The GSS is an unreliable indicator of biological sciences postdoc population trends

Christopher L. Pickett[a,b,1], Adriana Bankston[c] and Gary S. McDowell[c,d]

Keywords: biomedical workforce, graduate training, postdoctoral training

Abbreviations: NSF: National Science Foundation; NCSES: National Center for Science and Engineering Statistics; GSS: Survey of Graduate Students and Postdoctorates in Science and Engineering

[a]Rescuing Biomedical Research
1200 New York Avenue, Room 767
Washington, DC 20005

[b]Princeton University
Lewis-Sigler Institute
Princeton, NJ 08544

[c]The Future of Research
848 Brockton Avenue
Abington, MA 02351

[d]Manylabs
1086 Folsom Street
San Francisco, CA 94103

[1]Corresponding author
Email: clp3@princeton.edu

Abstract

The postdoctoral research position is an essential step on the academic career track, and the biomedical research enterprise has become heavily dependent on postdoctoral scholars to conduct experimental research. Monitoring the employment trends in the postdoc population is important for crafting and evaluating policies that affect this critical population. The primary survey for understanding the trends of the biological sciences postdoc population is the Survey of Graduate Students and Postdoctorates in Science and Engineering (GSS) administered by the National Center for Science and Engineering Statistics. Here, we analyzed the yearly changes in the biological sciences postdoc population at institutions surveyed by the GSS. We find that institutional variability in reporting their biological sciences postdoc populations, which sometimes varies by more than 2-fold over consecutive years, masks larger trends in the employment of biological sciences postdocs. Universities indicated the most common cause for the changes was improving institutional policy and tracking of postdocs. We propose the adoption of a unified definition of a postdoc, consolidation of postdoc titles and the creation of an index to better assess biological sciences postdoc trends.

Introduction

Postdoctoral scholars are pivotal for the process of discovery in the biomedical research enterprise. Over the past several decades, the expansion of university research capacity and the availability of federal research grants have created a strong demand for highly skilled postdocs (Heggeness et al, 2016; Heggeness et al, 2017; National Institutes of Health 2016a, 2016b; Stephan, 2012). Despite this population growth and their importance to the research enterprise, comprehensive population-level data on postdocs is lacking. Within a single university, some postdocs may be categorized as employees whereas others are contractors, and some universities have multiple designations for postdocs that reflect differences in departmental practice, seniority and funding source (Ferguson et al, 2014). The variety of titles, while intending to clarify human resources policies or confer status on the postdoc, create inconsistencies that introduce significant difficulties in collecting data on postdoc populations (Committee to Review the State of Postdoctoral Experience in Scientists and Engineers, 2014; McDowell, 2016). This is not a new problem: a 1969 report from the National Research Council states, "Although postdoctoral appointees were present on many campuses, their numbers and functions were not known nationally and, in many instances, were not even known to the host universities," (Curtis, 1969).

That's not to say the research community has not begun to tackle this problem. Spurred by National Academies reports and postdoc advocacy groups such as the National Postdoctoral

Association and Future of Research, universities have instituted offices dedicated to supporting postdocs, created institutional definitions of postdocs and crafted institutional policies to track them. However, these efforts are not uniform across the country: definitions of who is a postdoc as well as titles, benefits, compensation and other aspects of the postdoc experience vary across institutions.

At the national level, the National Science Foundation's National Center for Science and Engineering Statistics conducts several surveys of the nation's scientists and engineers to compile population data (http://www.nsf.gov/statistics/surveys.cfm), including the Survey of Graduate Students and Postdoctorates in Science and Engineering (GSS). The GSS provides the most comprehensive assessment of trends in the postdoc population, yet the GSS collects information on postdocs only from Ph.D. granting institutions and omits free-standing research centers and federal institutions that employ postdocs. The NCSES' forthcoming Early Career Doctorates Survey may clear up some of these issues, and these data are expected in 2018.

While the GSS is generally considered the best tool available to measure the U.S. postdoc population, the GSS data collection practices are fluid. From 2007 to 2010, the GSS altered its methods of postdoc data collection (Einaudi et al., 2013). And in 2014, the number of institutions surveyed in the GSS, known as the survey frame, increased so as to include more institutions with postdocs than had been evaluated before (Arbeit et al., 2016). These improvements in the GSS are welcome and enhance the accuracy of the survey, but they

present significant problems when trying to evaluate the size of the postdoc population and examine long-term trends. Beyond this, the quality and consistency of the GSS data is dependent on the quality and consistency of postdoc information reported by universities. Policy changes at institutions that alter the accounting of postdocs could have significant effects on the reported trends in the postdoc population.

