The Duke Health Data Science Internship Program: Integrating the Educational Mission into Real-World Research

Shelley Rusincovitch, MMCi1, Lisa Wruck, PhD2, Ricardo Henao, PhD1,2,3, Larisa Rodgers, MSW2, Allison Dunning, MS2, Peter Merrill, PhD2, Hillary Mulder, MS2, Robert Overton, MS2, Matthew Phelan, MS2, Erich Huang, MD, PhD1,3, Lawrence Carin, PhD2,4, Michael Pencina, PhD2,3

1Duke Forge; 2Duke Clinical Research Institute; 3Duke Department of Biostatistics and Bioinformatics; 4Duke University; Durham, NC

Introduction

Data science has great potential to impact and improve human health. The Duke Forge Demonstration Program was developed to “demonstrate the art of the possible” with vanguard projects in health data science, convening transdisciplinary teams including clinician leaders, quantitative experts, operations personnel, administrators, and informaticists. We designed the Duke Health Data Science Internship Program to incorporate a strong educational mission into the demonstration program and to foster the hands-on effort of quantitative students.

Contact Information
Shelley Rusincovitch, MMCi
Associate Director of Operations
Duke Forge (Center for Health Data Science)
shelley.rusincovitch@duke.edu
@rusincovitch

Design of the Program

Duke graduate students pursuing master’s-level studies in any quantitative discipline are eligible to apply to the program. Two unique aspects of the internship program are:

1. A strong tie to the demonstration program, allowing us to embed students into real, hands-on experiences
2. The matrixed organizational structure

Discussion

Since the internship program launched in May 2017, we have accepted two cohorts of students: the 2017–2018 cohort (6 students), and the current 2018–2019 cohort (5 students). We have found few details in the literature about the tactical approach of other data science internship programs. In this poster we share our experiences in the hope of initiating further discussion, especially with other CTSA institutions committed to integration of the educational mission with cutting edge health data science.

Acknowledgements

The Health Data Science Internship Program is supported by the Duke Clinical Research Institute and Duke Forge. Health Data Science at Duke is supported by the Center For Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR002553. We gratefully acknowledge and thank the students who have been the heart of our journey together.

Table 1. Components of the Duke Health Data Science Internship Program.

<table>
<thead>
<tr>
<th>Table 1. Components of the Duke Health Data Science Internship Program.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Matriced Organizational Structure</strong></td>
</tr>
<tr>
<td><strong>Program Processes</strong></td>
</tr>
<tr>
<td><strong>Technical and Regulatory Infrastructure</strong></td>
</tr>
<tr>
<td><strong>Cadence</strong></td>
</tr>
<tr>
<td><strong>Project Selection</strong></td>
</tr>
</tbody>
</table>

Table 2. Learning objectives for students in the Duke Health Data Science Internship Program.

1. **Comprehend the big picture clinical question.** Interns will demonstrate they understand the translation of a clinical question or problem into an analyzable scope and be able to articulate the objectives of the project clearly.
2. **Understand the data, including preparation and processing stages.** They will demonstrate a thorough understanding of the end-to-end process of working with and preparing the data. This includes consideration of the underlying population and data generation process; understanding potential sources of bias and confounders; and assessment of data quality and cleanliness.
3. **Apply appropriate methodologies for data exploration, description, analysis, and modeling.** Interns will demonstrate comprehension of the methods applied, including assumptions and limitations of a given approach.
4. **Ensure reproducibility of their results through good documentation and code provenance.** They will carefully manage their code and develop accurate descriptions of what they are doing, and why, in a way that others can understand.
5. **Use best practices in programming.** Interns will demonstrate tool-specific skills and apply common software development practices such as modularizing code, use of code repositories, and version control.
6. **Engage with their teams at a professional level and develop collegial relationships.** They will engage as an active contributor with their transdisciplinary team, communicate clearly and regularly, bring issues forward, and work with the team in developing solutions.
7. **Tailor their communications to the appropriate audience, including clinician and quantitative team members.** Beyond giving a static presentation, interns will have the ability to listen to others and the confidence to speak up professionally. They will demonstrate the ability to discuss and converse in many situations with many audiences, and tailor their use of language and concepts to the people they are engaging with.
8. **Cultivate self-awareness and use knowledge of their strengths and weaknesses to guide continuous learning.** They will be able to evaluate areas where they don’t have adequate knowledge and be able to discuss and actively develop this understanding.