Original Report:
Big Data
and its Application
in Health Research

Enhancing Diversity in Biomedical Data Science

Judith E Canner, PhD[1]; Archana J. McEligot, PhD[2];
María-Eglée Pérez, PhD[3]; Lei Qian, PhD[4]; Xinzhi Zhang, MD, PhD[5]

The gap in educational attainment separating underrepresented minorities from Whites and Asians remains wide. Such a gap has significant impact on workforce diversity and inclusion among cross-cutting Biomedical Data Science (BDS) research, which presents great opportunities as well as major challenges for addressing health disparities. This article provides a brief description of the newly established National Institutes of Health Big Data to Knowledge (BD2K) diversity initiatives at four universities: California State University, Monterey Bay; Fisk University; University of Puerto Rico, Río Piedras Campus; and California State University, Fullerton. We emphasize three main barriers to BDS careers (ie, preparation, exposure, and access to resources) experienced among those pioneer programs and recommendations for possible solutions (ie, early and proactive mentoring, enriched research experience, and data science curriculum development). The diversity disparities in BDS demonstrate the need for educators, researchers, and funding agencies to support evidence-based practices that will lead to the diversification of the BDS workforce. *Ethn Dis.* 2017;27(2):107-116; doi:10.18865/ed.27.2.107.

Keywords: Big Data; Health Disparities; Health Inequities; Workforce Diversity

[1]California State University, Monterey Bay
[2]California State University, Fullerton
[3]University of Puerto Rico, Río Piedras Campus
[4]Fisk University
[5]National Institutes of Health, National Institute on Minority Health and Health Disparities

Address correspondence to Judith E Canner, PhD; California State University, Monterey Bay; 100 Campus Center, Seaside, CA 93955; 831.582.4403; jcanner@csumb.edu

Introduction

Biomedical Data Science (BDS) is defined by the volume, variety, velocity, and veracity of complex data with the potential to create new knowledge and innovative solutions relevant to health disparities and disease prevention.[1] BDS is pivotal to identify and address factors associated with health disparities in disadvantaged populations. For example, African Americans (AA), and Native Hawaiians and other Pacific Islander (NHOPI) populations have significantly higher cancer morbidities and mortalities compared with non-Hispanic Whites.[2-5] Further, NHOPI suffer from a higher prevalence of the leading health disparity indicators compared with non-Hispanic Whites and other ethnic groups.[6] A multitude of interrelated variables, such as socioeconomic status, access to health care, genetics, obesity, tobacco use, and alcohol use contribute to health disparities in diverse ethnic populations. The depth and breadth of large and varied data sources, technologies, and platforms provides a clear opportunity to address health disparities via BDS.

However, the advancement of BDS is plagued by the lack of well-trained and empowered biomedical researchers and trainees who have access to the technology, tools, and financial resources necessary to address the challenges of BDS. Despite an increase in the racial and ethnic diversity of the general population, there remains a substantial gap in the number of underrepresented college students trained in biomedical sciences, with African Americans, Hispanics, and Native Americans accounting for

…the advancement of Biomedical Data Science (BDS) is plagued by the lack of well-trained and empowered biomedical researchers and trainees who have access to the technology, tools, and financial resources necessary to address the challenges of BDS.

only 7.1% of all employed biological/biomedical and life sciences workforce.[7] In addition, educational and research opportunities are not equally available to all. The gap in educational attainment separating underrepresented minorities from non-Hispanic Whites and Asians remains wide, including mathematics, statistics, and computer sciences, all of which are critical for BDS.[7] The reasons for the racial and gender gap in attainment are expansive and include family pressures, stereotype threat, institutional biases in recruitment, institutionalized racism and sexism in academic research culture, lack of financial support, lack of mentoring, and the list could continue. The diversity disparities in BDS demonstrate the need for educators, researchers, and funding agencies to support evidence-based practices that will lead to the diversification of the BDS workforce.

