

This is a “preproof” accepted article for *Journal of Clinical and Translational Science*.

This version may be subject to change during the production process.

10.1017/cts.2025.10170

Empowering Professionals: An Intensive Short Course on Fundamentals of Clinical Data Science

Richard F. Ittenbach,¹ Brian J. McCourt,² Maurizio Macaluso¹

¹Richard F. Ittenbach, Division of Biostatistics and Epidemiology, Cincinnati Children’s Hospital, and University of Cincinnati College of Medicine, Cincinnati, OH 45229-3039 (email: richard.ittenbach@cchmc.org)

²Brian J. McCourt, Duke Clinical Research Institute, Duke University, Durham, NC27701 (email: brian.mccourt@duke.edu)

³Maurizio Macaluso Division of Biostatistics and Epidemiology, Cincinnati Children’s Hospital, and University of Cincinnati College of Medicine, Cincinnati, OH 45229-3039 (email: maurizio.macaluso@cchmc.org)

Disclosure Statement: The authors have no conflicts of interest to declare.

Corresponding author: Rick Ittenbach, PhD, Division of Biostatistics and Epidemiology, Cincinnati Children’s Hospital, University of Cincinnati College of Medicine (richard.ittenbach@cchmc.org; phone: 513.803.0245), ORCID: 0000-0002-7914-4298.

This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use or in order to create a derivative work.

Abstract

Clinical data science, like the broader discipline of all data science, has quickly grown from obscurity only a few decades ago to one of the fastest growing specialties in biomedical research today. Yet, the education and training of the workforce has not kept pace with the growth of the field, the complexity of science, or the needs of the profession. The purpose of this paper is to provide a template for an intensive short course on fundamentals of clinical data science that meets the needs of working professionals in academic, industry, and government research settings. Care will be taken to introduce students to essential roles, responsibilities, and practice patterns within the field, the foundational components from which they come, and many of the soft skills needed for professional practice and advancement in the field today. The course is designed as an evidence-based, immersive learning experience taught over a 5-day period on a university campus, taught using principles of best educational practice and multiple modalities, to assure optimal interaction and engagement throughout the week. This template may be reproduced by any institution interested in and capable of offering such a program.

Keywords: clinical data science, clinical data management, curriculum, data science, executive training

Empowering Professionals:

An Intensive Short Course on Fundamentals of Clinical Data Science

Clinical data science, like the broader discipline of all data science, has quickly grown from obscurity only a few decades ago to one of the fastest growing specialties in biomedical research today.¹ Yet, the education and training of the workforce has not kept pace with the growth of the field, the complexity of the science, or the needs of the profession. A new approach to education is desperately needed for the field, and its dedicated professionals, to reach their full potential today.²

The pace, complexity, and sophistication of clinical research continues to expand well beyond what was imagined only a few decades ago.^{3,4} With a typical Phase III clinical trial now having 263 procedures, 22 endpoints, and 3.6M data points (up from 7 endpoints, 100 procedures, and 0.5M data points in 2005)⁵ and with 60% of clinical sites now using more than 20 overlapping software applications at any one time,⁶ even the more rigorously trained professionals find it challenging to keep up. Concomitant rises in ethical issues, cost of healthcare, and advancing technologies have raised the stakes even further, necessitating that today's clinical data scientists' understanding goes well beyond simply managing the data to now include the usability of the data, their merits as a scientific tool, and their role in the scientific process.⁷⁻⁹ The scientific community is now expecting these professionals to be true “scientists.”

While there are many programs training data scientists, there are currently no programs training “clinical” data scientists—at either the graduate or undergraduate levels. Today's clinical data scientists have generally received their training through a collection of efforts and opportunities.² The most rigorous cases are traditional or rebranded statistics or informatics programs, which are strong with respect to analytics, systems development, and technology, but often lack formal training in the data itself and most certainly lack the ability to manage the flow of data through a clinical research study. The more typical cases are employer-driven continuing education programs focusing on the processing of data rather than the theory, rationale, or scientific attributes of the data. Professional associations have offered specialized courses and programs just as individuals have created unique individually-tailored programs; however, these are not reliable mechanisms to train an entire workforce.

