LEVERAGING ADMINISTRATIVE DATA TO STRENGTHEN NCES POSTSECONDARY SAMPLE SURVEYS

A Paper Commissioned by the National Postsecondary Education Cooperative

LEVERAGING ADMINISTRATIVE DATA TO STRENGTHEN NCES POSTSECONDARY SAMPLE SURVEYS

A Paper Commissioned by the National Postsecondary Education Cooperative

Author:
Matthew Soldner
American Institutes for Research

August 2017

This project has been funded, either wholly or in part, with federal funds from the U.S. Department of Education under Coffey Consulting, LLC's Contract No. ED-IES-12-D-0016. The contents of this publication do not necessarily reflect the views or policies of the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement of same by the U.S. Government.

National Postsecondary Education Cooperative

The National Postsecondary Education Cooperative (NPEC) was established by the National Center for Education Statistics (NCES) in 1995 as a voluntary organization that encompasses all sectors of the postsecondary education community including federal agencies, postsecondary institutions, associations, and other organizations with a major interest in postsecondary education data collection. In 2007, NCES assigned NPEC the responsibility for developing a research and development agenda for the Integrated Postsecondary Education Data System (IPEDS). IPEDS is the core postsecondary education data collection program for NCES. NPEC also occasionally produces products of value to postsecondary data providers, users, and institutional representatives.

NPEC publications do not undergo the formal review required for standard NCES products. The information and opinions published in them are the products of NPEC and do not necessarily represent the policy or views of the U.S. Department of Education or NCES.

August 2017

The NCES Home Page address is http://nces.ed.gov
The NCES Publications and Products address is http://nces.ed.gov/pubsearch
The NPEC Home Page address is http://nces.ed.gov/ipeds/InsidePages/JoinIn?pageid=41

This publication is only available online. To download, view, and print the report as a PDF file, go to the NCES Publications and Products address shown above.

Suggested Citation
Soldner, M. *Leveraging Administrative Data to Strengthen NCES Postsecondary Sample Surveys*. (NPEC 2017). U.S. Department of Education. Washington, DC: National Postsecondary Education Cooperative. Retrieved [date] from http://nces.ed.gov/pubsearch.

Content Contact: Sean Simone at (202) 245-7631 or Sean.Simone@ed.gov

TABLE OF CONTENTS

ABBREVIATIONS

Below is a list of the most frequently used abbreviations in this report:

B&B	Baccalaureate & Beyond Longitudinal Study
BPS	Beginning Postsecondary Students Longitudinal Study
COD	Common Origination and Disbursement System
CPS	Central Processing System
DoD	Department of Defense
ED	Department of Education
FAFSA	Free Application for Federal Student Aid
FSA	Federal Student Aid
FTB	First Time Beginning Student
IPEDS	Integrated Postsecondary Education Data System
NCES	National Center for Education Statistics
NPSAS	National Postsecondary Student Aid Study
NSC	National Student Clearinghouse
SSA	Social Security Administration
UI	Unemployment Insurance
VA	Department of Veterans Affairs

INTRODUCTION

The U.S. Department of Education's (ED's) National Center for Education Statistics (NCES) National Postsecondary Student Aid Study (NPSAS) is the nation's preeminent source of information on how students and their families pay for college. Its spinoff studies, the Beginning Postsecondary Students Longitudinal Study (BPS) and the Baccalaureate and Beyond Longitudinal Study (B&B), provide additional, critical insight into the undergraduate experience with an emphasis on students' persistence and completion decisions, and the choices made by students who graduate with a bachelor's degree, respectively. Taken together, the NPSAS family of studies has grounded higher education research for nearly three decades.

The studies currently draw upon data from three primary sources, including (1) one or more student interviews; (2) administrative data gathered from institutions' student information systems; and (3) matches to a variety of ED's Federal Student Aid (FSA) data systems. NCES also conducts additional matches to augment its surveys, including with data maintained by testing companies like ACTthe College Board, and the National Student Clearinghouse (NSC), a third-party vendor that assists colleges and universities in complying with FSA reporting requirements.

In this paper, we consider whether additional administrative data sources might further benefit the NCES postsecondary sample surveys program, the institutions and students it surveys, and the education researchers it supports. We begin by exploring how NCES can leverage linkages to three federal data sources to better understand the price of college, students' post-college wage outcomes, and the educational experiences of veterans, active duty military, and the beneficiaries of their educational benefits. Then, we consider how non-federal linkages might provide additional data about students' acquisition of industry-recognized certifications and major life events. Our exploration is motivated by three factors.

The first reason to consider using additional administrative data in the NPSAS family of studies is burden reduction. Although the resulting analytic datasets are an incredibly powerful tool for researchers and policymakers, they can represent a significant investment of time for two key populations: the students asked to complete the interview, and the institutions that enroll them[1]. Students explicitly signal that burden by failing to respond: in the 2011–12 administration of NPSAS, 31 percent of the eligible sample failed to complete the survey instrument, despite an incentive offer from NCES for completing it. Institutions display a similar behavior, albeit at a lower rate, with 12 percent of NPSAS-sampled institutions failing to provide student enrollment lists for sampling which classifies them as nonrespondents for purposes of subsequent administrative data collections.

[1] Students took 28 minutes to complete the NPSAS interview, on average (Wine, Bryan, and Siegel 2014).

Second, enhanced use of administrative data can result in operational benefits for NCES, particularly if decreasing burden improves response to the student interview. Improved student-level response can improve data quality, and can decrease costs associated with nonresponse conversion and imputation. When improved student-level response occurs on the NPSAS interview within key subsamples—most notably those that form the foundation of BPS or B&B—those benefits are multiplied.

Finally, more intentionally leveraging extant administrative data within the NPSAS family of studies holds the potential to opening new lines of inquiry and for the exploration of both new and existing research questions using higher quality data. During each collection cycle, difficult decisions must be made about what *not* to include on the student survey in the name of reducing its length. Making better use of administrative data may allow NCES to free up valuable space on existing questionnaires for new purposes when old content can be acquired via matching, or it may add new information to the dataset at no burden to respondents whatsoever. To the extent that administrative data contributes less to total survey error than self-reported data by students, which may introduce error due to incorrect recall or socially-desirable response, data quality is improved.

These reasons notwithstanding, other efforts across government make now a good time for NCES to explore expanding the scope of administrative data linkages in the NPSAS family of studies. The Commission on Evidence-Based Policymaking, established by a 2016 Act of the same name, is charged with "[developing] a strategy for increasing the availability and use of data in order to build evidence about government programs, while protecting privacy and confidentiality."[2] Since the Commission's inaugural meeting, researchers have highlighted how improved linkages between administrative data sources can strengthen policymaking for the public good (Chetty 2016).

This paper begins with a brief review of the primary purposes of the studies considered here, including NPSAS, BPS, and B&B. That review is followed by a discussion of the ways in which administrative data are already used to inform each. This includes the use of administrative data in shaping student-facing and institution-facing instrumentation, sampling, data collection, and data processing.

After exploring how NCES currently uses administrative data in its postsecondary sample surveys, we look to new federal and nonfederal data sources that may add value to the NPSAS family of studies. For each, we consider its potential benefits, caveats, and challenges to use.

[2] For more information about the Commission, visit its website at https://www.cep.gov/

Finally, we provide a summary of the potential of administrative data to strengthen NCES efforts to understand students' pathways through postsecondary education and the workforce.

