Being Research Active in Teaching-Focused Colleges

August 2, 2017
The Joint Statistical Meetings
Baltimore MD

Amy Wagaman, Ming-Wen An, Paul Roback, Steve Wang, Gideon Zamba
Chair: Jo Hardin
Panelists

- Amy Wagaman (Amherst College)
- Ming-Wen An (Vassar College)
- Paul Roback (St. Olaf College)
- Steve Wang (Swarthmore College)
- Gideon Zamba (University of Iowa)
- Jo Hardin, Chair (Pomona College)
Be efficient with your time

- Teaching, service, and research expectations differ by institution, but you expect to have some research expectation
- How many of us get as much research done as we’d like in a year?

- Where does our time go?
  - Teaching - including class prep, and student demands on time - office hours, individual appts
  - Meetings for service
  - Administrative work
- What’s the first thing to go off the calendar when the semester gets busy?
Align your Time Spent with Your Priorities

If you need to meet a research expectation, you need to allocate appropriate TIME for that.

As an example, imagine you are thinking research 30%, teaching 60%, and service 10% while the semester is in session.

Now think about your calendar. Is 30% of your time spent on research?

What can you do to carve out time for research and keep it a priority?
Useful Skills/Mechanisms

- National Center for Faculty Diversity and Development (NCFDD) has a Faculty Success Program (FSP) that runs every semester and in the summer.
  - Check to see if your institution is a member!
- Different skills/mechanisms work better for different individuals
- Briefly, I’ll highlight four skills/mechanisms that work well for me:
  - Daily writing (or research) time
  - SMART goals
  - Weekly plan (Semester and summer plans too!)
  - Accountability group
Daily Writing (or Research) Time

- Little by little, you can get research done!
- Research won’t fall by the wayside if you spend a little bit of time every day doing something related to it.
- Examples:
  - Read one paper you want to in your field.
  - Work on one figure for an upcoming publication
  - Edit one section (but don’t edit endlessly!) of a paper
  - Check over code for a simulation
  - Enter a table into your draft
- How much time? 30 minutes to an hour is enough to see progress
- If you only have 15 minutes, work on a literature review or enter one item into a bibliography!
Tracking Daily Writing\Research Time

● Lots of online tracking systems exist.
  ○ Some are parts of programs and allow community engagement
  ○ Some are FREE for individual use (a Google search brought up several, but I haven’t tried any of those)
● Make a Google Form for yourself and check your progress every day/week.
● Mark it on your calendar.
SMART Goals

● SMART Goals are...
  ○ Specific
  ○ Measureable
  ○ Achievable
  ○ Realistic
  ○ Time-based

● Examples of SMART goals:
  ○ Create a caption for Figure 3
  ○ Write 2 paragraphs of my discussion
  ○ Look up 4 references for the introduction
  ○ Send draft to a collaborator

● Examples of non-SMART goals:
  ○ Write the entire introduction (worse: write entire paper)
  ○ Complete all figures, etc.

● Why are these useful? You can complete them in a reasonable time frame and feel accomplished that you met your goals! And, you make progress!
Weekly Plan

- List all your tasks for the week - research, teaching, and service
- Sketch out a plan to accomplish them all within the week
- Priorities! Organize carefully.
- Be realistic with your time.
- Many ways to do this: daily task list, calendar, etc.
- Creating the weekly plan takes about 30 minutes.
- I’ve learned that it can be helpful to schedule email checking IN your plan. You spend more time on email than you think you do!
Accountability Groups

● A variety of accountability mechanisms are available:
  ○ Buddy system - just one person you check in with (could be monthly, weekly, daily, before/after some set time period)
  ○ An online/not-in person group - my small group from the FSP (years ago!) still calls to check in every week
  ○ Individuals do not need to be at your institution - it can help if they are NOT at your institution

● What do you check in about?
  ○ Whatever you need to (and what is appropriate for the group)
  ○ Give your goals for the week; report on how last week went
  ○ Ask for advice related to some issue you are facing (navigating a grant, tenure, etc.)

