Journal of the Society for Clinical Data Management

Ittenbach RF, Neubauer A, Cameron S. Internships in Clinical Data Science: A Brief Report. *Journal of the Society for Clinical Data Management*. 2024; 4(1): 4, pp. 1–5. DOI: https://doi.org/10.47912/jscdm.316

EDUCATION AND PROFESSIONAL DEVELOPMENT

Internships in Clinical Data Science: A Brief Report

Richard F. Ittenbach*, Amy Neubauer† and Stephen Cameron‡

Field-based learning opportunities are considered to be an important and increasingly relevant part of professional learning in the biomedical sciences. The purpose of this study was to document internship opportunities in the combined fields of clinical data management and clinical data science among those attending the Society for Clinical Data Management 2022 annual conference. At the time of the survey there were 273 internship positions staffed across 35 organizations representing pharmaceutical, academic, clinical research, software development, and medical device companies. The greatest percentage of internships reported were in the United States, but with a wide range of opportunities in the aforementioned areas worldwide.

Keywords: clinical data management; clinical data science; internship; practicum; research education

Field-based learning opportunities are considered to be an important and increasingly relevant part of professional learning in the biomedical sciences today. In a field such as clinical data science, where there are few – if any – formal educational programs worldwide, the importance of rigorous field-based training opportunities is critical not only for a trainee's learning but also for workforce development.[1] Defined simply as learning activities that offer practical, experiential training in a specific area of professional practice, the purpose of field-based educational experiences in clinical data science is generally twofold: to offer learning opportunities that cannot easily be obtained in the classroom while at the same time, developing a pipeline of well-trained talent for employers needing the expertise.

With only a modestly developing literature base regarding field-based learning opportunities in the more established fields of computer and data science,[2,3] comparable information in the clinical data sciences is nonexistent. This irony exists despite a rapidly growing industry and a workforce that could easily benefit from the information.[4] Such experiences go by many different labels, including practica, internships, apprenticeships, traineeships, or cooperative learning experiences. The advantages of having such information can be tremendous to both the trainees and research organizations. For the aspiring professional, these experiences can introduce trainees to problems they are likely to encounter on the job, as well as exposure to real-life clients who may be as challenging as the problems themselves. Trainees can see first-hand how they will interact with professionals from other specialized areas.[5,6] Learning how to navigate new administrative systems is another important benefit worth acquiring early on. Strong training programs with strong highly motivated trainees will serve as good advocates when recruiting others irrespective of the organization (e.g., academe, governmental, industry, and the nonprofit sectors).

Benefits to the hiring organization are many, as well. Hiring organizations are able to employ staff at a considerably lower cost while instilling core knowledge in a potential employee. Internship periods have generally reduced onboarding costs relative to other hires. In the case of clinical data science, that often means training across the areas of computer science, statistics, regulatory affairs, and the medical sciences. Rawlings-Goss (2019) refers to this phenomenon as "learning-oriented employment," and describes the targeted approach to professional development as particularly well-suited for professionals who enter a field without substantial knowledge and/or fragmented preparation in a specific area of interest.[7] As practice patterns in industry have shifted from an emphasis on data quality to data integrity, there has been an equally prominent shift in the skill sets needed, from one of managing the research data to risk-based critical thinking and scientific scrutiny. Yet, today's data professionals must be proficient at both. For purposes of this paper, we will refer to this collection of practical, experientially driven, field-based learning programs as "internships."

* Division of Biostatistics and Epidemiology, Cincinnati Children's Hospital and the University of Cincinnati College of Medicine, Cincinnati, OH, United States of America

† Global Clinical Data Management, Alexion/AstraZeneca Rare Disease Unit, Boston, MA, United States of America

‡ Data Management, ICON plc, Victoria, BC, Canada

Corresponding author: Rick Ittenbach, PhD (richard.ittenbach@cchmc.org)

While many organizations in the biomedical sciences offer internships, many do not. Without a definitive resource for prospective professionals to turn to, it is extremely hard for trainees to know where to obtain the relevant skills, particularly with respect to the core competencies of the profession.[8–11] At a recent educational session devoted to career opportunities in clinical data science, 85 of 141 participants (60%) in India indicated that they would be looking for an internship within the coming year – but needed help finding points of entry into the profession.[12] From an organization's standpoint, trying to find interns in a world without formal training programs can be extremely challenging. Hiring organizations have to do both the vetting and the training, which puts a heavier burden on them. As a result, the Society for Clinical Data Management has begun tracking internship opportunities among participating members and institutions and offers consultative support for both. Hence, the purpose of this report is to document and report on internship opportunities within the combined field of clinical data management and clinical data science today.

