Facilitating Research in Physician Assistant Programs: Creating a Student-Level Longitudinal Database

Perri Morgan, PhD, PA-C; Katherine M. Humeniuk, MPH; Christine M. Everett, PhD, MPH, PA-C

Abstract As physician assistant (PA) roles expand and diversify in the United States and around the world, there is a pressing need for research that illuminates how PAs may best be selected, educated, and used in health systems to maximize their potential contributions to health. Physician assistant education programs are well positioned to advance this research by collecting and organizing data on applicants, students, and graduates. Our PA program is creating a permanent longitudinal education database for research that contains extensive student-level data. This database will allow us to conduct research on all phases of PA education, from admission processes through the professional practice of our graduates. In this article, we describe our approach to constructing a longitudinal student-level research database and discuss the strengths and limitations of longitudinal databases for research on education and the practice of PAs. We hope to encourage other PA programs to initiate similar projects so that, in the future, data can be combined for use in multi-institutional research that can contribute to improved education for PA students across programs.

INTRODUCTION

As physician assistant (PA) roles expand and diversify in the United States and around the world, there is a pressing need for research that illuminates how PAs may best be selected, educated, and employed in health systems to maximize their potential contributions to health.[1] A historical barrier to this research has been the relative lack of high-quality data on PAs.[2] Physician assistant education programs are well positioned to contribute to the solution to this problem by collecting data on their applicants, students, and graduates and organizing these data into formats that support research. Research based on these data can then be used by PA educators to improve and adapt educational approaches to meet evolving workforce realities.

Physician assistant educators are interested in how individual students change over time and the influence that education interventions have in bringing about these changes. Each student is a complex individual with physical, social, psychological, intellectual, and life experience characteristics that interact with educational and other influences over the course of a lifetime. Interventions that lead to a particular outcome in one student might produce a very different outcome in another. Therefore, education research needs rich information about individual students over an extended time period. Student-level longitudinal research databases can provide this information in a way that is relatively easy for researchers to access.[3–5]

Although PA programs currently collect large amounts of data, these data often are not organized and maintained in a manner that is easily accessible to researchers. When faculty members initiate a research project, time and effort are required to either gather existing data for a cross-sectional or retrospective study design or to conduct a prospective study and wait for data to be generated over the study period. Both approaches often benefit only one evaluation or study since storage and future use of the data are not typically a priority at the time the original study is undertaken.

In contrast to this study-by-study approach, we are initiating proactive collection of extensive student-level data into a permanent longitudinal education database for research. Others have used this approach successfully. For example, Jefferson Medical School started a similar database in 1970 and continues to collect data on its students and graduates.[6] Over 150 articles have been published using the Jefferson Medical School longitudinal database.[7] Many of these evaluations have correlated attributes that a student had during medical school with outcomes many years—even decades—later.[8–11]

A longitudinal database will allow us to conduct research on all phases of PA education, from admission processes through the professional practice of our PA graduates. Inclusion of rich data on each student will support evaluation of how education interventions impact students with varying characteristics. Over time, trends among different PA cohorts can be compared. By merging other data into our longitudinal database, we may be able to evaluate quality and efficiency outcomes of the care that our graduates provide. Table 1 provides samples of research questions that can be addressed using this database.

The purpose of this article is to describe our approach to constructing a longitudinal student-level database for research purposes. We also consider the strengths and limitations of longitudinal databases for research on education and practice of PAs. We recognize that generalizability of research from only one program will always be limited. By publishing this article, we hope to encourage other PA programs to initiate similar projects so that, in the future, data can be combined for use in multi-institutional research that can contribute to improved education for PA students across programs.

