Photo credit: Dr. Gilda Jones, CDC

The Microbiome Quality Control Project:
Implications for Reproducibility in Microbiome Science

David Clausen
Graduate Student
Department of Biostatistics
School of Public Health
University of Washington

In collaboration with

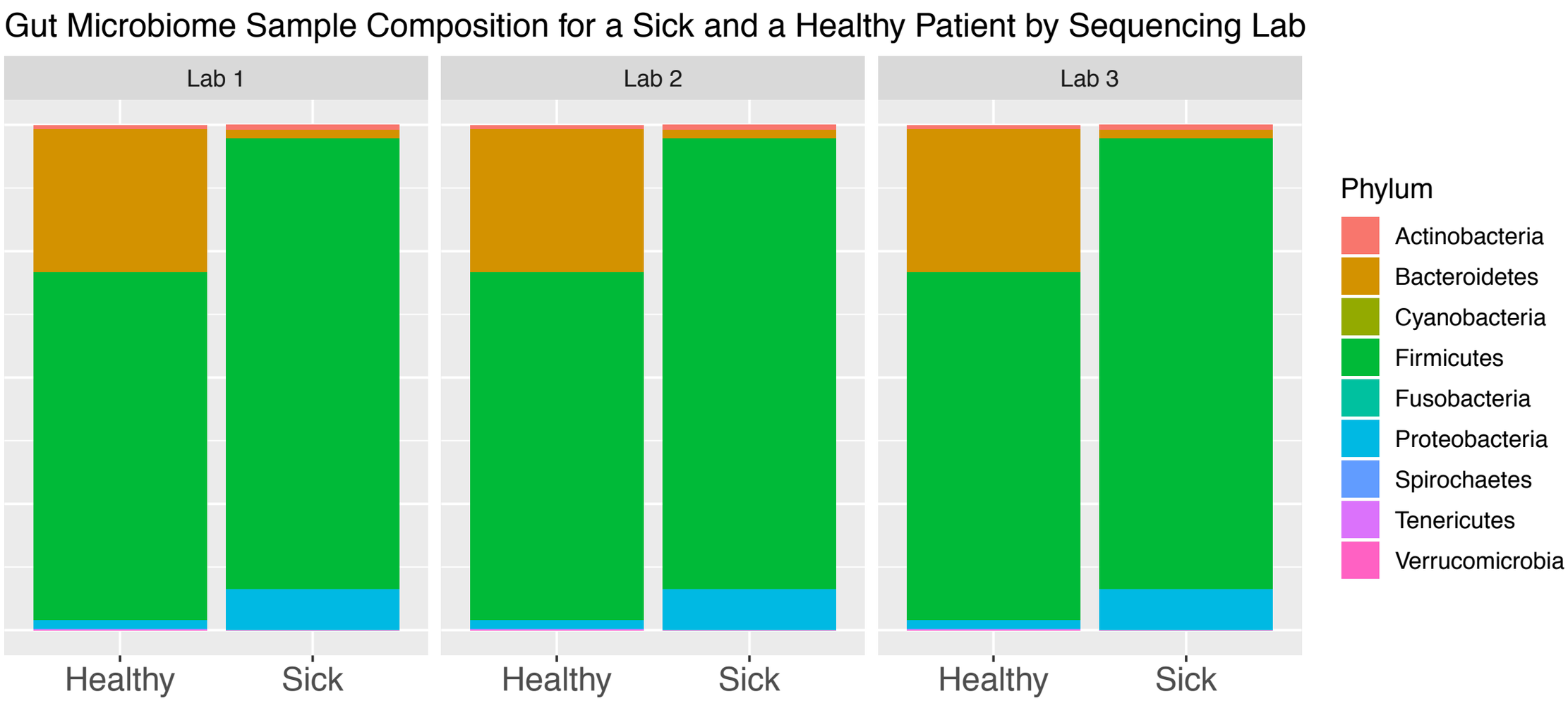
Amy Willis, PhD
Assistant Professor
Department of Biostatistics
School of Public Health
University of Washington

Reproducibility: A Refresher

- What it is
  - The ability of researchers working independently to repeat each other's experiments and reach similar conclusions
- Why you should care about it
  - Objective truth
  - Time and money
  - Avoiding frustration and embarrassment

