



Published in final edited form as:

Public Health Rep. 2026 ; 141(4): 596–600. doi:10.1177/00333549261427670.

Estimating Potential Supply of Newly Trained Data Scientists for Government Public Health Employment

Robert D. Kirkcaldy, MD, MPH [Public Health Workforce Branch Associate Chief for Science],

Division of Workforce Development, National Center for State, Tribal, Local, and Territorial Public Health Infrastructure and Workforce, Centers for Disease Control and Prevention, Atlanta, Georgia

United States Public Health Service, Rockville, Maryland

Hunter Doyle, MPH [Data analyst],

Center for Public Health Systems, Division of Health Policy and Management, University of Minnesota School of Public Health, Minneapolis, Minnesota; (Since completing the research, Mr. Doyle moved to the College of Pharmacy, University of Minnesota, Minneapolis, Minnesota)

Sarah Gusman, MPH, CHES [Senior Data Analyst],

Association of Schools and Programs of Public Health, Washington, D.C.

Emily Burke, EdD, MPH, CPH [Senior Director of Workforce Development and Applied Practice],

Association of Schools and Programs of Public Health, Washington, D.C.

Kyle T. Bernstein, PhD, ScM [Public Health Workforce Branch Chief],

Division of Workforce Development, National Center for State, Tribal, Local, and Territorial Public Health Infrastructure and Workforce, Centers for Disease Control and Prevention, Atlanta,

Corresponding Author: Robert D. Kirkcaldy, MD, MPH, Division of Workforce Development, National Center for State, Tribal, Local, and Territorial Public Health Infrastructure and Workforce, Centers for Disease Control and Prevention, Atlanta, Georgia, United States Public Health Service, Rockville, Maryland, 1600 Clifton Road V24-3, Atlanta, GA, USA, rkirkcaldy@cdc.gov.

AUTHOR CONTRIBUTIONS

R.D. Kirkcaldy: study concept and study design, wrote manuscript; H. Doyle: contributed to study design, data management and analysis, critical review of the manuscript; S. Gusman: contributed to study design, data management, interpretation of results, and critical review of the manuscript; E. Burke: interpretation of results and critical review of the manuscript; K.T. Bernstein: study concept and study design, interpretation of results, and critical review of the manuscript; J.P. Leider: contributed to study design, interpretation of results, critical review of the manuscript.

Declaration of conflicting interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Disclaimer: The findings and conclusions in this report are those of the authors and do not necessarily represent the official positions of the Centers for Disease Control and Prevention.

STATEMENTS AND DECLARATIONS

Ethical considerations

This study received a non-human subjects research determination by the CDC's Human Research Protections Office (August, 2024) because it used available data without personal identifiers and the lack of interaction with or collection of identifiable information about human subjects (Project ID 0900f3eb8240cfd0).

Consent to participate

Not applicable. This activity did not involve primary data collection and informed consent was not required.

Consent for publication

Not applicable. This submission did not contain data from individuals.

Georgia; (Since completing the research, Dr. Bernstein moved to the San Francisco Department of Public Health)

Jonathon P. Leider, PhD [Associate Professor]

Center for Public Health Systems, Division of Health Policy and Management, University of Minnesota School of Public Health, Minneapolis, Minnesota

Abstract

Objectives. In 2024, the Centers for Disease Control and Prevention and the Council of State and Territorial Epidemiologists suggested approaches, including upskilling and recruitment/hiring, to strengthen public health data science workforce capacity. We estimated the number of recently graduated data scientists who might be eligible for and potentially hireable into government public health jobs as a step toward filling data gaps in workforce planning.

Methods. We used data from the National Center for Education Statistics (NCES) to calculate the number of data science graduates in 2023. As a proxy for interest in government public health among graduates, we used data from the Association of Schools and Programs of Public Health (ASPPH, graduation years 2015-2022). We multiplied the number of data science graduates from US academic institutions (from NCES data) by the percentage of public health graduates who entered government public health employment (from ASPPH data) to estimate the number of data science graduates who might be eligible for and interested in government public health employment.

Results. In 2023, 467 435 graduates were awarded a data science degree from a US institution. Depending on the government employment criteria, 8.3% to 15.7% of 96 578 public health graduates reported first-destination employment in government public health. The number of data science graduates who might be eligible for and interested in government public health employment ranged from nearly 29 000 to more than 57 000.