Table 1: Examples of Research Questions that Could Be Answered Using a Longitudinal Research Database

Type of Data	Sample Research Questions
Pre-PA school	Does veteran status affect chance of admission to our program?
	What are the main themes included in essays of admitted vs non-admitted students?
During PA school	Are volunteerism and leadership positions associated with academic performance?
	Is student-reported stress associated with professionalism problems in PA students?
After PA school	Does participation in a loan repayment program predict whether graduates work in primary care 5 years after graduation?
	Is job satisfaction associated with specialty choice? If so, which specialties have the highest and lowest job satisfaction scores?
	Are graduate salaries correlated with race and gender?
From across time frames and/or with external sources added	What proportion of graduates work in underserved areas? (combine practice zip code data with national data for underservice categories)
	Do professionalism problems in PA school predict medical board actions against practicing PAs?
	Does expressed desire to work in primary care in the application essay correlate with practice specialty 5 years after graduation?
	Are graduates who participated in a primary care track more likely to work in primary care 5 years after graduation?
	Do graduates with more pre-PA school clinical experience provide higher quality of care?
	Is class rank associated with the delivery of high-quality care?
	Do PAs with public health training deliver lower cost care?
	Does student debt at the time of entering PA school predict specialty choice?
	What factors are associated with a graduate working in an underserved area (eg, socioeconomic status of the student before PA school, altruism demonstrated by volunteerism during training, personality traits)?
With supplemental survey data added	Surveys on any topic can be designed and the results merged onto the individual level database so that many control variables are available for use in the analysis

EDUCATION RESEARCH DATABASE

Our Education Research Database (ERD) is organized around the conceptual model shown in Figure 1. Pre-PA school factor information is derived mostly from our admission data, supplemented with a survey that we ask students to complete upon matriculation into our program. Information added to the database during PA school includes student academic performance, participation in extracurricular and leadership activities, data from at least 2 surveys (midpoint and graduation surveys), and performance on the Physician Assistant National Certifying Exam (PANCE). If a student is placed on probation for academic or professionalism reasons, this is noted as well. After graduation, we plan to follow these individuals with graduate surveys to track their practice specialty and location, earnings, leadership activities, and professional satisfaction throughout their lives. Finally, it might be possible to include their National Provider Information numbers so that we can link medical claims data to our graduates to study the cost and quality of care that they provide as practicing PAs.

Data Not Included in the Education Research Database

Some student-level information is not appropriate for inclusion in the ERD. This includes information that requires anonymity, such as course and program evaluations by students. We are careful to indicate to students what information they supply is completely anonymous and what information will be included in the ERD.

Evaluation versus Research

Since longitudinal databases can be constructed and used for evaluation or research purposes, it is worthwhile to consider the differences in these 2 uses.[12] Evaluation is typically done to inform a specific education program and to impact future planning of that program. Findings are not disseminated beyond the specific program, and human subjects review is rarely required. However, research is designed with the intent of creating generalizable knowledge that will be disseminated to the wider education community. For research, human subjects review is required, although expedited review and waiver of formal consent requirements are often granted. The database that we describe in this article is for research purposes.