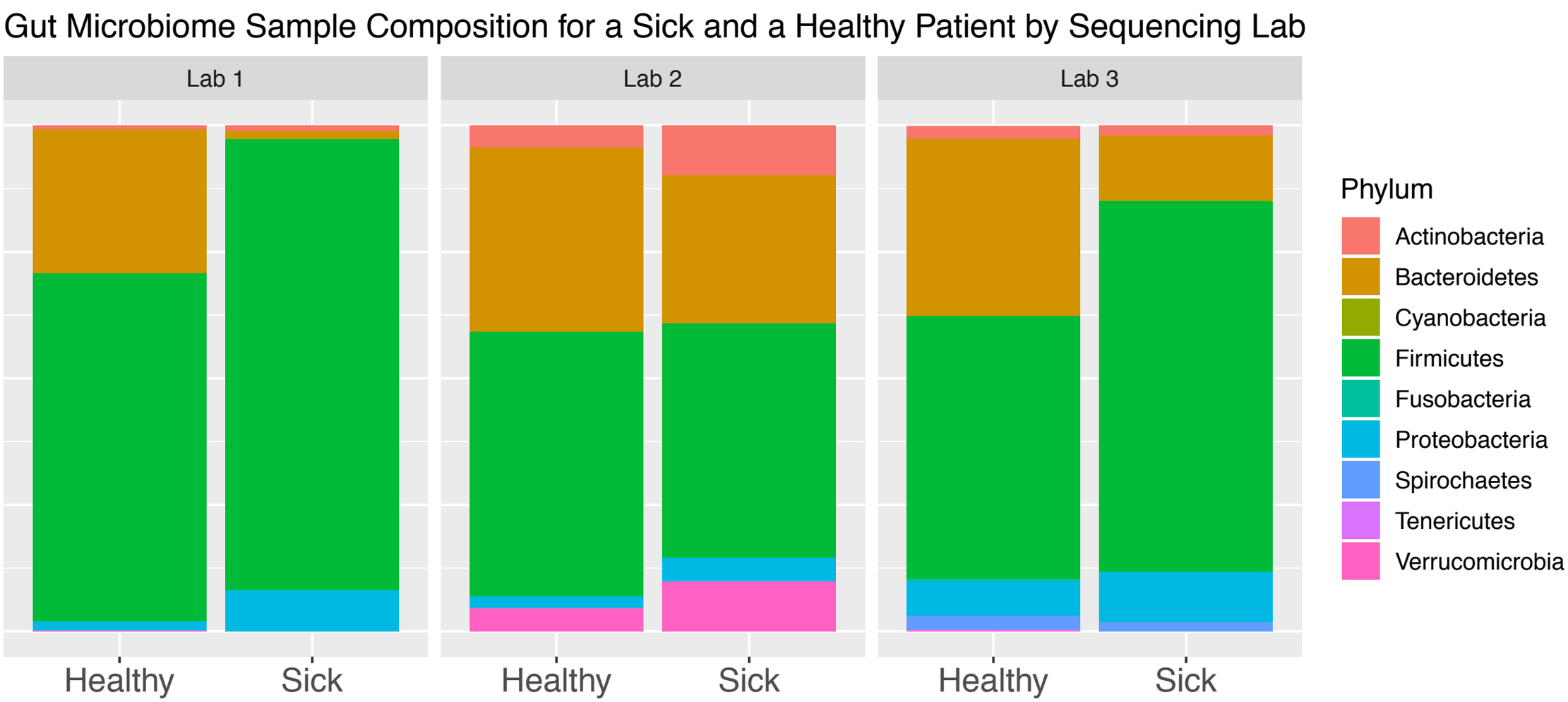
Slides: github.com/ailurophilia

What Does Reproducible Data Look Like?



Slides: github.com/ailurophilia

Here’s What Data Actually Looks Like



Slides: github.com/ailurophilia

Does Measurement Error Affect Results?

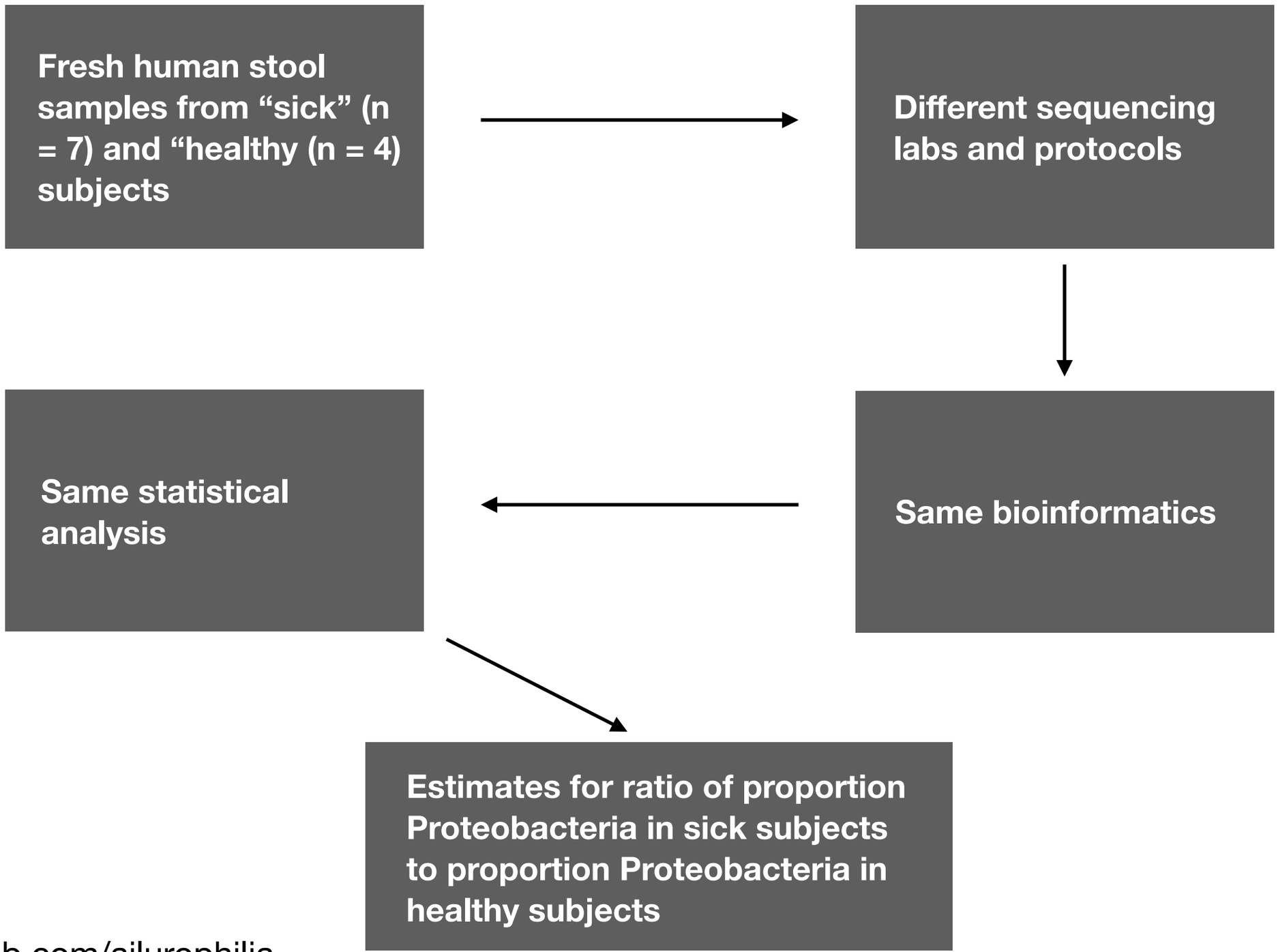
- Assuming no measurement error is a model assumption
- Significant results may be due to incorrect modeling assumptions

