

# TOWARD A BUSINESS CASE FOR SUSTAINED INVESTMENT IN STATE LONGITUDINAL INTEGRATED DATA SYSTEMS

Submitted to:

U.S. Department of Labor  
Employment and Training Administration  
Office of Policy Development and Research  
Division of Strategic Planning and Performance  
200 Constitution Avenue NW  
Washington DC 20210

Authors:

David Stevens, Director and Research Professor  
The Jacob France Institute  
University of Baltimore  
[dstevens@ubalt.edu](mailto:dstevens@ubalt.edu)  
410-837-4729

and

Ting Zhang, Research Assistant Professor  
The Jacob France Institute  
University of Baltimore  
[tzhang@ubalt.edu](mailto:tzhang@ubalt.edu)  
410-837-6551

January 2014

This report is the third in a series of reports completed under a 2010 Workforce Data Quality Initiative award of federal funds to the Maryland Department of Labor, Licensing and Regulation by the Employment and Training Administration of the U.S. Department of Labor. Jing Li contributed superb help throughout the late stages of report production. Institute colleagues John Janak and Treva Stack performed many behind the scene tasks that required their subject matter expertise. At a critical point in the project design process, Jay Pfeiffer offered invaluable perspectives based on his unique understanding of all aspects of longitudinal integrated data system development, operations, and most importantly value through use by decision makers that impact the effectiveness and efficiency of education and training investments. We are indebted to the agency executives that offered the necessary authorizations to conduct the research that is described here. Of course, we accept full and sole responsibility for the content of the report. Agreement or disagreement with the views expressed here should not be attributed to any other person or organization.

# Table of Contents

	<u>PAGE</u>
<b>A. INTRODUCTION</b> .....	1
<b>B. ORGANIZATION OF THE REPORT</b> .....	1
<b>C. LIMITED SCOPE</b> .....	1
<b>D. A COHORT RESEARCH DESIGN</b> .....	2
<b>E. FINDINGS</b> .....	7
1. Community college program assignment of the cohort members.....	9
2. University System of Maryland enrollment and degree counts, 1985-1997....	10
3. Appearance of cohort members in employment administrative records .....	11
4. Frequency of cohort member appearance in the Maryland UI wage record data .....	13
5. Median and 25th and 75th percentile bands of real annual earnings amounts.....	14
6. Five-year segments of the full time span covered .....	16
7. Comparison of 1986-1990 and 2011-2012 time segment actual earnings quintiles.....	19
8. Cohort member movement up or down in the time segment earnings distributions.....	20
9. Comparison of 1985 fourth quarter and 2012 4th quarter industry affiliations.	22
<b>F. CONCLUSIONS AND NEXT STEPS</b> .....	24
1. The business case relevance of our longitudinal visualization tools .....	24
2. Steps states can take to replicate or adapt parts of our research design .....	27
3. Our own future research agenda and priorities .....	28
4. Toward improved clarity of communication .....	29

# List of Figures

	<u>PAGE</u>
<i>FIGURE 1</i> Number of cohort members by fall 1984 community college program assignment and gender	9
<i>FIGURE 2</i> University System of Maryland enrollment and degree counts, 1985-1997 only	10
<i>FIGURE 3</i> Appearance of cohort members in employment administrative records	12
<i>FIGURE 4</i> Frequency of annual appearance in the Maryland UI wage record file, 1985-2012	13
<i>FIGURE 5</i> <i>Panel A:</i> Median real earnings amounts by administrative data source	15
<i>Panel B:</i> 25 <sup>th</sup> and 75 <sup>th</sup> percentile bands of real earnings amounts by administrative data source	16
<i>FIGURE 6</i> Segment earnings—median, mean, 25 <sup>th</sup> percentile, and 75 <sup>th</sup> percentile amounts for those with reported earnings in defined years	18
<i>FIGURE 7</i> MD earnings quintile snapshots 1986-1990 <i>versus</i> 2011-2012	20
<i>FIGURE 8</i> <i>Panels A &amp; B:</i> The 1984 community college program assignments of the cohort members that moved to a higher/lower earnings distribution quintile between 1986-1990 average and 2011-2012 average	21
<i>FIGURE 9</i> Industry sector changes 1985 quarter four to 2012 quarter four	23

## **A. INTRODUCTION**

This report responds to an urgent need. The flow of federal funds invested in the development of state education and workforce integrated data systems is scheduled to end state by state between now and 2016. Prospects for future federal investment in state data system integration will depend upon the business case that is made for doing so, justifying this allocation over popular competing alternatives.

There is scant evidence that a strong business case has been made for continued federal funding and successful promotion of state funding. However, essential components needed to make a strong business case for future investment can be drawn from existing administrative data sources. Illustrative new evidence of state education and workforce integrated data system value appears in the following pages; evidence that can help to inform and improve the effectiveness of future high impact education and workforce expenditure decisions.