In response, the National Institutes of Health (NIH) has encouraged institutions to diversify their student and faculty populations to enhance the participation of individuals from groups identified as underrepresented in the biomedical, clinical, behavioral, computational, and social sciences.[8] In 2012, the NIH Advisory Committee to the Director (ACD) Working Group on Diversity in the Biomedical Research Workforce recommended that under the leadership of the National Institute on Minority Health and Health Disparities (NIMHD), in coordination with other Federal government agencies, NIH should undertake a bold, well-funded, multi-year, incentive-based, competitive process to support infrastructure development in under-resourced Minority Serving Institutions (MSIs) that train and produce underrepresented scientists,[9] including the Building Infrastructure Leading to Diversity (BUILD) Initiative, in the emerging areas of BDS. In addition, the Group recommended that the NIH should support creative partnerships among these institutions and, where appropriate, with more research intensive institutions.[10]

The BD2K-Enhancing Diversity in Biomedical Data Science Program

The Working Group recommendations led the NIH Big Data to Knowledge (BD2K) initiative to request applications for the Enhancing Diversity in Biomedical Data Sciences Grant[11] to encourage collaboration between MSIs and NIH BD2K Centers of Excellence. The key to develop a skilled and sustained diverse pipeline in BDS is to create research learning communities that integrate ongoing peer-group collaboration, concentrated mentored research projects, and research ownership.[12] MSIs are uniquely positioned to implement high-impact training programs in BDS to prepare, educate and impact a readily available population of underrepresented student groups.[9] Here we define underrepresented students as students from traditionally underserved, under-resourced, or under-educated communities; this may include subgroups and intersections of African American, Hispanic and Asian-Pacific Islander populations in addition to first-generation students, low-income students, women, and students with disabilities. In 2015, four MSIs, each with a unique profile, were awarded the NIH BD2K Enhancing Diversity in Biomedical Data Science Grant:

California State University, Fullerton (CSUF) is a large (~41,000 students) public four-year comprehensive institution in Orange County, Calif.; CSUF is federally classified as a Hispanic Serving Institution (HSI) and an Asian Pacific-Islander Serving Institution. CSUF collaborates with the University of Southern California (USC) Center for Big Data for Discovery Science. **California State University, Monterey Bay** (CSUMB) is a smaller mid-sized (~7000 students), public four-year comprehensive university and HSI located in Seaside, Calif. CSUMB collaborates with the University of California, Santa Cruz (UCSC) Center for Big Data in Translational Genomics. **Fisk University** is a small (~850 students), private liberal arts Historically Black University (HBCU) in Nashville, Tenn. Fisk University collaborates with the University of Illinois, Urbana-Champaign (UIUC) NIH BD2K Center of Excellence for Big Data Computing, which hosts the Knowledge Engine for Genomics (KnowEnG) project. **University of Puerto Rico, Río Piedras Campus** (UPR-RP) is a public mid-sized HSI (~13,000 undergraduates) classified as an Intensive Doctoral/Research University by the Carnegie Foun-

Table 1. Characteristics of student population in four BD2K Diversity R25 Universities

Ethnicity	Sex					
	Male n, %		Female n, %		Total n, %	
CSUF						
American Indian	20	0	22	0	42	0
Black	236	1	404	1	640	2
Hispanic	5432	16	7804	24	13236	40
Asian	3328	10	3665	11	6993	21
White	3349	10	3956	12	7305	22
Hawaiian/ Pacific Islander	120	0	123	0	243	1
Multi-race	642	2	826	2	1468	4
Other/ decline	1646	5	1508	5	3154	10
Total	14773	45	18308	55	33081	100
CSUMB						
American Indian	9	0	21	0	30	0
Black	141	2	240	4	381	6
Hispanic	868	14	1666	27	2534	41
Asian	146	2	172	3	318	5
White	809	13	1115	18	1924	31
Hawaiian/ Pacific Islander	10	0	15	0	25	0
Multi-race	138	2	228	4	366	6
Other/ decline	275	4	378	6	653	10
Total	2396	38	3835	62	6231	100
FISK						
American Indian	0	0	0	0	0	0
Black	226	31	405	56	631	87
Hispanic	7	1	10	1	17	2
Asian	6	1	0	0	6	1
White	3	0	2	0	5	1
Hawaiian/ Pacific Islander	0	0	1	0	1	0
Multi-race	2	0	7	1	9	1
Other/ decline	29	4	31	4	60	8
Total	273	37	456	63	729	100
UPR-RP						
American Indian	4	0	11	0	15	0
Black	6	0	9	0	15	0
Hispanic	4163	34	6735	54	10898	88
Asian	7	0	4	0	11	0
White	38	0	54	0	92	1
Hawaiian/ Pacific Islander	1	0	0	0	1	0
Multi-race	15	0	20	0	35	0
Other/ decline	513	4	782	6	1295	10
Total	4747	38	7615	62	12362	100
Nationwide[a]						
American Indian	56117	0	82385	0	138502	1
Black	908128	5	1492130	8	2400258	13
Hispanic	1333893	7	1825075	10	3158968	18
Asian	472018	3	523469	3	995487	6
White	4148569	23	5155815	29	9304384	52
Hawaiian/ Pacific Islander	27088	0	33405	0	60493	0
Multi-race	233450	1	315894	2	549344	3
Other/ decline	687208	4	732872	4	1420080	8
Total	7866471	44	10161045	56	18027516	100