MBQC: An Opportunity to Quantify How Measurement Error Affects Results

- The Microbiome Quality Control Project (MBQC)
  - Identical samples sent to different sequencing labs
  - OTU count data available online (www.mbqc.org)
  - Only 16S
    - But similar patterns may hold for other technologies

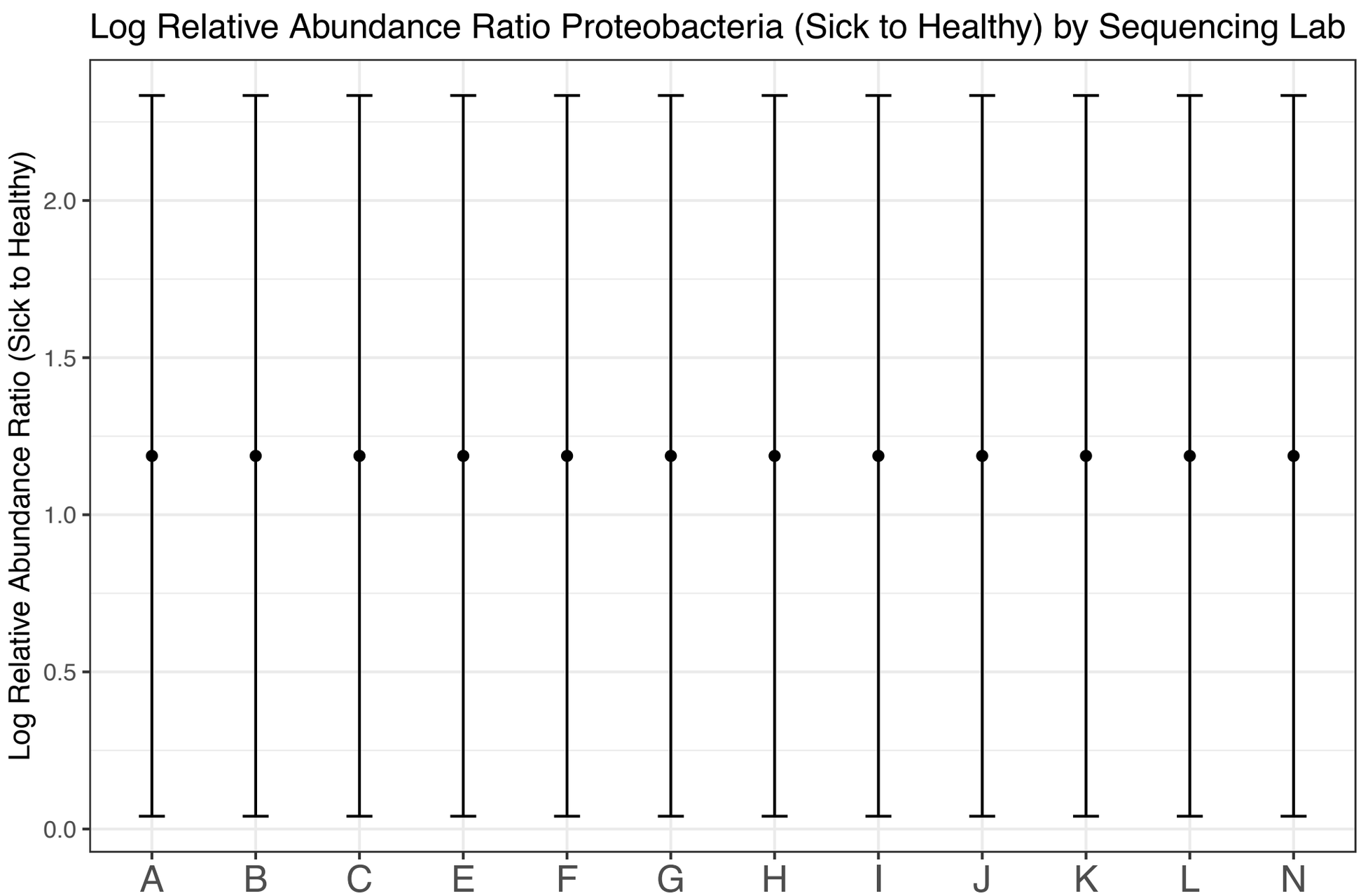
Slides: github.com/ailurophilia

MBQC Data: Does Measurement Error Affect Inference at the Phylum Level?