Here, we provide evidence that inconsistent reporting of institutional postdoc populations is a significant source of error in the GSS measurements. We focused on the population of postdocs in the biological sciences as defined by the GSS (National Science Foundation, 2012). We highlight several institutions with widely fluctuating biological sciences postdoc populations: the reported postdoc population of some institutions suddenly drops to 0 for one year only to be followed immediately by a rapid increase in the next year. Conversely, the biological sciences postdoc population at some institutions more than doubled over a single year only to return to near baseline levels one or two years later. Finally, we highlight instances where changes in a single institution's policies mask contractions of the overall biological sciences postdoc population. These data demonstrate the GSS biological sciences postdoc population data are unreliable when examining biological sciences postdoc population trends. The unreliability is the result of the biomedical research enterprise failing to create a unified definition of a postdoc or put in place sufficient methods to account for and track postdocs. This has real consequences for those attempting to understand and change policies affecting

postdocs. We call for a unified definition of a postdoc, simplified job titles and the creation of an index to more accurately follow trends in the biological sciences postdoc population.

Results

When quantifying the biological science postdoc population of an institution, policy changes can affect the postdoc population beyond simple job gains and losses. For example, an institution redefining who is considered a postdoc may have profound effects on the number of biological sciences postdocs reported. We examined the biological sciences postdoc populations at individual institutions as reported to the GSS between 1980 and 2015. To detect the largest changes, which are more likely to have a significant effect on the overall tally of biological sciences postdocs, we examined the 82 institutions that reported 100 or more biological sciences postdocs at least once between 1980 and 2015 (See Data collection and limitations).

We began by defining a large change in an institution's biological sciences postdoc population as a 2-fold increase or decrease over a single year. We identified 37 occurrences of an institution reporting 2-fold more biological sciences postdocs than the prior year and 25 instances of an institution reporting 2-fold fewer postdocs than the prior year (Fig. 1; Table S1). These 62 occurrences happened at 28 institutions (Table S1).

Ten institutions had a single two-fold change in the biological sciences postdoc population between 1980 and 2015—eight institutions reported a two-fold or more increase and two institutions reported a two-fold or more decrease (Table S1). Of the institutions registering multiple two-fold changes between 1980 and 2015, most are reciprocal changes

(Table S1). For example, the University of Florida reported 108 biological sciences postdocs in 2000, 232 in 2001, and 106 in 2002 (Table S1). Similarly, the University of California, Riverside reported 92 biological sciences postdocs in 2009, 27 in 2010 and 89 in 2011 (Table S1). Other institutions followed these patterns, sometimes stretching the increase or decline across two or more years.

Of note, five institutions reported 0 biological sciences postdocs after reporting more than 75 at least two years before. Northwestern University (1997), the University of Maryland, Baltimore (2007) and the University of Tennessee Health Science Center (2009) each reported 0 postdocs for a single year (Table S1). Brown University declined from 96 postdocs in 1999 to 0 in 2000, and it remained at 0 until 2005 when it increased to 88 postdocs (Table S1).

The GSS defines whether a postdoc is a biological sciences postdoc or a health fields postdoc based on the department or "organizational unit" they work in (National Science Foundation, 2012). One possible explanation for the 2-fold changes in the biological sciences postdoc population could be that some were reclassified as health fields postdocs. To determine if reclassification as health fields postdocs could account for the observed changes, we analyzed the numbers of health fields postdocs at the institutions analyzed above. Of the 62 2-fold changes identified in the biological science postdoc population between 1980 and 2015, there were 18 instances where the biological sciences and health fields postdocs changed in opposite directions, but the changes were of approximately equal magnitude in only three of

these instances (Table 1). These data indicate that reclassifying biological sciences postdocs as health fields postdocs cannot explain for many of the large changes in the biological sciences postdoc population.