THE NPSAS FAMILY OF STUDIES

NATIONAL POSTSECONDARY STUDENT AID STUDY

Historically conducted about every 3 to 4 years, the primary purpose of NPSAS is to understand how students and families pay for postsecondary education both at the undergraduate and graduate levels[3]. Secondarily, NPSAS provides population-level estimates for demographic and academic characteristics of students enrolled in a given federal financial aid year.[4] Finally, each new NPSAS serves as the base-year of data collection for one of two NCES postsecondary longitudinal studies—B&B and BPS—fielded in alternating administrations. The most recent NPSAS, for example, was fielded in 2015–16 and served as the base year of data collection for B&B:16/26. The NPSAS that preceded it, fielded in 2011–12, was the base year of data collection for BPS:12/17.

NCES achieves this goal, with the help of its data collection contractors, by collecting and processing data from multiple data sources, including approximately 1,700 postsecondary institutions and approximately 125,000 students, of which about 16,000 are graduate students (Wine, Bryan, and Siegel 2014). Although that effort involves hundreds, if not thousands, of discrete steps, those that support four major activities are important to understand as part of any discussion about how administrative data are already used in the NPSAS family of studies: (1) instrumentation development, (2) sampling, (3) data collection, and (4) data processing, imputation, and weighting. Many of these activities are mirrored in the development of BPS and B&B, with administrative data playing a similarly important role.

[3] In late 2016, NCES announced a new initiative designed to gather data on aided students every 2, rather than every 4, years: the 2017-18 National Postsecondary Student Aid Study, Administrative Collection (NPSAS:18-AC). This study will collect administrative data from Federal Student Aid and postsecondary institutions to create nationally and, for most states, state-representative financial aid estimates for students enrolled in the 2017-18 financial aid year. For more information, visit https://nces.ed.gov/surveys/npsas/

[4] It should be noted that NPSAS is not the only NCES study that explores these issues. NCES Integrated Postsecondary Education Data System (IPEDS), for example, relies upon an annual census of Title IV participating institutions to produce institution-level statistics on a small subset of data elements collected by NPSAS, including information on tuition and fees, average grant and loan amounts, net price (cost of attendance net of grant aid), and the demographic characteristics of fall enrollees. While IPEDS' periodicity and coverage are notable strengths, its unit of analysis—individual institutions—is limiting. Because NPSAS collects data at the student level, its estimates can be disaggregated by any number of policy-relevant student or institution characteristics, providing a more nuanced portrait of the key questions it is meant to answer.

DATA COLLECTION INSTRUMENTS DEVELOPED

An early activity in the development of NPSAS and its longitudinal follow-ups is instrumentation design, which proceeds along two parallel tracks. The first track develops instrumentation for the student interview, which will ultimately be respondent-administered via the Web or administered via telephone led by a trained interviewer. The second develops a web-based portal through which institutions submit both institution-level and student-unit record-level data from campus administrative data systems. The role of planned administrative data linkages in shaping both student-facing and institution-facing instrumentation cannot be understated, because a common principle guides the development of each: do not ask students or institutions for information that can be gleaned elsewhere later by linkages to other data sources.

INSTITUTIONAL AND STUDENT SAMPLING

As instrumentation development draws to a close, the NCES data collection team turns its attention to sampling. NPSAS employs a two-stage sampling design. In the first stage, the team samples eligible institutions from the universe of Title-IV awarding, postsecondary institutions in the United States. Once institutional participants have been identified and their participation secured, institutions are asked to submit lists of all enrolled, eligible students. In addition to including contact information for each student, enrollment lists include a variety of student characteristics that are used throughout the sampling effort to refine the respondent cohort. Of special note are indicators of whether the student is, to the institution's knowledge (1) a first-time, beginning student (FTB), meaning they are potentially eligible to participate in BPS; or (2) a graduating bachelor's degree student, meaning they are potentially eligible to participate in B&B. These data from institutions, as well as other administrative data linkages which are described in more detail below, are critical to maximizing the efficiency of sampling and the quality of the estimates that can eventually be produced using study data.

STUDENT DATA COLLECTION

Once sampling is underway, primary data collection with institutions and students begins. Over the course of several months, hundreds of thousands of discrete contacts are made with students to incent study participation: NCES data collection contractor reports contacting each sample member an average of eight times during the data collection period (Wine, Bryan, and Siegel 2014). Tens of thousands of web-based interviews are completed, and a smaller number of respondents complete the student survey with the assistance of a trained telephone interviewer. At the conclusion of the data collection period, about 70 percent of sampled students will have actively participated in a given NPSAS study (Wine, Bryan, and Siegel 2014). Response rates to subsequent BPS and B&B longitudinal follow-ups meet or exceed that rate, at around 70 and 85 percent, respectively (Cominole, Shepherd, and Siegel 2015; Hill et al. 2016).

Data Processing, Imputation, and Weighting

At the conclusion of data collection, NCES data collection contractor engages in three activities that each leverages administrative data. First, the contractor augments data collected from institutions and students with data from a variety of other sources. In the case of NPSAS, because of its emphasis on how students and families pay for college, the most critical administrative data sources leveraged by NPSAS are those collected through FSA systems, such as ED's National Student Loan Data System (NSLDS). Other sources, and their use in other surveys, are described in more detail below. Second, the contractor augments missing data using a variety of imputation approaches to maximize the amount of information available to analysts. Finally, the contractor's sampling statisticians weight the survey data to account for the study's sampling design as well as survey nonresponse, post-stratify to known totals in the population, and implement other weight adjustments to help ensure data quality.

Beginning Postsecondary Students Longitudinal Study

Since its inception in 1990, BPS has provided education researchers information about the persistence and completion behaviors of FTB students. Fielded in alternate NPSAS years, new iterations of BPS are begun approximately every 8 years, following cohort members for up to 6 years after their entry to postsecondary education. When it was developed, BPS was the only research tool that allowed analysts to track students' movements between institutions, first by use of a student interview and, later, for aided students through matching to NSLDS. For the first time, the calculation of national estimates for postsecondary persistence, stop-out, and completion became possible.

Since then, additional data resources—most notably the NSC—have been developed. In contrast to the fewer than 40,000 students included in a typical BPS cohort, NSC maintains regularly updated enrollment and completion data on 97 percent of all enrolled students nationally, more than 20 million (Dundar and Shapiro 2016; Hill et al. 2016). These data, a byproduct of member institutions' use of NSC as an enrollment reporting service to ED's FSA system via NSLDS, eclipse BPS in scope and coverage. Because of this, NCES has relied upon NSC as an important source of data across the NPSAS family of studies in recent study cycles. Note that BPS is still an important postsecondary collection to retain because it allows study of how students transition into and out of postsecondary education.

Baccalaureate and Beyond Longitudinal Study

The second of NCES postsecondary longitudinal studies, B&B, follows a cohort of between 15,000 and 20,000 bachelor's degree earners for up to 10 years after graduation (Cominole, Shepherd, and Siegel 2015). Like BPS, it uses NPSAS for its base-year data collection, with a new B&B beginning about every 8 years. Over the ensuing decade, B&B follows recent college graduates' employment experiences, decisions they make about returning to graduate school

to further their education, choices about family formation and other life milestones, and student loan repayment histories.

The breadth of topics addressed in B&B suggests that, among all NCES postsecondary longitudinal studies, it may have the most to gain through linking to high-quality administrative data. As is discussed below, this includes governmental and nongovernmental data about wages and employment histories, student loan repayment, and re-enrollment in educational institutions after completing the baccalaureate.