## **B. ORGANIZATION OF THE REPORT**

The next section places the content of this report within the complex array of today's state integrated education and workforce data systems. We then describe the cohort research design that underlies the findings that follow. In a final section, we turn to how our approach can be replicated and used to quickly pull together the strong business case that is needed to justify continued investment in appropriate state longitudinal integrated data systems. Here, 'appropriate' means relevant for improvement of future high impact education and workforce expenditure decisions. We end with identification of communication challenges that impede rapid progress.

## **C. LIMITED SCOPE**

To understand how the scope of this report aligns with the current status of federal investments in state integrated education and workforce data systems, it is important to know that:

- Eleven 2010 three-year Workforce Data Quality Initiative (WDQI) awards to states by the U.S. Department of Labor will end this year. Another twelve 2012 three-year WDQI awards will continue into 2015 and six third wave 2013 three-year WDQI awards will end in 2016.
- Forty-six currently active 2009 State Longitudinal Data System (SLDS) and SLDS ARRA awards to states by the U.S. Department of Education will end in 2014. Twenty-two 2012 awards will end in 2015, and two will end in 2016.

The federal WDQI and SLDS awards to states have not been prescriptive of common architecture, content or use requirements or expectations. The U.S.

Department of Education investment in state longitudinal data systems began in earnest in 2006, building out from a k-12 core to include postsecondary, workforce, and most recently early childhood coverage. The U.S. Department of Labor WDQI investment began in 2010, working from a workforce-to-education linkage perspective that continued in the 2012 and 2013 investment cycles.

An intentional result of these non-prescriptive investment strategies is that states are now at very different points along a continuum of segments integration and reporting capacity. The timing and intensity of workforce data integration remains uneven among the states. Relative inattention to the workforce component has retarded the timing and quality of the business case for continued investment. It is noteworthy that the state integrated data systems that are often described today as the Nation's leaders—Florida, Texas and Washington—began decades ago with balanced workforce and education content and priorities for use that aligned with state funding sources and amounts.

The content of this report is intended to speak to the federal, state and local decision makers charged with making prudent investments in public education and training to achieve complementary economic growth and citizen well-being goals. Decision-relevant new information responds to a basic longitudinal integrated data system investment question: What value is added by extensive time-span coverage?

We believe that the types of information presented in this report can, and should, be used to time and target future high impact actions that will produce favorable education and training effectiveness and efficiency results. Rapid progress on this front is needed.

If the flow of federal investment funds ends and is not replaced by state investments, opportunities to realize positive effectiveness and efficiency gains will be lost. A strong business case to ensure this does not happen is needed.

#### **D. A COHORT RESEARCH DESIGN**

A *cohort* is a defined group that has something in common. A cohort study design often relies upon data that have already been collected for reasons unrelated to the research study.

Our opportunity to undertake a cohort study began with authorized access to retained administrative information that had been collected about fall 1984 first-time enrollees in Maryland's public community colleges. The action of simultaneous first time enrollment in a state's public community colleges is the common element that defines these individuals as a cohort.

Leap forward 28 years, from fall 1984 to fall 2012. Again based on authorized access to administrative records that had already been collected for purposes other than our research study, we gained a capacity to analyze selected cohort member

descriptors *circa* 2012 for the former students found in the available 2012 administrative records.

Up to this point, we have described two snapshots of selected information about cohort members—a fall 1985 administrative record dataset that includes all the defined cohort members, and a fall 2012 administrative record dataset that includes new information, but only for those found in the available 2012 administrative data files. None of the information in either of these snapshots was originally collected for our research study. All the data elements had been collected for other administrative purposes.

In addition to the two ‘bookend’ snapshots recorded nearly three decades apart, in 1985 and 2012, we received authorized access to the following administrative records for all or some of the intervening years:

