years [F96 to F01 cohorts] S
- % Median Time to PhD Degree [F96 to F01 cohorts] S
- % PhD’s with Academic Positions [2001-2005] S
- PhD Student Work Space [-1, +1] [Fall 2005]
- PhD Student Health Insurance [-1, +1] [Fall 2005]
- Average Number of Student Support Mechanisms [Fall 2005]
NRC Methodology

- The 20 variables measured for each program are in different units and scales, making it hard to compare them, since we desire to combine them into one overall score.
- Variables were standardized according to program field so eventual importance/“weight” of variables could be easily modeled and compared.

NRC Methodology

- Standardize using $z_{i,j} = \frac{x_{i,j} - \text{Avg}_j}{\text{SD}_j}$
  
  where
  - $x_{i,j}$ = Value recorded for variable $j$ by program $i$ in field
  - $\text{Avg}_j$ = Average for variable $j$ over all programs in field
  - $\text{SD}_j$ = Standard deviation for variable $j$ over all programs in field
NRC Methodology

- A positive z-score indicates the value for a program is above average, negative indicates below average
- If the overall distribution of scores for a variable (over all programs in a field) is roughly mound-shaped, then about 2/3 of z-scores are in the range [-1,+1], about 95% are within [-2,+2], and values outside of [-3,+3] are very rare

NRC Methodology

- For the “S” rankings, all faculty (86% response rate) were asked to identify up to 4 “important” variables (score of 1), and of those, up to 2 “most important” variables (score of 2) in each of three categories: Faculty Characteristics, Student Characteristics, and Program Characteristics
- They were also asked to assign weights of importance totaling 100 to each of the 3 categories
NRC Methodology

• The final weight each faculty member assigned to a variable was the variable’s score times the weight of importance of its category, divided by the sum of that faculty member’s 20 weights (so weights would sum to 1)

• Within each field, the weight \( \bar{x} \) for a variable was the average final weight taken across all faculty members who responded

\[
\text{Survey score} = \bar{x}_1 z_1 + \bar{x}_2 z_2 + \ldots + \bar{x}_{20} z_{20}
\]

• If no randomization is incorporated, ordering survey scores for all programs provides rankings (larger survey score implies lower-numbered rank)

• The NRC incorporated randomness into both the z-scores and the \( \bar{x} \) weights to yield a range of possible ranks
NRC Methodology

• For the “R” rankings, a stratified random sample of faculty members were given program information and asked to rate 15 programs in their field (30 or 50 programs included over all raters, depending on size of field, also a stratified random sample).

• For programs that did not receive ratings from all faculty raters (58% response rate), additional faculty members were chosen until almost every program in the field had been rated by 40 or more faculty members.

• A regression equation was then calculated in each field to relate the program variables for these programs to their ratings (principal components accounted for correlation between standardized variables $z$, and weights $\bar{r}$ were adjusted so absolute values of weights summed to 1):

$$\text{Adjusted predicted rating} = \bar{r}_1 z_1 + \bar{r}_2 z_2 + \ldots + \bar{r}_{20} z_{20}$$

• Ordering adjusted predicted ratings for all programs provides ranks if no randomness is incorporated.
NRC Methodology

- Stopping here would provide one rank for each program within its field by both R and S methods.
- To account for variability in program data (i.e., year-to-year fluctuation) and in R and S weights (since results would vary according to who answers the surveys involved), the NRC decided to incorporate random variation.

NRC Methodology

- To account for variation in R and S weights, half of the faculty raters or survey respondents were randomly selected for the weight creation process, and this was repeated 500 times.
- To account for variation in program data, recorded values (before standardization) for each of the 20 variables were perturbed by adding a random error term from a normal distribution with a zero mean and a standard error that depended on the variable; this was also repeated 500 times.
NRC Methodology

- The 500 sets of R and S weights were used with the 500 sets of perturbed program data to produce 500 simulated ranking orders in each field with each method.
- For each program, the lowest and highest 5% of rankings (i.e., 25 lowest and 25 highest) achieved in this process were removed, and the lowest and highest of the remaining 450 rankings were reported as the 5th and 95th percentile rankings for both R and S methods.