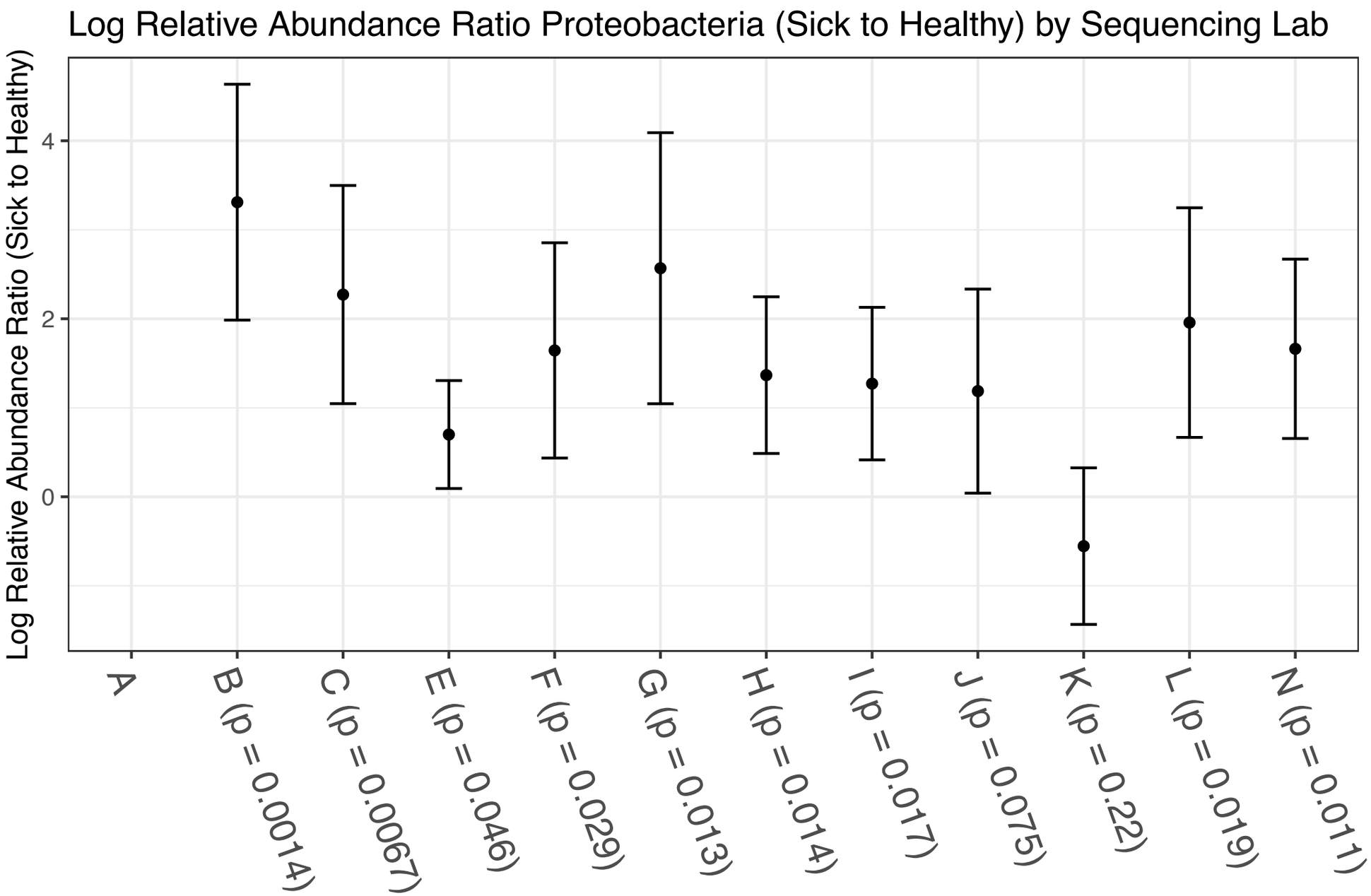
Slides: github.com/ailurophilia

Results: What Would No Lab-Induced Variability Look Like?