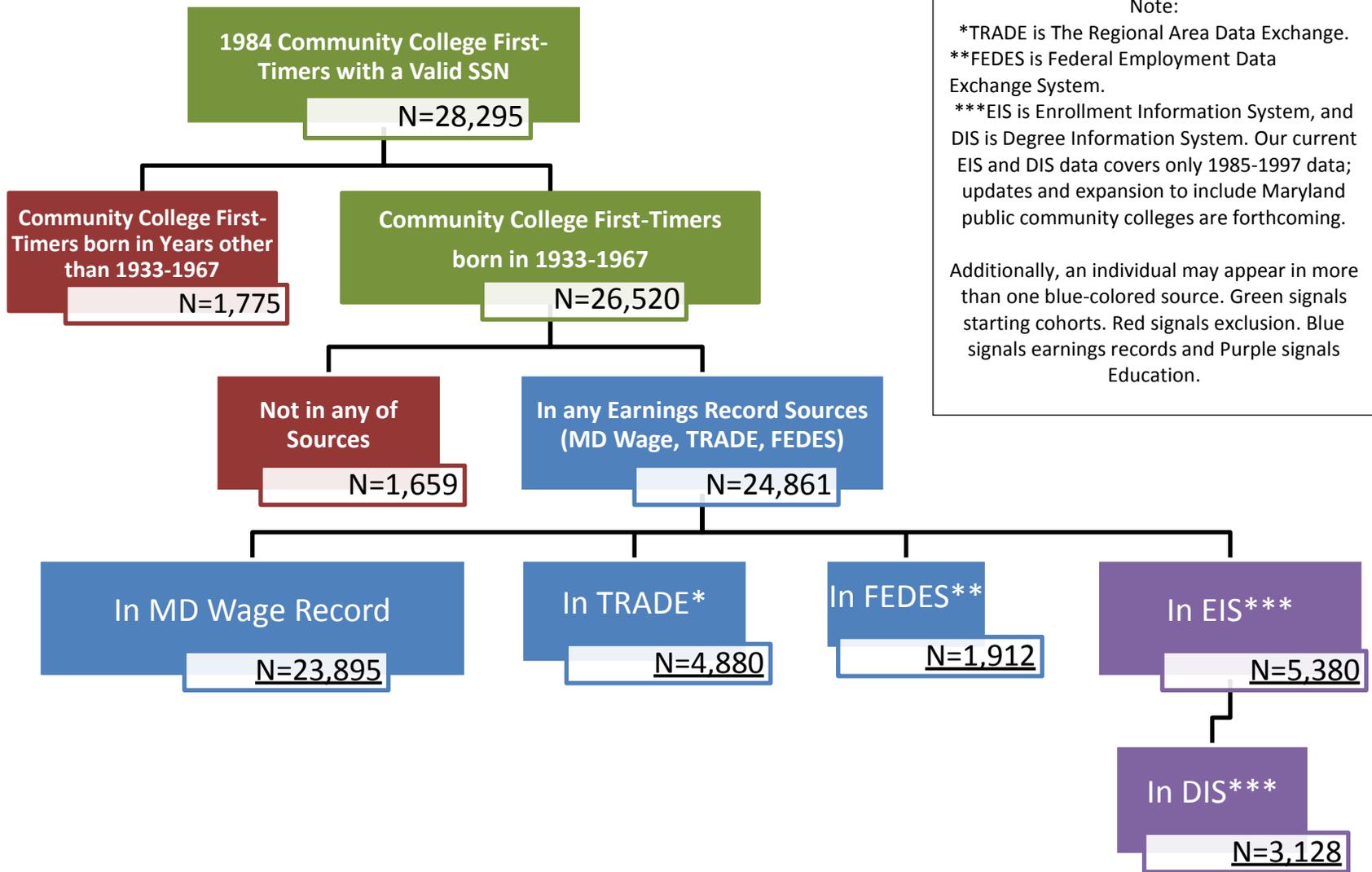
- Maryland UI wage records covering the full time span from April 1985 through December 2012. These records offer in-state quarterly employment status and earnings amount information for cohort members. While only cohort members are included in any of the more than one-hundred year/quarter data draws from 1985 quarter 2 through 2012 quarter 4, the cohort member composition can differ from reference year/quarter to year/quarter as individuals move in and out of covered employment.
- Maryland Quarterly Census of Employment and Wages extract of industry classification information for the employment affiliations found in the Maryland UI wage record file for cohort members.
- UI wage records from the District of Columbia, New Jersey, Ohio, Pennsylvania, Virginia and West Virginia, beginning in 1999 and continuing through the end of 2012.
- Federal civilian employee personnel record extracts beginning in 2003 and continuing through the end of 2012.
- *Enrollment Information System* (EIS) and *Degree Information System* (DIS) administrative record extracts received from the University System of Maryland, temporarily covering only 1985-1997.

The cohort ‘tree’ on page 5 shows what we were able to draw upon to create the single consolidated longitudinal analytical data file that was then relied upon exclusively to report the findings that appear in the next section of this report.

- The cohort N=28,295 includes fall 1984 first-time enrollees in Maryland’s public community colleges that have a valid Social Security number included in the

administrative data file created by the Maryland State Board for Community Colleges.