Other institutional changes, however, could account for these changes. For example, institutions may have engaged in mass hirings or layoffs of biological sciences postdocs, and sometimes both within a few years of each other. Second, institutions may have changed their policies on how postdocs are classified and counted, resulting in dramatic changes in the reported biological sciences postdoc population. We contacted the institutions having a 2-fold or more change in the biological sciences postdoc populations between 1980 and 2015 to assess whether these possibilities are responsible for our observation. Of the 62 2-fold or more changes, only seven could be confirmed as being due to an institutional policy change, and none were attributed to possible job gains and losses (Table S2; Supplemental Text). Twelve institutions, accounting for 23 2-fold or more changes, had no information on the population changes as they occurred before the institution instituted reliable tracking techniques (Table S2; Supplemental Text). Eleven institutions, accounting for another 27 of 62 2-fold or more changes, did not recognize the postdoc population data we presented them (Table S2; Fig. S1; Supplemental Text). This is likely due to differences in classification: institutions collect information on postdocs across all biomedical departments, while the GSS splits these postdocs

into biological sciences and health fields postdocs. Three institutions, accounting for five observed changes, did not respond to our request for information.

How likely are the 2-fold or more changes at individual institutions to affect the overall population trends of biological sciences postdocs? To answer this, we summed the change in the postdoc population due to the reported changes from schools that had a 2-fold or more change in their biological sciences postdocs over the previous year and compared it to the change in the overall biological sciences postdoc population. For most years, the change in the postdoc population derived from schools with a 2-fold or more change accounts for only a small fraction of the overall change. However, 2004, 2005, 2006 and 2007 are exceptions. In 2004, the overall biological sciences postdoc population increased by 91 postdocs, but a policy change at UCSF resulted in an increase of over 500 biological sciences postdocs over the previous year, a 4.8-fold increase (Table 2; Table S2). Thus, the policy change at UCSF masked a decline of 400 postdocs in the overall postdoc population. In 2005, the overall population increased by only 31 postdocs, but the Brown University biological postdoc population increased from 0 to 88 postdocs, suggesting the change at Brown masks a 57-postdoc decline in the rest of the enterprise (Table 2). In 2006, the overall population increased by 60, but three institutions reported a 2-fold or more change accounting for a decline of 194 postdocs (Table 2). This indicates those institutions reporting a 2-fold or more change masked much larger growth in the population (Table 2). In 2007, the number of postdocs accounted for in the institutions

reporting a 2-fold or more change was nearly equal to the overall change in the postdoc population suggesting that population growth could have been flat (Table 2). These data points are important because the general sense was that the postdoc population increased across these years (Garrison et al., 2016). Yet it appears this small expansion in the biological sciences postdoc population may have been driven by policy changes rather than true job growth.

Data collection and limitations

We queried the NCSES's GSS through WebCASPAR (https://ncsesdata.nsf.gov/webcaspar/). We collected the number of postdocs at all institutions in all Broad Academic Disciplines from 1980 to 2015.

Data sorting and analysis was done in Microsoft Excel. First, we isolated the Biological Sciences values for each institution in the survey frame. We then used standard Excel formulas to identify institutions reporting 100 or more postdocs at least once between 1980 and 2015. Some institutional data was manually corrected: We combined the UMDNJ and Rutgers data because of their 2013 merger, the Georgia Health Sciences Center and Augusta University data because of their 2013 name change, and the Texas A&M University and Texas A&M Health Sciences Center data because of their 2014 merger. The City of Hope, Cold Spring Harbor Laboratory, Icahn School of Medicine, the Mayo Clinic Graduate School, Sanford-Burnham Medical Research Institute and The Scripps Research Institute were not included in the survey frame for the entire 35 years between 1980 and 2015, and they were censored from our analysis.

To determine fold-change, the postdoc population reported by an institution in a specific year was divided by the postdoc population in the year immediately prior for all values reported between 1980 and 2015 for the dataset described above. Values ≥ 2.0 were a 2-fold increase and values ≤ 0.50 were a 2-fold decrease of the postdoc population over a one year

period. Postdoc populations that declined to 0 were classified as a >2-fold reduction and any increase from a 0 value was classified as a >2-fold increase.