A trend that has become apparent in the business community over the past several decades is to bring working professionals together for targeted, brief, intensive, immersive educational experiences.¹⁰ Such courses, often referred to as executive education courses, have made in-roads in the business world and traditional university communities, alike.¹¹⁻¹³ According to Mena-Guacas et al. (2023), short courses offer benefits not typically available through other instructional formats, such as being taught by professional educators used to delivering the content, relocating students from the typical office and work-day routines to a more typical learning environment, and a holistic integration of content.¹⁴ The benefits of using an evidence-based framework add to the impact of the experience through an enriched appreciation of content, higher levels of trust in the instructors, and instructors' ability to package the information in a more usable format for learners.^{11,15}

As with all investments, efficacy and potential for a return on one's investment remains a concern. Beginning with the managerial revolution of business schools in the United States and Harvard Business School's Advanced Management Program as a model for other programs in 1945, programs are now taught worldwide across a wide range of formats.^{16,17} In a survey of 52 corporations conducted by the Consortium for University-based executive education programs, 92% of companies reported using participant feedback data sources to evaluate their executive education programs and 49% use it in career tracking and promotions, with 60% connecting executive education to their overall strategy.¹⁸

Given the pace and complexity of clinical research today, the lack of instructional programs globally, and the general lack of educational opportunities for working professionals, the purpose of this paper is to provide a template for an intensive short course on fundamentals of clinical data science that meets the needs of working professionals in academic, industry, and government research settings.

Materials and Methods

Expectations of the Course

This week-long course was developed with several expectations in mind: first, to map the coursework onto existing curricula and competency frameworks in the professional literature, specifically drawing from the foundational domains of biostatistics, biomedical informatics,

biomedical science, regulatory science, , and the clinical research literature, more generally (see Figure 1).^{8,19-24} The course, while short, needs to be a snapshot of content shared in a graduate level program in clinical data science “*devoted to the measurement, acquisition, care, treatment, and inferencing of clinical research data.*” Second, the content must be scaled to the abbreviated nature of the program and the accelerated pace of the instruction, but still prioritize weaving the information into a coherent whole, with each module building on the ones that came before it and laying the foundation for the ones that come after it (vertical articulation).²⁵ Third, for it to truly be a clinical data science course, it needs to show fidelity to “science” and to the scientific method. Furthermore, executive education faculty are encouraged to incorporate experiential learning activities to better integrate knowledge with applications of the material in typical work settings.

As with all scientific disciplines, there is core content to help define the discipline and supporting content that connects the new knowledge base to the other sciences. In the case of clinical data science, what began as an operational subspecialty to help get drugs and devices to market faster and with better reliability, effectiveness, and safety has evolved into a scientific specialty in its own right, one that draws from the scientific knowledge bases of related, foundational disciplines.^{26,27} Consequently, a short course in clinical data science should consist of core knowledge representing the essential components of the field such as first principles, roles and responsibilities, and fundamentals of design and implementation, while also orienting learners to core tenets of the disciplines on which the new field is based (biostatistics, biomedical informatics, biomedical science, and regulatory science).²

Pedagogy

Implicit within the course is a commitment to adhering to principles of best educational practice. That is, to meet the students where they are professionally, to engage them using multiple instructional formats (lecture, case study, group discussion), and to introduce them to exemplary models of literature and professional practice. While all instructional modules are organized around specific learning objectives, not all learning will come from structured exercises. Some will come from the less structured parts of the course by design—the breaks, meals, and networking. The course should offer an educational experience that is greater than the sum of its parts and stimulate learning and development once the students leave the course.

To be consistent with the educational characteristics of a graduate program, this intensive, abbreviated course strives to relay content from a full graduate program, just in a more condensed form. If the course is successful, it will move the students toward a deeper understanding of clinical data science. To help guide the curriculum and provide a point of reference for what students will be learning in the short course, we offer seven learning objectives—one from each of our three core modules and one from each of the four foundational disciplines of biostatistics, biomedical informatics, biomedical science, and regulatory science.

With respect to the core modules, the learning objectives begin with mapping a very simple data-flow diagram and process onto the scientific method (Objective 1), to recalling from memory and describing elements of a data management plan (Objective 2), to actually connecting data fields and variables from the study protocol to the statistical analysis plan, to the data management plan—and, then finally, to the study database (Objective 3). Implicit in these core modules is a developmental sequence that takes the learner from the scientific method to the plan for the data to how the fields and variables progress through all phases of the study from both an operational and a scientific standpoint.

Regarding the research courses that support the core content, the learner moves from a recognition and understanding of the impact of measurement bias and imprecision on the scientific process (Objective 4, biostatistics); to formatting data elements using globally-recognized standards (Objective 5, biomedical informatics); to distinguishing among various health conditions of patients and research participants along with the role of diagnosis and intervention in responding to those conditions (Objective 6, biomedical science); and, finally, an understanding and reliance of influential guidelines protecting human subjects in clinical research (Objective 7, regulatory science). Please see Figure 2 for a list of the seven learning objectives for this short course.