- Introduction of a date of birth criterion—1933 through 1967—resulted in deletion of 1,775 individuals from the starting cohort. This limits the 2012 age range to include cohort members from age 45 to age 79.
- Based on Social Security number queries of the five defined administrative data files, 1,659 individuals from the starting cohort were not found in any of these five administrative data files for the available time coverage.
- The cohort subpopulation counts in the five available administrative data files are:
  - Maryland UI wage records N=23,895.
  - The regional area data exchange UI wage records N=4,880.
  - Federal Employment Data Exchange System records N=1,912.
  - University System of Maryland Enrollment Information System N=5,380 (through 1997 only).
  - University System of Maryland Degree Information System N=3,128 (through 1997 only).



In sum, we were able to assemble a truly *longitudinal* integrated analytical data file comprised of confidential administrative record extracts drawn from six education and workforce sources covering up to nearly three decades between 1984 and 2012.

This enabled us to study two important topics that are relevant for future decisions about education and training investment priorities and strategies:

- Whether the mix of observed cohort member 2012 industry affiliations and annual earnings amounts aligns with the industry affiliation and earnings distribution information recorded on a defined baseline date.

Based on available late 1980's administrative information, could decision makers at that time have predicted cohort member 2012 industry affiliations and relative earnings amounts, and included this future-cast information in their investment actions? If so, the decision-relevance and value of aggressively marketed short-term employment and earnings measures is bolstered. But, if not, there is cause to rethink whether and how these measures should be incorporated into decisions of consequence.

- Improving public understanding of the limits of what can be learned and acted upon based on longitudinal integrated data access alone, accompanied by specific recommendations about gaps in understanding that will remain unless and until additional relevant data elements are collected.

Throughout this section we have referred to decision makers without further delineation. Now, as we prepare to present our findings in the next section, we urge reader attention to a relevance criterion that we imposed on ourselves in deciding what to present—what role-based decision maker(s) can be expected to recognize the relevance of this information for their own education investment decisions?

Pertinent decision making roles include governors and key legislators that hold the keys to future state education and training investments, future postsecondary students, postsecondary leadership teams, career counselors, and public employment and training program administrators. Individually and together, these investment decisions will impact how the U.S. economy fares and how the benefits and costs associated with this activity will be allocated among us.

## E. FINDINGS

The research findings reported in this section are intended to accomplish two immediate goals:

- Deliver evidence-based advances in the development of a strong business case for continued investment in actionable<sup>1</sup> longitudinal state integrated data systems.
- Offer states a practical research design that can quickly add power to the business case through expanded geographic and time span coverage.

After extensive trial-and-error testing of many visual presentation possibilities, we selected nine graphics that constitute our message of longitudinal integrated data system value. The order of presentation is:

- *Figure 1*—the nine Higher Education General Information Survey (HEGIS) programs that account for 93 percent of the fall 1984 cohort assignments.
- *Figure 2*—the number of cohort members that appear in University System of Maryland Enrollment Information System and Degree Information System administrative records for each of the years 1985-1997.
- *Figure 3*—the number of cohort members that appear in retained Maryland UI wage record files for each of the years 1985-2012, and the count of cohort members that appear in The Regional Area Data Exchange and Federal Employment Data Exchange System files covering only more limited segments of the full 1984-2012 time span.
- *Figure 4*—a summary of how the cohort members are distributed on a continuum defined by how often each individual is found in the Maryland UI wage record database. There are 28 possible appearances at the annual level of granularity—1985-2012.
- *Figure 5*—a two-panel presentation. Panel A shows median inflation adjusted (real) annual earnings for the cohort members that were found in two relevant administrative data sources—Maryland UI wage records, and other states UI wage records for the limited time span coverage of The Regional Area Data

---

<sup>1</sup> While we and others have referred to the importance of *actionable* information for many years, the Actionable Intelligence for Social Policy (AISP) initiative, funded by the John D. and Catherine T. MacArthur Foundation and led by University of Pennsylvania professors Dennis Culhane and John Fantuzzo, has honed the power of this word to a fine art, resulting in a continuing flow of compelling findings that carry into budget allocation debates and decisions in their chosen scope of influence. Planning has begun for an April 2014 research workshop that will concentrate on examples of decision-relevant research based on successful longitudinal integration of education, workforce and social services administrative data

Exchange initiative. Panel B adds the 25<sup>th</sup> percentile and 75<sup>th</sup> percentile inflation adjusted earnings bands around these annual inflation adjusted earnings levels.

