A DATA BASED ASSESSMENT OF RESEARCH DOCTORATE PROGRAMS

A Short History

- Data collection in 2007—using 2005-6 data for the most part
- 2008-reviewing and validating data and the methodology
- 2009—A Guide to the Methodology is published.
- 2010—The final report and accompanying spreadsheets will appear.

This Talk

- A little background

- The data that will become available and how they

- can be used.

- Thinking about quality of doctoral programs and program characteristics: How do we assess quality using data from programs?

- A variety of measures—overall and in different dimensions

- Study release—and after

CGS and the Study

- A special group who will be responsible for explaining and using the results.

- A great deal of information to absorb and interpret.

- NRC will try to give you early notice of the study release.

- Details of the release are not yet available.

Committee*

- **Jeremiah P. Ostriker (NAS)**
  Committee Chair
   (Astrophysics)
  Princeton University

- **Virginia S. Hinshaw,** *Vice-Chair**
  (Biology)
  University of Hawai'i Mano'a

- **Elton D. Aberle**
   (Agriculture)
  University of Wisconsin-Madison

- **Norman Bradburn**
  (Statistics)
  University of Chicago

- **John I. Brauman (NAS)**
   (Chemistry)
   Stanford University

- **Paul Holland**
  (Statistics)
  Educational Testing Service

- **Eric W. Kaler (NAE)**
  (Engineering)
  SUNY-Stony Brook

- **Earl Lewis ***
  (history)
  Emory University

*Several members of Committee are present or former Deans of graduate schools.

Committee (2)

- **Joan F. Lorden***
  (Neuroscience)

  University of North Carolina, Charlotte

- **Carol B. Lynch***
  (Biology)

  University of Colorado, Boulder

- **Robert M. Nerem (NAE)**
  (Bioengineering)

  Georgia Institute of Technology

- **Suzanne Ortega***
  (Sociology)

  University of New Mexico

- **Catharine R. Stimpson***
  (English)

  New York University

- **Richard P. Wheeler***
  (English)
  University of Illinois

What will be released?

- The Report
  - A "slim volume" discussing what was done in the study, the data, and two *illustrative* methodologies for data based rankings.

- Online spreadsheets
  - Data for 4838 programs for the 20 variables used in the ratings calculation, and for 9 additional variables.
  - Range of rankings for 5 types of illustrative rankings: 2 overall and 3 dimensional
  - Ability to "click through" to get detail of ranking calculations

Release materials (cont'd)

- Demos to show how to query the spreadsheets

- Related effort on PhDs.com will permit calculations with user determined weights

- Press release and FAQ's

- Press conference

- Revised Methodology Guide

Later

- Ranges of rankings for computer science
- Public use database
- Release of all questionnaire data (with individual identities masked) to researchers who request it and sign a confidentiality agreement

Six Months Later

- Conference on analytic uses of the data

What can you do with the spreadsheets?

- Pick out programs to compare with programs at your institution along many lines:
  - Research activity variables
  - Student support and outcomes variables
  - Diversity of students and faculty
- And, oh yes, ranges of *illustrative* rankings along those lines and across all 20 variables.

Key points

- It is helpful to compare programs that are doing similar things by collecting the same data from all the programs in a field.

- It is possible to compare data values by forming a ranking.

- There are many ways to develop rankings—the NRC did it in two ways—there are many others.

- It is important to know what goes into a ranking.

Some things that will change from the July 2009 Methodology Guide

- The rankings and their ranges
  - 1 overall ranking range $\Rightarrow$ 2 separate ranking range calculations as *illustrations* of data-based ranking schemes
  - Change in length of range from covering 50% of a program's rankings to covering 90%

- Emphasis
  - Ranges of rankings are *illustrative*. You could get different results with different assumptions.

What is the Assessment?

- Collection and dissemination of data on important aspects of doctoral programs
  - Programs
  - Students
  - Faculty

- Development of a benchmarking/rating methodology
  - Compare **doctoral programs** in a **single field** across universities

- 212 Universities, 59 fields with ratings

Audiences

1) Prospective graduate students. Give them better information about the various programs to make more informed decisions re where to apply.

2) Faculty in the programs to better evaluate their own strengths and weaknesses.

3) Those responsible for the health of graduate programs to enable them to better assess the programs under their charge and compare more objectively to those in other institutions.

4) Those with more global interests (legislators, boards of trustees, the US government, other nations) to provide more transparency in assessing a vital US national institutional resource.

5) During the "Recession of 2008-2010" Those responsible for resource allocation decisions.

Where do the data come from?