a. Nationwide data for fall 2014 enrollment obtained from the National Center of Education Statistics (NCES), http://nces.ed.gov

dation for the Advancement of Teaching. UPR-RP collaborates with several NIH BD2K Centers of Excellence including the University of Pittsburgh; the University of California, Santa Cruz; and Harvard University.

The four MSIs awarded the NIH BD2K Enhancing Diversity in Biomedical Data Science Grant serve nearly triple the proportion of underrepresented students (Hispanics, APIs, and African Americans at respective institutions) compared with national enrollment rates (Table 1). Each program involves underrepresented students from across disciplines (Table 2), such as mathematics and statistics, computer science, and various biologically related fields, who are recruited to participate in professional development, new or redesigned courses relevant to BDS, and summer research experiences at leading research institutions in BDS. Each program, though, is unique in how it leverages its size, institution type, and resources to build each unique BDS program to address the barriers to diversity in biomedical data science.

Barriers to Diversity in Biomedical Data Science

The four awarded MSIs have a long-standing history of educating underrepresented students. In general, MSIs reach the largest proportion of underrepresented students though they constitute only a small fraction of US colleges and universities,[9] but that does not exempt MSIs or their students from facing tremendous barriers in their pursuit of biomedical data science careers. In addition, it is critical to address barriers to diversity in BDS as they affect all institutions of higher education, regardless of minority serving status. Though we could consider the long-term negative effects of stereotype threat, racism, and sexism that act as barriers to underrepresented students to enter biomedical data science fields, we choose instead to focus on barriers that can be addressed at the institutional and programmatic level, whether it be at an undergraduate institution, a research institution, or a funding agency. Therefore, we will assess just three specific barriers to BDS careers for traditionally underserved and underrepresented student groups: preparation, exposure and access to resources. It is these three barriers that the four NIH-funded programs at MSIs seek to overcome through their BDS programs.

Preparation

The roots of the barrier of preparation begin well before post-secondary education, perhaps even to kindergarten. Primary and secondary education, though, is beyond the scope of this article, so we will maintain our focus on the impact of preparation on the post-secondary education experience of the underrepresented student. Preparation to study fields relevant to BDS includes rigorous coursework in mathematics, statistics, computer science, and biology. In addition, many careers in BDS require, at minimum, a graduate degree. Many underrepresented and low-income students are placed into remediation mathematics courses that often only extend the time to completion, increase the cost of education with little success in improving graduation rates, and carry significant stigma.[13] In addition, student lack of identity as a "math person" and confidence/self-efficacy in mathematics can greatly affect students' perseverance, especially in female students, in mathematics courses such as calculus,[14,15] which is a vital and required course for most majors relevant to BDS.

Preparation for graduate education in BDS cannot be achieved through computational preparation alone. Adequate development of the professional and cognitive skills necessary for entrance to, and success in, graduate school is also necessary. Underrepresented students are often unaware of the necessary requirements for potential careers in data science, let alone graduate school, and the competition for entrance into graduate school is fierce. Professional and personal preparation are also integral to the success of a student in BDS.[16] Underrepresented students must understand the rigors of research culture, especially as it may conflict with personal cultural identity. There is evidence to indicate that a lack of understanding of research culture may lead to a lack of diversity in STEM and discouragement for underrepresented groups to pursue biomedical research.[17] Therefore, we must consider methods and programs to mitigate the academic, professional, and cultural barriers for underrepresented students to BDS to diversify the workforce.