As important as the roadmap for learning is, the educational process would not be complete without an evaluation plan. As such, the evaluation mechanism used for the short course will be similar to the one used for our graduate program and includes a two-phase approach to program evaluation. Phase I will include a straightforward pretest-posttest approach to evaluation in which the students are asked a series of questions related to typical work products, immediately prior to and immediately following delivery of the content (see Figure 2

for examples of the work product evaluation). In each case, the assessments will be designed to be succinct, focused, time limited, and based on content used in the course. Phase II will consist of components from the Environment,

Pedagogy, Institution, and Course (EPIC) system, which is a standardized (NSF funded) evaluation system of surveys for university-level data science programs—but tailored to our short course rather than that of an entire semester long course.²⁸

One challenge will be to find a meaningful but efficient way of incorporating homework into the intensive short course. One strategy that is available to instructors is to have the students bring specific examples of work products with them to the short course, that they can use in the break-out sessions following content-related discussions (e.g., Data Management Plan, abbreviated protocol, data flow diagram). Foundational concepts can be effectively demonstrated using data science tools that faculty employ in real-world applications. With the widespread availability of online data, online data-collection methods, databases, and analytical tools, these elements can be integrated into case studies, enabling participants to gain a deeper understanding of the material encountered in traditional classroom discussions. Finally, the course must be developed in keeping within the fiscal expectations and limitations of host institutions

Budgetary Considerations

As with all innovative programs, this one must be funded in a way that is not only affordable but is also sustainable and does not deplete resources from other programs and initiatives. As such, we have provided a preliminary budget that details typical costs of such a program (see Table 1).

The proposed budget is divided into four categories: facilities, catering, speaker travel/honoraria, and project management support. As one might imagine, implementing an intensive, 5-day course such as this one requires meticulous planning and coordinated efforts as well as time for follow-up communication and post-course processing. The higher the quality, the more likely students are to attend from great distances, and, as a result, the greater the need for communication, planning, and organizational support. We have proposed the use of a skilled project manager at 30% time for six months. The underpinning value of this program lies in the carefully curated and coordinated content and opportunities for interaction with leaders in the

field. Unlike traditional university courses based on a single instructor, a multi-disciplinary program such as this one requires the input and synthesis of multiple experts from many areas, driving a

significant portion of the budget. Facilities and catering are the most easily quantifiable, and the most easily scaled to align with the expectations and needs of the program. Cost estimates provided in Table 1 are drawn from actual examples of conferences offered at two university medical centers and experienced with on- as well as off-campus continuing education courses and meetings.

Results

The clinical data science course is designed to be a 50-hour course consisting of 19 educational modules taught over an intensive, five-day period. The 19 modules are organized as follows:

- 5 primary modules associated with core content (overview, roles & responsibilities, design & implementation, field placements, good clinical data management practice)
- 5 supporting modules associated with foundational research content (biomedical informatics, biostatistics, biomedical science, regulatory science, , and clinical research ethics)
- 4 supporting modules on soft skills needed for successful professional practice (leadership, critical thinking, team science, communication)
- 5 supporting modules devoted to challenging cases studies.

Day 1 of the program begins with an opening session, which includes a light breakfast, an introduction to the program, faculty, course objectives, housekeeping, and news and notes for the day. The remaining days will begin similarly but be much more focused, beginning with a light breakfast, a recap of the prior day, an introduction to the current day and any news or notes that may be needed. The instructional portion of the day is divided into morning and afternoon sessions, with each half-day beginning with a substantive content session (e.g., Overview of Clinical Data Science) followed by a supporting case study or instruction in a “soft skill” deemed

important for professional practice (e.g., critical thinking). A number of different instructional methods will be used to keep delivery fresh and of interest to students with different areas of interest, ranging from the traditional didactic method to hands-on exercises and small group discussions.

When people are brought to campus for an intensive and demanding short course, especially when brought to campus from great distances, having meals available and aligned with the conference are crucial to keeping the students present, focused, and on schedule. As alluded to previously, even the meals are designed to offer the students time to network with faculty, staff, and other students—to be fun and relaxing but still immerse them in the field of clinical data science. Three of the five days will end with a working dinner and a professional speaker from the community (see Figure 3).

The most beneficial component of a short, intensive program such as this one is what it can do for current, working professionals who may not have the time or resources to enroll in a more traditional, multi-year program. While the benefits of the course will be shaped by the instructors as much as each student's commitment to learning, this short course will be designed to equip the students with an integrated knowledge base that extends the foundational content of the parent disciplines. Provided below are examples of ways in which the short course will benefit working professionals, organized by the four respective foundational areas:

Biostatistics

Data science as a distinct discipline is rooted in statistics (or biostatistics) and computer science. As such, the new discipline draws heavily from the analytical side of clinical inquiry—one that is premised upon searching for certainty in a world of unknown, but probabilistic events. Appreciation of the uncertainty that necessarily affects data capture and the importance of a rigorous statistical approach to harnessing the uncertainty of the inferences and the point estimates on which the estimates are based will require a critical understanding of study design and the ways in which the data are measured and collected. The connection among fundamentals of measurement, the statistical tests used to analyze the data, and other supporting principles discussed in most introductory statistics courses will highlight the importance of the rationale, validity, and need for the precision of the biomedical sciences. Such relationships are generally detailed in the core content, Statistical Analysis Plan, and its companion document, Data

Management Plan, noted above(Objectives 2, 4).