EXISTING NCES USE OF ADMINISTRATIVE DATA

FEDERAL STUDENT AID DATA SYSTEMS

FSA maintains more than two dozen data systems to manage ED's FSA system. Of those, three are currently used by NCES and its data collection contractors in the development of the NPSAS family of studies.

1. **NSLDS.** Consisting of more than 30 billion records, NSLDS is the workhorse of FSA's current data systems. It contains detailed information about every Title IV federal student loan, including those loans' origination amount, current balance, and repayment status (Soldner and Campbell 2016).
2. **Central Processing System (CPS).** Annually, 20 million student-aid seekers file the Free Application for Federal Student Aid (FAFSA), used by ED, states, and institutions to determine students' eligibility for a wide range of grant and loan programs (Federal Student Aid, n.d-a.). Data provided on the FAFSA, along with a series of calculations used to determine a student's aid eligibility, are stored in CPS.
3. **Common Origination and Disbursement System (COD).** Each year, ED awards more than $30 billion in Pell Grants to undergraduate students (Soldner and Campbell 2016). COD stores information about the disbursement of these grants, as well as disbursements from other aid programs operated by FSA.

NCES uses NSLDS, CPS, and COD data in five ways.

1. **Verification of** FTB **status.** As noted above, NCES postsecondary sample surveys rely on a two-stage sampling design. In the first stage, the data collection contractor samples postsecondary institutions. Those institutions then submit enrollment lists, which the contractor uses in a second stage to sample potential student respondents. When a given NPSAS study is designed to serve as the base year data collection cycle for BPS, those enrollment lists include an indicator of whether the institution believes that a student is an FTB student and therefore eligible for longitudinal follow-up. In 2011, NCES began matching enrollment lists to FSA data systems in an effort to confirm the validity of the FTB indicators reported by students, out of a growing concern for false-positives. In that 2011 match, NSLDS was used to attempt to verify the FTB status of more than

2.1 million enrolled students, revealing a false-positive rate of nearly 20 percent (Hill et al. 2016).

2. **Verification and augmentation of enrollments and enrollment spells.** Because Pell Grant and loan data maintained in NSLDS and COD include information about the institution that authorized the disbursement of aid, matches to FSA data systems allow NCES to identify periods of enrollment for federally-aided students even when they are not reported by students in the interview. Enrollment spell data are augmented by information provided by students during the student interview.
3. **Verification of graduation for loan holders.** Institutions have always been responsible for advising FSA when a student borrower was no longer enrolled. As part of those enrollment reporting responsibilities, institutions were also to have indicated whether a student was no longer enrolled because he or she graduated. However, results from a 2011 NCES analysis suggested that graduation flags found in NSLDS were often inaccurate. In 2012, this triggered FSA to issue Dear Colleague Letter GEN-12-06, admonishing institutions to improve data quality. That letter, coupled with the transition away from student loans authorized under the Federal Family Education Loan program, has improved NCES capacity to use NSLDS data to verify completion among aid recipients.
4. **Verification of loan and grant amounts.** FSA data systems aggregate data from institutions, loan servicers, and other financial partners to catalog the types and amounts of federal aid students receive. As ED has promulgated new regulations governing student aid programs, institutions have been required to provide additional data about federal aid recipients beyond the characteristics of their aid packages; the potential of that new data to augment NCES datasets is discussed below.
5. **Addition of repayment history.** Once a student loan has been originated, FSA data systems track its status until it is satisfied through repayment or discharge. By matching to NSLDS, NCES data collection contractors are able to capture data that allow analysts to learn how, and how quickly, borrowers are repaying their federal student loans. This includes information about students' selection of repayment plan and whether students experienced delinquency (for Direct Loans) or default.

NATIONAL STUDENT CLEARINGHOUSE

For more than 20 years, NSC has supported institutions that participate in Title IV FSA programs by helping them comply with FSA enrollment reporting requirements (Dundar and Shapiro 2016). These requirements result in complex data needs as changes in enrollment intensity, withdrawal, and completion trigger loans to enter repayment. Therefore, accurate and timely data are critically important to students, loan servicers, and the federal government. To do so, NSC asks institutions to provide lists of *all* enrolled students, both aided and un-aided, on a monthly basis.

As a result, NSC has amassed a record system that includes the complete enrollment histories of a significant majority of students enrolled in degree-granting Title IV participating institutions nationwide. Public reports from the Clearinghouse suggest that about 84 percent of U.S. colleges and universities participate in its services, covering about 97 percent of all postsecondary enrollees (Dundar and Shapiro 2016). While NCES most recent match to NSC, conducted as part of NPSAS:12, suggests a lower coverage rate—80 percent of sampled students were found in NSC databases, including only 48 percent of students enrolled at less-than-2-year institutions—NSC remains a critical source of administrative record data for NCES (see Table 45 in Wine, Bryan, and Siegel 2014).

Over time, NSC has augmented data on dates of students' enrollment with other student-level data elements that support their own products and research and for regulatory compliance purposes. Those supporting research include race, gender, high school code, current major, and a series of flags indicating a student's status as (1) degree-seeking, (2) first-time, full-time, (3) a veteran, (4) a Pell Grant recipient, and (5) participation in remedial coursetaking. Those supporting ED regulations, motivated by limitations on the maximum length of time students are eligible to borrow subsidized student loans, includes detailed information for up to six programs of study, including each program's name, defined by NCES Classification of Instructional Program code, and credential level (e.g., certificate, associate's degree), length, date it was begun by the student, and the student's enrollment status in that program (NSC 2014).[5,6]

NCES uses NSC data in three ways.

1. **Verification of FTB status**. As noted above, NCES and its data collection contractor have redoubled their efforts to ensure the accuracy of institutions' identification of students as FTBs. In addition to linking to FSA data systems for disconfirming evidence of a student's status as an FTB, a similar match is conducted against NSC data systems, looking for evidence of historic enrollments.
2. **Verification of enrollment and enrollment spells.** Although NPSAS, BPS, and B&B all ask students to provide detailed enrollment histories for the purpose of asking detailed questions about specific enrollment spells, those data are supplemented by information provided by schools to NSC.
3. **Verification of awards conferred.** NCES data collection contractor uses NSC information about students who have been awarded certificates and degrees to augment longitudinal data about completion outcomes.

[5] National Student Clearinghouse. (2014). *Enrollment Reporting Programming & Testing Guide*. Retrieved from http://www.studentclearinghouse.org/colleges/files/EnrollRept_ProgrammingandTestingGuide.pdf.
[6] For more information about the Classification of Instructional Programs (CIP), visit the NCES CIP site at https://nces.ed.gov/ipeds/cipcode

NCES use of multiple data sources for these and other data elements sometimes results in the need to resolve inconsistencies between data providers. Over time, NCES has developed variable-by-variable trumping rules to ensure consistency of data processing, based upon their expert judgment of which provider (e.g., student interview, institutional data, federally-held data) is most likely to yield the highest quality data. Data File Documentation released with each postsecondary sample survey details the priority of sources for each derived variable found in the data file.

COLLEGE TESTING SERVICES

Historically, NCES and its data collection contractor have sought to use data from college testing services, such as the College Board and ACT, to augment data found in the NPSAS family of studies. Concepts of interest have included two domains: the description of students' precollege characteristics and experiences, and of postbaccalaureate outcomes. Only the former have routinely been included in postsecondary sample surveys, including students':

- high-school coursetaking behaviors, including whether the student took honors courses and broad indicators of the content of their high school curriculum (e.g., years of math taken);
- high-school GPA; and
- SAT and ACT scores.