- *Figure 6*—a four panel display of both nominal and real (inflation adjusted) annual earnings for five five-year segments and the most recent two-year segment of the full 1986-2012 time span<sup>2</sup> for two cohort subgroups—those that appear in any one or more of the five years of a reference time segment; and those that appear throughout the twenty-eight years.<sup>3</sup>
- *Figure 7*—a matrix showing the annual earnings quintile assignment of cohort members that appear in the Maryland UI wage record file for both the 1986-1990 five-year and 2011-2012 two-year time span segments. This matrix addresses the key topic of whether short-term earnings measures have predictive value about anticipated earnings over a longer working life-cycle time span.
- *Figure 8*—a two-panel pie charts presentation of the HEGIS code mixes for cohort members that moved up, or down, in the time segments earnings distributions shown in *Figure 8*. Refinement of this to include full Enrollment Information System and Degree Information System updates will add important new insights and demonstration of longitudinal integrated data system value.
- *Figure 9*—a matrix showing the fourth quarter 1985 and fourth quarter 2012 industry affiliations of cohort members that appeared in the Maryland UI wage record file in both snapshot quarters and corresponding assigned industry affiliation codes.

Following this presentation of findings to date, the final section of this report includes our plans for building upon this foundation of new knowledge. The Enrollment Information System and Degree Information System data covering most of the full time span 1985-2012 for all Maryland public community colleges and University System of Maryland institutions will be our highest immediate priority for refinement of this research agenda. Inclusion of cohort member participation in Job Training Partnership Act, Workforce Investment Act, Wagner-Peyser, Temporary Cash Assistance and Supplemental Nutrition Assistance Program activities will also be pursued.

---

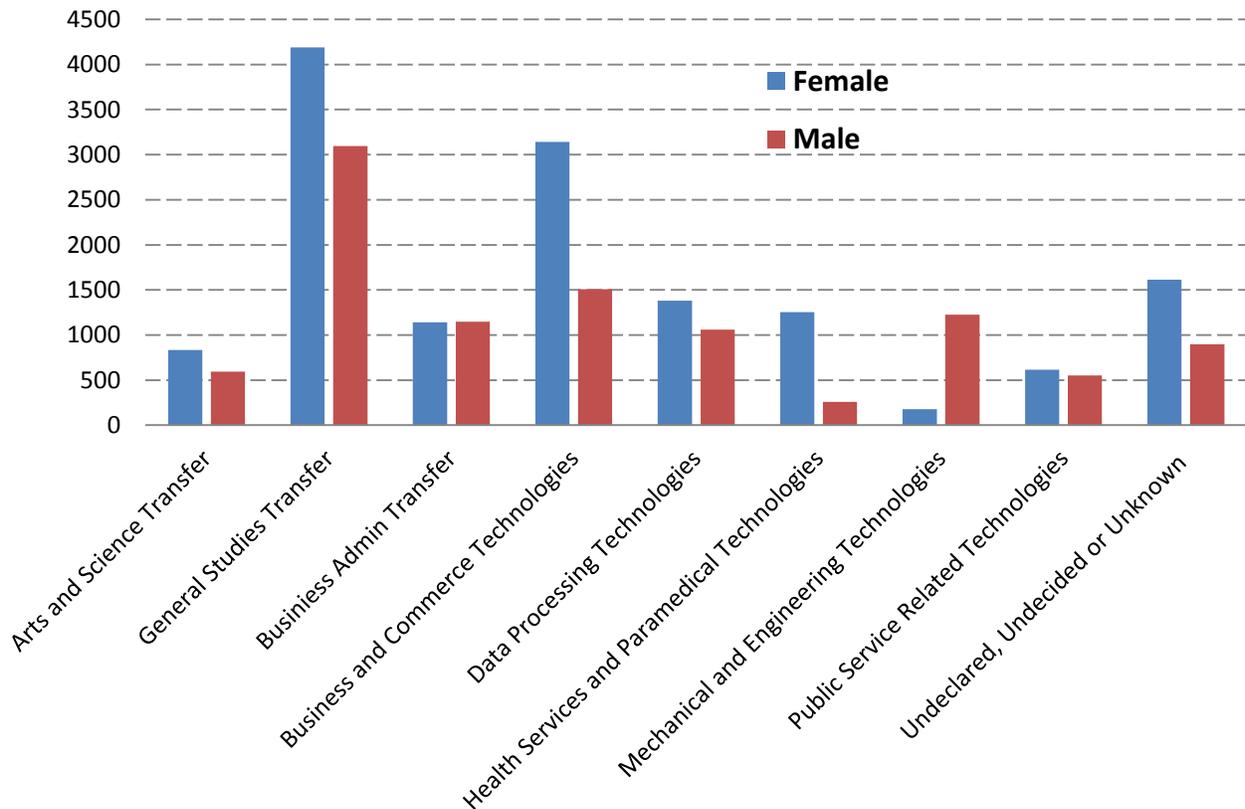
<sup>2</sup> Retained Maryland UI wage records begin in April 1985, so the initial five-year interval begins in 1986; thus what is referred to here as the 'full' time span coverage is 27 years—1986-2012.