Biomedical Informatics

Information science and the technology that supports it provides the means to ensure scientific rigor to a degree previously deemed unachievable—across all areas of science. Mastery of the data capture software and the programming techniques that support it may allow presenting survey questions with branching logic in a user-friendly format, representing the data entry modules in a table that reproduces a table of events, displaying summary tables reporting which forms have been completed or how many data elements are missing. Alignment of data formats to existing standards will ensure that the data produced are “findable, accessible, interoperable and reusable” (FAIR), enhancing the impact of the clinical research study. The Clinical Data Interchange Standards Consortium (CDISC) has developed standards that cover most aspects of clinical research, from protocol development to data collection, presentation, and analysis. Adherence to CDISC standards is mandatory for submission of data to the U.S. Food and Drug Administration (FDA)(Objectives 3, 5)

Biomedical Science

Understanding the research questions that motivate a clinical study, the students will verify how the study protocol operationalizes the scientific objectives into a schedule of events and outlines the procedures needed to capture interventions and endpoints in a synthetic study flow diagram. It is also important, however, for the clinical data scientist have some basis for understanding the health condition being studied, diagnostic strategies used, and avenues of intervention. The mechanism through which the health-related data will be collected, processed, and stored should be specified in a Data Management Plan²⁹ and its corresponding Manual of Operations, the roles of key clinical research staff, who in turn will interface with the clinical data managers (Objectives 3, 6)

Regulatory Science

The technical aspects of designing and deploying a clinical research study must be embedded in an overarching ethical/legal framework that reinforces the principles of respect for persons, beneficence, and distributive justice specified in the Belmont Report and further developed in the current body of rules and regulations governing research with human subjects.³⁰

Understanding these principles is indispensable for clinical data scientists, whose technical expertise must be put at the service of the ethical treatment of research participants. Thus, the research protocol must not only reflect the scientific objectives of the study but must align the study procedures and objectives in such a way that review boards can make an independent assessment of the risks and benefits to the participants. The work of the data management team must be inspired by the respect for the autonomy and privacy of the participants and must adhere to all applicable data confidentiality regulations (Objectives 1, 7)

Integration of principles and concepts from the four foundational areas of biostatistics, biomedical informatics, biomedical science, and regulatory science transforms a clinical data manager into a “clinical” data scientist.^{31,32} A clinical data scientist does not simply use the data as a tool but rather treats the clinical research data as a scientific object in its own right, worthy of study and further scientific development. The clinical data scientist strengthens the precision of the data to achieve the objectives of a research study and enhances the usability of the data and its power to better serve clinical and translational science, in general.⁴

Discussion

Executive education courses offer a new and potentially exciting option for filling the gap in continuing education opportunities for working professionals. Extensive short courses such as the one presented here can delve more deeply into selected domains and still put the information into a broader context. Executive education programs remain a vastly under-utilized form of continuing education fully capable of having an impact at both the individual and institutional levels.¹² This paper provides a template for an intensive short course on fundamentals of clinical data science that meets the needs of working professionals in academic, industry, and government research settings.

The most notable feature of this course is instruction in the core tenets of clinical data science—the measurement, acquisition, care, treatment, and inferencing of clinical research data.² The data are not simply the product of research but the basis for it. This is the material that distinguishes the content from other content received in other settings. It is the material that distinguishes this profession from all others. With respect to the core content noted previously, following is a snapshot of the content covered during the morning sessions:

- Overview of the field, focusing on the data as the basis for scientific hypotheses and the inferences that come from them Module 1
- Roles and responsibilities of today's data scientists Module 2
- Design and implementation, from development of the protocol to database lock and closeout, and every step in between Module 3
- Simulation of field-based work Module 4
- Good Clinical Practice/Data Management Practice Guidelines Module 5

All of the above are critical components of a clinical data scientists' world and routines. Modules 1-5 will be the featured content in the mornings, respectively (see Figure 3).

Research Courses

The field of clinical data science is little more than 50 years old. Yet, the field as it is known today actually evolved out of other related fields, initially at the intersection of statistics and computer science, but now with heavy influence from the biomedical and regulatory sciences. The fact that today's professionals must now deal with complex issues not fully appreciated in years past, requires that today's clinical data scientists receive instruction in the tenets of clinical research ethics. Whereas core content will be featured in the morning, with the research courses featured in the afternoons:

- Principles of biomedical informatics, the linking of computational tools and algorithms with biomedical information and data Module 6
- Principles of biostatistics, inferencing, and clinical trial design. Module 7
- Principles of biomedical science, including that which directly relates to healthcare Module 8
- Federal regulatory guidelines and principles that shape biomedical research today, particularly those that pertain to the protection of research subjects Module 9
- Clinical research ethics, a topic that is often passed over for the more familiar and comfortable analytical content Module 10