Leveraging data held by college testing services to gain a better understanding of student outcomes after college has proven difficult. In recent years, NCES has considered using administrative data matches to capture graduates' scores on a variety of examinations, including the GMAT, GRE, LSAT, MCAT, and PRAXIS, so that they might be included in B&B. Unfortunately, the incidence of these tests is sufficiently small within the context of a 16,000 person sample survey as to make any match of little analytic value. As a result, these efforts have been largely abandoned.

LEVERAGING LINKAGES TO FEDERAL DATA SOURCES

Although there is already extensive use of administrative data in the NPSAS family of studies, there are several opportunities for NCES to make better use of federally held administrative data to improve data quality, reduce respondent burden, and increase the analytic capacity of its postsecondary sample surveys.

Below, we consider three linkages that have already been the subject of extensive conversation in the higher education data policy community, spurred largely by the work of David Bergeron, Senior Fellow at the Center for American Progress and a former Acting Assistant Secretary for Postsecondary Education at ED. Additional information about each can be found in Bergeron's (2016) *Leveraging What We Already Know: Linking Federal Data Systems*, recently

commissioned by the Institute for Higher Education Policy. They include: (1) using IRS data to better understand the price students pay for college, (2) leveraging tax or Unemployment Insurance data to gain insight into students' post-college wage outcomes, and (3) collaborating with the Department of Defense of Department of Veterans Affairs to study the effects of educational benefits afforded veterans, active duty military, and their beneficiaries.

EXAMPLE 1: NET PRICE AND IRS FORM 1098-T

One of the most fundamental concepts in the NPSAS family of studies is the price that students pay for college. Unfortunately, as most analysts already know, there are a variety of price measures, each with distinctly different meanings and uses. The most common include

- total cost of attendance, which comprises tuition and fees, room and board, books, supplies, transportation costs, and costs associated with select personal and other activities;
- tuition and fees alone; and
- net price, defined as total cost of attendance minus grant aid.

Of these three concepts, net price is typically thought of as being most analytically useful. Tuition and fees, for example, fails to include a variety of other important expenses associated with attending college. Total cost of attendance, also known as "sticker price," includes these other expenses but fails to account for often significant offsets to price due to federal grant aid, grants from states, and institutional need or merit-based scholarships. In contrast, net price reflects the actual, immediate price that a student must meet—be it through work, assistance from friends or family, or student loans—to attend a postsecondary institution.

Currently, NCES derives an estimate of each student's net price by collecting two series of data from institutions, matches to FSA data systems, and students: individual price components (e.g., tuition, fees, books) and individual offsets to that price (e.g., federal grants, institutional grants, outside scholarships). Once collected, NCES data collection contractor sums the price components (BUDGETAJ), subtracts all grant aid (TOTGRT), and produces an estimated net price (NETCST3). These data figure prominently in NPSAS and are collected on a longitudinal basis in BPS.

Bergeron (2016) correctly identified that there is another source for net price data, one that is updated annually and includes virtually all students enrolled in postsecondary education: filings made by colleges and universities to the Internal Revenue Service (IRS) that support the administration of the American Opportunity, HOPE, and Lifetime Learning tax credit programs. These data, recorded on IRS Form 1098-T, include:

- identifying information for the institution and the enrolled student;

- the total amount an individual student was billed (price) and the total payments received by the institution for qualified tuition and related expenses; and
- the total amount of scholarships and grants processed by the institution for an individual student (offsets).

Using the 1098-T, net price can be calculated as the difference between the amount a student was billed and what an institution reports as his or her total grant aid. As we discuss below, there are arguments in favor and against such an approach.

Implications for Data Quality

No publicly available study has compared calculated net price from NPSAS to institutionally reported net price via Form 1098-T. As a result, it is impossible to determine the extent to which these two measures vary and under what circumstances. However, it is known that the Department of the Treasury's Office of Tax Analysis (OTA) has identified a number of shortcomings with the 1098-T as currently implemented. Some of these may negatively impact data quality, including a lack of clarity about the inclusion or exclusion of some costs and offsets on the form itself and challenges aligning tax years with institutions' academic calendars. OTA has also noted that not all students are required to receive the 1098-T, most notably nonresident aliens and exclusively noncredit students (Ackerman, Cronin, and Turner 2014).

IMPLICATIONS FOR BURDEN REDUCTION AND ANALYTIC CAPACITY

If NCES were to gain access to 1098-T data for the purposes of calculating net price as part of the NPSAS family of studies, institutions could accrue a small reduction in burden. Students would likely be unaffected.

As noted above, net price is a function of two series of data: itemized price and itemized offsets. Data on prices are typically provided by institutions during the student record collection process. Although they are collected at a fairly granular level, they are also readily known to institutions due to their important role in the student aid packaging process. These price data could be replaced by information included on a 1098-T, reducing institutional burden.

Data on offsets, however, are a very different story. They are provided primarily by institutions and can vary widely from student to student. Although there are relatively few federal grant programs, states and institutions can have dozens of waiver, grant, and scholarship programs in support of their students. Each of these programs must be enumerated by institutions and categorized by type (e.g., merit-based) prior to data collection, and then reported separately to NCES for each sampled student. Replacing these disaggregated data elements with a more generic "total amount of scholarships and grants" from the 1098-T would not likely be acceptable to most analysts.

Student respondents to NPSAS are not expected to report on federal, state, or institutional grants. They are, however, asked about grant aid they receive from other, outside sources (e.g., a church scholarship paid directly to the student). Because these data are known only to the student, related survey items would not likely be dropped were 1098-T data suddenly available.

Importantly, simply because 1098-T data are not immediately valuable within the context of NCES sample surveys does *not* mean they are not important to ED or NCES more generally. As Bergeron (2016) notes, these data would make it possible for ED to improve consumer information tools related to the cost of college-going, a decade-long Department priority.

EXAMPLE 2: WAGES, INCOME, AND EMPLOYMENT DATA FROM W-2S AND UNEMPLOYMENT INSURANCE REPORTING

Wages, income, and employment data are closely related in the NPSAS family of studies. For the sake of simplicity, wages are considered to be individuals' base compensation from work. For some individuals, this may differ from income, which can include both earned wages and income from non-work sources like investments. This distinction notwithstanding, both can reflect an important outcome of postsecondary education: education is a human capital investment that, hopefully, is exchanged for monetary rewards in the labor market. While data on wages and income take the form of dollars, data on employment are contextual and include occupation, industry, and work intensity. Perhaps not surprisingly, both wages/income and employment are central to B&B. Because its cohort is based on beginning students and follows respondents for only 6 years, BPS's capacity to explore employment outcomes for all but short-cycle degree-seekers is limited.

Although there are a variety of federal agencies that receive information about the wages of workers in the United States, the overwhelming majority of data have their source in one of two primary data collections: employers' annual submission of workers' W-2s to the Social Security Administration (SSA) or their quarterly submissions of Unemployment Insurance (UI) wage data to state workforce agencies. The former, effectively tax filings, are maintained exclusively by the SSA and the IRS. The latter, UI wage records, are transmitted by states to federal partners including: (1) the U.S. Department of Labor for use in its Wage Record Interchange System (WRIS/WRIS2), (2) the Department of Health and Human Services as part of its National Directory of New Hires (NDNH), and (3) the Bureau of the Census and its Longitudinal Employer-Household Dynamics (LEHD) project. A detailed review of the strengths and weaknesses of each federal source of wage and employment data, including caveats for federal and military employees, can be found in a recent review by the Workforce Data Quality Campaign (Zinn 2016).

