Session Number: T153
Session Title: Creating a Longitudinal Education Database: Conceptual and Methodological Issues

Perri Morgan, PA-C, PhD
Christine Everett, PA-C, MPH, PhD
Duke University
Creating a Longitudinal Educational Database for Research: Conceptual and Methodological Issues
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 go longitudinal?

• To move toward the dream studies

• Dream question
  How do we select and educate students to produce PAs with the best and largest impact on the health of the nation?
Why go longitudinal?

• Longitudinal analysis allows analysis of *changes* at both the *group* and *student* level.

• 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:
An example: **cross-sectional vs. longitudinal data**

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

<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>Student</td>
</tr>
<tr>
<td>Anne</td>
</tr>
<tr>
<td>Sue</td>
</tr>
</tbody>
</table>
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>Student</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Mean student response 5.5 5.5

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>Student</th>
<th>1&lt;sup&gt;st&lt;/sup&gt; year response</th>
<th>Response at graduation</th>
<th>Change in student response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>1</td>
<td>10</td>
<td>+9</td>
</tr>
<tr>
<td>X</td>
<td>10</td>
<td>1</td>
<td>-9</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.
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?
- 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 facilitate use of previously-collected data into research on new questions
  – This might lead to shorter surveys and
  – *might* reduce survey fatigue among your students
• You might be able to combine your program data with that 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
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


Examples of longitudinal research in PA education


Duke PA Program

Education Research Database
Our conceptual approach: The big picture

- PRE-PA SCHOOL
- DURING PA SCHOOL
- POST-PA SCHOOL
Data Sources

- Admissions data
- New student survey
Data Sources

- Midpoint student survey
  - Repeats select items from new student survey
- Academic data during PA training
- PANCE pass/fail
- Graduation student survey
  - To be developed
Data Sources

- Alumni survey
  - To be developed
- Practice-related data
  - Claims data
  - State medical board sanctions data
- Others TBD
Examples of Research Questions

• 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)
Data that is 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 it 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
- Human subjects review
- Privacy protection
- Choosing software
- Maintenance of database
- Linking data
- ??
Student participation

• Program leadership emphasizes the contribution that students can make to knowledge about the profession by participating

• 2013 entering class 85/90 consented
  – Midpoint survey 89/90 completed the survey—snacks might have helped.

• 2014 entering class 89/90 consented
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. These are expedited, with 2-3 day turnaround

• Any research using the database will require individual protocols
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 possibly illegal.
Secure storage

• Data on a protected server
• Access to identifiable data limited
• De-identified datasets 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

• Many PA programs combining data for research?
Discussion

• Suggestions
• Questions
References