Discussion

Postdocs are a vital part of the research enterprise, and understanding the dynamics of this population is essential to developing policies that promote a vibrant research enterprise. Despite this, data on the postdoc population are unreliable. When considering the NSF surveys that gather data on postdocs, the authors of the 2014 Postdoc Experience Revisited report remarked, "[The committee] has little confidence in the accuracy of the absolute number of postdoctoral researchers, and it is particularly dubious about the quality of the information about postdoctoral researchers who are temporary residents and earned their Ph.D.'s in other countries. Nevertheless, the committee considers the available data to be a reliable indicator of trends over time. The gaps and flaws that exist are the same gaps and flaws that have existed for decades, so at least it may be supposed that the data possess some internal consistency," (Committee to Review the State of Postdoctoral Experience in Scientists and Engineers et al., 2014).

Our findings challenge the notion that the Survey of Graduate Students and Postdoctorates in Science and Engineering is internally consistent and "a reliable indicator of trends over time." Dramatic changes in the biological sciences postdoc population due to institutional policy changes have demonstrably distorted the trends of this population in the mid-2000s. Rather than a continuous increase from 1980 to 2010, our data suggest policy

changes at a few institutions in specific years masked declines in the overall biological sciences postdoc population.

Concerns over the quality of data on the postdoc population are not new. Reports since the mid-1990s recommended improving data collection on postdocs (National Academy of Sciences, 2000; National Research Council, 1994, 1998, 2005). We note that 50 of 62 observed 2-fold increases or decreases in an institution's postdoc population occurred after 1995 possibly reflecting institutions attempting to follow these recommendations and better track their postdocs. In addition to these recommendations, focused advocacy on behalf of postdocs by organizations like the National Postdoctoral Association began in the early 2000s (Sreenivasan, 2003). The recommendations made by these reports and organizations may be partly responsible for the improved tracking of postdocs we observed.

We commend the institutions that have changed policies to better track their postdocs. Similarly, the expansion of the survey frame in 2014 and the encouragement of the designation of officers responsible specifically for reporting postdoc information in 2010 are positive developments for these datasets. These improvements in data collection are essential to better understand this population that is so critical to the biomedical research enterprise. However, as long as institutions individually make policy changes with regard to how they count and report their postdocs, the GSS will remain an imperfect source of information on the trends of the overall postdoc population.

One way to circumvent this problem is widespread adoption of simplified Human Resources job classifications and a common postdoc definition across institutions. In lieu of this, we recommend the NCSES develop an institutional index, similar in concept to the Dow Jones Industrial Average or the S&P 500 indices (Investopedia, 2015, 2017). This index would be defined, at least in part, by selecting institutions whose postdoc populations have changed within a reasonable percentage up or down over a specified time. For example, an index could be comprised of institutions that (1) for the past 15 years, (2) have reported 100 postdocs or more and (3) have had a yearly change in its biological sciences postdoc population within a specified range that excludes large additions or subtractions due to policy changes. The exclusion of artificial changes in the biological sciences postdoc population would make an index an important tool to better understand the population's trends.

Drawing conclusions from continuous datasets, such as the surveys conducted by the NCSES, should be done with caution. We proposed the introduction of an index that is sensitive to the trends in the biological sciences postdoc population while eliminating artefacts introduced by institutional policy changes. This is not an ideal solution to the problem of understanding the trends of the overall postdoc population. However, it appears to be one of the best options the community has until a unified definition of a postdoc is adopted by all institutions and other complicating factors of counting postdocs, such as the proliferation of job titles, are removed. Beyond improved data reporting, transparency in career outcomes and

greater inclusion of junior scientists in policy discussions are required to better understand the factors affecting the Ph.D. and postdoc populations (Dolan et al., 2016; McDowell et al., 2014; Pickett et al., 2015; Polka et al., 2015).

Acknowledgements

We thank Shirley Tilghman for helpful comments on an earlier version of this manuscript. C.L.P. and G.S.M. are supported by the Open Philanthropy Project and G.S.M. is supported at Manylabs by the Gordon and Betty Moore Foundation.

Figure Legends

Figure 1: The number of 2-fold or more changes in the number of biological sciences postdoc populations of individual institutions reported in a given year between 1980 and 2015. Bars are the summation of the number of occurrences of a 2-fold increase or decrease.

Figure S1: Graphs of the biological sciences postdoc population at individual institutions from 1980 to 2015. The number of postdocs reported by an institution is on the y-axis and the year is on the x-axis. A blue line indicates a 2-fold or more increase in the postdoc population, a yellow line indicates a 2-fold or more decrease, and a black line indicates the year-to-year change was less than 2-fold.