Conclusions. These data contribute to the public health workforce planning evidence base but are likely to be overestimates of supply. If the estimated supply of data scientists falls short of demand projections, staff data science upskilling and changes to academic curricula could be emphasized.

Keywords

data scientist; data science; public health; workforce; government

INTRODUCTION

Increasing computing and data storage capacity, evolving data science and artificial intelligence tools, and expanding availability of complex public health data hold promise for transforming the data and analytic capacity of government public health agencies. However, public health officials may be unable to harness these resources because the government public health workforce may lack sufficient data science skills and competencies.¹ A workforce that lacks data science skills and competencies may also not be able to optimally modernize outdated and siloed public health data infrastructures.

In 2024, the Centers for Disease Control and Prevention and the Council for State and Territorial Epidemiologists emphasized the importance of strengthening public health data science workforce capacity and suggested approaches such as recruitment/hiring, workforce exchanges with the private sector, and on-the-job training of the existing public health workforce (i.e., upskilling).^{2,3} During the past several years, recruitment and hiring of data scientists into government public health had to contend with barriers such as complex application processes and the lure of lucrative private industry careers.^{3,4} At least at the federal level, data science positions often require that candidates have data science-related education and degrees. Recent changes in the federal government employment and funding landscape have also introduced uncertainties and challenges.⁵

Approximately 15 000 epidemiologists and data scientists are in the state and local workforce, about 42% of whom have been in public health for 5 years or less.⁶ Among this subset of epidemiologists, about 84% have a public health degree, compared with 30% of other data scientists who have worked in public health for 5 years or fewer.⁶ Other than these general measures, few data sources are currently available to inform governmental strategic workforce planning approaches that support public health capacity-building and goal-setting. Although data from the Public Health Workforce Interests and Needs Survey provide insight into workforce training needs and general numbers of staff,^{6,7} we are not aware of an evidence base for data science workforce planning, such as estimates of the number of data scientists that are needed, the number of data scientists with data science-related education and degrees who may be available to fill government public health positions, and the number of public health workforce members in other disciplines who could be upskilled to fill data science positions.

Because these numbers are not readily available, the use of novel data sources can allow incremental progress toward an evidence base for workforce planning. We provide an example. To provide an upper-bound estimate of the size of the recruitment pool of data scientists with data science-related education (but far higher than the size we expect to actually be recruitable by public health), we combined 2 existing data sources not previously used for this purpose.

METHODS

To first calculate the number of recent graduates with data science degrees, we used data from the National Center for Education Statistics Postsecondary Education System (IPEDS) to calculate the number of graduates from US academic institutions who were awarded data science–related degrees in 2023.⁸ IPEDS is a system of interrelated surveys conducted annually by the US Department of Education’s National Center for Education Statistics. Every college, university, and technical and vocational institution that participates in federal student financial aid programs is mandated to report data, such as enrollment, program completion rates, graduation rates, and number and types of degrees conferred, to IPEDS. We predefined program codes (Classification of Instructional Programs [CIP])⁹ for data science disciplines that would likely qualify someone for government employment as a data scientist (eTable 1 in the Supplement). Using the CIP codes, we calculated the number of 2023 graduates who were awarded a bachelor’s, master’s, or doctoral degree.

As a proxy for interest in government public health employment among graduating public health students, we used first-destination employment data of recent public health graduates (graduation years 2015–2022) as reported by members of the Association of Schools and Programs of Public Health (ASPPH).^{10,11} ASPPH collects data on the academic public health student pathway through its annual institutional reporting, which includes information about public health graduates' first-destination employment, such as employer type (employment sector) and workplace setting (employment sector detail). We used these data because, to our knowledge, they are the only available data source on postgraduation government public health employment. We used first-destination employer type and employment setting, which included whether the workplace was in a federal, tribal, state, or local setting, to calculate percentages of public health graduates whose first-destination employment was in government public health. To allow for a range of sensitivity and specificity, we classified government public health employment in 3 ways. The broadest category included first-destination employment by a government entity or within a government setting. The second category included employment in a government setting. The narrowest category included employment in a government public health setting, such as a US Department of Health and Human Services agency or state health department. When possible, we stratified setting by federal, tribal, state, or local government public health settings. We anticipated that graduates in the ASPPH data would be more likely to be employed in government public health than those who graduated with non-public health degrees. As such, we used these data to help construct the upper bound of first-destination employment in government public health for other non-public health graduates.