Methods

Procedures

The current survey was a subset of a much larger task analysis survey administered to attendees at the 2022 Society for Clinical Data Management (SCDM) meeting in San Antonio, TX. Surveys were administered by the survey designers and support staff in sessions throughout the conference. When possible, conference attendees were asked to complete and return the surveys while in session, or to complete them on their own and return them to a staff member as soon as possible after completion. A total of 108 surveys were completed and returned (16% of the conference attendees). Response time was approximately 15 minutes for the entire survey, including 2 minutes for the internship portion.

Statistical Analyses

All data were analyzed using traditional descriptive statistics. Frequency counts and percentages were used for binary and categorical data, while measures of central tendency and variability were used for continuous data. A single Spearman-rho correlation coefficient was used to estimate the relationship between organization size and number of interns hired. The criterion for statistical significance was set at the nominal $\alpha = 0.05$ level. The SAS code and data to corroborate these results are provided in Supplemental Files 1 and 2, respectively.

Results

Forty-seven (43.5%) of 108 participants reported that their organizations hired interns, while 29 (26.8%) reported that their organizations did not (32 [29.6%] did not respond to this question). Of the 47 respondents who reported that their organization routinely hires interns, 35 reported the actual number of internship opportunities staffed or available in their organization at the time of the survey. At the time of the survey, there were 273 staffed internships (7.8 interns/organization ($\sigma = 10.9$), ranging from a low of 0 to a high of 50 interns (median = 3) within an organization. A respondent could report zero interns if their organization did not have any interns at the time of the survey. That was the case for one respondent in this survey.

Survey respondents represented small (< 1,000 employees; $n = 45$ [41.7%]), medium (1,001 to 10,000 employees; $n = 18$ [16.7%]), and large organizations (> 10,000 employees; $n = 45$ [41.7%]). The number and pattern was relatively similar for organizations reporting internship opportunities, in favor of the larger organizations: 16 small (34.0%), 9 medium (19.2%), and 22 large (46.8%) organizations (see **Table 1**). There was no statistically significant relationship between organization size and number of interns ($r = –0.31$, $p = 0.07$).

A broad range of organizations was represented among the survey participants, with pharmaceutical ($n = 56$, 51.8%) and contract research organizations ($n = 21$, 19.4%) having the most interns, followed by software providers ($n = 9$, 8.3%), medical device ($n = 8$, 7.4%) and biological research and development companies ($n = 4$, 3.7%). Other respondents reported having three or fewer internship positions. With respect to organizations offering internship opportunities, the patten was similar: pharmaceuticals ($n = 29$, 61.7%), contract research organizations ($n = 7$, 14.9%), and medical device companies ($n = 5$, 10.6%). See **Table 1** for a complete listing of internships by organization type.

Among all survey participants, most respondents were located in the United States ($n = 92$, 85.2%). Other regions were represented as follows: Asia Pacific ($n = 3$, 2.8%), Canada ($n = 3$, 2.8%), Europe ($n = 2$, 1.8%), India ($n = 4$, 3.7%), and South America ($n = 4$, 3.7%). Among those offering internship opportunities, the United States ($n = 42$, 89.4%), Asia Pacific ($n = 2$, 4.3%), South America ($n = 2$, 4.3%), and Canada ($n = 1$, 2.1%) were the only regions with identified internship opportunities (see **Table 1**). There were no opportunities reported in Europe or India. Among those with internship opportunities, three sectors predominated: Private or publicly traded companies ($n = 43$, 91.5%), Academe ($n = 2$, 4.3%), and Sole Proprietorship companies ($n = 1$, 2.1%); one respondent failed to report a market sector.

Discussion

Clinical data management and clinical data science represent a critical and emerging component of the biomedical science enterprise. Despite the importance of clinical data management to public health over the past 50 years, formal educational opportunities have been greatly lacking.[13] Without the benefit of degree-granting programs in clinical data science at the undergraduate and graduate levels to train the workforce, less formal, employer-driven programs like internships have been used as a principal training mechanism. For those trying to enter the profession, or trying to move from one sector

Table 1: Organizational Characteristics.