As NCES considers potential linkages to any federal data source on wages and employment, they should consider balancing two factors: relative ease versus additional analytic utility.

- **Relative ease.** ED currently maintains three linkages to administrative data on wages that operate at significant scale. The first, with the SSA, was developed in support of ED's Gainful Employment regulations and is used to calculate program-level median wages for students enrolled in certificate programs at all institutions and all programs at for-profit colleges and universities (see https://ifap.ed.gov/GainfulEmploymentInfo/indexV2.html). Importantly, that agreement does not allow SSA to return record-level data that could be directly matched to a specific B&B or BPS respondent (Bergeron 2016). NCES attempted to establish an agreement with SSA that sought to generate plausible values for wages when those data were not available from the student interview. However, because express legislative authority to do so has not been found, the effort to date has failed. A similar agreement with the Department of the Treasury produces institution-level wage estimates for Education's College Scorecard (see https://collegescorecard.ed.gov/data/).

  The second linkage is ED's relationship with the IRS around the FAFSA-IRS Data Retrieval Tool (DRT). When students complete the web-based version of the FAFSA, they (and their parents, if applicable) are asked to give consent for the DRT to automatically import tax data directly into their current year's FAFSA (FSA, n.d-b.). This includes both wages from work and income from other sources. One might imagine that a similar tool—one not tied to filing for federal student aid—could be built directly into the NPSAS, B&B, or BPS interview and that, over time, some students' familiarity with the FAFSA's IRS DRT might smooth the adoption of a similar tool within the context of NCES sample surveys.

- **Additional analytic utility.** Although ED does have agreements with some agencies that maintain UI wage records, none of these existing arrangements operate at scale. (An example includes the Secretary of Education's statutory authority to match to NDNH, but only for the purpose of debt collection [Administration for Children and Families 2015].) Prior efforts, most notably an attempt to merge UI wage records with B&B and make the resulting data available within a Federal Statistical Research Data Center, administered by the U.S. Census Bureau (www.census.gov/fsrdc), failed due to administrative hurdles. However, whereas tax records *only* contain information about wages, UI wage records often include contextual information about workers' employment circumstances. Currently, UI wage records include much of what would be required to create a reasonably detailed work history, as they typically include a worker's employer, the period of time covered by a wage report, the employer's industry, and, in some cases, measures of work intensity. In the future, additional data may be added to UI wage records, including workers' occupations, a boon to those who

wish to understand the relationship between employment outcomes and the alignment between programs of study and specific occupations (Zinn 2016).

Implications for Data Quality

There is no doubt that increased reliance on administrative sources of data for wages, income, and employment—if those data could be returned on an individual level—would yield higher quality estimates than the self-reported data currently found in the NPSAS family of studies (see Moore, Stinson, and Welniak 2000) . Leveraging aggregate data to improve imputation may also improve data quality for survey nonrespondents, though that is an empirical question that merits further study. As discussed in Zinn (2016), there is little difference in data quality between federal sources, though IRS, SSA, and LEHD have slightly higher levels of worker coverage than WRIS or NDNH.

Implications for Burden Reduction and Analytic Capacity

Currently, data on wages and employment found in the NPSAS family of studies are provided by survey respondents. Although providing that data for a single year is not necessarily onerous, doing it over a longer period of time—including the multiple years that may fall between B&B administrations—certainly could be. In addition to reducing the surveys' length and level of cognitive demand, moving to a records-based approach on wages and income may reduce the potential for the interview process to be perceived as intrusive.

The potential impact on the analytic capacity of NCES sample surveys varies based upon the final linking approach used. UI wage records already contain more "information" than do wage records, and all indications are that the information value of UI wage records are poised to grow (Zinn 2016). Additionally, UI wage records are produced on a quarterly, rather than an annual basis, making it possible to create more detailed time series than would be possible with tax records.

Example 3: Veterans, Active Duty Military, and Veterans' Benefits Users

Although a student' status as a veteran or active duty member of the military is not an explicit focus of either NPSAS, B&B, or BPS, understanding the experiences of these two groups is an important public policy priority. Similarly, although veterans' education benefits are not a part of the Title IV FSA program, they play a critical role in helping veterans' (or their dependents) finance their education. It continues to be a priority of NCES (1) to identify active duty military and veterans, by student records collection or other means; and (2) to gather accurate data about the use of veterans' education benefits, most typically those offered under the Post-9/11 GI Bill.

IMPROVED DATA ON VETERANS AND ACTIVE DUTY STATUS

The proportion of all students enrolled in postsecondary education who are veterans is small. According to NPSAS:12, approximately five percent of all students, both graduate and undergraduate, are veterans, active duty, reservists, or members of the National Guard. Furthermore, between three and four percent of all students use veterans' benefits and military tuition grants to help pay for college (NCES 2016a, b). Because these students represent a relatively small share of any NPSAS sample, and a particularly small one in B&B and BPS cohorts, calculating precise estimates for these important subgroups is difficult.

Increasing the precision of estimates for veterans and service members requires larger absolute sample sizes within NPSAS, B&B, and BPS. To achieve that goal, NCES has sought to oversample these groups. In other words, NCES would intentionally sample them into the study at a higher rate than their natural occurrence in the population in order to have sufficient numbers in the sample to provide stable estimates. To do so effectively, a student's status as a veteran or service member must be provided by sampled institutions on enrollment lists, much like institutions currently do to indicate FTBs for oversampling into BPS or graduating seniors for oversampling into B&B.

Oversampling this population on the basis of indicators on enrollment lists has proven difficult. According to anecdotal feedback provided to NCES from Technical Review Panels, institutions have incomplete knowledge of students' current or prior military status because many veterans and service members opt to not disclose their status for fear of undesirable special attention. As a result, NCES has set its sights on other potential sources for information about who is, and who is not, a veteran or member of the military: going directly to the Department of Veterans Affairs (VA) and the Department of Defense (DoD).

As recently as the 2015-16 NPSAS, NCES has made substantial progress in improving its ability to collect data about veteran status and veterans' benefit use. NCES managed to negotiate a memorandum of understanding with the VA for NPSAS:16, allowing them to identify education benefit applicants and recipients. Because NPSAS:16 has not yet been released, what other data may be available to analysts cannot be known. What is clear, however, is that this match cannot provide information about non-applicants and active-duty military.

Collecting additional data about veterans who are not benefits applicants may be possible through matching to the VA's DoD Identify Repository, also known as VADIR. According to the VADIR System of Records Notice (SORN; 74 FR 37093), it is permissible that:

> The name(s) and address(es) of a veteran may be disclosed to another federal agency or to a contractor of that agency, at the written request of the head of that agency or

> designee of the head of that agency for the purpose of conducting government research necessary to accomplish a statutory purpose of that agency.

ED is statutorily mandated to conduct NPSAS (20 U.S.C. § 1015), suggesting that this clause in VADIR's SORN may reflect an opportunity for more robust record matching.

Finally, at least in principle, identifying who is an active member of the military should be relatively simple. ED currently maintains an agreement with the DoD's Defense Manpower Data Center (DMDC), for the purpose of identifying the children of service members who died 7performing military service in Iraq and Afghanistan after 2001 (77 FR 38610). Matching opportunities also exist beyond ED's existing agreements. As part of the Servicemembers Civil Relief Act (SCRA; 50 U.S.C. §§ 501), DMDC maintains a publicly facing website for financial service providers that allows them to verify individuals' active duty status, either one at a time or via file upload, after providing those individuals' social security numbers and last names (DMDC 2014). It is not hard to imagine how NCES and its data collection contractor might leverage something like the SCRA website to generate its own military status flag prior to sampling.