To this end, we combined results from the previously mentioned calculations to estimate upper bounds of the number of recent data science graduates who may be both interested in and potentially employable by government public health. We multiplied the number of data science graduates with each degree type (from the IPEDS data) by the percentage of students whose first-destination employment was government (from the ASPPH data) across the 3 government employment categories. For percentages, we used degree-specific and setting-specific values. We used Stata version 18 (StataCorp LLC) and Microsoft Excel version 2402 (Microsoft Corp) to conduct all analyses.

Ethical Considerations

This study received a non-human subjects research determination by CDC's Human Research Protections Office (August 2024) because it used available data without personal identifiers and because of the lack of interaction with or collection of identifiable information about human subjects (Project ID 0900f3eb8240cfd0). This activity did not involve primary data collection; as such, informed consent was not required.

RESULTS

In 2023, a total of 467,435 graduates were awarded a data science degree from U.S. institutions (Table 1). Of these graduates, 286,561 (61%) received a bachelor's degree, 169,916 (36%) received a master's degree, and 10,958 (2%) received a doctoral degree. The

number of degree conferrals increased substantially, more than doubling the nearly 230 000 students that graduated with data science-related degrees in 2013.

ASPPH has first-destination outcome data for 96 578 students who graduated from an ASPPH-member school or program of public health during the 2015–2016 and 2021–2022 academic years (Table 2). Of these graduates, 23 062 (23.9%) were awarded a bachelor's degree, 66 296 (68.6%) were awarded a master's degree, and 7220 (8.0%) were awarded a doctoral degree. Overall, 18411 (19.1%) of 96 578 graduates received a biostatistics, epidemiology, or health informatics degree; 18 307 (24.9%) of those who graduated with a master's or doctoral degree (n = 73 516) received a biostatistics, epidemiology, or health informatics degree.

In total, 15,201 (15.7%) of all 96,578 graduates were either employed by a government entity or worked in a government workplace setting, the broadest category. First-destination employment in a government workplace setting was reported by 12,073 (12.5%) graduates. First-destination employment in a government public health workplace setting, the most specific category, was reported by 8,045 (8.3%) graduates. Those with a bachelor's degree tended to be employed at the local level, those with a master's degree at the local and state levels, and those with doctoral degrees tended to be employed at the federal level.

With cautious extrapolation from the first-destination outcomes of public health graduates to interest in government public health employment among data science graduates, we estimated data science applicant pools for government public health jobs to be between nearly 29 000 and more than 57 000 (eTable 2 in the Supplement).

DISCUSSION

We combined 2 existing data sources as an initial incremental step toward establishing an evidence base to inform strategic public health workforce planning. We found a substantial number of data science-trained students graduating from US colleges and universities—more than 450 000 in 2023 alone. Based on our assumptions and calculations, we estimated that approximately 29 000 to 57 000 data science graduates might be in the government public health employment pool. Clearly, however, government public health is not drawing this level of interest from data science graduates, inside or outside public health programs.^{3,4} Why is this?

It's about money, but also more.³ Fundamentally, when attempting to hire data scientists, hiring officials in government public health face competition from lucrative private sector employers, who can often offer higher salaries, greater resources, and career advancement opportunities.^{3,4} Private sector applicants report better job applicant experiences, faster hiring periods, and, increasingly, better benefits.³ The number of data scientist positions in the private sector is likely to grow, further exacerbating the competition for well-trained data scientists.¹²

We lack estimates for the number of data scientists who are needed in government public health, as well as the number of workforce members who could be upskilled to build sufficient data science capacity. Without such estimates, it is difficult to put our findings

regarding the potential supply of data scientists in proper context. However, we can draw a few conclusions. We expect that our findings represent overestimates and that the actual supply of recent data science graduates potentially interested in government public health positions during the period of data collection was lower, likely substantially lower. When spread across government public health settings and considering a previously described workforce deficit¹³, the number of recently trained data scientists might not be enough to meet workforce needs.

If this is the case, efforts to strengthen data science capacity could emphasize upskilling of existing staff and consider development of data science career pathways for upskilled staff. Programs such as CDC's Data Science Upskilling (DSU) (for CDC employees) and CSTE's CDC-supported Data Science Team Training (DSTT) are effective upskilling models.^{14,15} Upskilling can complement hiring by equipping staff with foundational data science skills and thus enabling better integration of newly hired data scientists into existing teams. The potential benefit of upskilling is reinforced by recent changes to the size of the federal workforce.