References

Arbeit, C. A., Einaudi, P., Green, P., & Kang, K. H. (2016). *Assessing the Impact of Frame Changes on Trend Data from the Survey of Graduate Students and Postdoctorates in Science and Engineering.* (No. Special Report NSF 16-314). Arlington, VA: National Science Foundation, National Center for Science and Engineering Statistics. Retrieved from http://ttp://www.nsf.gov/statistics/gradpostdoc/

Committee to Review the State of Postdoctoral Experience in Scientists and Engineers, Committee on Science, Engineering, and Public Policy, Policy and Global Affairs, National Academy of Sciences, National Academy of Engineering, & Institute of Medicine. (2014). *The postdoctoral experience revisited*. Washington (DC): National Academies Press (US). doi:10.17226/18982

Curtis, R. B. (1969). *Invisible University: Postdoctoral Education in the United States. Report of a Study Conducted Under the Auspices of the National Research Council. [Richard B. Curtis, Study Director]*. Washington, D.C.: National Academies Press. doi:10.17226/18693

Dolan, K. T., Pierre, J. F., & Heckler, E. J. (2016). Revitalizing biomedical research: recommendations from the Future of Research Chicago Symposium [version 1; referees: 2 approved with reservations]. *F1000Research*, *5*. doi:10.12688/f1000research.9080.1

Einaudi, P., Heuer, R., & Green, P. (2013). *Counts of Postdoctoral Appointees in Science, Engineering, and Health Rise with Reporting Improvements*. Arlington, VA: National Science

Foundation.

Ferguson, K., Huang, B., Beckman, L., & Sinche, M. (2014). National Postdoctoral Association Institutional Policy Report 2014: Supporting and Developing Postdoctoral Scholars. *National Postdoctoral Association*. Retrieved from http://www.nationalpostdoc.org/?policy_report_databa

Garrison, H. H., Justement, L. B., & Gerbi, S. A. (2016). Biomedical science postdocs: an end to the era of expansion. *The FASEB Journal*, *30*(1), 41–44. doi:10.1096/fj.15-280552

Heggeness, M., Gunsalus, K., Pacas, J., & McDowell, G. (2016). Preparing for the 21st Century Biomedical Research Job Market: Using Census Data to Inform Policy and Career Decision-Making Version 1. *Self-Journals of Science*. Retrieved from http://sjscience.org/article?id=570

Heggeness, M. L., Gunsalus, K. T. W., Pacas, J., & McDowell, G. (2017). The new face of US science. *Nature*, *541*(7635), 21–23. doi:10.1038/541021a

Investopedia (2015). What does the S&P 500 index measure and how is it calculated? Retrieved July 24, 2017, from http://www.investopedia.com/ask/answers/040215/what-does-sp-500-index-measure-and-how-it-calculated.asp?lgl=rira-baseline-vertical

Investopedia (2017). How Now, Dow? What Moves The DJIA? Retrieved July 25, 2017, from http://www.investopedia.com/articles/02/082702.asp?lgl=rira-baseline-vertical

McDowell, G. S. (2016, April). Four reasons we don’t need 37 names for postdocs. Retrieved

July 8, 2016, from http://www.asbmb.org/asbmbtoday/201604/Education/Postdoc/

McDowell, G. S., Gunsalus, K. T. W., MacKellar, D. C., Mazzilli, S. A., Pai, V. P., Goodwin, P. R., … Polka, J. K. (2014). Shaping the Future of Research: a perspective from junior scientists. [version 2; referees: 2 approved]. *F1000Research*, *3*, 291. doi:10.12688/f1000research.5878.2

National Academy of Sciences; National Academy of Engineering; Institute of Medicine; Committee on Science, Engineering, and Public Policy. (2000). *Enhancing the Postdoctoral Experience for Scientists and Engineers*. Washington, D.C.: National Academies Press. doi:10.17226/9831

National Institutes of Health. (2016a). Kirschstein-NRSA Stipend History. Retrieved July 19, 2016, from https://researchtraining.nih.gov/sites/default/files/pdf/NRSA_Stipend_History_Graph.pdf

National Institutes of Health. (2016b). NIH Data Book. Retrieved May 31, 2016, from https://report.nih.gov/nihdatabook/index.aspx

National Research Council (US) Committee on Bridges to Independence: Identifying Opportunities for and Challenges to Fostering the Independence of Young Investigators in the Life Sciences. (2005). *Bridges to Independence: Fostering the Independence of New Investigators in Biomedical Research*. Washington (DC): National Academies Press (US).