Improved Data on Military Education Benefits

Notwithstanding what may come of NCES newest match to the VA, information about VA and DoD benefits are primarily collected by the NPSAS student interview, supplemented by institutional records (NCES, 2016a, 2016b). As of fiscal year 2013, about 70 percent of VA education program beneficiaries were accessing their benefits through the Post-9/11 GI Bill, with the bulk of the remaining 30 participating in the Montgomery GI Bill and Vietnam Era programs (National Center for Veterans Analysis and Statistics n.d.). For the purposes of data collection, the primary differences between these programs is the payee. Prior to the Post-9/11 GI Bill, payments went directly to veterans' beneficiaries, making survey respondents the primary source of information on how those benefits are being used. With the advent of Post-9/11 GI Bill, at least some of the benefit—that which is to be applied to tuition and fees—is paid directly to the postsecondary institution (VA n.d.) As a result, NCES has only partial information about the value of these benefits from existing administrative data resources. The remainder must come from the students themselves.

Implications for Data Quality

Improved information about military and veteran status benefits NCES in two ways. First, when these data are available during sampling, they provide NCES and its data collection contractor a mechanism to oversample, creating more precise estimates for this subgroup. Alternatively, were these data not available until the data processing phase they still could help to ensure that status flags found in NPSAS, BPS, and B&B were accurate. Similarly, improved information about the *amount* of benefits received under education programs administered by the VA could

be used to verify information provided by institutions about tuition and fee payments on behalf of beneficiaries and information provided by the beneficiaries themselves about other payment received. The extent to which the new NCES VA match can provide useful information on this latter point remains to be seen.

IMPLICATIONS FOR BURDEN REDUCTION AND ANALYTIC CAPACITY

Reporting veteran and military status information may not be particularly burdensome for students or institutions, but, as noted above, it may also not be accurate. Relying upon existing FSA matches, augmented with queries against SCRA and VADIR, would reduce whatever burden it does pose. Were information about the *amount* of benefits provided to beneficiaries provided administratively, it would represent one less data element in the student records collection. It would also spare survey respondents the challenge of calculating the total value of their education benefits.

Both the DoD and the VA have substantially more information about members of our military than simply their identities and their use (including use by their beneficiaries) of education benefits. This includes detailed information about their service, including rank, dates of service, pay and benefits, and military occupational specialty or equivalent. However, it is still unclear whether the availability of these data would meaningfully advance the analytic capacity for researchers.

POTENTIAL BARRIERS: LEGAL FRAMEWORKS AND LINKING FEDERAL DATA

Although there are potential benefits to the postsecondary survey program, leveraging potential linkages is not without challenges. If the obstacles were merely technical, such as mismatched student identifiers or computer systems that lacked a high degree of interoperability, they might be relatively easy to resolve. Unfortunately, the largest barriers to improved data linking are legal and regulatory. A patchwork of statutes and interpretation guide what data can and cannot be shared among government agencies and for what purposes.

Among the most well-known laws governing the privacy of student data is the Family Educational Rights and Privacy Act of 1974 (FERPA; 20 U.S.C. § 1232g). FERPA prevents educational institutions from sharing students' educational records without consent. To the extent that FERPA bears upon the NPSAS family of studies, it does so when institutions provide personally-identified student records to NCES data collection contractor. However, federal regulations provide a specific exemption from consent when those records are provided to authorized representatives of the U.S. Secretary of Education [34 CFR § 99.31(a)(3)]. FERPA does not, therefore, represent a barrier to intragovernmental data linkages *per se*, so long as the NCES linking effort did not re-disclose student data to another party.

Instead, it is the Privacy Act of 1974, as amended (5 U.S.C. § 552a), that dictates how government entities may share identified data, once collected from (or about) the public (Grama 2016). To safeguard individual privacy and help prevent the misuse of information contained in Federal records, the Privacy Act (1) restricts what personally-identified data the government may share, and with whom; provides a mechanism for individuals to (2) access data the government has collected about them and (3) amend those data when they believe it to be in error; and (4) establishes "fair information practices … for collection, maintenance, and dissemination of records" (U.S. Department of Justice, Office of Privacy and Civil Liberties 2015).

A key element of the Privacy Act is its requirement that the government disclose in the Federal Register all *systems of records*, defined as virtually any assemblage of data by a government agency that is personally-identified (see 5 U.S.C. § 552a(e)(4) and 5 U.S.C. § 552a(a)(5)). Those notices, as well as any legislation that specifically authorizes a collection of personally-identified data, detail the conditions under which those data can be shared without consent of the individual, both inside and outside of government. It is here that many attempts at administrative data linkages fail.

The Privacy Act's requirement that agencies receive consent before sharing personally-identified records, either inside or outside of government, is absolute unless one of twelve conditions are met (see 5 U.S.C. § 552a(b)). These include sharing compelled by Congress or a court order, requests from law enforcement, and, importantly, what is referred to as occurring during *routine use* (see 5 U.S.C. § 552a(b)(3)). The Department of Justice's Office of Privacy and Civil Liberties (2015) noted that routine use can be established when (1) the public is informed of a use through a Systems of Record Notice published in the Federal Register and (2) the use envisioned is "compatible" with original purposes of a collection.

What does and does not constitute routine use has been the subject of significant litigation and debate (U.S. Department of Justice, Office of Privacy and Civil Liberties 2015). Although the purpose of this paper is not to present a legal analysis of the Privacy Act and its consequences for the types of intergovernmental linkages that might strength postsecondary survey data, it seems unlikely on face that most of the uses envisioned above would pass the routine use test.

Bergeron (2016) has called for a new exception to the requirements of the Privacy Act of 1974, to "specifically provide for the exchange of federal data on individuals for purposes of improving service to students and families and to permit a better understanding of the effectiveness of the federal student aid system and our nation's higher education institutions" (p. 16). An amendment to the Privacy Act, consistent with the spirit of Bergeron's recommendation and making a positive statement that identifiable data can and should be used for the purposes of improving education and training programs that receive federal

support, would no doubt go a long way toward removing barriers to NCES future administrative data linking plans.

Amendments to the Privacy Act would likely be insufficient to remove all legal barriers to the types of intragovernmental data sharing that could ultimately improve NCES sample surveys. As noted above, other federal (and state) law set forth restrictions on how specific data can and cannot be shared, and it seems likely that, at least initially, those laws would take precedence over the Privacy Act.

Particularly relevant for NCES, given its interest in wages, earning, and employment, are federal laws that govern tax and unemployment insurance filings. At the federal level, these are largely governed by the Social Security Act of 1935, as amended (42 U.S.C. § 1306), and the Internal Revenue Code (26 U.S.C. § 6103). Both craft narrowly tailored rationales for data release, permitting disclosure of personally-identified data without affirmative consent to other governmental agencies only when it is directly related to the administration of specific programs. A notable example is a specific disclosure allowing the Secretary of the Treasury to provide the Secretary of Education information about tax filers' adjusted gross income, for the purpose of administering income-driven repayment programs (26 U.S.C. § 6103(i)(13)), and mailing information, for the purposes of remediating over-awards in the Pell program and collecting on defaulted student loans (26 U.S.C. § 6103(m)(4)).