Schools and programs of public health are integral to providing current and future professionals with essential skills, including data science, to address emerging public health challenges. The upcoming 2026 Council on Education for Public Health criteria revisions present an opportunity to embed data science training into existing public health curriculum. Embedding data science, and particularly the data science skills needed in government public health, into schools and programs of public health curricula might strengthen skills among students already showing an interest in public health. Further, embedding foundational public health coursework within undergraduate core curriculum can foster a broader understanding of public health, spark early career interest, and equip students in fields like data science with foundational public health knowledge. Students and graduates who are interested in public health employment can find job postings on platforms such as Indeed, USAJobs, and public health-specific job boards.^{16–18}

Limitations

Our methods and analyses had several limitations. First, we excluded non-US data science graduates and people with data science skills who lacked a data science degree. Although inclusion of these people would increase supply estimates, some of them might be ineligible for government employment because of immigration, work authorization, and hiring policies. Second, we did not consider changes over time in first-destination employment data because interest in and availability of public health employment is anticipated to change over time. Third, availability of public data such as IPEDS for future updated estimates might be subject to changes in government employment and funding.

CONCLUSIONS

Robust estimates of the number of data scientists required to fill government public health workforce gaps are needed to guide evidence-based decision making for workforce planning and to effectively address these gaps. We intend these findings as a step towards defining supply. Future research can refine these estimates; possible approaches include

assessing the number of data science graduates who enter government public health jobs as first-destination employment or surveys of data science graduates' interest in public health and factors that may increase their interest. Projections of demand (which can incorporate attrition to other industries and the current dynamic landscape) are also needed. If the number of estimated graduates falls short of projected demand attributable to a lack of interest or other factors, strategies such as upskilling of existing staff, strengthening the public health-specific curricula, and embedding public health education into broader undergraduate training might be needed.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

REFERENCES

1. Martin LT, Chandra A, Nelson C, et al. Technology and Data Implications for the Public Health Workforce. *Big Data* 2022;10(S1):S25–S29. doi: 10.1089/big.2022.0208 [PubMed: 36070510]
2. Data Modernization Initiative: Strategic Implementation Plan. Centers for Disease Control and Prevention. Available at: <https://www.cdc.gov/surveillance/data-modernization/index.html>. Accessed November 12, 2024
3. Driving Public Health in the Fast Lane: The Urgent Need for a 21st Century Data Superhighway. Council of State and Territorial Epidemiologists. Available at: https://debeaumont.org/wp-content/uploads/2019/09/DSI-White-Paper_v15-Spreads.pdf. Accessed November 12, 2024.
4. Locke R, McGinty M, Ramirez GG, Sellers K. Attracting New Talent to the Governmental Public Health Workforce: Strategies for Improved Recruitment of Public Health Graduates. *J Public Health Manage Pract* 2022;28(1):E235–E243. doi: 10.1097/PHH.0000000000001336.
5. Government Executive. RIF Watch: See which agencies are laying off federal workers. February 27, 2025, Updated October 24, 2025. Accessed January 10, 2026. <https://www.govexec.com/workforce/2025/02/rif-watch-see-which-agencies-are-laying-federal-workers/403342/>
6. de Beaumont Foundation. PH WINS 2024 findings. 2024. Accessed October 8, 2024. <https://debeaumont.org/phwins/2024-survey/>
7. Rajamani S, Leider JP, Gunashekar DR, Dixon BE. Public health informatics specialists in state and local public health workforce: insights from public health workforce interests and needs survey. *J Am Med Inform Assoc* 2025;32(4):748–754. doi: 10.1093/jamia.ocaf019. [PubMed: 39936849]
8. National Center for Education Statistics. Integrated Postsecondary Education Data System. 2024. Accessed November 5, 2024. <https://nces.ed.gov/ipeds>
9. National Center for Education Statistics. The Classification of Instructional Programs (CIP). Accessed January 10, 2026. <https://nces.ed.gov/ipeds/cipcode/Default.aspx?y=56>.
10. Association of Schools and Programs of Public Health. Research. 2024. Accessed November 12, 2024. <https://aspgh.org/our-work/research/>
11. Plepys CM, Krasna H, Leider JP, Burke EM, Blakely CH, Magaña L. First-Destination Outcomes for 2015–2018 Public Health Graduates: Focus on Employment. *Am J Public Health* 2021;111(3):475–484. [PubMed: 33476234]
12. US Bureau of Labor Statistics. Occupational Outlook Handbook: data scientists. 2024. Accessed October 6, 2024. <https://www.bls.gov/ooh/math/data-scientists.htm#tab-6>
13. de Beaumont Foundation. Staffing up: workforce levels needed to provide basic public health services for all Americans. October 2021. Accessed June 6, 2025. <https://debeaumont.org/wp-content/uploads/2021/10/Staffing-Up-FINAL.pdf>