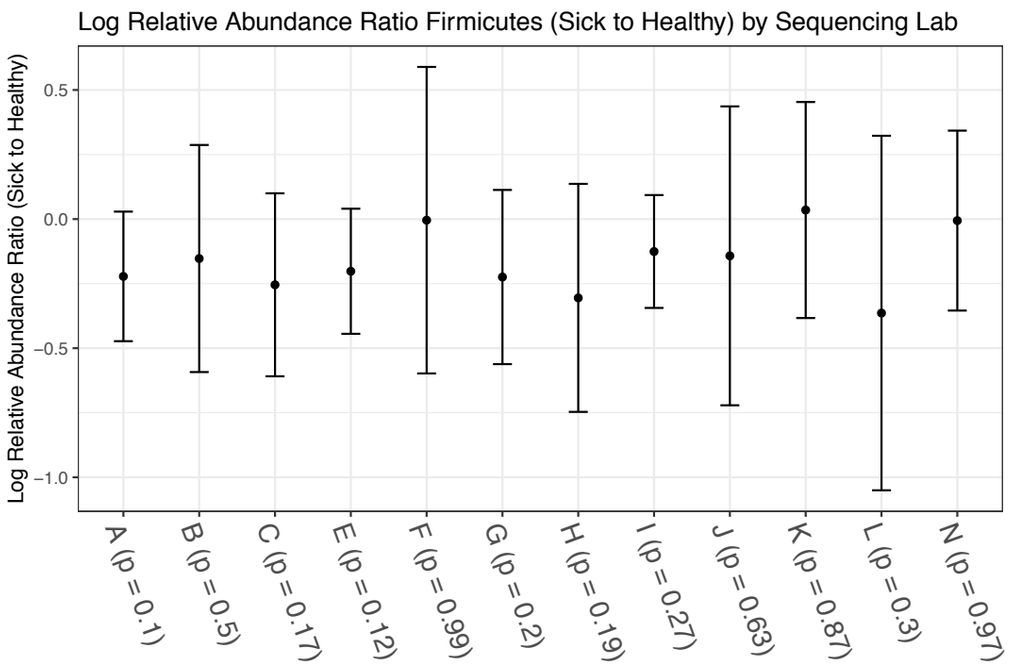
Slides: github.com/ailurophilia

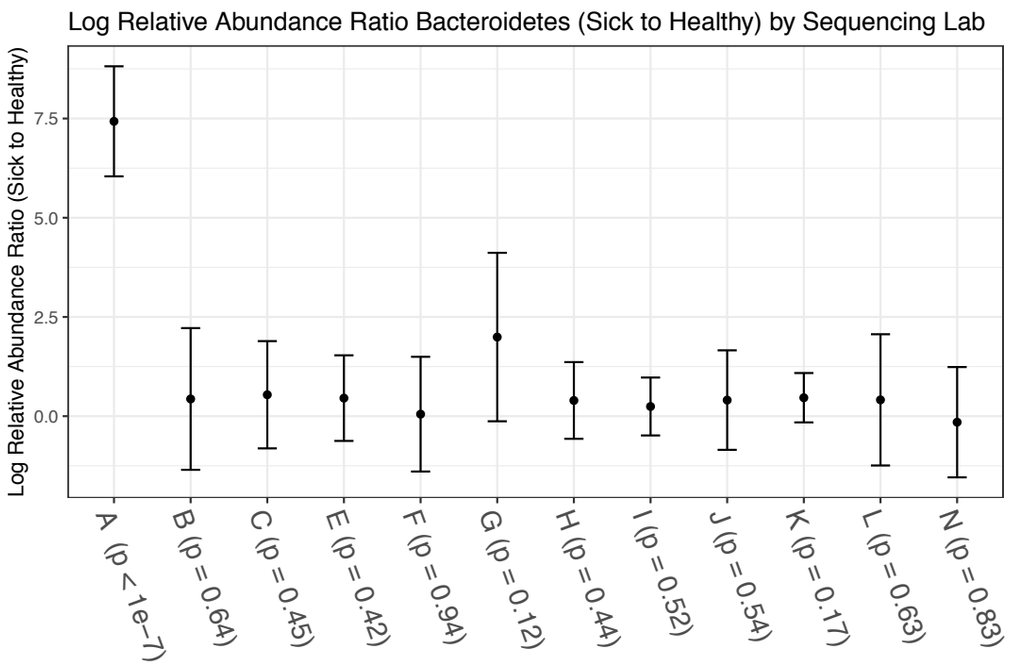
Results: What Does Real Data Look Like?