Supplemental Courses

Recognizing that science is no longer practiced in a vacuum, scientists are now expected to work closely and collaboratively with investigators from multiple disciplines. Consequently, the closing hour of each morning and afternoon will be devoted to either a challenging case study reinforcing principles presented in that day's session, or supplemental content on one of four soft skills deemed critical to the practice of clinical data science and advancement in the field: critical thinking, communication with colleagues, leadership fundamentals and, finally, team science skills (Modules 11 through 14), with challenging case studies designed to illustrate the challenges in today's intricate and fast-paced world of biomedical research (Modules 15 through 19). Sessions featuring soft skills and challenging case studies are designed to be offered in alternating sessions (see Figure 3).

Benefits of Short Courses

For individuals whose training has become outdated, those trained in a different area of expertise, or even those whose supervisors simply want the staff member to have additional training, intensive short courses can provide the education needed for strengthening one's professional practice. Whereas it is often difficult to devote time to a semester-long course that may be disruptive to a person's life or work schedule, most colleagues understand dedicating several days to continuing education. And most will agree that 40 to 50 hours of formal instruction by university faculty is substantial enough to improve one's knowledge base irrespective of level. Whether through supplemental readings, repetition of concepts, increased exposure time, or in-depth dialog between the students and instructors, intensive short courses are likely to produce returns on investment well beyond what employer- and learner-driven webinars can typically offer.

Not surprisingly, students are not the only ones who benefit from these short, intensive courses. Courses such as these can also have tremendous appeal for university faculty. Staffing the courses with experienced university faculty who are familiar with the content makes sense from an administrative standpoint. They know the material and the pitfalls—and, in many cases, have dealt with the barriers to practice. This is their world. What university faculty have not always encountered, however, are the problems of managers and employees in the trenches. Here faculty get to see the problems and pitfalls of practice through a different lens, by seasoned

administrators and employees, which can be both exciting and challenging for all involved. Because the courses are short, university faculty are often willing to invest their time in the area and with students who have an acute need to know. The faculty can then return to their normal schedules. In short, university faculty can often draw from both theory and practice to go deeper, quicker than others for whom the teaching is an add-on to their normal work routine.

In addition, universities often appreciate the inclusion of short courses to their portfolio of classes. These courses often bring non-traditional students to campus, expanding the exposure of the program and the university to new groups of potentially interested students. In addition, these courses have the potential to put university faculty and staff in direct contact with industry professionals, opening opportunities for other forms of collaboration, consultation, and engagement. And why not expand the portfolio of learning opportunities to these new groups of students? The best data available suggest that universities currently account for less than 1% of the executive education market—a market that universities are most qualified to compete in—and, one for which there is ample room for growth in this profoundly important area of professional development.^{10,33} Intensive short courses and the opportunities for interaction mean that industry professionals also have access to new faculty with whom they can interact with when needed.

As noted previously, faculty should strive to incorporate a number of different instructional models in their teaching. As working professionals know, professional education classes can often be long and boring. But, lessons learned from the best elementary and secondary teachers suggest that active engagement is the best way to learn and retain information. For this reason, it is recommended that short course faculty balance traditional lectures with small group discussions, as well as some activity-based assignments to allow the students to shape the direction of the content as well as be actively engaged with the material. An example of an activity-based assignment is to create a data flow diagram, data collection form, example code or output using public datasets and freely accessible tools. This blend will likely need to be adjusted frequently based on the faculty's strengths and the needs of the content. Asking students to bring examples from their prior work experience into activities can be a highly valuable way of engaging students and alignment to individual learning goals.

Budgetary Considerations

Education costs. More importantly, though, investments in education represent the values of a person, the community, and the profession. Intensive short courses such as the one presented here begin to make the content available to a much wider audience than ever before, working professionals from a broad range of backgrounds. The week-long commitment requires time away from family, work, and typical routines, but elevates the knowledge and skills of those who engage.

The course brings together and packages information that is not otherwise available to the professional community who can directly benefit. If the students leave the course being able to achieve the objectives listed here and begin operating at levels substantially above their previous levels of performance, everyone will benefit—the student, the studies they are working on, the employees they supervise and work with, the organizations in which they work, and science more generally. Investment in education pays dividends well beyond the cost of the instruction. But education does indeed cost; the more essential the education, the greater the cost. This program is an example. Organizations committed to these processes, and their outcomes may replicate this model, its purpose, and scope.

The budget provided in Table 1 is generated from conference expense sheets designed to increase professional education in biomedical research at the authors' institutions. We are now applying that model to clinical data science. The short course presented here proposes a budget of approximately \$140,000, which not all organizations will be able or even interested in offering. It is not a formal prescription for a one-size-fits-all course but may be scaled up or down to meet the needs of the students, sponsors, and institutions involved. There are always ways to trim expenses, and to scale the information to the budgets and resources available, either by using more local, less experienced instructors, offering the program in less expensive venues, and/or not providing snacks and meals. A piece-meal, fragmented system of webinars and short courses has been used with this segment of the workforce for decades, it is now time to offer them more. This is a profession that oversees increasingly complex biomedical data without any formal training—it is time to change that model—and offer them more training that is commensurate with their importance to the field.