Factor	All Organizations† *n* (%)	Organizations with Internships *n* (%)
Organizational Size (Number of Employees)		
Small (1,000 or less)	45 (41.7)	16 (34.0)
Medium (1,001 to 9,999)	18 (16.7)	9 (19.2)
Large (10,000 or more)	45 (41.7)	22 (46.8)
Organizational Type		
Academic Research Organization	3 (2.8)	1 (2.1)
Biologics Research & Development	4 (3.7)	3 (6.4)
College or University	3 (2.8)	1 (2.1)
Contract Research Organization	21 (19.4)	7 (14.9)
Medical Device Research & Development	8 (7.4)	5 (10.6)
Healthcare Facility	1 (0.9)	
Pharmaceutical Research & Development	56 (51.8)	29 (61.7)
Software Provider	9 (8.3)	
Other	1 (0.9)	1 (2.1)
No response	2 (1.8)	
Geographic Location		
Asia Pacific	3 (2.8)	2 (4.3)
Canada	3 (2.8)	1 (2.1)
Europe	2 (1.8)	
India	4 (3.7)	
South America	4 (3.7)	2 (4.3)
United States	92 (85.2)	42 (89.4)

Note. †Sample sizes are as follows: all organizations (n = 108), organizations with internships (n = 47).

to another, internship programs are considered to be a viable route to continuing education and development. Without a definitive resource in place to document and connect interns with organizations, trainees and organizations must search on their own for the right match. For the trainees, that could mean several years of searching and career progression before achieving the desired role of clinical data manager or clinical data scientist.

The purpose of this report was to document internship opportunities within the combined field of clinical data management and clinical data science. Findings presented here offer initial but convincing evidence of internship opportunities worldwide for those looking to get first-hand experience in pharmaceutical, academic, clinical research, software development, and medical device research and development organizations. While this report is only preliminary, it does suggest that many opportunities exist for those wishing to pursue internships in the newly emerging field of clinical data science. While the survey itself appeared weighted in favor of organizations located within the United States (n = 92), the number of actual internship opportunities was more balanced, with 178 of the 273 (65%) internship opportunities located in the United States, and 95 of 273 (35%) internship opportunities located outside of the United States. Given the comparability of organizations between those with and without internships reported here, it is very likely that the number of internship positions reported in this study is an underestimate. Participants at other professional meetings may report other internship opportunities, as well. Equally promising, the internships described here reflect small, medium, and large research organizations, allowing for variation in preferences for interested applicants.

The major strength of this study is that it formally documents and describes the pattern of internship opportunities available to clinical data management and clinical data science professionals available today—providing a call to action with respect to more formalized training opportunities in the biomedical sciences. Like the professionals currently in the workforce, the internship positions are spread across many types of organizations throughout many regions of the world. There should

therefore be opportunities for everyone, it is just a matter of finding the right fit for the intern and for the hiring organization. Recommendations for further research include doing a deeper dive into which organizations have the openings, the number of openings available, and what levels of training and experience are needed for the respective internship positions. Participating organizations should also provide institutional contact information so that potential interns know whom to contact for follow up.

There are a number of limitations to this study. First, it was conducted at a single, albeit international, conference in one location and may not fully represent the pattern of internship opportunities industry-wide. Second, responses were embedded within a longer and more complex survey on practice patterns and competencies with a fairly low participation rate of 16%. No attempt was made to control for multiple responses from the same organization. Future research should have a brief but targeted survey, designed to obtain internship availability information at the organizational level. Finally, it is recommended that professional organizations make obtaining this information a common practice at annual meetings so that the field can track its progress in this important area of professional development.

Conclusion

Findings presented here offer convincing evidence of internship opportunities worldwide for those looking to get first-hand experience in the clinical data sciences, specifically when it relates to pharmaceutical, academic, clinical research, software development, and medical device development organizations.

Additional Files

The additional files for this article can be found as follows:

- **Appendix 1.** SAS Code. DOI: https://doi.org/10.47912/jscdm.316.s1
- **Appendix 2.** Data File Internship Study. DOI: https://doi.org/10.47912/jscdm.316.s2

Funding Information

Portions of this report were made possible by a sabbatical award to RFI by the Division of Biostatistics and Epidemiology, Cincinnati Children's Hospital. The authors would like to extend a heartfelt thank you to Dr. Meredith Zozus and the Clinical Research Informatics team at the University of Texas Health Science Center, San Antonio for their work on the survey, from which the internship data were obtained.