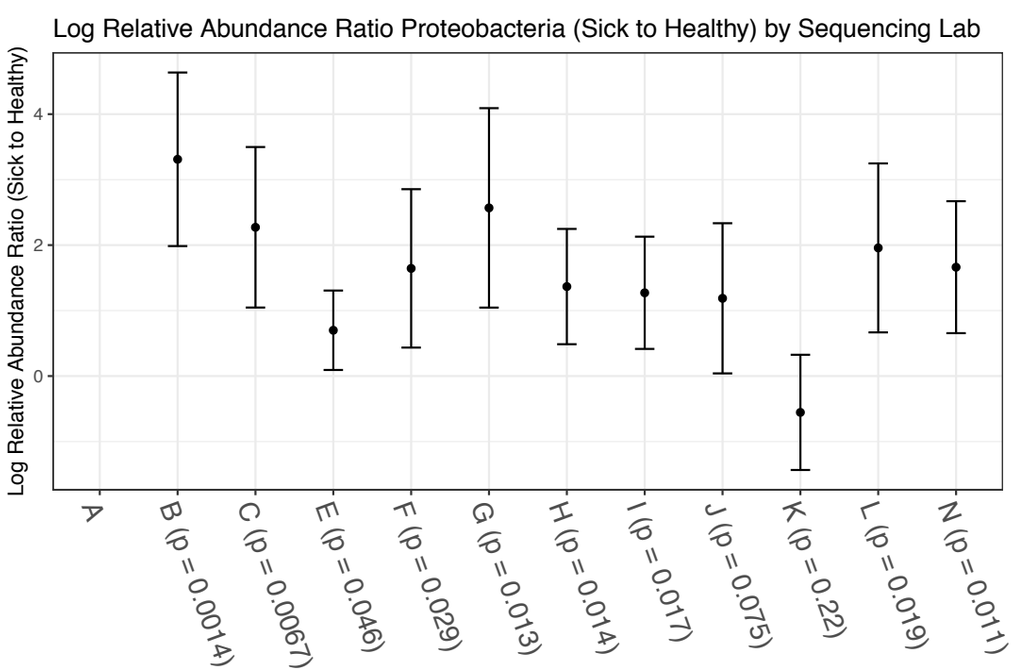


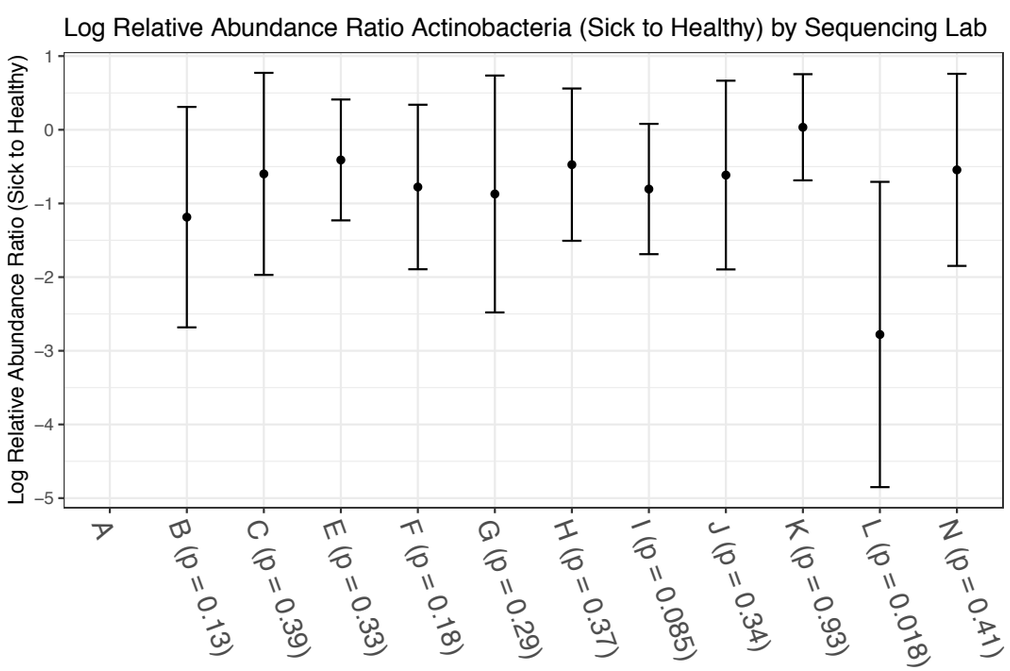
Slides: github.com/ailurophilia

What About Other Phyla?



