Creating a Trainee-Level Longitudinal Education Database: Conceptual and Methodological Considerations

Perri Morgan, PA-C, PhD
Brandi Leach, PhD
Objectives

• Discuss potential uses of a longitudinal educational database.

• Describe examples of education studies using longitudinal databases.

• Summarize a conceptual approach to creating education databases.

• Identify existing sources of information for inclusion into a database.

• Describe processes associated with development and maintenance of a longitudinal database.
Why longitudinal?

• Longitudinal analysis allows analysis of changes at both the group and the student levels.
  – As educators, we are interested in changes in our students/graduates over time.

• This is the example we give our students when explaining our Education Research Database (ERD):
Cross-sectional vs. longitudinal data:
An example

Anne and Sue both respond to a survey about their attitudes toward working in surgery.

| How likely are you to choose a career as a surgical PA? (1-10 scale with 1= very unlikely and 10=very likely) |
|-------------------------------------------------|-------------------------------------------------|
| Student                  | 1st year response | Response at graduation |
| Anne                     | 1                  | 10                  |
| Sue                      | 10                 | 1                   |
An example: **Cross-sectional data**

<table>
<thead>
<tr>
<th>How likely are you to choose a career as a surgical PA? (1-10 scale with 1= very unlikely and 10=very likely)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Student</strong></td>
</tr>
<tr>
<td>------------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Mean student response</strong></td>
</tr>
</tbody>
</table>

Conclusion: Student attitudes toward working in surgery do NOT change over the course of their PA education.
An example: **Longitudinal data**

<table>
<thead>
<tr>
<th>How likely are you to choose a career as a surgical PA? (1-10 scale with 1= very unlikely and 10=very likely)</th>
<th>Absolute value of change in student response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student</td>
<td>1st year response</td>
</tr>
<tr>
<td>Y</td>
<td>1</td>
</tr>
<tr>
<td>X</td>
<td>10</td>
</tr>
</tbody>
</table>

**Conclusion:** Student attitudes about working in surgery change during their PA education.

**For the longitudinal analysis, we have to be able to link each student’s first response to their later response.**
Reasons we chose a student-level, longitudinal approach

• We want to avoid the potential fallacy of equating group changes with individual change.

• We will be able to limit some research to students/trainees with certain characteristics.

• We collect data at the individual level anyway—why not use them?
An essential question

• Will our student-level education database be used for program evaluation or for research, or both?
Research vs. Evaluation

**Research**
- Produces generalizable knowledge
- Uses scientific methods
- Requires human subjects review (IRB)

**Evaluation**
- Intent is to improve a *specific* program
- Findings are expected to directly impact a program and to identify potential improvements
- Geared toward program decision-making
- Sometimes does not require human subjects review (IRB)
Why might you want a longitudinal database for **evaluation** purposes?

- To help organize your data
- To use for program improvement
- To analyze issues specific to your students or your program
  - Ex: Does a specific admissions factor predict a specific problem in your program?
  - Ex: Does a specific educational intervention work better for a particular type of student in your program?
- You do not want to deal with human subjects review and informed consent (but we think this is a weak excuse!)
Why would you want a longitudinal database for research?

• To share your findings with other programs and the education community
• To help your faculty produce research
• To facilitate use of previously collected data into research on new questions
  – This might lead to shorter surveys and
  – This might reduce survey fatigue among your students
• You might be able to combine your program data with those of other institutions in the future
When does evaluation NOT require human subjects review?

• When the activity does not involve non-standard interventions
• The intent is to only provide information for and about the setting in which it is conducted
• The activity is part of standard operating procedures
Human subjects review: Our experience

• We have a separate protocol approved for creation of the database.
  – Each new survey that is added to the database requires IRB approval. So do alterations to existing surveys.
  – These are expedited, with 2-3-day turnaround.
• Any research using the database will require individual protocols.
Examples of education research using longitudinal databases

• Jefferson Medical School started a longitudinal database in 1970.
  – Over 150 articles have been published based on it.


Our current project

• The Education Research Database (ERD) is a permanent database that contains extensive longitudinal student-level data from the Duke PA Program (DPAP).

• Data collection starts with the admissions process and will continue throughout DPAP graduates’ professional lives.

• Supports research on PA selection, training, and practice.
Duke PA Program
Education Research Database (ERD)

Conceptual Approach – Big Picture

PRE-PA SCHOOL

DURING PA SCHOOL

POST-PA SCHOOL
Pre-PA School Data Sources

- Admissions data
- New student survey
During PA School Data Sources

- Midpoint student survey
  - Repeats select items from new student survey
- Academic data during PA training
- PANCE (certification exam) pass/fail
- Graduation student survey
Post-PA School
Data Sources

- New graduate survey
- Practice-related data
  - Claims data
  - State medical board sanctions data
Examples of research questions with ERD

• What student characteristics predict admission into the Duke PA program?
  – Pre-PA School → PA School

• Which PA program experiences are associated with post-graduate leadership positions?
  – PA School → Post-PA School

• What PA program experiences are associated with the delivery of high-quality care?
  – PA School → Post-PA School
A PCTE longitudinal database?