NRC Methodology

- To create the dimensional rankings, the variables from the survey method were analyzed separately for each dimension in each of the 500 simulations.
- The dimensional rankings process for each dimension was otherwise similar to the overall survey ranking process.
Philosophical Observations

• Incorporation of randomness
• R vs. S rankings
• Effects of Program Size
• Politically Incorrect Results

NRC Excel File

• The Excel workbook, downloadable as “ResDocTableWin.xls” (if using a Windows computer), contains 15 sheets
The first sheet, “Start”, has basic instructions for using the Excel workbook. Make sure to enable active content, or the workbook will not be fully functional!

The second sheet, “Guide”, contains more detailed information about using the spreadsheet, as well as basic information about the data and the methodology.
NRC Excel File

- The third sheet, “Master”, contains information about all programs in the study

<table>
<thead>
<tr>
<th>ID</th>
<th>Title</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Program Name 1</td>
<td>University of California Davis</td>
</tr>
<tr>
<td>2</td>
<td>Program Name 2</td>
<td>University of California Los Angeles</td>
</tr>
<tr>
<td>3</td>
<td>Program Name 3</td>
<td>University of California Irvine</td>
</tr>
<tr>
<td>4</td>
<td>Program Name 4</td>
<td>University of California Santa Barbara</td>
</tr>
<tr>
<td>5</td>
<td>Program Name 5</td>
<td>University of Chicago</td>
</tr>
<tr>
<td>6</td>
<td>Program Name 6</td>
<td>University of Illinois at Urbana-Champaign</td>
</tr>
<tr>
<td>7</td>
<td>Program Name 7</td>
<td>University of Kentucky</td>
</tr>
<tr>
<td>8</td>
<td>Program Name 8</td>
<td>University of Maryland at College Park</td>
</tr>
<tr>
<td>9</td>
<td>Program Name 9</td>
<td>University of Michigan at Ann Arbor</td>
</tr>
<tr>
<td>10</td>
<td>Program Name 10</td>
<td>University of New Mexico</td>
</tr>
</tbody>
</table>

- The “Master” sheet is the link to all remaining sheets
- Each row is associated with one program
NRC Excel File

- Clicking on that row will link all remaining sheets to information specific to that program (if active content is enabled!)

NRC Excel File

- “Variables” contains the actual program data, as well as the standardized values and descriptive ranges of field-specific R and S weights
NRC Excel File

- The next 10 sheets give 5th and 95th percentile information for both the R and S methods of overall ranking, as well as for the three dimensional rankings.

- Each sheet gives the actual program data, along with the perturbed standardized data and random R or S weights that achieved that percentile ranking and how they combined.

NRC Excel File

- The final sheet, “Emerging Programs”, provides brief data on 14 fields in 4 broad fields.
- These programs are unranked.
Interpreting Results

Regression (R) rankings

- This school did well in its field with its .05 R ranking
- Large positive numbers in (Col 6) show why program received a high (low-numbered) ranking

For this same school, .95 R still ranked highly
- Same variables counted less, however, due to variation
Interpreting Results
Survey (S) rankings

- This school did not compete as well as the previous one
- Large negative numbers in (Col 6) demonstrate why the school received a low (high-numbered) ranking in its .05 S ranking

For this same school, the .95 S ranking was affected by similar areas as the .05 S ranking, but random variation resulted in a higher-numbered ranking.
Interpreting Results

- To find the variables in R or S rankings that positively influenced your program’s rankings, look for large (in absolute value) standardized program values and large coefficients of the same sign (+/-).
- To find the variables that negatively influenced your program’s rankings, look for large (in absolute value) standardized program values and large coefficients of the opposite sign.
- “Important” areas are those with large coefficients.
- Compare your program to others in your range of rankings, and others that received ranges outside of your own (both lower and higher).

Dimensional Rankings

- Same basic idea as S rankings, but more specific concentration of variables.
- High (positive) values in (Col 6) show where a program did well.
Interpreting Results

- Of the dimensional rankings, Research Activity had the most impact on overall rankings, and Diversity of the Academic Environment had the least.
- However, programs shouldn’t disregard these areas—while the NRC’s R and S rankings may not have depended highly on certain variables, students may value them highly when deciding on a program.

Questions?