14. Bertulfo MC, Kirkcaldy RD, Franzke LH, Papagari-Sangareddy SR, Reza F. Advancing Data Science Among the Federal Public Health Workforce: The Data Science Upskilling Program, Centers for Disease Control and Prevention. *J Public Health Manage Pract* 2024;30(2):E41–E46. doi:10.1097/PHH.0000000000001865.
15. Council of State and Territorial Epidemiologists. Data science team training. 2024. Accessed November 12, 2024. <https://www.cste.org/page/dstt-webpage>
16. Public Health Careers.Org. Search public health job opportunities across the nation. 2025. Accessed December 29, 2025. <https://www.publichealthcareers.org/>
17. American Public Health Association. Career center. 2025. Accessed December 29, 2025. <https://careers.apha.org/>
18. Emory University Rollins School of Public Health. Public health employment connection. 2025. Accessed December 29, 2025. <https://apps.sph.emory.edu/PHEC>

POLICY IMPLICATIONS

Novel data sources may provide initial steps towards addressing workforce planning data gaps. If the supply of data science graduates who are potentially interested in public health falls short of projected demand, workforce strategies such as upskilling of existing staff and strengthening of academic curricula may be indicated.

TABLE 1—
Number and Types of Data Science Graduates: National Center for Education Statistics Postsecondary Education System, Graduating Years 2013–2023

Year	Bachelor's	Master's	Doctoral	Total
2013	162 064	58 919	6 922	227 905
2014	173 261	62 428	7 281	242 970
2015	184 951	71 431	7 281	263 663
2016	196 607	83 983	7 358	287 948
2017	210 202	94 183	7 836	312 221
2018	225 373	100 308	8 277	333 958
2019	243 724	105 971	8 649	358 344
2020	261 895	117 689	9 221	388 805
2021	278 060	127 606	9 421	415 087
2022	283 590	131 321	10 596	425 507
2023	286 561	169 916	10 958	467 435

Refer to Supplemental Table 1 for definition of programs included in these estimates

Source: National Center for Education Statistics. Integrated Postsecondary Education Data System^{7,10}

TABLE 2—

Number and Percentage of Public Health Graduates by Degree Level and Known First-Destination Employment: Association of Schools and Programs of Public Health Members, Graduating Years 2015–2022

First-destination employment	Bachelor's (n = 23 062) No. (%)[*]	Master's (n = 66 296) No. (%)[*]	Doctoral (n = 7 220) No. (%)[*]	Total (n = 96 578) No. (%)[*]
Government employer or workplace setting				
Government	1 966 (8.5)	12 091 (18.2)	1 144 (15.8)	15 201 (15.7)
Other	21 096 (91.5)	54 205 (81.8)	6 076 (84.2)	81 377 (84.3)
Government workplace setting				
Total government	1 602 (6.9)	9 582 (14.5)	889 (12.3)	12 073 (12.5)
Federal (United States)	444 (1.9)	2 570 (3.9)	507 (7.0)	3 521 (3.6)
Tribal	9 (0.0)	58 (0.1)	7 (0.1)	74 (0.1)
State	455 (2.0)	3 534 (5.3)	234 (3.2)	4 223 (4.4)
Local	694 (3.0)	3 420 (5.2)	141 (2.0)	4 255 (4.4)
Other	21 460 (93.1)	56 714 (85.5)	6 331 (87.7)	84 505 (87.5)
Government public health workplace setting				
Total	869 (3.8)	6 572 (9.9)	604 (8.4)	8 045 (8.3)
Federal HHS agency	192 (0.8)	1 363 (2.1)	316 (4.4)	1 871 (1.9)
Tribal government	9 (0.0)	58 (0.1)	7 (0.1)	74 (0.1)
State HD	234 (1.0)	2 484 (3.7)	174 (2.4)	2 892 (3.0)
Local HD	434 (1.9)	2 667 (4.0)	107 (1.5)	3 208 (3.3)
Other	22 193 (96.2)	59 724 (90.1)	6 616 (91.6)	88 533 (91.7)

Note: HD = public health department; HHS = U.S. government Department of Health and Human Services

* Percentages in the cells of these columns were used for results in Supplemental Table 2.