Human Subjects Issues and Review

Our institutional review board (IRB) approved the database as a stand-alone resource for use in future projects. The IRB application included a list of information that would be included and the sources of this information. Copies of existing surveys and the student consent form were included. Expedited review is required when new surveys or other data
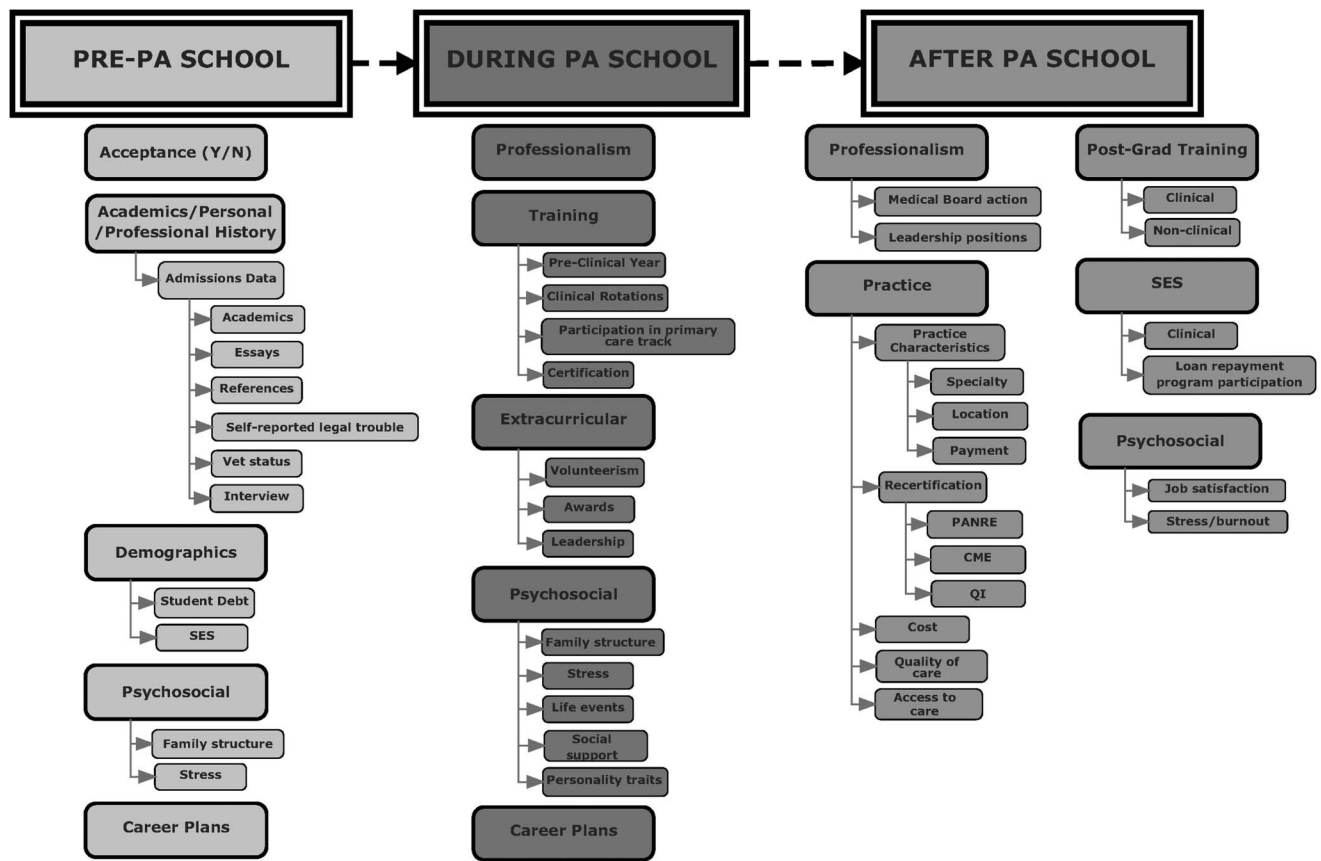
Figure 1. Conceptual model for Educational Research Database

sources are added to the ERD. The IRB documents specify the means and location of data storage and other methods to protect participant information, to be described below.

Each research project using the ERD requires a separate IRB proposal, similar to any proposal for research using existing data. Most of these proposals will qualify for waivers of formal informed consent procedures and will often be granted a waiver from full human subjects committee review.

Informed Consent

For PA program applicants, a short statement indicating their consent to participate in the database is included on our program's supplemental admissions application, along with a statement explaining that the applicant's decision about whether to participate will not affect his or her chance of admission to our program. For admitted students, this step becomes unnecessary since all admitted students are invited to participate in the ERD shortly after matriculation. However, inclusion of all applicants will allow us to undertake future research, examining differences in admitted and not admitted students. For matriculating students, a full consent form is provided that explains the purpose of the database, procedures that will be followed to protect confidentiality, and risks and benefits of participation.