<sup>3</sup> UI wage record information is collected quarterly, so appearance in a reference *year* can range from a single quarter to all four quarters. This means, for example, that a cohort member that worked only in a single quarter in each of the five years in a five-year segment—say, part-time employment during the end-of-year holiday season—would be assigned to the subgroup defined as having worked in each of the five years of a five-year time segment.

## 1. Community college program assignment of the cohort members

Figure 1 includes 93 percent of the cohort members, showing the concentration of their programmatic assignments in just nine programs.<sup>4</sup> This information will become a baseline in time, fall 1984, from which we will be able to trace subsequent gender-specific Enrollment Information System and Degree Information System sequences.

**FIGURE 1. Number of cohort members by fall 1984 community college program assignment and gender**

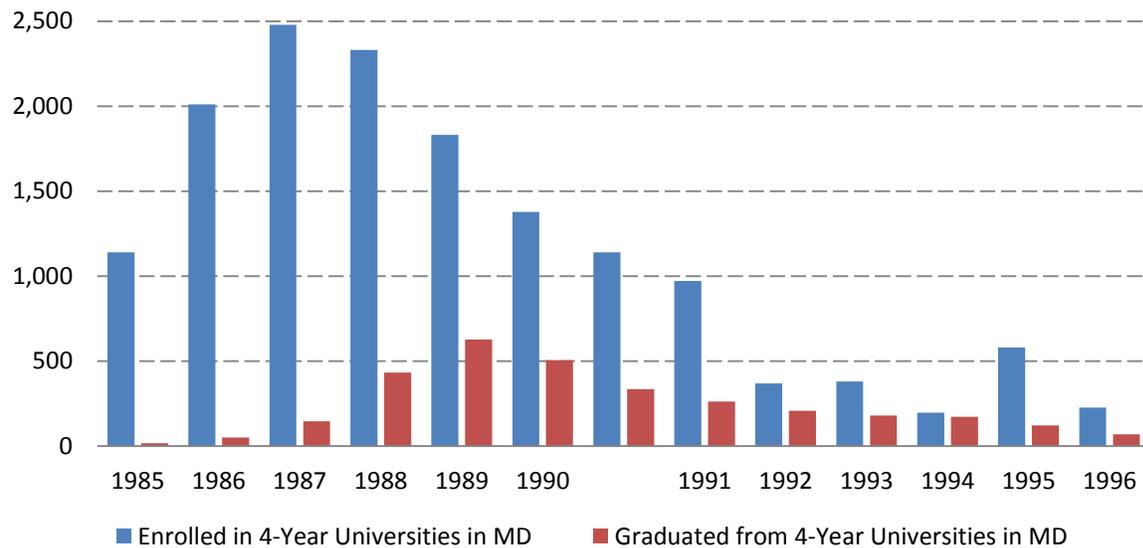


<sup>4</sup> We adopted a program-specific minimum cohort member N of 1,000 criterion to improve the clarity of the presentation.

## 2. University System of Maryland enrollment and degree counts, 1985-1997

We see in *Figure 2* why longitudinal data integration is valuable. With the limited time span and institutional coverage constraints, 11.8 percent of the cohort members are known to have received at least one degree from a University System of Maryland institution.

**FIGURE 2. University System of Maryland enrollment and degree counts, 1985-1997**



### 3. Appearance of cohort members in employment administrative records

*Figure 3* shows the timing of 24,861 cohort members that are found in three administrative data sources—Maryland UI wage records, The Regional Area Data Exchange (TRADE) UI wage records, and the Federal Employment Data Exchange System (FEDES). This is the third step in our integration of longitudinal administrative data—advancing from documented fall 1984 first time enrollment in a Maryland public community college, through an overlay of subsequent in-state public postsecondary enrollments and awarded credentials, to this alignment of documented employment statuses over the same extensive time span.