Exposure

The exposure of underrepresented students to potential careers is another avenue to recruit students into

Table 2. Proportion and count of biology, mathematics and computer science by race/ethnicity in four BD2K Diversity R25 Universities, 2015-2016

Ethnicity	Biology		Mathematics		Computer science	
	n	%	n	%	n	%
CSUF						
American Indian	1	0	.	.	.	.
Black	31	2	5	1	11	1
Hispanic	546	43	151	45	258	25
Asian	356	28	58	17	386	37
White	339	17	65	19	246	23
Hawaiian/ Pacific Islander	16	2	1	0	9	1
Multi-race	50	4	16	8	48	5
Other/ decline	64	5	41	9	92	9
CSUMB						
American Indian	1	0	.	.	.	.
Black	74	11	7	6	8	7
Hispanic	259	39	51	46	31	28
Asian	57	8	6	5	12	11
White	211	31	33	29	42	38
Hawaiian/ Pacific Islander	4	1	.	.	2	2
Multi-race	50	7	9	8	15	13
Other/ decline	16	2	6	5	2	2
FISK						
American Indian	.	.	.	.	.	.
Black	136	91	19	95	16	89
Hispanic	4	3	.	.	.	.
Asian	1	1	.	.	2	11
White	.	.	.	.		
Hawaiian/ Pacific Islander	.	.	.	.	.	.
Multi-race	1	1	.	.	.	.
Other/ decline	8	5	1	5	.	.
UPR-RP						
American Indian	1	0	.	.	.	.
Black	3	0	.	.	.	.
Hispanic	1273	93	166	95	118	100
Asian	.	.	.	.	.	.
White	8	1	1	1	.	.
Hawaiian/ Pacific Islander	.	.	.	.	.	.
Multi-race	.	.	.	.	.	.
Other/ decline	82	6	8	5	.	.

BDS, but many students are not exposed to BDS research in primary and secondary education. If underrepresented students were exposed to how BDS directly impacts health disparities, for example, underrepresented students may consider related majors and career options. Exposure must be considered even earlier than in post-secondary education.[14] Universities and industry should reach out to the community, including middle and high schools, community colleges, and public four-year universities with high underrepresented student populations to provide exposure to BDS. Outreach must go beyond "cool science" or the sale of a high paying job and should show students how BDS can support the reduction of health disparities in their communities and beyond.[16] In addition, exposure through presentations by biomedical data scientists that are also from underrepresented groups in the field will serve to break down internalized stereotypes and provide role models for students.

For institutions already serving underrepresented students, there is often an absence of exposure to innovative research in undergraduate level curricula that develop the skills

and concepts relevant to the world of big data, while also allowing students to focus on specific sub-disciplines of this broad field.[16,18] Biomedical data science curriculum must extend in focus beyond the areas of informatics, statistics, and biology. It must enhance the breadth of understanding of the students of how these disciplines can be integrated, which can be difficult to achieve since many undergraduate programs are not interdisciplinary in a true sense.[18] Early exposure to the potential careers available in BDS[15] and clear articulation of the necessary academic preparation for those careers is paramount to recruiting underrepresented students into BDS fields.

Access to Resources

Finally, and perhaps most challenging, is the issue of access to education, resources and experiences for underrepresented students. Access for underrepresented students is significantly different from their non-underrepresented counterparts, even when controlling for college-readiness and income.[19] Many underrepresented students will start in community college as nationally, 50% of Hispanic students, 31% of African American students start in community college, compared with 28% of White students.[19] Equally qualified underrepresented students attend community colleges and open-access four year universities at a much higher rate than their White counterparts and this leads to unequal access to resources.[19] Though African American and Hispanic students attend college at the same rate as similarly qualified White students, they are less likely to graduate with a Bachelor's degree, contributing to an estimated talent loss of 111,000 African American and Hispanic students annually from the top half of the nation's high school graduating classes who do not obtain a college degree within eight years of graduation.[19]

MSIs are in precarious financial situations given their dependence on diminishing federal and state funding and an inability to increase tuition and fees due to the socioeconomic challenges of their students.[9] The high proportion of underrepresented students attending under-resourced universities leads to limited access to opportunities for undergraduate research prior to graduation at home institutions; these resources are vital to research preparation and a successful application to graduate school. This challenge is particularly acute at undergraduate focused four- and two-year institutions, where faculty-led opportunities to engage in research are limited.[16] Faculty members at MSIs must update their training in the era of Big Data in order to properly train their students. However, most faculty at MSIs have very heavy teaching loads and lack access to funding[20] to develop these new skills and/or research programs to benefit their students. If underrepresented students are to be successful in BDS, they must have access to faculty prepared to teach and guide them in addition to research experiences.