A few days after matriculation, the faculty members responsible for the database give a 10-minute presentation to the students that describes the purpose of the database and highlights important aspects such as the voluntary nature of

participation and procedures to protect the confidentiality of the data. Students are given a chance to ask questions and are informed that a short session to sign actual consent forms will occur the next day. At the conclusion of this session, students are sent an electronic version of the consent form and asked to review it. The next day, in our classroom, paper copies are distributed by program staff, and students are invited to sign them. To minimize the perception of coercion, no faculty members are present at this time. All students are asked to return the consent forms, whether they are signed or not, so that they do not have to reveal their participation decision to other students who may be sitting nearby.

Privacy Protection Procedures

A program research staff member assigns an identifier code to each student. All information that is included in the ERD is identified by this code, so that no student names are identified in the database. The code linking the identifier to the student name is kept under lock and key by the research staff and is never seen by any faculty member.

Students may have identifying characteristics that are included in the database. Faculty members who know the students well could identify them by these characteristics, such as their state of origin, their gender and race characteristics, or their veteran status. Attempting to identify students in this manner constitutes a breach of professional ethics and could even reach the level of a federal crime. Therefore, all investigators using the database are required to participate in

research ethics training. In addition, no faculty members are given access to the entire database. When faculty investigators initiate a research project, they are required to stipulate in their IRB application which pieces of data their study will require. If the proposal is approved by our IRB, the research staff creates a limited data set for the study that includes only the approved variables. Access to this data set will be limited in accordance with IRB requirements, with stringency of privacy protections based on the sensitivity of the data being used in the particular study. The ERD is kept on a secure server maintained by our university and approved by our IRB. Only the research analyst and the ERD primary investigator have access to the ERD server location.

THE PROCESS

Participation

During program orientation, our program leadership mentions the ERD as a strength of our program and tells students that they will be invited to participate. The study is explained in detail as part of the informed consent process. Students have responded positively, with 94% and 99% agreeing to participate in each of our first 2 enrolled classes.

Long-term participation of alumni will be critical to the success of our project. We will use strategies based on established means of maintaining participant engagement, including:

- Engaging students and alumni in planning retention strategies;
- Showing alumni results of the research based on the database;
- Branding the database with a study logo;
- Developing and maintaining personal relationships between study staff and students and alumni; and
- Using special alumni tracking methods including social media that we hope will help to engage our graduates long term.[13,14]

During students' time at our program, we express gratitude for their participation, remind them of the value that we expect to come from the project, and build relationships with our alumni coordinator. To foster a sense of engagement with the research, we will complete a research project designed to be of interest to students, using their own data, and present the results of this project to each class before it graduates. We will keep graduates apprised of research findings based on the ERD through program publications and social media.

Software

Choice of software is determined by a number of factors, including which packages are provided and supported by the institution and analyst preferences regarding interface and usability issues. The software must be capable of exporting output into formats that can be manipulated by other statistical and presentation software. Table 2 reviews some software options and attributes of each.

Data Sharing for Multi-Institutional Research

Since generalizability of research from any single program will be limited, we hope to partner with other institutions to conduct multi-institutional research. Human subjects committee approvals and data use agreements will be required, and extensive data protection procedures will need to be developed. Data from multiple institutions must be in matching formats before they can be combined for analysis. To facilitate this process, we will make our ERD documents, including the data dictionary and data structure, available on our research Web page for use by other programs.

DISCUSSION

When deciding whether to create a longitudinal database for research, PA programs will want to consider the benefits and costs of such a project. Benefits include creating capacity for faculty to more easily complete research and providing a source of data capable of supporting high-quality projects due to their validity, rich information, and accessibility.