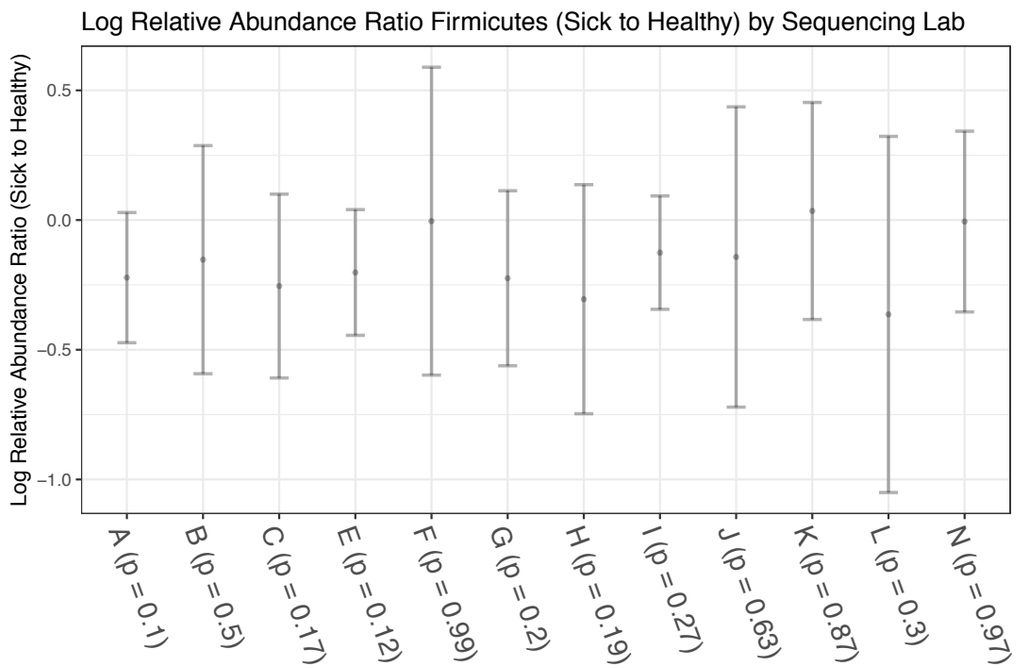


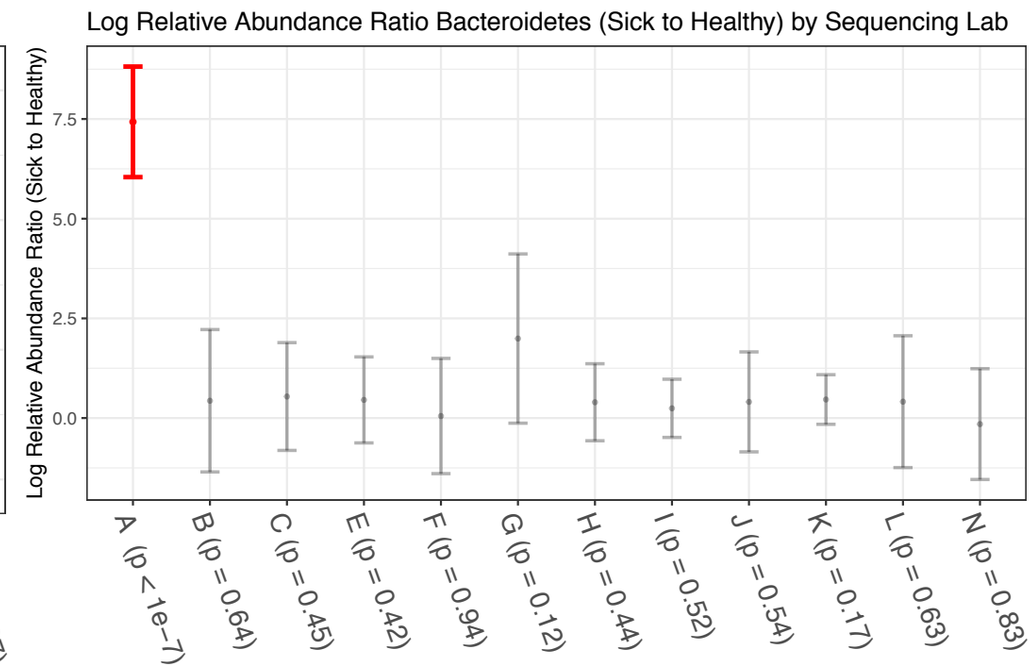


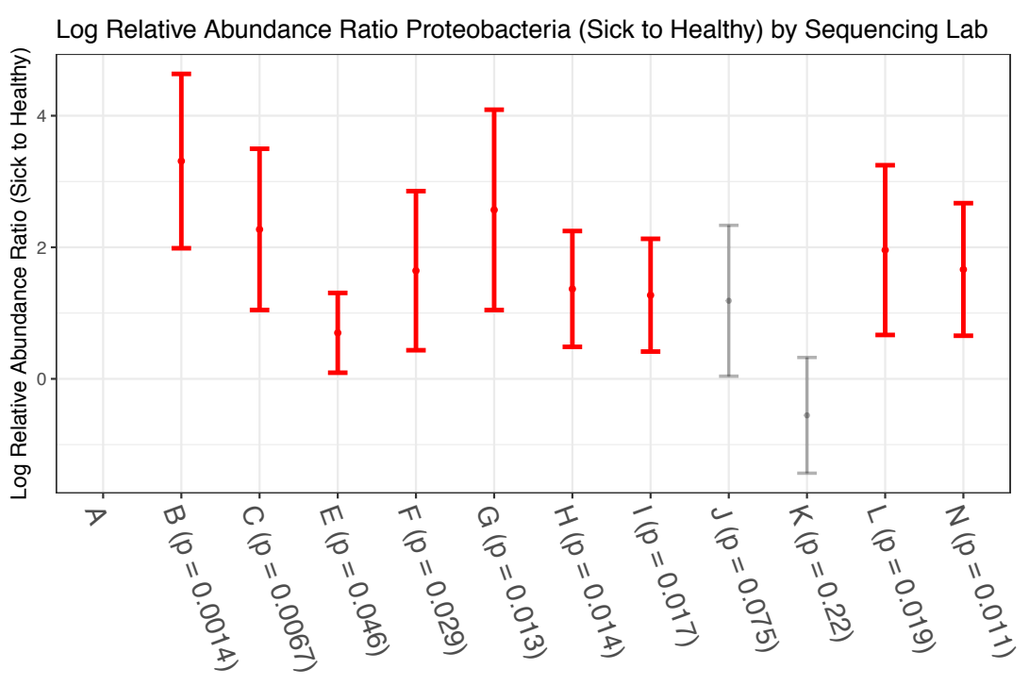


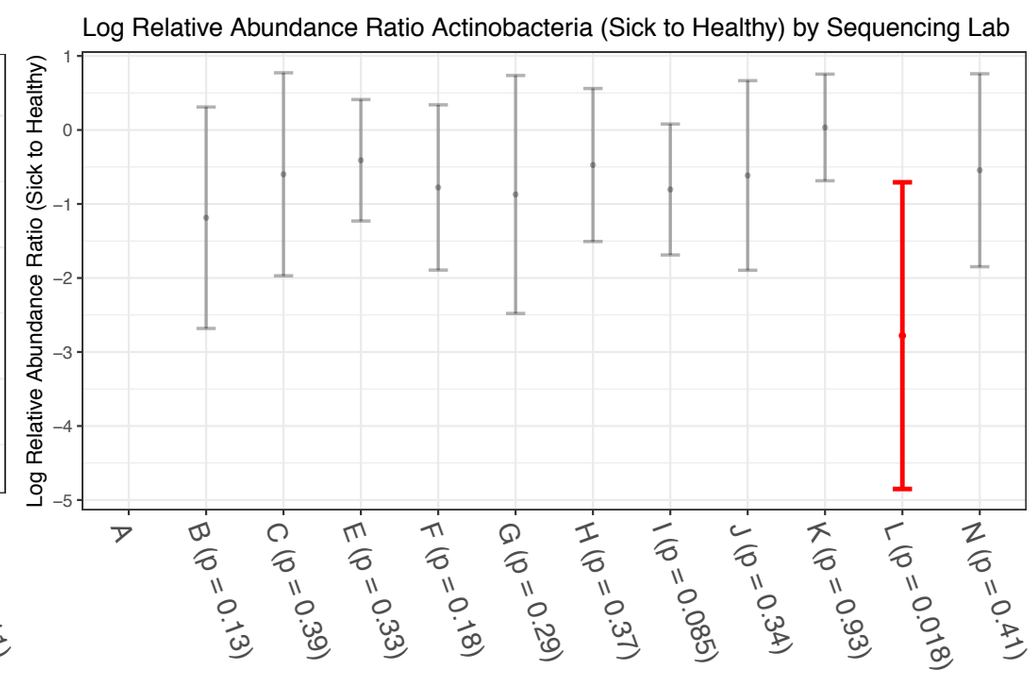
Slides: github.com/ailurophilia

Who’s Getting Published?