Recommendations to Overcome the Barriers

Though there are many recommendations that could be considered to support underrepresented students, we will focus on three evidence-based practices that any agency or institution may adopt: Mentoring; Research Experiences; and Curriculum Development. Here we outline the recommendations and provide examples from the BD2K programs of their implementation at various types of institutions (private/public, large/small, undergraduate/doctoral). Note that though we use specific examples from each institution, all four BD2K programs incorporate the provided recommendations in some capacity.

Mentoring

Future biomedical data scientists may not begin their postsecondary education prepared to be scientists or even aware they are scientists; we (the scientists) must mold them into scientists regardless of their starting point. Early advising and mentoring have proved to be an extremely effective method to overcome the barriers of exposure and preparation.[14,21] Mentoring can provide a student with confidence, guidance on professional behavior, networking, writing skills and support, information on academic culture, and ethics training.[21] Mentoring programs can be established at any institution type. Industry and academia can provide mentoring programs or join mentoring networks to reach into and beyond their local communities. In addition to one-on-one mentoring experiences, cohort programs have been shown to increase retention and reduce stereotype threat among underrepresented students. The Meyerhoff Scholars Program demonstrated that African American Meyerhoff

students in the 1989–2005 entering cohorts were 5.3 times more likely to enter STEM graduate programs than equally talented declined sample students (41.1% vs 7.8%).[22] Each of the BD2K programs incorporates cohorts in some capacity in their respective program as cohort programs can work at any institution type and serve to increase success of underrepresented students in BDS.

As many of the available mentors in BDS are not a part of an underrepresented group, mentoring an underrepresented student may provide challenges in understanding cultural barriers to success. Research mentors need to understand how racial/ethnic/cultural differences interact with academic culture.[23] Therefore, proactive mentoring is key to a successful mentoring relationship until the student has the knowledge and confidence to steer the mentoring relationship themselves. Especially in academia, research mentors must not view undergraduates as "free labor" or burdens, but they must focus on the educational aspect of the research experience. Mentors must provide individualized attention, personalized letters of recommendation, and regular interaction despite their busy research agendas. For example, faculty members at both Fisk University and UIUC participate in a "Mentoring the Mentor" workshop to improve faculty mentoring skills. In addition, the UCSC summer research mentors, which include graduate students, postdoctoral fellows, and research staff serving as first-time mentors, will participate in a mentoring workshop led by CSUMB staff and faculty to prepare them to mentor underrepresented undergraduate students. Mentoring workshops for research mentors are a simple and low-cost solution to make research culture more inclusive to underrepresented students, especially as many underrepresented students report a disconnect in the community and support received at their MSI compared with the isolation and competitive nature of their graduate programs, which are often predominately White.

Research Experiences

The National Science Foundation (NSF) has already seen the success of investment into MSIs[24] where NSF programs emphasized the need for student support and access to undergraduate research. Research experiences have been shown to increase student performance and retention in addition to providing valuable preparation and a competitive edge for graduate school. Therefore, research experiences should not be thought of as a reward for the already "elite student," but as a training ground to cultivate the "elite student." The research experience must extend beyond just a lab experience, though. Students must receive support through mentoring, professional development, and personal engagement beyond the lab.[16] For example, the student researchers at UPR-RP, on return from their summer research experiences, will continue to participate in interdisciplinary undergraduate research projects with local researchers. Multi-year undergraduate research experiences allow students to develop both intellectual skills and a deep understanding of research culture.[25] In addition, continued funding of research experiences (including conference attendance and presentations) throughout the academic year provide students with meaningful work experience that supports their education and their academic aspirations,[14,17] as many underrepresented students support their own education.[11,14] Many undergraduate-inclusive conferences already offer needs-based travel scholarships for undergraduate students (eg, SACNAS), but if professional conferences relevant to biomedical data science provided needs-based travel scholarships to undergraduates, it would provide an invaluable experiences for underrepresented students.

In addition, socioeconomically disadvantaged students often attend resource-limited colleges and universities where research experiences are unavailable to them at their home institution.[20] CSUMB, which is only just beginning its programs in biomedical data science, takes advantage of its proximity (~40 miles) to its partner, UCSC, to allow students to continue their summer research projects throughout the academic year. Therefore, research institutions, industry and funding agencies must consider the importance of providing summer research programs, on-campus research experiences, or satellite collaborations with students at local community colleges and MSIs.