Although current law provides an initial statutory basis for collaboration between the Secretary of the Treasury and the Secretary of Education for the effective administration of Department of Education programs, what it does not do is make an explicit case for the use of taxpayer data for the *evaluation* of such programs—and the NPSAS family of studies is motivated by a statistical purpose, not an administrative one. This suggests that an expansion of Bergeron's (2016) recommendation for an amendment to the Privacy Act might be extended to the Internal Revenue Code. Borrowing from his language, any such amendment might "specifically provide for the exchange of taxpayer identity and data concerning income earned from work and other sources, for purposes of better understanding of the effectiveness of educational programs eligible to participate in federal student aid programs administered by the Department of Education."

OPPORTUNITIES FOR LINKAGES TO NONFEDERAL DATA

Because of their emphasis on student financial aid and students' academic experiences, linkages to federal and institutional systems are the sources of data that are most central to the purposes of the NPSAS family of studies. Notable exceptions include additional data about progression and completion available through the NSC. Linkages to nonfederal data, such as those data that may be housed on social media or maintained by non-institutional providers of education and training, present opportunities for adding ancillary data that may broaden the

analytic capacity of studies like NPSAS. Any of those linkages, however, have distinct shortcomings.

LINKAGES TO REPOSITORIES OF INDUSTRY-RECOGNIZED CERTIFICATIONS AND MICROCREDENTIALS

For the past five years, the federal statistical community has placed a greater emphasis on the role of industry-recognized certifications in building human capital; NCES involvement in the Interagency Working Group on Expanded Measures of Enrollment and Attainment is but one example. This interest has been echoed by those who support workforce development reforms, with projects like the Lumina Foundation-sponsored Credential Transparency Initiative, led by George Washington University, which is working to deepen our knowledge base of their role in preparing adults for work and careers.[7] It is not surprising that the most recent iterations of both B&B and BPS have asked respondents about whether they held industry-recognized certification in addition to their educational credentials. In the most recent of those two administrations, BPS:12/14, approximately one quarter of respondents indicated that they did hold an industry license of certification (Hill et al. 2016). What the current NPSAS family of studies has yet to capture are more detailed data about the certifications respondents hold, such as their issuer or specific type.

Unfortunately, there is no singular registry of certifications or microcredentials, also known as badges, in the United States. Individual credentialing authorities maintain databases of who holds what certifications and credentials, but it is typically the responsibility of the learner to maintain documentation of his/her work. Two approaches for doing so appear common: (1) the use of credentialing authorities' proprietary transcripting systems, and (2) open-source, publicly-facing services that allow learners to "post" their badge and describe its contents.

Certification authorities that offer proprietary approaches toward documenting certification include Apple, Cisco, CompTIA (e.g., A+ or Network+ certification), Microsoft, and PMP (i.e., project management). Each offers some form of verification service that allows others, typically employers, to establish the validity of a certification with only an individual's name and a unique credential number.

In contrast, at least some microcredentialers appear to be relying on an open technical standard known as OpenBadges, developed by Mozilla with the support of the MacArthur Foundation, to make it possible for credential holders to easily post their badges and achievements on-line in a consistent format.[8] But any standard is only as useful as it is *used*. Although it is too soon to judge the extent to which OpenBadges will reach scale, there is evidence it has the support of important stakeholders in the credentialing ecosystem. The

[7] For more information about the Credential Transparency Initiative, visit https://www.credentialtransparencyinitiative.org/

[8] For more information about OpenBadges, visit http://openbadges.org/

Badge Alliance, a membership organization that supports the standard, counts among its members Acclaim/Pearson, BlackBoard, Mozilla Backpack, and DigitalMe, all leaders in digital badging.

How NCES might best leverage extant, but diffuse, administrative data on certifications and microcredentials is unclear. If the Credential Transparency Initiative gains momentum, its goal of forming a registry of credentials (but not credential holders) may provide a useful frame from which NCES could select a subset of certification and microcredentialers for exploratory conversations about whether and how administrative data could be shared. Alternatively, if OpenBadges-based platforms proliferate, it may be possible for NCES to simply "scrape" the resulting public registries. Due to their use of a standardized framework and common meta-data, we might assume that OpenBadges-based registries of any type—public or proprietary—would make any such effort easier, though there can be no guarantee.

LINKAGES TO SOCIAL MEDIA FOR EMPLOYMENT AND LIFE MILESTONE DATA

Increasingly, working adults use social media platforms, such as LinkedIn™, to make public information about their current employment and their employment history. Others use networks like Facebook to announce engagements, partnerships, marriages, the birth of children, and the death of loved ones. Current estimates place the size of LinkedIn's current user base in excess of 138 million adults in the United States (LinkedIn 2017). Facebook currently has 163 million users in the United States (eMarketer 2016). It is not unreasonable to consider whether, in lieu of enhanced state UI wage records, extracts from LinkedIn might not supplement the kind of employment histories that are invaluable in B&B and, to a lesser extent, BPS. NCES could also consider whether Facebook might begin to provide useful data about family formation and other life milestones, in lieu of the current student interview.

Some (if not many) in the research community may question the validity of data available from social media. This is a fair concern. However, there is no reason to assume that respondents to NPSAS, B&B, or BPS are more honest when responding to a web-based interview than they are when building their social media profile. Indeed, the social nature of LinkedIn profiles may make it less likely that they are wildly incorrect. Because others can view what one posts to social media, this opens the door to peer fact-checking. In contrast, data provided to NCES as part of a statistical data collection are confidential.

Notwithstanding any technical, legal, or financial challenges in executing linkages to data like those maintained by LinkedIn or Facebook, the greatest impediment to the use of social media for federal statistical purposes is likely optical. The notion that ED is somehow harvesting social media for governmental purposes would likely be be off-putting to many and stoke privacy concerns for others. LinkedIn and Facebook are foremost social networks, and many people have concerns about government intrusion into their personal lives.

LOOKING AHEAD

There are possibilities in the administrative data matching space that NCES and NPEC-S might want to pursue that, for various reasons, remain elusive. Some of the most compelling are not currently feasible because data are held in 50 (or more) discrete state data repositories. They include:

- precollege characteristics of recent college graduates held in state P-20 longitudinal data systems;
- vital statistics databases, including official registries of marriages, divorces, births and deaths; and
- data on unemployment benefit and social service usage, including the Supplemental Nutrition Assistance Program and Temporary Assistance to Needy Families.

Because many of these data are maintained at the state level, help accessing them may come in the form of a national, although perhaps not federal, student-unit record system. Spurred by the work of philanthropies and national higher education advocacy groups, there is an active and elevated level of discussion surrounding how a national student-unit record system might be developed through nonfederal means (Cubarrubia and Perry 2016).

One approach, the "National Federated" model, expands upon a proof of concept developed by the Western Interstate Commission on Higher Education (Prescott and Lane 2016). In this approach, states, either individually or in multiple consortia, would work with a third party to develop the capacity for data exchange between existing, state-held P-20 data systems. To the extent these federations expand the breadth of data they exchange, such as those that might help postsecondary researchers understand workforce and other outcomes, new data may eventually become available to agencies like NCES.

In other cases, data that might be of great interest to postsecondary researchers are not maintained in data systems that are easily accessible for research purposes. Prior survey work has, for example, demonstrated the relationship between greater levels of educational attainment and improved wellness. However, legal and technological barriers make more systematic linkages across relevant administrative record systems practically impossible – effectively cutting off a large array of possible research studies that could greatly improve our understanding of how health and education work in tandem.