No revenue projections were provided to allow institutions to gauge their own rates of return due to differences in what they are willing and able to offer. The intensive short course presented here is designed for industry-leading organizations and workforces willing to invest in training at the highest levels. For purposes of completeness, a registration fee of \$3,500 for 40 students would cover the costs of the course described here, while a registration fee of \$4,370 for the same 40 students would return a 25% margin for the host institutions.

Additional factors will also need to be considered such as student versus practicing professional rates as well as early bird and/or regular registration rates. Even deadlines for drop-outs can impact one's budget markedly after expenses have been committed, so establishing firm drop-out deadlines can provide a cushion should drop-outs come too late for others to register. In addition, host institutions will need to decide if they wish to pursue sponsorship from educational and/or commercial organizations and whether they will want to offer continuing education credit. Such factors are usually institution-specific and were not factored into this model. As with all continuing education, not all organizations will be able or interested in undertaking this type of training for their community. This entire model can be scaled based on the needs of an organization.

Evaluation

Care should be taken to evaluate not only the delivery of information (pedagogy) but the course's timeliness and receptivity by the students to keep the material fresh and useful for the learners. The specifics of the evaluation tool should be up to the course organizers; however, one key component should be whether the students can accurately and effectively master the seven objectives presented in Figure 2. Other process-oriented questions on planning such as content, meals, and overall costs may be included, as well. The evaluation plan described here should be designed a priori, prior to the launching of the program, to help in determining its effectiveness and planning for subsequent instruction.³⁴

Conclusion

The purpose of this paper was to provide a template for an intensive short course on fundamentals of clinical data science that meets the needs of working professionals in academic, industry, and government research settings. The 50-hour, 19 module course is divided into three

sets of instructional modules: core content, supporting research content, and supplemental modules. Whereas the first set of modules is designed to instruct on the core content of the profession, the research modules are designed to convey information that serves as the basis for the new field, and, finally, the supplemental modules are designed to guide instruction on challenging case studies as well as many of the soft skills needed for practice as a clinical data scientist today.

Acknowledgement

The authors thank Ms. Karen Whyte, MSLIS, MA of Cincinnati Children's Hospital, for her insight and help assisting us with the literature search.

Funding Statement

This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Author Contributions:

Richard F. Ittenbach: Conceptualization, Methodology, Project administration, Supervision, Writing - original draft, Writing - review & editing; Brian J. McCourt: Conceptualization, Investigation, Methodology, Writing - original draft, Writing - review & editing; Maurizio Macaluso:

Conceptualization, Investigation, Methodology, Writing - original draft, Writing - review & editing.