As previously mentioned, the preparation of students by faculty is directly influenced by faculty training. We must consider ways to support faculty at MSIs if we are to also support underrepresented students. Faculty members at the universities that serve underrepresented populations need the opportunity to expand

their training in BDS through: 1) training/workshop support at research institutions; 2) collaborative learning and research experiences at research institutions; 3) summer support to advance training and research programs, and 4) access to grants and funding sources they might not traditionally qualify for due to their focus on teaching and only some research at their institutions. Each of the BD2K programs provides support in these areas for the MSI faculty. For example, at CSUMB, CSUF, and Fisk University, where most of the academic year is focused on teaching, the BD2K Program supports faculty research and training in the summer so that faculty are prepared to teach data science courses and guide students in BDS research. UPR-RP supports short sabbaticals for its faculty to develop research programs in biomedical data science. The four grantees have the luxury of support for this training currently, but most MSIs do not. Therefore, research institutions, industry, and funding agencies must consider that to create a diverse workforce in biomedical data science, they must train the people directly responsible for training that diverse workforce.

..given the disparities in educational access, graduate programs must recognize that many underrepresented students will have nontraditional educational pathways.

Curriculum Development

Data science is an evolving interdisciplinary field and as such undergraduate programs in the fields related to BDS are only just beginning.[26] New courses and course lesson plans embedded in traditional courses that expose students to BDS are vital to the recruitment and training of students. At many under-resourced universities, though, training in data science needs further support. Therefore, training to develop new curriculum and tools for BDS must be a part of the dissemination of new courses in BDS. For example, at CSUF, BDS curricula is being integrated across colleges and a new Big Data and Health course has been established that explores: an overview of brain health and neuroimaging; an understanding of next generation Big Data workflow technologies; and further develops a knowledge of modeling, visualizing, and the interactive exploration of large-scale imaging data. In situations where it is not possible to offer a new course or there are no faculty with the appropriate expertise to provide the course, creative solutions might be necessary. For example, students at the UPR-RP program take a two-course sequence on Biomedical Big Data based on the massive open online classes (MOOC) developed by Rafael Irizarry (Harvard University).

Regardless of the approach (new courses, revision of old courses, MOOCs), resourced universities can serve as mentor institutions or provide mentor teachers for the faculty at institutions serving underrepresented students that want to develop biomedical data science related curriculum. For example, faculty at UCSC provide resources and mentorship to faculty at CSUMB as they develop new data science and bioinformatics courses. The BD2K Program at CSUMB also includes plans to increase accessibility to courses and create interdisciplinary majors/minors using existing courses across disciplines. The interdisciplinary approach is appropriate given the interdisciplinary nature of BDS and to take advantage of the limited resources (such as faculty time) across many currently small programs. In addition, CSUMB is working with local community colleges to create lower-division computationally intensive statistics, biology, and computer science courses to prepare students for transfer into data science programs and bioinformatics. As undergraduate-level curriculum relevant to BDS is still in its infancy, it will be critical to disseminate evidence-supported and cost-effective curriculum design widely in addition to training on the implementation of the curriculum.

Recruiting and Serving Underrepresented Students

The four NIH supported programs provide models for what various types of institutions might consider to increase diversity in BDS. Dissemination of the programs developed will encourage the creation of other similar programs at MSIs and will serve

to increase underrepresented students in BDS in the long-run. In the interim, though, the institutions and agencies already training biomedical data scientists must evaluate their institutional practices and how they can increase retention and recruitment of underrepresented students, specifically in graduate programs in BDS. There are several considerations graduate programs must consider in the recruitment of underrepresented students into BDS. First, given the disparities in educational access, graduate programs must recognize that many underrepresented students will have nontraditional educational pathways. Unlike students who attend well-resourced and research-oriented institutions, underrepresented students may have less research experiences, either because they chose paid work over research or research experience was unavailable to them at their home institutions.[20] Second, the recruitment practices and program culture at graduate-degree granting institutions and industry may directly impact the matriculation and success of the underrepresented students in their programs. For example, Brown University evaluated its culture of recruitment and retention and found that its own culture impacted the success of underrepresented students in its biomedical programs. Brown University then took a systematic approach to addressing this systemic problem, including a change in the recruitment culture by creating partnerships with MSIs to directly recruit their students into Brown's programs.[27] In comparison, if diversity is important to the BDS community as a whole, we must make it a priority and adjust the current culture so that the future of BDS is supported by a thriving and well-trained workforce.