Table 2: Selection of Database Programs that Institutions Might Use in Designing a Research Database

Software	Web Site	Where Is Database Located?	Data Export Options	Cost
Microsoft Access	http://office.microsoft.com/en-us/access/	On user's computer	Excel, txt, Word, XML	Office 365—individual license $70
REDCap	http://www.project-redcap.org/	On Internet; need user rights to access	Excel, PDF, SPSS, SAS, Stata, R	Institutional partnership required; no cost
Medrio	http://medrio.com/	On Internet; need user rights to access	Excel, SAS, SPSS, STATA	Free for investigator-initiated trials; $1200 per year once you hit 100,000 data points
StudyTrax	http://www.sciencetrax.com/studytrax/	Hosted on own server or ScienceTrax secure servers	Excel, CSV, SAS, SPSS, Word	$99 student license
OpenClinica	https://www.openclinica.com/	On user's computer (after free download)	HTML, tab-delimited, Excel, SPSS	Open source; no cost
QuesGen	http://www.quesgen.com/	On Internet; need user rights to access	Stats packages and Excel	Pay as you use, with per-user, per-month charge as set-up fee

Weaknesses of this approach include the labor costs of creating and maintaining the database and the challenge of maintaining strong long-term participation of graduates.

Education Research Database Strengths and Benefits

Prospective Data Collection
Data will be entered into the ERD close to the time that they were generated. Therefore, investigators will not need to rely on information that subjects must recall from the remote past. This is expected to result in more accurate data.

Ability to Control for Many Student Factors
Survey researchers must balance the need for information with the need to keep surveys brief. The quest for brevity can lead to a limited number of variables that can be examined for potential confounding or interaction with the main exposures and outcomes of the study. Many variables needed for these purposes, such as student demographic and socioeconomic factors, previous academic and health care experience, and PA training experiences, will be available in the ERD.

Reduced Survey Fatigue
Research projects using the ERD can be based solely on the information contained in the database or can collect new data to be merged with the existing data in the ERD. For example, an investigator might wish to study student attitudes toward interprofessional practice (IPP). After obtaining IRB approval for the survey in attitudes toward IPP, he or she might administer a survey on these attitudes to students, but the investigator would not need to ask respondents for other pertinent information that is already in the ERD, such as demographic or clinical rotation information. The investigator would ask for his/her data to be merged into the existing ERD and for an analytic data set to be created that includes the survey plus the requested demographic and clinical rotation variables. In theory, this could result in surveys that are briefer, possibly minimizing survey fatigue for students.

Improved Outcome Measures
Because the ERD will follow students over the course of their careers, long-term outcomes will be available for analysis. For example, rather than evaluating the intent of students to choose a primary care specialty, or the likelihood of students choosing primary care positions at the time of graduation, ERD long-term follow-up of specialty choice will allow researchers to analyze the total portion of PAs' careers spent in primary care practice. In addition, if the ERD is linked to other data sources, such as insurance claims data, it will be possible to evaluate the quality or efficiency of care provided by PAs. Many researchers favor these ultimate outcomes over intermediate outcomes, such as whether graduates pass the national certification examination.

Facilitating Physician Assistant Faculty Research
Barriers to PA faculty research productivity include the difficulty of collecting data for research projects. If a rich data set is available to faculty for use, program directors could potentially be more supportive of faculty participation in research, and faculty may have more success in completing research.[1,15,16]

Barriers to Education Research Database Implementation

Initiating and Maintaining the Education Research Database
Since our PA program already collects most of the data that will be included in the ERD, data collection will be minimal; however, significant labor will be required to initiate and maintain the ERD. Modern software will facilitate database construction.

Extended Data Collection Period
It will take many years to collect the data necessary to evaluate some of the questions discussed above, such as the portion of time spent in various specialties over a PA's career. Interim analyses, such as at 5 or 10 years, will be possible.

Incomplete Follow-up
Some students and graduates will drop out of the database or will respond to surveys only sporadically. Intensive efforts will be necessary to keep graduates engaged.

CONCLUSION

Many voices in the PA education community have identified a need to develop research capacity and to produce more high-quality research on PAs and PA education.[1,15–17] This research is needed to inform education strategies to produce high-quality graduates and workforce strategies to use PAs to their best possible effect on the nation's health.

Longitudinal ERDs provide a promising source of data for research on PAs and PA education. Their greatest potential is in their ability to improve research quality and support evaluation of long-term outcomes of PA education. Because the generalizability of any research from a single institution will be limited, PA programs should embrace a long-term goal of longitudinal student-level multi-institutional research.