- Standardized source providers (e.g.citations from ISI, NSF for post-graduate student plans)

- New surveys (e.g. faculty & students)

- *US institutions of higher education:*
  - *A HUGE TASK*

What Data will become Available?

Research Activity

- **Publications per faculty member going back to 1981**

- **Citations per publication (except for humanities fields) in 2005-6 with pubs going back to 1981**

- **Percent of faculty with grants (from NRC faculty questionnaire)**

- **Honors and awards per faculty member (from honorary and scholarly societies)**

Student Support and Outcomes

- **Number of PhDs**

- **Percent receiving financial support in first year**

- **Median time to degree**

- **Percent of entering cohort(s) completing within six years (eight for the humanities)**

- **Percent of graduates with definite employment or postdoc plans (from NSF)**

Summary Descriptive Information for Each Program

Program Diversity

- **Faculty:**
  - Gender diversity
  - Racial/ethnic diversity

- **Students**
  - Gender diversity
  - Racial/ethnic diversity
  - International diversity

Program Interdisciplinarity

- **Percent of faculty associated with other programs**

- **Identification of "umbrella" programs**

Questions

- A prospective student
  - What do I want to do when I finish and does the program seem to support that aim?
  - Am I likely to get funding?
  - How long will it take to complete?
  - How likely is it that if I start in a program that I will complete in a reasonable amount of time?
  - Will I be the only (woman, minority)?

More questions

- A department chair
  - What are the strengths and weaknesses of the program?
  - How does my program compare to peer programs?

- A provost
  - Where can additional resources result in the most improvement?
  - What programs could benefit from being combined with similar programs?

And Yet More Questions

- A state board of higher education
  - Do we have too many doctoral programs in a given field?
  - Which programs are strong nationally and deserve more support?

Characteristic	Program A	Program B	Program C	Program D	Program E
Publications per Allocated Faculty	4.993	4.328	4.448	2.937	2.379
Cites per Publication	3.573	3.401	2.782	2.819	2.386
Percent Faculty with Grants	88.6%	100.0%	95.5%	90.5%	73.4%
Percent Faculty Interdisciplinary	71.4%	0.0%	38.1%	18.8%	0.0%
Percent Non-Asian Minority Faculty	0.0%	5.0%	0.0%	2.9%	3.1%
Percent Female Faculty	16.2%	13.6%	8.0%	17.9%	8.8%
Awards per Allocated Faculty	1.929	7.291	1.896	0.640	0.424
Average GRE-Q	712	772	767	703	673
Percent 1st yr. Students w/Full Support	100.0%	100.0%	100.0%	100.0%	86.0%
Percent 1st yr Students with External Funding	0.0%	0.0%	22.2%	0.0%	0.0%
Percent Non-Asian Minority Students	2.8%	1.9%	3.2%	8.0%	13.6%
Percent Female Students	39.3%	39.1%	39.8%	42.2%	37.3%
Percent International Students	23.0%	42.7%	37.2%	45.1%	31.3%
Average PhDs 2002 to 2006	31.6	17.4	20.2	11.400	19.800
Percent Completing within 6 years	49.3%	77.8%	67.6%	41.6%	54.0%
Time to Degree Full and Part Time	5.7	5	4.9	4.3	5.000
Percent students in Academic Positions	17.2%	32.1%	25.6%	20.0%	12.2%
Student Work Space	1	1	1	1	1
Health Insurance	1	1	1	1	1

THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

How will the ratings/rankings work?

Two Approaches

- Asked faculty what they thought was important to the quality of a doctoral program and developed weights (S-weights).

- Asked a sample of faculty in each field how they would rate a sample of programs. Related those ratings to 20 program characteristics through a regression (R-weights)

- Calculated ratings using each approach for all programs in a field, based on program values for the 20 characteristics.

- The rankings will be *illustrative*.

Overall Rating AND Dimensional Measures

- Student Treatment and Outcomes

- Diversity of the Academic Environment

- Research Activity of Program Faculty

The Twenty Key Variables used in the Rankings

- Publications per allocated faculty
- Citations (exc. Humanities) per publication
- Percent faculty with grants
- Awards per faculty
- Percent 1st Yr. Full Support
- Percent Completing in 6 yrs. or less (8 yrs. for humanities)
- Median Time to degree
- Students with Academic Plans
- Collects Outcomes data

- Percent Faculty Minority
- Percent Faculty Female
- Percent Students Minority
- Percent Students Female
- Percent Students International
- *Percent Interdisciplinary*
- *Average GRE-Q*
- *Number of PhDs 2002-2006*
- *Student Workspace*
- *Student Health Insurance*
- *Student Activities*

Ratings: What measures "Quality of PhD Program"?