- Additional post-graduation data collection could facilitate assessment of HRSA PCTE outcomes of interest including:
  - Rate of program graduates practicing in primary care or underserved areas at least 1 year after program completion
  - Type/amount of patient services provided by program graduates
  - Quality of care provided by program graduates
  - Care delivery by trainees and faculty at PCTE clinical training sites including the quality and cost of care, and patient service

- What additional variables would be required?
- What potential issues might arise?
## ERD Data Points: Post-Graduation

### Professionalism
- Medical board action
- Leadership positions

### Practice
- In active practice?
- Practice characteristics
  - Specialty
  - Location
  - Payment

### Recertification
- PANRE
- CME
- QI

### Care outcomes
- Cost of care
- Quality of care
- Access to care

### Post graduate training
- Clinical
- Non-clinical

### SES
- Clinical income
- Loan repayment program participation

### Psychosocial
- Job satisfaction
- Stress/burnout

### Potential Additional Variables
- X
- Y

## HRSA PCTE Outcomes of Interest

<table>
<thead>
<tr>
<th>Research Question</th>
<th>Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate of program graduates practicing in primary care</td>
<td>Post-graduation survey that asks for practice specialty</td>
</tr>
<tr>
<td>Quality of care provided by program graduates</td>
<td>Use provider ID to link to claims data and regulatory board actions.</td>
</tr>
<tr>
<td>Patient services provided by program graduates</td>
<td>Use provider ID to link to claims data: types of services provided (e.g., do primary care physicians deliver babies?), productivity numbers (# visits per year, etc).</td>
</tr>
</tbody>
</table>

Use provider ID to Quality of care link to claims data provided by program and regulatory graduates board actions.
• Let’s take a break for questions and discussion

• But hold your questions about nuts and bolts, because we will discuss them next.
Nuts and bolts
Data that are NOT included

• Data not included because anonymity is necessary
  – Student evaluations of courses
  – Other student evaluations of the program (exit survey, etc.)

• Data not included because we consider them mandatory for every student
  – Data required for reporting to HRSA for grant applications and progress reports (data for determining # of disadvantaged students, etc.)
Practical issues

• Student participation
• Obtaining informed consent
• Privacy protection
• Choosing software
• Maintenance of database
• Linking data
Student participation and retention in longitudinal research

While students are in your program
• Program leadership emphasizes the contribution that students can make to knowledge about the profession by participating
  – Reiterate importance of student contributions before each new survey
• Provide incentives, such as snacks

After students leave your program
• Identify a student to act as a “champion” for your research database after graduation
• Offer incentives for survey completion
• Share results of any research using the database with students
Informed consent

• We give a 10-minute presentation to new students about the database and distribute the consent forms electronically.

• The next day, in the classroom, staff distributes paper consent forms and collects them. Faculty are not present.

• In order to obtain application data for all applicants (including those not admitted), we added a one-paragraph consent statement to our supplemental application.
Privacy protection

• Faculty does not know which students consented to participate.
• Staff assign a database identifier to each student and keep the code with student names under lock and key.
• Faculty who wish to use the database will be issued limited datasets by staff that include only the variables required for their project.
• Even without student names, faculty could identify many students using other variables. However, this would be a breach of research ethics and would violate institutional and/or federal guidelines.
Secure storage

• Data on a protected server

• Access to identifiable data limited
  – For example, researchers are only given access to variables necessary for their project

• De-identified datasets are created for individual research projects
Choosing software

- Institutional resources
  - Any existing programs available through institution (e.g., REDCap)?
  - Support readily available?
- Interface preferences – overall usability, security issues
  - Desktop-based (e.g., Microsoft Access, FileMaker Pro)
  - Server-based (e.g., MySQL)
  - Web-based (e.g., REDCap, Medrio)
- Import/export file type options (e.g., SAS, Stata, SPSS, Excel, others)
- Cost
# Database software options

<table>
<thead>
<tr>
<th>Software</th>
<th>Website</th>
<th>Where is database located?</th>
<th>Data export options</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>REDCap</td>
<td><a href="http://www.project-redcap.org/">http://www.project-redcap.org/</a></td>
<td>On Internet; need user rights to access</td>
<td>Excel, PDF, SPSS, SAS, Stata, R</td>
<td>Institutional partnership required; no cost</td>
</tr>
<tr>
<td>Medrio</td>
<td><a href="http://medrio.com/">http://medrio.com/</a></td>
<td>On Internet; need user rights to access</td>
<td>Excel, SAS, SPSS, STATA</td>
<td>Free for investigator-initiated trials; $1200/year once you hit 100k data points</td>
</tr>
<tr>
<td>OpenClinica</td>
<td><a href="https://www.openclinica.com/">https://www.openclinica.com/</a></td>
<td>On user’s computer (after free download)</td>
<td>HTML, tab-delimited, Excel, SPSS</td>
<td>Open source; no cost</td>
</tr>
<tr>
<td>QuesGen</td>
<td><a href="http://www.quesgen.com/">http://www.quesgen.com/</a></td>
<td>On Internet; need user rights to access</td>
<td>Stats packages and Excel</td>
<td>Pay as you use, with per-user, per-month charge as set-up fee</td>
</tr>
</tbody>
</table>
Linking data

- Format matters
- IRB issues
- Data use agreements
- Data cleaning
The future

One big database for all of our programs?
References