● Takes < 1 hour of time a week
Partner with research institutions

- Build networks as much as possible, keeping in mind previous connections
- Accept invitations (or Offer) to give talks at conferences and in departmental seminars
- Seek existing programs or fellowships
- Approach larger institutions in your geographic area for possible collaboration
Partner with research institutions

- Build networks as much as possible, keeping in mind previous connections
- Accept invitations (or Offer) to give talks at conferences and in departmental seminars
- Seek existing programs or fellowships
- Approach larger institutions in your geographic area for possible collaboration
Partner with research institutions

- Build networks as much as possible, keeping in mind previous connections
- Accept invitations (or Offer) to give talks at conferences and in departmental seminars
- Seek existing programs or fellowships
- Approach larger institutions in your geographic area for possible collaboration
Partner with research institutions

- Build networks as much as possible, keeping in mind previous connections
- Accept invitations (or Offer) to give talks at conferences and in departmental seminars
- Seek existing programs or fellowships
- Approach larger institutions in your geographic area for possible collaboration
Partner with research institutions

- Build networks as much as possible, keeping in mind previous connections
- Accept invitations (or Offer) to give talks at conferences and in departmental seminars
- Seek existing programs or fellowships
- Approach larger institutions in your geographic area for possible collaboration
Apply for funding

- **Why?** Funds travel to conferences and research sites, student research, summer salary; requires you to write a research proposal
- **Institutional funding opportunities**
- **Federal funding opportunities**
  - e.g. NIH (R21, R15), NSF (REU)
  - Reach out to program officers early on - they are here to help!
  - As much as possible, include collaborators on your grant
- **ASA Committee on Funded Research**
  - Join the *Funding Opportunities* ASA Community Group
  - Check [https://www.amstat.org/ASA/Your-Career/External-Funding-Sources.aspx](https://www.amstat.org/ASA/Your-Career/External-Funding-Sources.aspx)
- **Funding Recommendation Resource (new and free!):**
Take a sabbatical

- If your institution offers sabbaticals, take them! (Seems obvious, but not always)
- Plan and prepare for your sabbatical
- If possible, arrange to spend part of your sabbatical at another institution.
  - Physical separation from home institution reduces “distractions”
  - Might help with developing new collaborations
Paul Roback
St. Olaf College
Collaborate with faculty at your institution

Judicious Use of Multiple Hypothesis Tests

PAUL J. ROBACK* AND ROBERT A. ASKINS††
*Department of Mathematics, Statistics, and Computer Science, Saint Olaf College, 1520 Saint Olaf Avenue, Northfield, MN 55057, U.S.A.
†Department of Biology, Connecticut College, 270 Mohegan Avenue, New London, CT 06320–4196, U.S.A.,
email raask@conncoll.edu
Add undergraduates to the mix

- Holy Grail of Undergraduate Research:
  Interdisciplinary collaboration with faculty colleagues and student co-authors in a peer-reviewed journal

- Our approach at St. Olaf College:
  Conduct research projects through the Center for Interdisciplinary Research

![Center for Interdisciplinary Research](CIR.png)
Center for Interdisciplinary Research (CIR): Overview

- Year-long CIR Undergraduate Fellowships, awarded on a competitive basis to qualified students (~18-24 per year)
- Interdisciplinary research teams - typically 3 students with statistics mentor and domain expert
- Weekly student seminar series on consulting, communication, team collaboration, ethics, and stat careers
- Weekly office hours: drop-in consultation services


NSF DMS-0354308 and DMS-1045015
Center for Interdisciplinary Research: Why this model?

- Statistics faculty can consolidate efforts
  - Mentoring multiple students in research projects
  - Satisfying campus need for statistical assistance, both larger collaborations and smaller consulting questions
  - Training students as researchers and collaborators
- Year-long format allows time for subject-area research, exploring new methods, communicating results

Some projects just end, but others continue to another year and/or lead to publications and/or lead to grant proposals

But... don’t need CIR framework to form collaborations with students and faculty colleagues or campus offices or outside organizations
A predicted operon map for *Mycobacterium tuberculosis*

P. Roback¹, J. Beard², D. Baumann², C. Gille³, K. Henry², S. Krohn², H. Wiste², M.I. Voskuil¹, C. Rainville⁵ and R. Rutherford⁵,⁎