National Research Council (US) Committee on Dimensions, Causes, and Implications of Recent

Trends in the Careers of Life Scientists. (1998). *Trends in the Early Careers of Life Scientists*. Washington (DC): National Academies Press (US).

National Research Council (US) Committee on National Needs for Biomedical and Behavioral Research Personnel. (1994). *Meeting the Nation's Needs for Biomedical and Behavioral Scientists*. Washington (DC): National Academies Press (US).

National Science Foundation. (2012). Worksheet for Survey of Graduate Students and Post doctorates in Science and Engineering Fall 2012.

Pickett, C. L., Corb, B. W., Matthews, C. R., Sundquist, W. I., & Berg, J. M. (2015). Toward a sustainable biomedical research enterprise: Finding consensus and implementing recommendations. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(35), 10832–10836. doi:10.1073/pnas.1509901112

Polka, J. K., Krukenberg, K. A., & McDowell, G. S. (2015). A call for transparency in tracking student and postdoc career outcomes. *Molecular Biology of the Cell*, *26*(8), 1413–1415. doi:10.1091/mbc.E14-10-1432

Sreenivasan, A. (2003). The National Postdoc Association Makes Its Debut. *Science*. Retrieved from http://www.sciencemag.org/careers/2003/03/national-postdoc-association-makes-its-debut

Stephan, P. E. (2012). *How economics shapes science*. Cambridge, Mass: Harvard University Press.

Table 1: Roughly reciprocal changes between biological sciences and health fields postdoc populations

Institution	Postdoc population					
	Year	Biological sciences[a]	Health fields[b]	Year	Biological sciences[a]	Health fields[b]
Augusta University/Georgia Health Sciences University	1998	27	43	1999	69 (+42)	3 (-40)
University of Florida	2001	232	90	2002	106 (-126)	188 (+98)
University of California, San Francisco	2011	652	439	2012	325 (-327)	689 (+254)

[a]Number of biological sciences postdocs reported by the institution in the indicated year. Figures in parentheses indicate the change in postdoc numbers from the previous year.

[b]Number of health fields postdocs reported by the institution in the indicated year. Figures in parentheses indicate the change in postdoc numbers from the previous year.

Table 2: Dramatic changes in the biological sciences postdoc populations of just a few institutions can significantly alter the perceived trends of the population.

Year	Change in >2-fold changing institutions[a]	Total population[b]	Difference[c]
2004[d]	+509	+91	-418
2005[e]	+88	+31	-57
2006[f]	-194	+60	+254
2007[g]	+287	+302	+15

[a]The sum of the changes at institutions that reported a change in the biological sciences postdoc population that was greater than or less than 2-fold over the previous year.

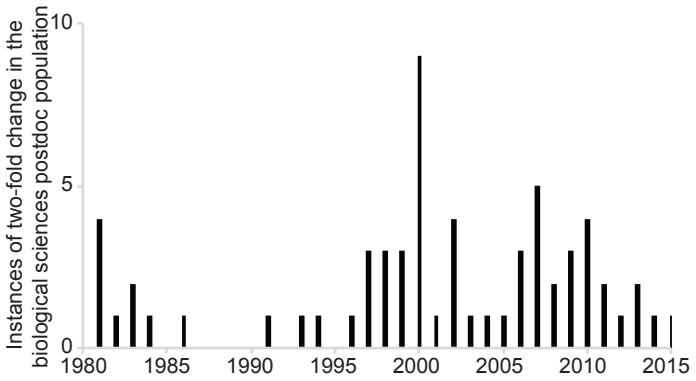
[b]The change in the total biological sciences postdoc population over the previous year.

[c]Total population change minus the change from >2-fold changing institutions.

[d]The only institution reporting a 2-fold or more change in 2005 was the University of California San Francisco.

[e]The only institution reporting a 2-fold or more change in 2005 was Brown University.

[f]The institutions that reported a 2-fold or more change in 2006 were Case Western Reserve University, Stony Brook University and the University of Maryland Baltimore.

[g]The institutions that reported a 2-fold or more change in 2007 were Stony Brook University, University of Maryland Baltimore, University of Massachusetts Medical School, The Univeristy of Tennessee Health Science Center.

Pickett and McDowell
Figure 1