Competing Interests

The authors have no competing interests to declare.

Author Contributions

RFI drafted the initial manuscript. AN and SC provided substantial comment, edits, and contributions to the initial draft. All authors approved the final manuscript for submission and publication.

References

1. **Ittenbach RF.** From clinical data management to clinical data science: Time for anew educational model. *Clinical and Translational Science*. 2023; 16(8): 1340–51. DOI: https://doi.org/10.1111/cts.13545
2. **Wu W.** (ed.) Investigating internship experiences of data science students for curriculum enhancement. *Proceedings of the 27th ACM Conference on on Innovation and Technology in Computer Science Education*. 2022; 1: 505–511 DOI: https://doi.org/10.1145/3502718.3524741
3. **Tang R, Sae-Lim W.** Data science programs in U.S. higher education: An exploratory content analysis of program description, curriculum structure, and course focus. *Education for Information*. 2016; 32(3): 269–90. DOI: https://doi.org/10.3233/EFI-160977
4. **U.S. Department of Labor.** Clinical Data Manager: 15-2051.02. Washington, DC: Employment and Training Administration; 2023.
5. **Freel SA, Snyder DC, Bastarache K,** et al. Now is the time to fix the clinical research workforce crisis. *Clinical Trials*. 2023 Oct; 20(5): 457–462. DOI: https://doi.org/10.1177/17407745231177885
6. **Wizemann T, Gee AW, Shore C.** Envisioning a transformed clinical trials enterprise for 2030. Washington, DC: National Academies of Science; 2022. DOI: https://doi.org/10.17226/26349
7. **Rawlings-Goss R.** Data science careers, training, and hiring: A comprehensive guide to the data ecosystem: How to build a successful data science career, program, or unit. Springer; 2019. DOI: https://doi.org/10.1007/978-3-030-22407-3
8. **Williams TB, Schmidtke C, Roessger K, Dieffenderfer V, Garza M, Zozus M.** Informing training needs for the revised certified clinical data manager (CCDMTM) exam: Analysis of results from the previous exam. *JAMIA Open*. 2022; 5(1): ooac010. DOI: https://doi.org/10.1093/jamiaopen/ooac010
9. **Sonstein SA, Jones CT.** Joint task force for clinical trial competency and clinical research professional workforce development. *Frontiers in Pharmacology*. 2018; 1148. DOI: https://doi.org/10.3389/fphar.2018.01148
10. **Valenta AL, Berner ES, Boren SA,** et al. AMIA Board White Paper: AMIA 2017 core competencies for applied health informatics education at the master's degree level. *JAMIA*. 2018; 25(12): 1657–68. DOI: https://doi.org/10.1093/jamia/ocy132
11. **Lamer A, Oubenali N, Marcilly R, Fruchart M, Guinhouya B.** Master's degree in health data science: Implementation and assessment after five years. *Studies in Health Technology Informatics*. 2022; 298: 51–5. DOI: https://doi.org/10.3233/SHTI220906

12. **Cameron S, Ittenbach RF, Mohammed A, Rozami R.** Careers in clinical data management and clinical data science and introduction to the Society for Clinical Data Management (SCDM). Brussels, Belgium (Virtual): Society for Clinical Data Management; 2023.

13. **Banach MA, Fendt KH, Proeve J, Plummer D, Qureshi S, Limaye N.** Clinical data management in the United States: Where we have been and where we are going. *Journal of the Society for Clinical Data Management*. 2021; 1(3). DOI: https://doi.org/10.47912/jscdm.61

How to cite this article: Ittenbach RF, Neubauer A, Cameron S. Internships in Clinical Data Science: A Brief Report. *Journal of the Society for Clinical Data Management*. 2024; 4(1): 4, pp. 1–5. DOI: https://doi.org/10.47912/jscdm.316

Submitted: 31 October 2023 **Accepted:** 17 February 2024 **Published:** 09 April 2024

Journal of the Society for Clinical Data Management is a peer-reviewed open access journal published by Society for Clinical Data Management.

OPEN ACCESS