¹Department of Mathematics, Statistics and Computer Science, ²Center for Interdisciplinary Research and ³Department of Biology St. Olaf College, Northfield MN 55057, ⁴Department of Microbiology, University of Colorado Health Sciences Center, Aurora, CO 80045 and ⁵Department of Biology, Seattle University, 901 12th Ave, Seattle, WA 98052, USA

Received October 18, 2006; Revised June 18, 2007; Accepted June 19, 2007

ABSTRACT

The prediction of operons in *Mycobacterium tuberculosis* (MTB) is a first step toward understanding the regulatory network of this pathogen. Here we apply a statistical model using logistic regression to predict operons in MTB. As predictors, our model incorporates intergenic distance and the correlation of gene expression calculated for adjacent gene pairs from over 474 microarray experiments with MTB RNA. We validate our findings with known examples from the literature and experimentation. From this model, we rank each potential operon pair by the strength of evidence for cotranscription, choose a classification threshold with a true positive rate of over 90% at a false positive rate of 9.1%, and use it to construct an operon map for the MTB genome.

Studies of MTB to date suggest that transcription is as complex and varied as it is in other prokaryotes. For example, the genome shows approximately 190 putative transcriptional regulators. Even in the best studied class, MTB’s 13 different sigma factors, complexity and unanswered questions are common (2,3) to date only five (SigA, SigC, SigE, SigF and SigH) have a defined putative promoter consensus sequence (1). In addition, the genome contains at least five anti-sigma factors, each of which conduct post-translational regulation of one or more sigma factors, and seven genes encoding anti-anti-sigma factors. Furthermore, the examples of transcripts studied to date demonstrate that even within a single well-characterized operon, transcriptional regulation can be complex in MTB: alternative internal promoters and competing promoters on the opposite strand have been identified (4); single genes may be regulated by multiple promoters (5); and, as in other prokaryotes, supercoiling plays a role in gene expression (6). Finally, little is known about transcription termination in MTB.
Sample CIR Projects: Going Vague

Going Vague: Ambiguity and Avoidance on House Candidate Websites

Submitted to: Political Behavior, July 2017

Authors: Chris Chapp, Paul Roback, Kendra Johnson-Tesch, Adrian Rossing, Jack Werner

ABSTRACT: While unambiguous communication of issues positions is often seen as a critical component of electoral campaigns, previous research has been offered limited empirical testing of the conditions that promote clear communication. We address this limitation, scoring website issues pages from 2014 House candidates. We accomplish this by using a supervised learning procedure that allows us to score a large corpus of text based on a smaller subset of hand-coded text. Second, while ambiguity research has focused on the ideological distance between a candidate and the median voter, we conclude that the district’s dispersion is ultimately just as important.
Collaborations with students and faculty colleagues: issues to consider

- CIR start-up involved “pro-bono” work until internal support materialized (e.g. floating research credit; credit for teaching seminar)
- Some funding support helpful – travel, speakers, etc.
- Must be willing to develop new areas of “expertise”
- Continuing writing process after graduation can be tricky (GitHub)
- Students may begin with limited experience in messy data, advanced modeling, and small group communication
- We have become better at identifying good projects (investment = cup of coffee plus one hour)
- Not every project leads to publication (must most well worth it!)
Steve Wang
Swarthmore College
Choose your research area wisely
Estimating the diversity of dinosaurs

Steve C. Wang*† and Peter Dodson†

*Department of Mathematics and Statistics, Swarthmore College, 500 College Avenue, Swarthmore, PA 19081; and †School of Veterinary Medicine and Department of Earth and Environmental Science, University of Pennsylvania, 3800 Spruce Street, Philadelphia, PA 19104

Communicated by David M. Raup, University of Chicago, Chicago, IL, July 19, 2006 (received for review June 14, 2006)

Despite current interest in estimating the diversity of fossil and extant groups, little effort has been devoted to estimating the diversity of dinosaurs. Here we estimate the diversity of nonavian dinosaurs at ~1,850 genera, including those that remain to be discovered. With 527 genera currently described, at least 71% of dinosaur genera thus remain unknown. Although known diversity declined in the last stage of the Cretaceous, estimated diversity was steady, suggesting that dinosaurs as a whole were not in decline in the 10 million years before their ultimate extinction. We also show that known diversity is biased by the availability of fossiliferous rock outcrop. Finally, by using a logistic model, we predict that 75% of discoverable genera will be known within 60–100 years and 90% within 100–140 years. Because of nonrandom factors affecting the process of fossil discovery (which preclude the possibility of computing realistic confidence bounds), our estimate of diversity is likely to be a lower bound.

