



Gastroenterology Fellowship and Postdoctoral Training in Omics and Statistics—Part II: How Can It Be Achieved?

Madeline Alizadeh¹ · Natalia Sampaio Moura² · Alyssa Schledwitz² · Seema A. Patil² · Jacques Ravel¹ · Jean-Pierre Raufman^{2,3,4,5}

Received: 13 August 2023 / Accepted: 27 September 2023 / Published online: 2 November 2023
This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2023

Abstract

Data are being generated, collected, and aggregated in massive quantities at exponentially increasing rates. This “big data,” discussed in depth in the first section of this two-part series, is increasingly important to understand the nuances of the gastrointestinal tract and its complex interactions and networks involving a host of other organ systems and microbes. Creating and using these datasets correctly requires comprehensive training; however, current instruction in the integration, analysis, and interpretation of big data appears to lag far behind data acquisition. While opportunities exist for those interested in acquiring the requisite training, these appear to be underutilized, in part due to widespread ignorance of their existence. Here, to address these gaps in knowledge, we highlight existing big data learning opportunities and propose innovative approaches to attain such training. We offer suggestions at both the undergraduate and graduate medical education levels for prospective clinical and basic investigators. Lastly, we categorize training opportunities that can be selected to fit specific needs and timeframes.

Keywords Medical informatics · Medical education · Big data · Gastroenterology · Hepatology

Introduction

Despite the increasing prevalence and use of enormous datasets in GI and hepatology research, training in the handling of ‘big data’ appears to be underutilized. As discussed in Part 1 of this two-part series, there is an evident need for comprehensive training in the acquisition, analysis, and comprehension of large datasets, which those interested

in pursuing these lines of investigation frequently lack [1]. While expansion of existing training pathways is necessary, multiple opportunities currently exist for those interested in performing big data GI and liver research. Here, we offer an overview of the current options for training in big data use and analysis and offer potential avenues for attaining this training during undergraduate and graduate medical education. We review these opportunities in the context of the time available, arbitrarily dividing options into those requiring less than 1 year, 1 to 2 years, and 3 or more years (Fig. 1).

✉ Jean-Pierre Raufman
jraufman@som.umaryland.edu

¹ The Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD 20201, USA

² Division of Gastroenterology and Hepatology, Department of Medicine, University of Maryland School of Medicine, Baltimore, MD 21201, USA

³ VA Maryland Healthcare System, Baltimore, MD 21201, USA

⁴ Marlene and Stewart Greenebaum Cancer Center, University of Maryland School of Medicine, Baltimore, MD 21201, USA

⁵ Department of Biochemistry and Molecular Biology, University of Maryland School of Medicine, Baltimore, MD 21201, USA

Existing Options for Trainees Who Wish to Pursue Experience and Develop Expertise in Big Data Analysis

Numerous options and pathways exist for trainees wishing to attain competence in working with or incorporating large datasets in their work (Table 1). Students enrolled in dual MD/PhD or DO/PhD programs (henceforth referred to as “MD/PhD”) can obtain a PhD in formal biostatistics or bioinformatics and rotate in laboratories that focus on big data generation and/or analysis. However, it is important