Conclusion

In order to use big data to address health disparities, we must resolve the diversity disparities in biomedical data science education. The more funding agencies like the NIH support diversity initiatives, promoting collaboration between 2- and 4-year MSIs and research institutions, the broader the impacts on developing the inclusive talent pool to advance BDS. Though many of the proposed best practices for increasing diversity in BDS may not necessarily be novel to STEM education in general, the fast pace of advancement in BDS and the interdisciplinary nature of BDS necessitates an immediate and well-executed implementation of the recommendations now, before we find ourselves without a diverse BDS workforce to continue the advance of BDS research. We hope that sharing our efforts to start the process of enhancing diversity in BDS will serve as a model for other institutions as we approach a solution together.

Acknowledgements

Funding for the BD2K Enhancing Diversity in Biomedical Data Science is supported by the Office of the Director (OD) of the National Institutes of Health. The program at CSUMB is funded by NIH 1R25MD010391. The program at CSUF is funded by NIH 1R25MD01397. The program at UPR-RP is funded by NIH R25MD010399. In addition, Dr. María-Eglée Pérez is funded by NIH 2 P20 GM103475-14A1 and NCI/NIH U54CA096297. The program at Fisk University is funded by NIH 1R25MD010396. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of Interest

No conflicts of interest to report.

Author Contributions

Research concept and design: Canner, McEligot, Pérez, Qian, Zhang; Acquisition of data: Canner, McEligot, Pérez, Qian; Data analysis and interpretation: Canner, McEligot, Pérez, Qian, Zhang; Manuscript draft: Canner, McEligot Pérez, Qian, Zhang; Statistical expertise: McEligot, Pérez, Zhang; Acquisition of funding: Canner, McEligot, Qian; Administrative: Canner, Qian, Zhang; Supervision: Canner, McEligot, Qian

References

1. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Information Science and Systems.* 2014;2(1):3.
2. Chu KC. Cancer Data for Asian Americans and Pacific Islanders. *Asian Am Pac Isl J Health*. 1998;6(2):130-139. PMID:11567423.
3. Look MA, Braun K. *Queen's Health S. A mortality study of the Hawaiian people, 1910-1990.* Honolulu: Queen's Health Systems; 1995.
4. Miller BA, Kolonel LN. National Cancer Institute, Cancer Control Research. *Racial/Ethnic Patterns of Cancer in the United States, 1988-1992.* Bethesda, Md.: U.S. Dept. of Health and Human Services, National Institutes of Health; 1998.
5. Tsark JU. Cancer in Native Hawaiians. *Asian Am Pac Isl J Health*. 1998;6(2):157-173. PMID:11567426.
6. Centers for Disease Control and Prevention. *Health Disparities among Native Hawaiians and other PIs Garner Little Attention.* Atlanta, GA: Chronic Disease Notes & Reports; 2002.
7. National Science Foundation. *Women, Minorities, and Persons with Disabilities in Science and Engineering.* Arlington, VA: Special Report NSF 15-311. Accessed on October 18th, 2016 at: https://www.nsf.gov/statistics/2015/nsf15311/digest/;2015.
8. McEligot AJ, Behseta S, Cuajungco MP, Van Horn JD, Toga AW. Wrangling Big Data Through Diversity, Research Education and Partnerships. *Calif J Health Promot.* 2015;13(3):vi-ix. PMID:27257409.
9. Cunningham A, Park E, Engle J; Institute for Higher Education Policy. *Minority Serving Institutions: Doing More with Less.* 2014. Available at: http://www.ihep.org/research/publications/minority-serving-institutions-doing-more-less. Last accessed: March 6, 2016.
10. Working Group on Diversity in the Bio-