Most previous work in paleontology has focused on computing diversity curves by using rarefaction or sampling standardization (1, 9). Our work differs in that we estimate the absolute number of genera, including those not yet discovered, whereas rarefaction and related methods compare relative counts if all samples had the same size as the smallest sample (9). Our method thus uses all available data without being limited by the size of the smallest sample.

It is important to note that the ACE estimates the total number of discoverable genera, including future discoveries. By “discoverable,” we mean that fossils of the genus could potentially be discovered by using techniques traditionally or currently used by paleontologists. A genus that left no fossils at all, because of lack of preservation, loss due to subduction or diagenesis, or other factors, can never be discovered, and its absence thus cannot be inferred by the ACE or any similar statistical method based only on abundance counts. Thus, our estimated diversity provides a lower bound on true original diversity (defined as the number of all dinosaur genera that ever existed, including those with no discoverable fossil record). Other authors have at-
Cope’s rule in the evolution of marine animals

Noel A. Heim, Matthew L. Knope, Ellen K. Schaal, Steve C. Wang, Jonathan L. Payne

Cope’s rule proposes that animal lineages evolve toward larger body size over time. To test this hypothesis across all marine animals, we compiled a data set of body sizes for 17,208 genera of marine animals spanning the past 542 million years. Mean biovolume across genera has increased by a factor of 150 since the Cambrian, whereas minimum biovolume has decreased by less than a factor of 10, and maximum biovolume has increased by more than a factor of 100,000. Neutral drift from a small initial value cannot explain this pattern. Instead, most of the size increase likely reflects differential diversification across classes, indicating that the pattern does not reflect a simple scaling-up of widespread and persistent selection for larger size within populations.

**Fig. 1.** Body size evolution across the past 542 million years. The distribution of fossil marine animal biovolumes across the Phanerozoic is shown. The colored horizontal lines show genus durations. The thick black line indicates the stage-level mean body size. The thin black lines demarcate the 5th and 95th percentiles. Cm: Cambrian; O: Ordovician; S: Silurian; D: Devonian; C: Carboniferous; P: Permian; Tr: Triassic; J: Jurassic; K: Cretaceous; Pg: Paleogene; N: Neogene.
<table>
<thead>
<tr>
<th>Activity Number:</th>
<th>176</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type:</td>
<td>Contributed</td>
</tr>
<tr>
<td>Date/Time:</td>
<td>Monday, July 31, 2017: 10:30 AM to 12:20 PM</td>
</tr>
<tr>
<td>Sponsor:</td>
<td>Section on Statistics and the Environment</td>
</tr>
<tr>
<td>Abstract #324744</td>
<td></td>
</tr>
<tr>
<td>Title:</td>
<td>Estimating the Duration of the Cambrian Explosion</td>
</tr>
<tr>
<td>Author(s):</td>
<td>Zijin Lin* and Steve Wang and Tuan Nguyen and Katrina Midgley and Heather Zhou and Daniel Wang and Linda Gai and Chengying Wang and Susannah Porter and John Moore and Adam Maloof</td>
</tr>
<tr>
<td>Companies:</td>
<td>Swarthmore College and Swarthmore College and Swarthmore College and Swarthmore College and Swarthmore College and Swarthmore College and Johns Hopkins University and Swarthmore College and UC Santa Barbara and UC Santa Barbara and Princeton University</td>
</tr>
<tr>
<td>Keywords:</td>
<td>paleontology; extinction; Signor-Lipps effect; fossils</td>
</tr>
</tbody>
</table>
Student research can be a win-win situation