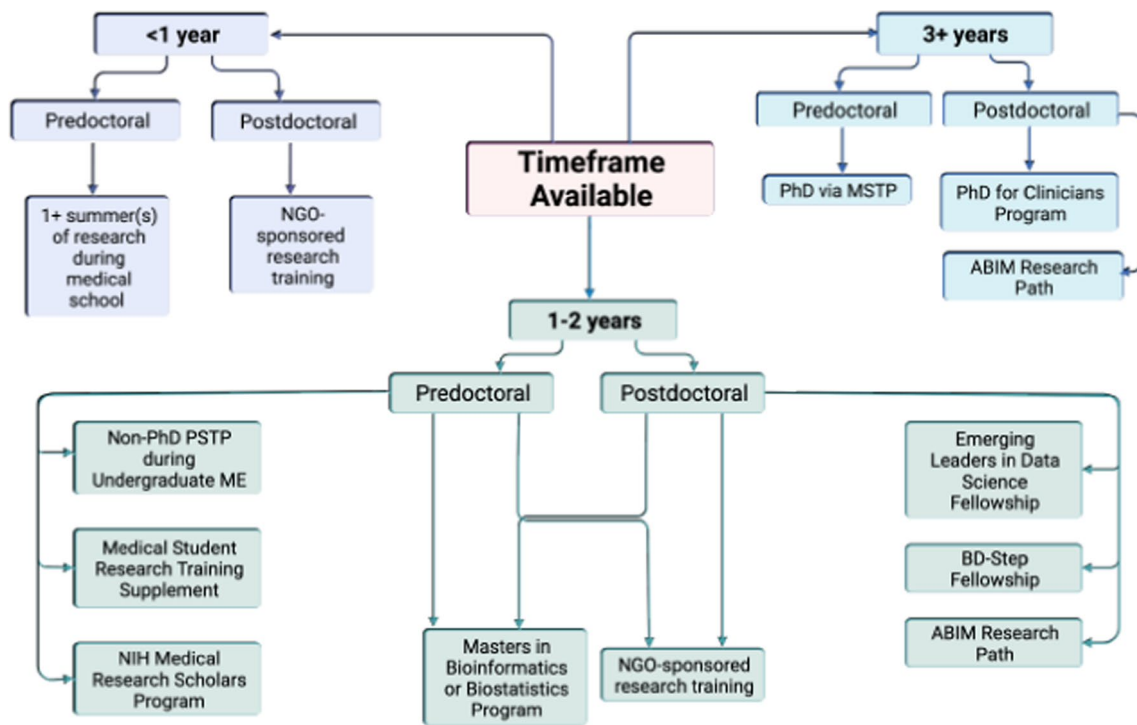


Fig. 1 Matching pre- and postdoctoral timeframes to opportunities for training in big data acquisition, analysis, and interpretation. The steep learning curve for proper big data analysis and interpretation can be mitigated by increasing the duration of training. Options to obtain even a rudimentary understanding of these methods are available

for those with limited time, i.e., less than 1 year. *ABIM* American Board of Internal Medicine, *BD-Step* Big Data Scientist Training Enhancement Program, *NGO* non-governmental organization, *ME* medical education, *MSTP* Medical Scientist Training Program, *PSTP* Physician Scientist Training Program. Created with Biorender.com

to highlight opportunities accessible to the larger group of MD-only aspiring physician-investigators to develop the skills needed to harness the research power of big data distinct from MD/PhD.

MD-only candidates with research-intensive career intentions have different research interests, practice plans, interests in advocacy and administration roles, time constraints, and foreseeable obstacles in career progression than MD/PhD or traditional MD candidates [2]. Support for their career development must be tailored accordingly, particularly in light of the declining percentages of R01-equivalent research grants awarded to MD-only physician-scientists [3], as well as the currently high attrition rate of physicians in academic medicine [4].

An option involving intensive and structured training in research at institutions with big data capabilities is “year-out” or “2-years-out” training programs for medical students, such as those offered by the NIH. Indeed, it is increasingly common for medical student interested in academic careers to take so-called ‘gap years’ to enhance their academic skills. Students typically apply for such training during their second year of medical school and, upon acceptance, take a leave of absence to complete research programs before starting the third year of medical school. The Medical Research Scholars

Program at the NIH offers accepted students a year-long research opportunity in a variety of areas, including big data generation and analysis, additionally providing networking opportunities, training in clinical protocol development, and exposure to emerging technologies. Alternatively, medical students with established research connections via a training grant (e.g., NIH T32 training program) can apply for “year-out” funding through a Medical Student Research Training Supplement offered by some NIH institutes (e.g., NIDDK).

Research support during medical school is also possible through programs like the 2-year Carolyn L. Kuckein Student Research Fellowship and non-governmental organizations (NGOs). Benefits of undertaking a “year-out” research program include focused enrichment of research techniques and skills without distractions that foster the analytical capabilities needed for impactful GI/liver-directed big data analysis. Nonetheless, these options require financial support for the training as well as the additional work required for successful research projects (e.g., time and resources required to satisfy manuscript revisions); the latter may pose substantial obstacles following the medical student’s return to clinical rotations. Another limiting consideration is the substantial debt owed by many students

Table 1 Overview of big data research training opportunities available at various stages of medical education and clinical training