medical Research Workforce, The Advisory Committee to the Director, NIH. *Draft Report of the Advisory Committee to the Director Working Group on Diversity in the Biomedical Research Workforce*: National Institutes of Health. Available at: https://acd.od.nih.gov/Diversity%20in%20the%20Biomedical%20Research%20Workforce%20Report.pdf. Last accessed: October 26, 2016.
11. National Institute of Minority Health and Health Disparities. National Institutes of Health. Available at: https://grants.nih.gov/grants/guide/rfa-files/RFA-MD-15-005.html. Last accessed: October 18, 2016. 2015.:
12. Corwin LA, Graham MJ, Dolan EL. Modeling course-based undergraduate research experiences: an agenda for future research and evaluation. *CBE Life Sci Educ*. 2015;14(1):es1. https://doi.org/10.1187/cbe.14-10-0167. PMID:25687826.
13. Complete College America. Remedial education's role in perpetuating achievement gaps. 2016.Available at: http://completecollege.org/remedial-educations-role-in-perpetuating-achievement-gaps/. Last accessed: March 6, 2017.
14. Wang XL. Why students choose STEM majors: motivation, high school learning, and postsecondary context of support. *Am Educ Res J*. 2013;50(5):1081-1121. https://doi.org/10.3102/0002831213488622.
15. Ellis J, Fosdick BK, Rasmussen C. Women 1.5 Times More Likely to Leave STEM Pipeline after Calculus Compared to Men: Lack of Mathematical Confidence a Potential Culprit. *PLoS One*. 2016;11(7):e0157447. https://doi.org/10.1371/journal.pone.0157447. PMID:27410262.
16. O'Donnell K, Botelho J, Brown J, González GM, Head W. Undergraduate research and its impact on student success for underrepresented students. *New Dir Higher Educ*. 2015;2015(169):27-38. https://doi.org/10.1002/he.20120.
17. Malcom LE, Dowd AC, Yu T. *Tapping HSI-STEM Funds to Improve Latina and Latino Access to the STEM Professions.* Los Angeles, CA: University of Southern California; 2010.
18. Greene AC, Giffin KA, Greene CS, Moore JH. Adapting bioinformatics curricula for big data. *Brief Bioinform*. 2016;17(1):43-50. https://doi.org/10.1093/bib/bbv018. PMID:25829469.
19. Carnevale AP, Strohl J. *Separate and Unequal: How Higher Education Reinforces the Intergenerational Reproduction of White Racial Privilege.* Washington, DC: Georgetown University; 2013.
20. Boyington JE, Maihle NJ, Rice TK, et al. A Perspective on Promoting Diversity in the Biomedical Research Workforce: The National Heart, Lung, and Blood Institute's PRIDE Program. *Ethn Dis*. 2016;26(3):379-386. https://doi.org/10.18865/ed.26.3.379. PMID:27440978.
21. Haeger H, Fresquez C. Mentoring for Inclusion: The Impact of Mentoring on Undergraduate Researchers in the Sciences. *CBE Life Sci Educ*.ar36. In Press. PMID:27543635.
22. Maton KI, Pollard SA, McDougall Weise TV, Hrabowski FA. Meyerhoff Scholars Program: a strengths-based, institution-wide approach to increasing diversity in science, technology, engineering, and mathematics. *Mt Sinai J Med*. 2012;79(5):610-623. https://doi.org/10.1002/msj.21341. PMID:22976367.
23. Jeste DV, Twamley EW, Cardenas V, Lebowitz B, Reynolds CF III. A call for training the trainers: focus on mentoring to enhance diversity in mental health research. *Am J Public Health*. 2009;99(S1)(suppl 1):S31-S37. https://doi.org/10.2105/AJPH.2008.154633. PMID:19246662.
24. Cullinane J. *Diversifying the STEM Pipeline: The Model Replication Institutions Program*: 2009. Available at: http://www.ihep.org/sites/default/files/uploads/docs/pubs/report_diversifying_the_stem_pipeline_report.pdf. Last accessed: March 6, 2017.
25. Thiry H, Weston TJ, Laursen SL, Hunter AB. The benefits of multi-year research experiences: differences in novice and experienced students' reported gains from undergraduate research. *CBE Life Sci Educ*. 2012;11(3):260-272. https://doi.org/10.1187/cbe.11-11-0098. PMID:22949423.
26. De Veaux RD, Agarwal M, Averett M, et al. Curriculum guidelines for undergraduate programs in data science. *Annu Rev Stat Appl*. 2016;4(1):ePub Ahead of Publication: https://doi.org/10.1146/annurev-statistics-060116-053930.
27. Thompson NL, Campbell AG. Addressing the challenge of diversity in the graduate ranks: good practices yield good outcomes. *CBE Life Sci Educ*. 2013;12(1):19-29. https://doi.org/10.1187/cbe.12-04-0054. PMID:23463225.