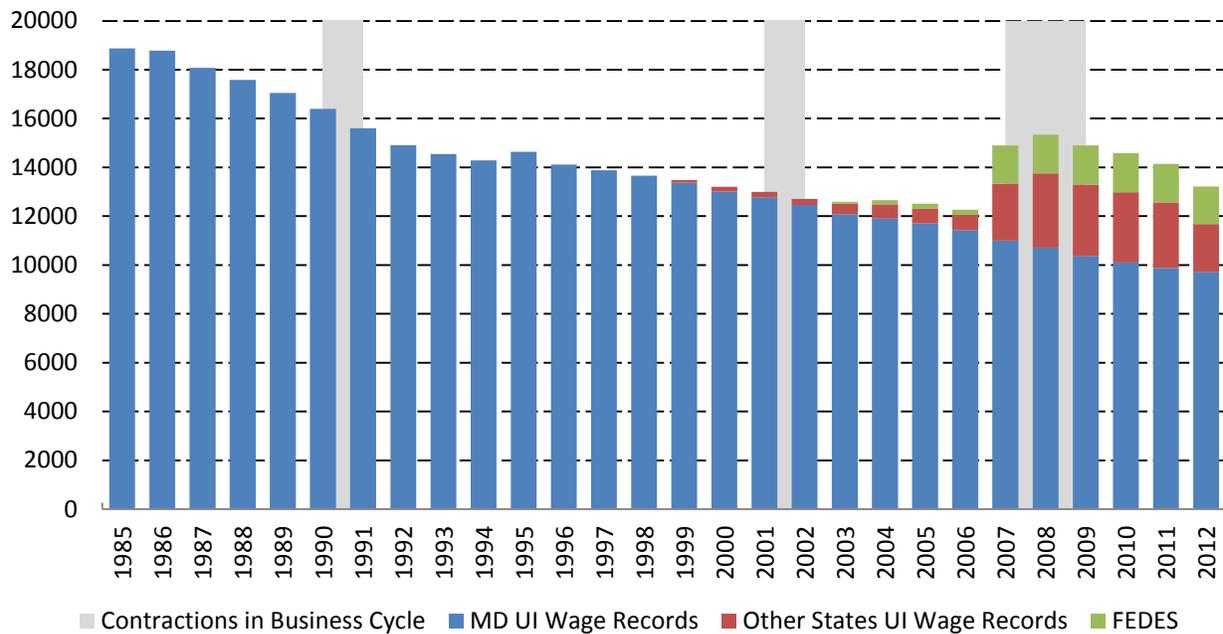
We urge exercise of three cautions to avoid misinterpretation of the color-coded information in *Figure 3*.

1. The mid-Atlantic regional exchange of UI wage record extracts for authorized performance reporting and research purposes did not begin until 1999. The Federal Employment Data Exchange System quarterly cycles of federal civilian employment status did not begin until 2003. So, the pop-up of these source-specific counts in the respective collection start year and continued appearance thereafter is an artifact of the data collection process alone, not indicating anything about cohort member employment transitions. We have a capability to study such transitions, and we have conducted preliminary diagnostics of this type, but further pursuit of this line of inquiry has been deferred for now.
2. Another *Figure 3* caution will be of increasing importance as our future research agenda develops. The availability of the federal civilian employee data began just before the December 2007-June 2009 national recession shown as one of three shaded bars in *Figure 3*. This could lead some to mistakenly conclude that there was a large increase in the number of cohort members that are federal civilian employees that might somehow be related to the recession. While this may be a correct conclusion in some cases, this should not be inferred from this graph alone. We do have a capability to study relevant transitions, as was acknowledged in the previous paragraph, and we do plan to include study of this topic in our broader commitment to use the longitudinal integrated data system capability to understand more about the causes and consequences of recession-based employment transitions. We share the substantial interest expressed by many policy and research colleagues about how state longitudinal integrated data systems can make a positive contribution to more informed, and therefore better, decisions about education investments when economic conditions worsen.
3. Finally, it is important to understand the person-specific composition of each reference year's source-specific count of cohort members is unique to that reference year. That is, while there may be only small changes in the source-specific count from one year to the next, there may be substantial differences in the person-specific composition within each of the years being compared.

This is not a panel study that is designed to include information about the same individuals in each reported observation. A panel study design can be thought of as a subset of cohort study designs.

The immediate value of *Figure 3* is its demonstration that longitudinal linkage of administrative data sources allows us to understand how many former enrollees in a state’s public postsecondary education system then continue to be employed in the state or in other defined locations over nearly three decades.<sup>5</sup>