Option	Training period	Phase of training	Examples of participating institutions
Biostatistics or bioinformatics PhD within MD/PhD programs	Typically, 4 years	Between 2nd and 3rd year of medical school	NYU, ISMMS, UMSOM, Columbia University, Harvard Medical School, University of Washington, University of Wisconsin-Madison, and others
Non-PhD-granting PSTP, as part of an MD program	Typically, 1–2 years	Between 2nd and 3rd year of medical school	University of Pittsburgh, Stanford University, Duke University, and others
NIH medical research scholars program	1 year	Typically, between 2nd and 3rd year of medical school	NIH
Medical student research training supplement	1 year	Typically, between 2nd and 3rd year of medical school	ISMMS, Washington University in St. Louis, UCSD, UMSOM, and others
Alternatively sourced funding year or 2 years out, such as NGO-sponsored research training, (e.g., Mach-Gaensslen Foundation, Canada) or Carolyn L. Kuckein Student Research Fellowship	Varies	Typically, the summer between 1st and 2nd year or year-out (or two) between 2nd and 3rd year of medical school; Postdoctoral opportunities also exist	
ABIM Research Pathway	3 years, with at least 80% research involvement in mentored	Integrated into residency and fellowship training	UMSOM, UCSF, BIDMC, JHU, University of Washington, Rutgers University, VUMC, and others
MS in bioinformatics or clinical research	1–2 years	Varies	Stanford University, Boston University, UMSOM, Penn State, UCSD, Harvard Medical School, Rush University, UVA, and others
PhD for clinicians	Varies	Postdoctoral	UMSOM, JHU, University of Wisconsin-Madison, University of Texas Southwestern, and others
BD-STEP fellowship	2 years	Postdoctoral; PhD and relevant research experience required	NCI/VHA
Emerging leaders in data science fellowship program	1–2 years	Postdoctoral; Bioinformatics experience required	NCI, NIAID

PSTP physician-scientist training program, *NGO* non-governmental organization, *BD-STEP* Big Data Scientist Training Enhancement Program, *ABIM* American Board of Internal Medicine, *NYU* New York University, *ISMMS* Icahn School of Medicine at Mount Sinai, *UMSOM* University of Maryland, School of Medicine, *UCSF* University of California, San Francisco, *VHA* Veterans Health Administration, *NIH* National Institutes of Health, *NCI* National Cancer Institute, *NIAID* National Institute of Allergy and Infectious Disease, *BIDMC* Beth Israel Deaconess Medical Center, *JHU* Johns Hopkins University, *VUMC* Vanderbilt University Medical Center, *UVA* University of Virginia, *UCSD* University of California, San Diego

who need to repay undergraduate- and graduate-level loans obtained to pay for their education.

Alternatively, some undergraduate medical education programs offer non-PhD-granting physician-scientist training programs (PSTPs), distinct from residency PSTPs, that provide dedicated time to gain research expertise in managing big datasets during undergraduate medical education. For example, the University of Pittsburgh School of Medicine offers a 5-year PSTP for aspiring MD-only clinical researchers, which aims to produce physician-scientist preclinical and translational researchers through career mentorship, PSTP-designed research courses, and one dedicated research year supported by institutional funding. The program's success is well documented; students and alumni of the PSTP had higher measures of research productivity and funding acquisition compared to classmates in the traditional MD-granting program [5]. At the Stanford University School of Medicine, all students are required to complete a “scholarly concentration” in an area of interest, traditionally spanning all years of medical school under the mentorship of a topic expert and culminating in a publishable product—to dedicate time to research, most students commit an additional year to medical school training. Likewise, at the Duke University School of Medicine, 9 months of the standard curriculum are devoted to an intensive research program requiring completion of a formal thesis.

Requiring a more modest time commitment, some medical school curricula offer “tracks” or “concentrations” in bioinformatics, which provide introductory but insufficient exposure for dedicated research or training in the use of big data. For example, the scholarly concentration in Biomedical Informatics at the Brown University Warren Alpert Medical School provides foundational curricula through a 3-week summer course and workshops; similar programs exist at several medical schools (UCLA David Geffen, the University of Michigan, the University of Texas Medical Branch at Galveston, and others).

A prominent talent development pathway available at the resident level is the American Board of Internal Medicine (ABIM) Research Pathway, a program available at many top-tier medical research institutions that integrates internal medicine residency and subspecialty fellowship training in GI or hepatology with several years of mentored postdoctoral research. This allows candidates committed to academic research careers to gain valuable experience while “fast-tracking” into their intended subspecialty. During the three dedicated research years afforded by the ABIM Research Pathway for those planning to subspecialize in GI or hepatology, trainees can collaborate with and learn from investigators experienced in big data and -omics methodologies.