Perri Morgan, PhD, PA-C, is the director of Research and associate professor in the Physician Assistant Division in the Department of Community and Family Medicine at Duke University Medical Center, Durham, North Carolina.

Katherine M. Humeniuk, MPH, is a research analyst in the Department of Community and Family Medicine at Duke University Medical Center, Durham, North Carolina.

Christine M. Everett, PhD, MPH, PA-C, is an assistant professor in the Physician Assistant Division in the Department of Community and Family Medicine at Duke University Medical Center, Durham, North Carolina.

Correspondence should be addressed to: Perri Morgan, PhD, PA-C, Physician Assistant Division, Department of Community and Family Medicine, Duke University Medical Center, DUMC 104780, Durham, NC 27710. Telephone: (919) 681-6791; Email: Perri.morgan@duke.edu

REFERENCES

1. Cawley JF, Ritsema TS. Where are the PA researchers? JAAPA. 2013; 26:13, 22.
2. Morgan P, Strand J, Ostbye T, Albanese M. Missing in action: care by physician assistants and nurse practitioners in national health surveys. Health Serv Res. 2007;42:2022–2037.
3. Ander-Peciva S. Construction of longitudinal databases—for flexibility, transparency and longevity. Paper presented at: Meeting of the International Commission for Historical Demography; July 6, 2005; Sydney, Australia.

4. Cook DA, Andriole DA, Durning SJ, et al. Longitudinal research databases in medical education: facilitating the study of educational outcomes over time and across institutions. Acad Med. 2010;85:1340–1346.

5. Triola MM, Pusic MV. The education data warehouse: a transformative tool for health education research. J Grad Med Educ. 2012;4:113–115.

6. Veloski J, Hojat M, Gonnella J. The Jefferson Longitudinal Study of Medical Education: five decades of outcomes assessment. Population Health Matters. 2013;26:2–3.

7. Gonnella JS, Hojat M, Veloski JJ. Abstracts: Jefferson Longitudinal Study of Medical Education. 3rd ed. Jefferson Longitudinal Study of Medical Education; 2005. Paper 1. http://jdc.jefferson.edu/jlsme/1.

8. Cooter R, Erdmann JB, Gonnella JS, et al. Economic diversity in medical education: the relationship between students' family income and academic performance, career choice, and student debt. Eval Health Prof. 2004;27:252–264.

9. Hojat M, Brigham TP, Gottheil E, et al. Medical students' personal values and their career choices a quarter-century later. Psychol Rep. 1998;83:243–248.

10. Papadakis MA, Teherani A, Banach MA, et al. Disciplinary action by medical boards and prior behavior in medical school. N Engl J Med. 2005;353:2673–2682.

11. Tamblyn R, Abrahamowicz M, Dauphinee D, et al. Physician scores on a national clinical skills examination as predictors of complaints to medical regulatory authorities. JAMA. 2007;298:993–1001.

12. Chen H. Designing Education Lab: Evaluation vs. Research—What's the Difference? Stanford University, 2013. http://web.stanford.edu/group/design_education/wikiupload/2/27/Helen_Evaluation.pdf. Accessed September 20, 2014.

13. Mychasiuk R, Benzies K. Facebook: an effective tool for participant retention in longitudinal research. Child Care Health Dev. 2012;38:753–756.

14. Robinson KA, Dennison CR, Wayman DM, et al. Systematic review identifies number of strategies important for retaining study participants. J Clin Epidemiol. 2007;60:757–765.

15. Hegmann TE, Axelson RD. Benchmarking the scholarly productivity of physician assistant educators: an update. J Physician Assist Educ. 2012;23:16–23.

16. Hegmann T, Dehn R. Physician assistant program director opinions regarding the importance of faculty research and publication. J Physician Assist Educ. 2006;17:5–10.

17. Dehn RW. Missing the mark: why is some research on PAs just wrong? JAAPA. 2014;27:9.