Confidence intervals for the duration of a mass extinction

Steve C. Wang, Aaron E. Zimmerman, Brendan S. McVeigh, Philip J. Everson, and Heidi Wong
Gideon Zamba
University of Iowa
A Step Toward Useful Research Collaboration with Heavy Teaching Load Institutions

Gideon K. D. Zamba

Department of Biostatistics, The University of Iowa, Iowa City, IA

Baltimore MD, August 2, 2017
Outline

Biostatistics at UI: Program and Mission
Undergraduate Research Opportunity (ISIB)
How to expand ISIB and include faculty from teaching institutions
Program Focus

Provide excellent education in Biostatistical theory and methods

Conduct outstanding biostatistical methodological and collaborative research

Use our skills to serve the CPH, the Carver College of Medicine, the UI, the State and the health science community
Program Focus

16 faculty strong with

**Expertise:** Computational Statistics, Bayesian Statistics, Statistical Genetics, Survival Analysis, Syndromic Surveillance, Time Series and Sequential Analysis, Causal Inference, Spatial Statistics, Model Selection, Clinical Trials, Network Analyses, ...

Currently training 40+ graduate students (MS, PhD)

Hosts centers such as CTSDMC, CPHS, BCC, feel the statistical needs for the HCCC
Other Opportunities

Staying true to our mission

In 2007 we faced some facts
Data showed that some demographics are not being trained for graduate work in Statistics or Biostatistics

These are students from:
Small Liberal Arts Colleges
Minority Serving Institutions
Financially disadvantaged and native backgrounds
Institutions with little to no exposure to the field of statistics
Outreach

Staying true to our mission

How do we reach the under-represented demographic?

A famous Japanese quote say:

.... A vision without action is daydream...
GENESIS

2007: Join the National Alliance for Doctoral Studies in the Mathematical Sciences
GENESIS: In 2008 others join the cause.
The Biostatistics Summer Initiative (BSI)

With support from the NSF (through the National Alliance) and some departmental assistance, recruited a small group of talented undergraduate math majors for a summer course and research opportunity in Biostatistics.

Design a curriculum on the basics of biostatistics.
Assign student on faculty mentored research projects.
After 8 weeks, research findings are presented at a joint symposium with the Math Alliance.
Build a proof of concept.
Iowa Summer Institute in Biostatistics (ISIB)

In 2009 we submitted an application to the NHLBI and NIH for a training grant to pursue our objective to
Train talented undergraduate math majors or biology majors with clear mathematical thinking in a summer
Provide them with an exposure to the field of biostatistics through teaching, research, field trip, seminars, invited speech, ...

Build a pipeline for recruitment into the field of Biostatistics
Focus on small Liberal Arts Colleges and on demographics less likely to have prior exposure to the field
Iowa Summer Institute in Biostatistics

Fortunate to be funded by the NIH/NHLBI
Program has been running over the past 8 years

Things we have been able to accomplish:
Impact

BSI and ISIB

From 08 to 17, 154 scholars have been trained in BSI/ISIB

83/154 (54%) of underrepresented ethnic minorities

97/154 (63%) females

More than 85% from small Liberal Arts Colleges and Minority Serving Institutions

More than 60% of the graduating students are pursuing graduate degrees in Stat/Biostat/Applied Math/STEM
Undergrad Summer Training

Impact

Scholars were able to present their research project in their home institution upon completion.

Scholars earn letter grades from the course—transferable to their home institutions.

Scholars present their research findings at national conferences (SACNAS, ABRCMS, SIDIM, etc..) Science day(s) and Science fair(s).

Scholars are groomed to embrace Statistics or Biostatistics as graduate majors.
Undergrad Summer Training

Things to improve on

Involve the home department mentor who recommend students to the summer program

Coordinate the summer research project with home institution mentor

Involve home-institution mentor in manuscript preparation for publication

Include small Liberal Arts faculty mentors in research involvements

Build a stronger pipeline with targeted institutions across the nation
Floor Discussion
Thank you!

Amy Wagaman    awagaman@amherst.edu
Ming-Wen An    mian@vassar.edu
Paul Roback    roback@stolaf.edu
Steve Wang     scwang@swarthmore.edu
Gideon Zamba   gideon-zamba@uiowa.edu
Jo Hardin      jo.hardin@pomona.edu