CONCLUSION

Increasing the extent to which administrative data, federal or otherwise, are leveraged in NCES postsecondary sample surveys offers three distinct benefits. First, it reduces burden on both institutions and students that NCES asks to respond to its surveys. Second, it allows NCES to benefit from operational efficiencies that may come with burden reduction, including (1) decreased costs associated with non-response conversion, and (2) improved data quality through lessening the need for imputation. Finally, leveraging extant administrative data holds the potential for opening new lines of inquiry and for the further exploration of existing research questions held by postsecondary researchers.

In this paper, we consider five distinct opportunities:

- better understanding the price of college through the IRS Form 1098-T;
- gaining more complete and accurate information about students' post-college wage outcomes through data collected through UI or tax filings;
- detailing the use of educational benefits and other experiences of veterans and active duty military through collaborations with the VA and DOD;
- collecting information about students' acquisition of industry-recognized certifications; and
- leveraging social media to capture data on students' major life events.

Each of these opportunities would represent some level of benefit to NCES and the communities it serves through some combination of burden reduction, decreased cost, improved data quality, and enhanced analytic utility. Their feasibility varies, however, as does their value to important stakeholder groups. As a result, their exploration is sure to proceed at an uneven pace and with uncertain results. As it does, NCES and NPEC-S can continue to consider how the NPSAS family of studies does what surveys do best: collect data from individual respondents around topics of interest in postsecondary education for which no other source of data—administrative or otherwise—exists.

REFERENCES

Ackerman, D., Cronin, J.A., and Turner, N. (2014). *Improving Form 1098-T: How a Revised Form Could Increase Take-Up, Improve Compliance, and Lower Taxpayer Burden.* Retrieved January 5, 2017, from https://www.irs.gov/uac/soi-tax-stats-2014-irs-tpc-research-conference.

Administration for Children and Families. (2015). *A Guide to the National Directory of New Hires*. Retrieved January 5, 2017, from https://www.acf.hhs.gov/sites/default/files/programs/css/a_guide_to_the_national_directory_of_new_hires.pdf.

Bergeron, D.A. (2016). *Leveraging What We Already Know: Linking Federal Data Systems*. Retrieved January 5, 2017, from http://www.ihep.org/postsecdata/mapping-data-landscape/national-postsecondary-data-infrastructure.

Chetty, R. (2016). *Improving Economic Opportunity in America: New Lessons from Administrative Data.* Retrieved May 26, 2017 from https://www.cep.gov/content/dam/cep/events/2016-07-22/chetty.pdf

Cubarrubia, A., and Perry, P. (2016). *Creating a Thriving Postsecondary Education Data Ecosystem*. Retrieved January 5, 2017, from http://www.ihep.org/postsecdata/mapping-data-landscape/national-postsecondary-data-infrastructure.

Cominole, M., Shepherd, B., and Siegel, P. (2015). *2008/12 Baccalaureate and Beyond Longitudinal Study (B&B:08/12) Data File Documentation* (NCES 2015-141). Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved January 5, 2017, from http://nces.ed.gov/pubsearch.

Defense Manpower Data Center. (2014). *Servicemembers Civil Relief Act Website Users Guide*. Retrieved January 5, 2017, from https://scra.dmdc.osd.mil/appj/scra/welcome.xhtml.

Dundar, A., and Shapiro, D. (2016). *The National Student Clearinghouse as an Integral Part of the National Postsecondary Data Infrastructure.* Retrieved January 5, 2017, from http://www.ihep.org/postsecdata/mapping-data-landscape/national-postsecondary-data-infrastructure.

eMarketer. (2016). *This Year, More than Half of Americans Will Use Facebook.* Retrieved January 5, 2017, from http://www.emarketer.com/Article/This-Year-More-Than-Half-of-Americans-Will-Use-Facebook/1013560.

Federal Student Aid. (n.d-a.) *FAFSA Data by Demographic Characteristics, 2014–2015 Application Cycle.* Retrieved January 5, 2017, from http://studentaid.ed.gov/sa/about/data-center/student/application-volume/fafsa-school-state.

Federal Student Aid. (n.d-b.) *Simple Steps to Transfer Tax Information Into Your FAFSA.* Retrieved January 5, 2017, from https://studentaid.ed.gov/sa/resources/irs-drt-text.

Grama, J. L. (2016). *Understanding Information Security and Privacy in Postsecondary Education Data Systems.* Retrieved January 5, 2017, from http://www.ihep.org/sites/default/files/uploads/postsecdata/docs/resources/information_security_and_privacy.pdf

Hill, J., Smith, N., Wilson, D., and Wine, J. (2016). *2012/14 Beginning Postsecondary Students Longitudinal Study (BPS:12/14): Data File Documentation* (NCES 2016-062). Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved January 5, 2017, from http://nces.ed.gov/pubsearch.

LinkedIn. (2017). *About Us.* Retrieved from https://press.linkedin.com/about-linkedin

Moore, J.C., Stinson, L. L., & Welniak, E. J. (2000). Income measurement error in surveys: A review. Journal of Official Statistics, 16(4), 331.

National Association of College and University Business Officers. (2003). *Colleges and Universities Subject to New FTC Rules Safeguarding Consumer Information* (NACUBO Advisory Report 2003-01). Washington, DC: National Association of College and University Business Officers. Retrieved January 5, 2017, from http://www.nacubo.org/documents/news/2003-01.pdf

National Center for Education Statistics. (2016a). *NPSAS:12 Graduate Students, All Variables by Subject.* Retrieved January 5, 2017, from http://nces.ed.gov/datalab/powerstats [login required].

National Center for Education Statistics. (2016b). *NPSAS:12 Undergraduates, All Variables by Subject.* Retrieved January 5, 2017, from http://nces.ed.gov/datalab/powerstats [login required].

National Center for Veterans Analysis and Statistics. (n.d.) *Department of Veterans Affairs Education Program Beneficiaries: FY 2000 to FY2013.* Retrieved January 5, 2017, from http://www.va.gov/vetdata/Utilization.asp.

National Student Clearinghouse. (2014). *Enrollment Reporting Programming and Testing Guide*. Retrieved January 5, 2017, from http://www.studentclearinghouse.org/colleges/files/EnrollRept_ProgrammingandTestingGuide.pdf.

U.S. Department of Justice, Office of Privacy and Civil Liberties. (2015). *Privacy Act of 1974.* Retrieved January 5, 2017, from https://www.justice.gov/opcl/policy-objectives

Prescott, B.T., and Lane, P. (2016). *Fostering State-to-State Data Exchanges*. Retrieved January 5, 2017, from http://www.ihep.org/postsecdata/mapping-data-landscape/national-postsecondary-data-infrastructure.

Soldner, M., and Campbell, C. (2016). *Using—and Improving—FSA Data Systems to Support Policy Analysis.* Retrieved January 5, 2017, from http://www.ihep.org/postsecdata/mapping-data-landscape/national-postsecondary-data-infrastructure.

U.S. Department of Veterans Affairs. (n.d.) *Comparison Tool/Payment Rates*. Retrieved January 5, 2017, from http://www.benefits.va.gov/gibill/comparison_tool.asp.

Wine, J., Bryan, M., and Siegel, P. (2014). *2011–12 National Postsecondary Student Aid Study (NPSAS:12) Data File Documentation* (NCES 2014-182). Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved January 5, 2017, from http://nces.ed.gov/pubsearch.

Zinn, R. (2016). *Classroom to Career: Leveraging Employment Data to Measure Labor Market Outcomes.* Retrieved January 5, 2017, from http://www.ihep.org/postsecdata/mapping-data-landscape/national-postsecondary-data-infrastructure.