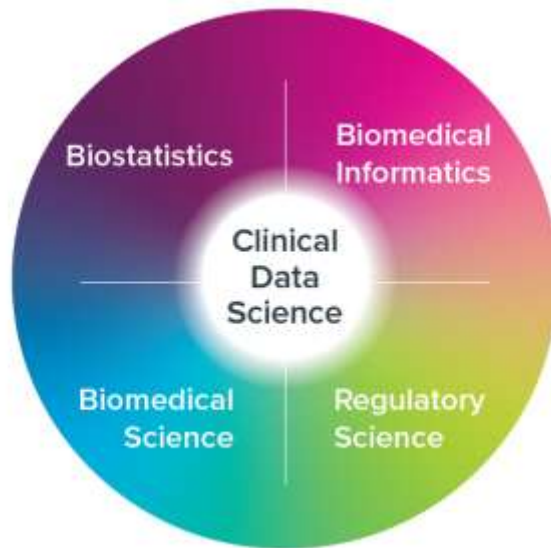
References

1. U.S. Department of Labor. Occupational Outlook Handbook. U.S. Government Printing Office. Accessed February 11, 2025, <https://www.bls.gov/ooh/fastest-growing.htm>
2. Ittenbach RF. From clinical data management to clinical data science: Time for a new educational model. *Clinical and Translational Science*. 2023;16(8):1340-1351. doi:10.1111/cts.13545
3. Zerhouni EA. The NIH roadmap. *Science*. 2003;302:63-72.
4. Austin CP. Opportunities and challenges in translational science. *Clinical and Translational Science*. 2021;14(5):1629-1647.
5. Smith Z, Bilke R, Pretorius S, Getz K. Protocol design variables highly correlated with, and predictive of, clinical trial performance. *Therapeutic Innovation & Regulatory Science*. 2022;56(2):333-345.
6. Hooker D, Johnson J, Bokus T, Silvester R, Curtis J. *Clinical Trial Technologies: Market Research Report*. 2025;34. January 22, 2025. <https://www.bourne-partners.com/wp-content/uploads/2025/01/Bourne-Partners-Clinical-Trial-Technology-Report-Jan-2025.pdf>
7. Beller E. Clinical data management: Challenges in training. *Health Information Management*. 1996;26(1):17-19.
8. Succar T, Terteryan A, Manrique K, Pacifici E. Optimizing Clinical Trial Education in Academia. *Clinical Researcher*. 2024;38(5):6-15.
9. Weil SA, Crumpler A, Medendorp SV, Weil SA, Medendorp SV. The implementation of a data learning series focused on clinical development teams in a contract research organization. *Journal of the Society for Clinical Data Management*. 2022;2(1):1-12. doi:10.47912/jscdm.39
10. Voller S, Honoré S. *Innovation in Executive Development: A Case-based Study of Practice in International Business Schools*. Ashridge; 2008.
11. Smith MA, Keaveney SM. A technical/strategic paradigm for online executive education. *Decision Sciences Journal of Innovative Education*. 2017;15(1):82-100.
12. Tushman ML, O'Reilly C, Fenollosa A, Kleinbaum AM, McGrath D. Relevance and rigor: Executive education as a lever in shaping practice and research. *Academy of Management Learning & Education*. 2007;6(3):345-362.

13. Nielson FA, Bittencourt JP, Presada WA, Cavalcanti C, Berardo BM. *Pedagogical Innovation: Best Practices Through the Perspectives of Some Major Business Schools Around the World*. 2019. February 18, 2025.
14. Mena-Guacas AF, Chacon MF, Munar AP, Ospina M, Agudelo M. Evolution of teaching in short-term courses: A systematic review. *Heliyon*. 2023;9(6)
15. Jessani NS, Hendricks L, Nicol L, Young T. University curricula in evidence-informed decision making and knowledge translation: Integrating best practice, innovation, and experience for effective teaching and learning. *Frontiers in Public Health*. 2019;7:313.
16. Amdam RP. Executive education and the managerial revolution: The birth of executive education at Harvard Business School. *Business History Review*. 2016;90(4):671-690.
17. Amdam RPS. Executive Education. *Oxford Research Encyclopedia of Business and Management*. Oxford University Press; 2020.
18. Cataldo P, Stilliard B, Topping P. ROI on Executive Education: Revisiting the Past and Looking to the Future. *UNICON Research Report*. UNICON; 2018.
19. Joint Task Force for Clinical Trial Competency. Domains and leveled core competencies. Multi-Regional Clinical Trials Center of Brigham and Women's Hospital and Harvard. Accessed March 6, 2022, 2022. <https://mrctcenter.org/clinical-trial-competency/framework/domains/>
20. 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. *Journal of the American Medical Informatics Association*. 2018;25(12):1657-1668.
21. Zozus MN, Lazarov A, Smith LR, et al. Analysis of professional competencies for the clinical research data management profession: Implications for training and professional certification. *Journal of the American Medical Informatics Association*. 2017;24(4):737-745.
22. Takata M, Miyaji T, Hayashi Y, Sanada S, Yamaguchi T, Takata M. Analysis of Core Competencies for the Clinical Data Management Profession in Japan. *Journal of the Society for Clinical Data Management*. 2024;4(1):1-9. doi:10.47912/jscdm.355
23. Sonstein SA, Jones CT. Joint task force for clinical trial competency and clinical research professional workforce development. *Frontiers in Pharmacology*. 2018:1148.

24. Read KB. Adapting data management education to support clinical research projects in an academic medical center. *Journal of the Medical Library Association: JMLA*. 2019;107(1):89.
25. Hlebowitsh PS. *Designing the school curriculum*. Allyn & Bacon; 2005.
26. 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:10.47912/jscdm.61
27. Meadows M. Promoting safe and effective drugs for 100 years. *FDA Consumer Magazine*. 2006;40(1)
28. Unfried A, Whitaker D, Batacki L, et al. Peters S, Zapata-Cardona LB, F, Fan A, eds. *The Big Picture: A Family of Instruments for Understanding University-Level Statistics and Data Science Attitudes*. 11th International Conference on Teaching Statistics (ICOTS 2022); National Science Foundation; 2022.
29. Lebedys E, Famatiga-Fay C, Bhatkar P, Johnson D, Viswanathan G, Zozus MN. Data Management Plan. *Journal of the Society for Clinical Data Management*. 2021;1(1):1-20. doi:<https://doi.org/10.47912/jscdm.X>
30. National Commission for the Protection of Human Subjects. *The Belmont report: ethical principles and guidelines for the protection of human subjects of research*. vol 2. U.S. Government Printing Office 1978.
31. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*. 2016;3(1):1-9.
32. Wilkinson M, Dumontier M, Jan Aalbersberg I, Appleton G, Axton M, Baak A. Erratum: addendum: the FAIR guiding principles for scientific data management and stewardship (Scientific Data (2016) 3 (160018)). *Sci Data*. 2019. *Scientific Data*. 2019;6(1):6.
33. Lloyd FR, Newkirk D. University-based executive education markets and trends. *Unpublished manuscript Retrieved from UNICON (University Consortium for Executive Education)* <http://192185>. 2011;166:2011.
34. Lockyer J, Ward R, Toews J. Twelve tips for effective short course design. *Medical teacher*. 2005;27(5):392-395.