Unfortunately, postdoctoral fellowship options catering specifically to physicians who recently finished GI fellowship training are scarce. General postdoctoral fellowships aimed at big data training tend to accept applicants who already possess substantial experience in computer science and/or bioinformatics. For example, the NIH offers Data Science Fellowship opportunities through the National Cancer Institute (NCI) and the National Institute of Allergy and Infectious Disease (NIAID), each provide training and hand-on experience in managing big datasets. The Big Data Scientist Training Enhancement Program (BD-STEP) is another valuable opportunity provided by NCI in partnership with the Veterans Health Administration to train postdoctoral fellows in data science processing while addressing patient health challenges, but this is currently available only to MD/PhDs or PhDs with relevant experience.

An additional complementary approach is to incorporate big data analysis into the medical school curriculum and subsequent fellowship training by means of expert presentation at divisional Grand Rounds and workshops and conferences sponsored by the American Gastroenterological Association (AGA, e.g., the AGA Academic Skills Workshop). At a minimum, the information offered must contain education on big data-type terminology, applications in healthcare and disease surveillance, available open-source big data platforms and utilities, and review recent big data research in the field.

Conclusions and Future Directions

Although there is much room to improve training opportunities, several avenues and pathways currently exist that can and should be leveraged by those interested in incorporating big data analysis to their work. These include tracks which may be pursued at the undergraduate medical level, such as pursuing an MD/PhD in bioinformatics/statistics, enrolling in a research-intensive non-PhD program or PSTP, and taking advantage of gap-year opportunities, such as the Medical Student Research Training Supplement or NIH Medical Research Scholars Program. During post-graduate clinical training, trainees can pursue the ABIM Research Pathway, obtain a masters level degree in bioinformatics or clinical research, or a more time-intensive PhD degree. MA, the lead author of the current manuscript, has witnessed firsthand how inadequate physician training in big data slows the advancement of medicine. She is currently leveraging the MSTP at the University of Maryland School of Medicine to complete the requirements for a PhD degree with a strong emphasis on multi-omics integration and interpretation.

Undoubtedly, including such instruction will contribute to the progressive expansion of the duration of scientific training; the average junior scientist can expect to receive

initial independent research funding (NIH R01 award or equivalent) no earlier than in their mid-forties [6]. To maximize return on this financial and time investment, this increased duration of training must, by necessity, result in a proportional increase in productivity and professional lifespan. Indeed, the former concept of retirement at or about age 65 years has been laid to rest by the observation that many in the scientific workforce, taking advantage of advances in human health and functional lifespan, work well past age 70 years. Rather than regard an aging scientific workforce as competitors for funding and research positions that might otherwise go to junior scientists, we urge our colleagues to view this as an opportunity to take the time to both broaden and deepen training in big data research and, therefore, expand our scientific workforce and capabilities to add important new knowledge while at the same time bolstering research rigor and reproducibility.

Key Messages

- Training GI and liver investigators in the appropriate application of big data in their work has become increasingly important, if not vital, for a successful academic career.
- Several training pathways that vary in intensity and duration are available for those wishing to acquire big data research skills.

Funding M.A. and N.S.M. were supported by an award from the National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases (T32 DK067872; J-P Raufman, PI).

Declarations

Conflict of interest The authors have no conflicts to disclose.

References

1. Alizadeh M, SampaioMoura N, Schledwitz A, Patil SA, Ravel J, Raufman JP. Big data in gastroenterology research. *Int J Mol Sci.* 2023;24:2458.
2. Kwan JM, Daye D, Schmidt ML et al. Exploring intentions of physician-scientist trainees: factors influencing MD and MD/PhD interest in research careers. *BMC Med Educ.* 2017;17:115.
3. Garrison HH, Deschamps AM. NIH research funding and early career physician scientists: continuing challenges in the 21st century. *Faseb J.* 2014;28:1049–1058.
4. Alexander H. The long-term retention and attrition of US medical school faculty. *AAMC Anal Brief.* 2008;8:1.
5. Steinman RA, Proulx CN, Levine AS. The highly structured Physician Scientist Training Program (PSTP) for Medical Students at the University of Pittsburgh. *Acad Med.* 2020;95:1373–1381.
6. Lauer M. Long-term trends in the age of principal investigators supported for the first time on NIH R01-equivalent awards. City; 2021.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.