- Usual Approaches:

  Those who design the study construct measures on an ad hoc basis.

    - Based on reputation
    - Based on refinements of scholarly productivity measures

- NRC Approach:

  Faculty input on a field by field basis determines the measures. Two estimators of faculty values to estimate best measures: direct (S) and regression-derived (R).

Sources of uncertainty for any rating

- Differences among raters

- Year-to-year variation in the data

- Range of error in any statistical estimation

➔ Every rating has a range, and so do the rankings derived from the ratings

- We settled on a broad range-one that covers 90% of the estimated rankings for a program

- Unincluded and unquantifiable factors may also matter—but the committee focussed on what could be quantified.

Changes to Encourage Use of the Study Data

- Make data easily available via web

- Disseminate through professional societies

- Permit customized comparisons by users

- Provide analytical tools and associated essays (later)

Does it matter that it's "late?"

- There is a trade-off between speed and accuracy
  - We spent a lot of time trying to get universities to provide comparable data and developing our model.
- In most fields, and especially now, doctoral faculty change relatively slowly, as do patterns of publication
- We would like to update the study in the next two years
  - Now that we have developed the statistical machinery, it is a data updating task, which could be carried out on-line.
 - But, we will need to obtain funding

Looking at the Rankings based on R's and S's

- How are the R's different from the S's?
  - R's are regression-based. Look at how ratings depend on the program characteristics.
    - More technically, carry out a backwards regression on the characteristics that have been transformed with a principle components transformation.
    - R's based on relatively small samples of faculty
  - S's are survey-based weights derived from an idealized question

Looking at the Rankings based on R's and S's (2)

- Why might rankings based on R's be different from those based on S's?
  - R's have a reputational component
    - What can go into reputation? Program visibility, age of program, halo effects
      - Example: size of program may be very important, even though faculty don't think that size matters to quality
  - S's reflect normative judgments by faculty of the components of perceived quality