Figure 1. Clinical Data Science Components



Source: Adapted from Ittenbach. Clin & Trans Sci. 2023; 16(8):1343. doi: 10.1111/cts.13545.

Figure 2. Short Course Learning Objectives and Evaluation

Number	Learning Objective	Evaluation Strategy	Disciplinary Source
1	Construct and interpret a data flow diagram within the context of the scientific method	Draw an abbreviated data flow diagram with 7 to 10 steps with corresponding phases of the scientific method	Core Content: Overview
2	Provide an outline and supporting text for essential components of a Data Management Plan	Construct an abbreviated outline of a DMP with 6 to 8 key sections, and a 2 to 3 sentence description of each section	Core Content: Roles & Responsibilities
3	Map key variables in a clinical study protocol to the data management plan, statistical analysis plan, case report forms, and study database	Identify inconsistencies for a given variable when presented with a study protocol, SAP, DMP, CRF, and database	Core Content: Design & Implementation
4	Recognize potential sources of bias and imprecision in the data and propose remedies	Identify and explain 3 ways in which bias and 3 ways in which imprecision can undermine the measurement of key outcomes in a clinical research study	Biostatistics
5	Format data elements according to acceptable global standards	Identify the correct formatting required for three data elements using an established global standard	Biomedical Informatics
5	Recognize the diagnostic and therapeutic interventions commonly utilized in the case of patients	Define and distinguish among a specific health condition, diagnostic tools used to assess the condition, and therapeutic interventions used to resolve the health condition among patients	Biomedical Science
7	Describe how the study protocol and DMP reflect the requirements of research ethics and regulatory guidelines	Identify 3 regulatory guidelines that influence how a clinical research study is conducted	Regulatory Science

Figure 3. Example Curriculum

Day 1	Day 2	Day 3	Day 4	Day 5
Breakfast				
Opening Session	News and Notes	News and Notes	News and Notes	News and Notes
Clinical Data Science I: Overview*	Clinical Data Science II: Roles & Responsibilities*	Clinical Data Science III: Design & Implementation*	Clinical Data Science IV: Field Placements	Good Clinical Practice Good Clinical Data Management Practices
Challenging Case Study	Critical Thinking	Challenging Case Study	Communication	Challenging Case Study
Lunch				
Biomedical Informatics	Biostatistics	Biomedical Science	Regulatory Science	Clinical Research Ethics
Leadership	Challenging Case Study	Team Science	Challenging Case Study	Closing Session
Class Dinner with Invited Speaker	Dinner on Own	Class Dinner with Invited Speaker	Class Reception and Dinner in Community	

Note. * Denotes core content.

Table 1. Example Budget

Cost Category	Cost	Cost with 20% Service Fee	Cost with 20% Fee, 7.5% Tax
Facilities			
Marketing and Materials	2,500		2,688
1 Breakout Room (\$800/day)	4,000	4,800	5,160
<u>AV, Internet, Onsite Support Staff (\$3,000/day)</u>	<u>15,000</u>		<u>16,125</u>
Facilities Total			23,973
Catering (Meals, Snacks, Reception)			
Breakfast (\$25 x 5 days x 50 participants [10 staff])	6,250	7,500	8,063
Lunch (\$35 x 5 days x 50 participants [10 staff])	8,750	10,500	11,288

Breaks (\$30 x 25 participants x 9 break periods)	6,750	8,100	8,708
Dinners (\$50 x 3 days x 55 participants [15 staff])	8,250	9,900	10,643
<u>Alcohol (\$10/glass x 50 participants x 2 glasses)</u>	<u>1,000</u>	<u>1,200</u>	<u>1,290</u>
Catering Total			39,992
Speaker Travel, Honoraria			
Honoraria (\$3,500 x 10 faculty, \$500 x 10 faculty)	40,000		40,000
<u>Travel and Expenses for Speakers (\$1,000 x 7 faculty)</u>	<u>7,000</u>		<u>7,000</u>
Honoraria and Travel Total			47,000
Project Management Support			
Project Manager (2080hrs x ½ yr x 0.30 effort x \$100/hr)	31,200		31,200
Total			142,165

Note. catering costs for instructional faculty and staff should be built into the overhead costs.