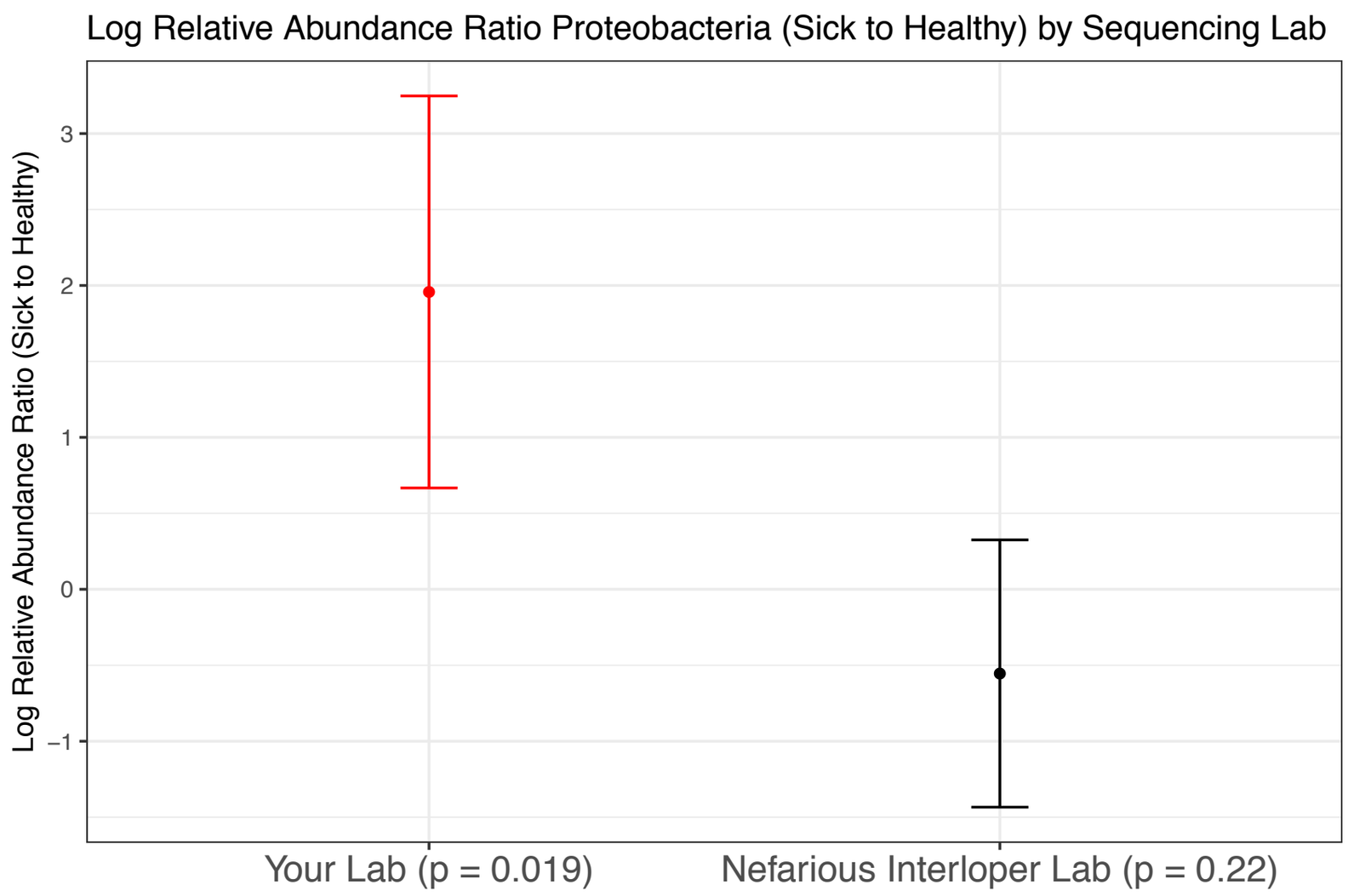




Slides: github.com/ailurophilia

“Okay, But My Lab Has Good Protocols!”

- That’s great!
  - … until another lab tries to reproduce your results



Slides: github.com/ailurophilia

What You Can Do

- Be aware of these problems at every step of research
  - writing and reviewing grants
  - reviewing articles
  - planning, conducting, and reporting your own research
    - take replicates
    - use controls
- Engage in sustained collaboration with statisticians

Slides: github.com/ailurophilia

Further Research

- Addressing between-lab variability not explained by protocol variables
- Developing statistical tools to make replicates and controls more useful

Slides: github.com/ailurophilia

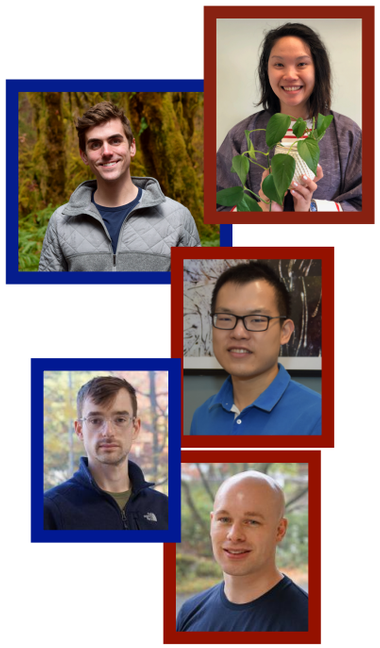
Thank You

Further Resources

- UW Statistics & Biostatistics Consulting
  - https://www.stat.washington.edu/consulting/
- Statistical Diversity Lab
  - adwillis@uw.edu
  - dsc24@uw.edu

statistical diversity lab

Prof. Amy Willis, PhD fearless leader



Pauline Trinh — bioinformatics, human-animal-microbe interactions

Bryan Martin — relative abundance, compositional data

Kendrick Li — networks, taxonomy

David Clausen — replication and data quality in microbiome science

Alex Paynter — diversity and regularised ML

Slides: github.com/ailurophilia