R and G Median Rankings for Programs in Chemistry

G Median Ranking
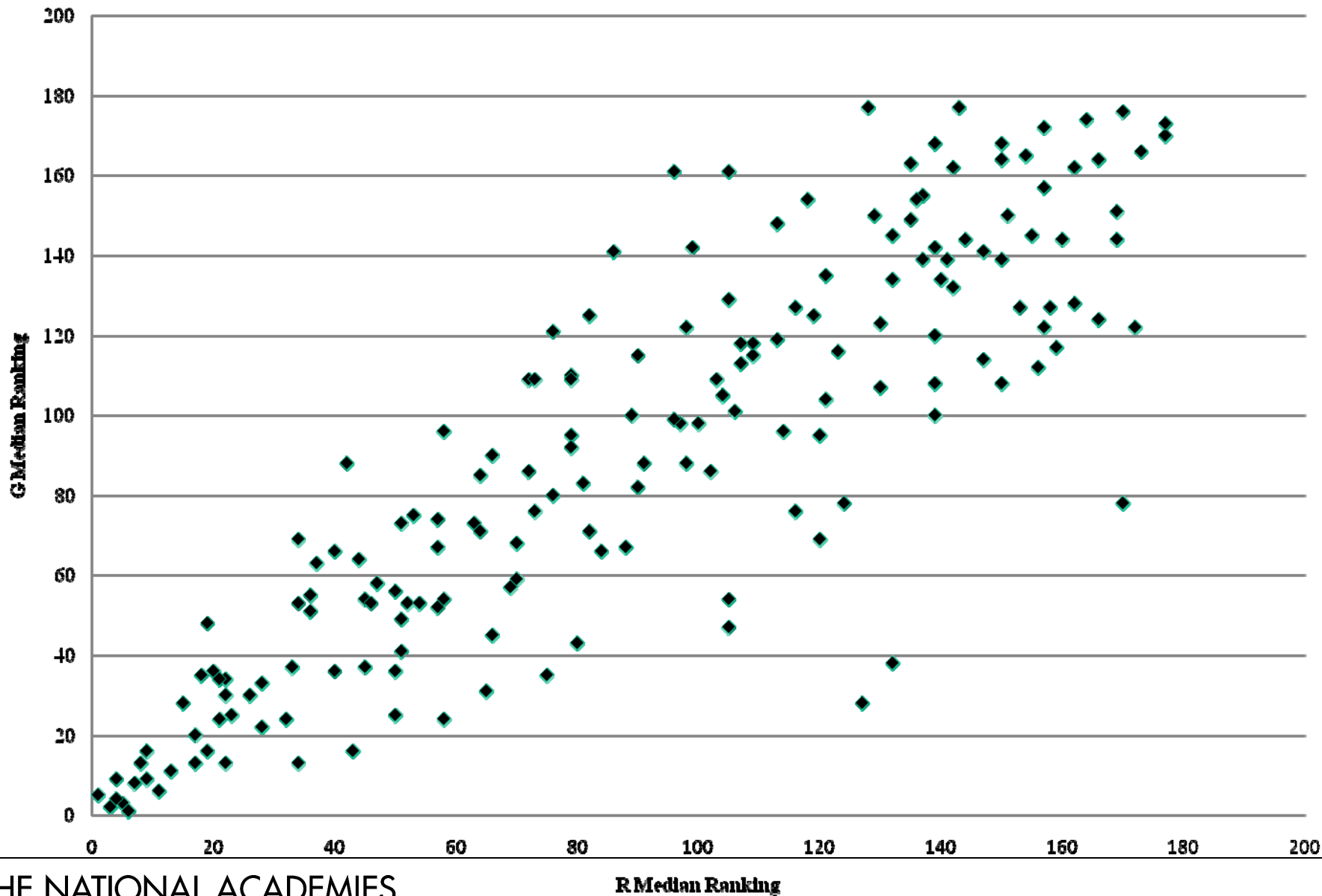
R Median Ranking

What do we make from this?

- Median R-rankings and S-rankings are close but nowhere near perfectly correlated.

- S-rankings show more programs with a broad range

- Need to look at the coefficients that go into the calculation and ask what the purpose of the ranking is.

A Sample Comparison

R and S-based Rankings for 5 Programs in a Field

Institution Name	R5	R95	S5	S95
Institution A	4	17	10	29
Institution B	4	27	3	10
Institution C	13	37	8	23
Institution D	31	79	31	86
Institution E	52	102	91	150

Dimensional Rankings for the Same Programs

Institution Name	RA5	RA95	SS5	SS95	D5	D95
Institution A	7	29	9	66	81	131
Institution B	3	12	31	110	97	147
Institution C	9	39	6	42	101	151
Institution D	21	85	21	93	42	97
Institution E	53	124	53	133	77	128

Coefficients for Chemistry Programs

Characteristic	R5	R95	S5	S95
Publications per Allocated Faculty	-0.011	0.144	0.146	0.151
Cites per Publication	0.037	0.086	0.125	0.130
Percent Faculty with Grants	0.066	0.118	0.163	0.167
Percent Faculty Interdisciplinary	-0.002	0.083	0.033	0.036
Percent Non-Asian Minority Faculty	-0.027	0.049	0.007	0.009
Percent Female Faculty	-0.061	0.011	0.011	0.013
Awards per Allocated Faculty	0.015	0.088	0.081	0.086
Average GRE-Q	-0.011	0.062	0.066	0.070
Percent 1st yr. Students w/ Full Support	0.045	0.101	0.053	0.057
Percent 1st yr Students with External Funding	-0.049	0.005	0.043	0.047
Percent Non-Asian Minority Students	-0.062	-0.007	0.015	0.017
Percent Female Students	-0.023	0.037	0.016	0.018
Percent International Students	-0.068	-0.022	0.007	0.009
Average PhDs 2002 to 2006	0.101	0.181	0.038	0.041
Percent Completing within 6 years	-0.025	0.026	0.045	0.048
Time to Degree Full and Part Time	-0.019	0.028	-0.025	-0.023
Percent students in Academic Positions	-0.026	0.055	0.067	0.069
Student Work Space	0.006	0.076	0.005	0.006
Health Insurance	0.022	0.082	0.003	0.004
Number of Student Activities Offered	0.062	0.117	0.022	0.024

Big Points

- Data-based ranking is not a simple task

- Rankings depend on the values of measures used and the weight that is put on them.

- The NRC is not endorsing any method as "best"

- The NRC study will be complex. We will try to make it useful—but that is also up to you.

Things to Remember

- The rankings come from the ratings of the programs arranged in numerical order.

- The ratings are calculated 500 times with the different half samples of raters and variation (in a small range) of the data values

- The database will show the rankings for the 5th and 95th percentile values on R, S, and three dimensional measures

More things to remember

- You will be able to access the values that went into the calculation of the 5th and 95th percentile values.

  - Note: the calculation uses standardized values. We will also show the actual values for the program and the standardized value in the rating calculation

- The dimensional rankings spotlight program characteristics not prominent in the overall rankings

To Learn More About the Study

http://sites.nationalacademies.org/pga/Resdoc/index.htm

Or contact

ckuh@nas.edu

Or

jvoytuk@nas.edu