**FIGURE 3. Appearance of cohort members in employment administrative records**



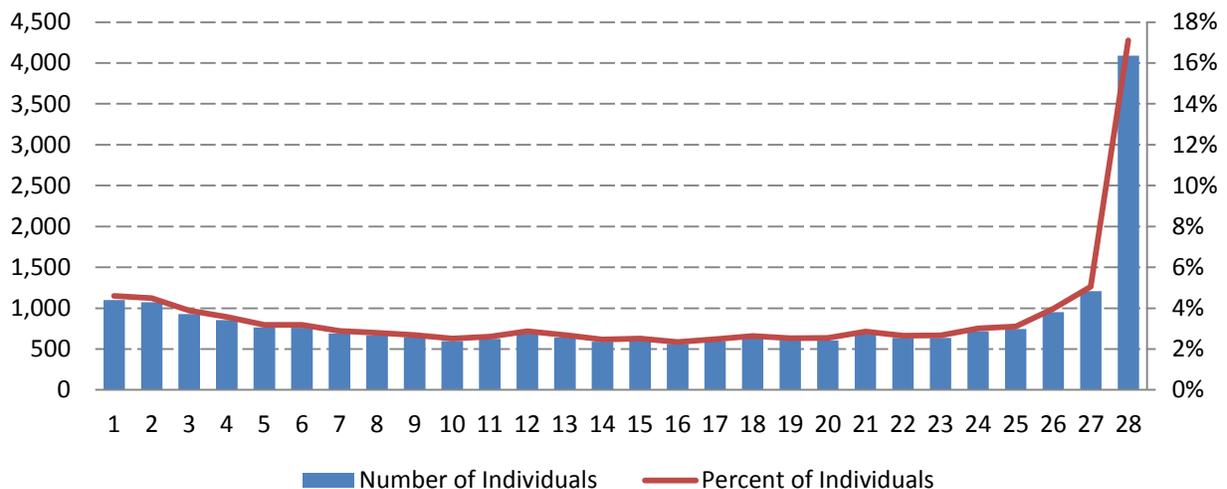
<sup>5</sup> Documented in-state *employment* bears no necessary relationship to a cohort member’s *residence* in a particular reference year. The Maryland UI wage record, regional UI wage record and federal civilian employee data sources tell us where covered employment was recorded, irrespective of where the employee lived at the time. We have chosen to leave mention of WRIS and WRIS2 capabilities to another time, except to point out that the continuing limitations on authorized access to unit-record information prohibit the kind of analysis that our study illustrates and that is replicable elsewhere.

#### 4. Frequency of cohort member appearance in the Maryland UI wage record data

We noted in the previous subsection that, while only cohort members appear in *Figure 3*, the mix of individuals included in a particular reference year's count almost certainly differs from the composition in other annual counts. Another way to understand the importance of this from a decision relevance perspective is to show the distribution of the cohort members based on how many times they appear out of the maximum 28 annual observations that are possible. Think of it this way—take 28 snapshots, each including only those cohort members that had reported earnings in any one, two, three or four of the particular reference year's quarters. Then sum the number of years that each person appears without regard to the particular years of appearance; a person may appear in 1985, 1986 and 1987 only, or in 1985, 1995 and 2005 only, but in either case they would be counted as having appeared in three of the twenty-eight possible annual snapshots.

*Figure 4* shows that 4,000, 16.7 percent, of the 23,895 cohort members that appear at any time in the Maryland UI wage record data file appear in each, that is all, of the 28 years from 1985 through 2012. This reinforces the point made in the previous subsection—the cohort member composition of a particular reference year's employment differs from the mix found in other reference years.

**FIGURE 4. Frequency of annual appearance in the Maryland UI wage record file, 1985-2012**



*Figure 4* delivers a clear message that extrapolation of presumed employment continuity beyond actual observed quarters and aggregated annual data should be avoided. We elaborate on this point later in this section.

## 5. Median, 25<sup>th</sup> and 75<sup>th</sup> percentile real annual earnings amounts

*Figure 5, Panel A*, clearly shows the impact of administrative data source on the calculation and reporting of median annual earnings amount. The data field content and definitions are the same in the Maryland UI wage records and selected other states UI wage records, so this is an apples-to-apples comparison.

The levels and slopes of the two 2007-2012 time span segments are striking. Until 2007, cohort members found in the Maryland UI wage records data file had consistently rising and higher median earnings relative to the cohort members that were found in the regional states UI wage records file beginning in 1999. Then, as the official national recession took hold beginning in December 2007 and continued through mid-2009, the two median earnings paths quickly converged so that in the most recent annual observation year, 2012, the out-of-state median actually exceeded the in-state median annual earnings level.

The six-year 2007-2012 trends shown in *Figure 5, Panel A*, illustrate an important point about the decision relevance of longitudinal integrated data system findings—the presentation of important relevant new information often triggers an outpouring of new questions, seeking an actionable explanation for the observed evidence. For example, did cohort members previously employed in-state in relatively high earnings jobs suffer from recession related involuntary terminations followed by acceptance of lower earning out-of-state opportunities? What is the gender mix of the pre- and post-recession employment locations? Did the industry composition of the in-state and out-of-state median earnings calculations change after 2007?

Our ability to answer important drill-down questions to improve the decision relevance of our studies depends upon the data field content of the available administrative data sources. For example, the quarterly cycles of mid-Atlantic regional sharing of UI wage record extracts for authorized purposes do not currently include industry affiliation information, so the last question posed in the previous paragraph cannot be answered unless the participating states agree to add this data field to the exchange, including recovery of the historical time span coverage that would be needed to seek a better and more policy action relevant explanation of the pre-recession post-recession convergence pattern observed to date.

**FIGURE 5. Panel A: Median annual real earnings amounts by administrative data source**

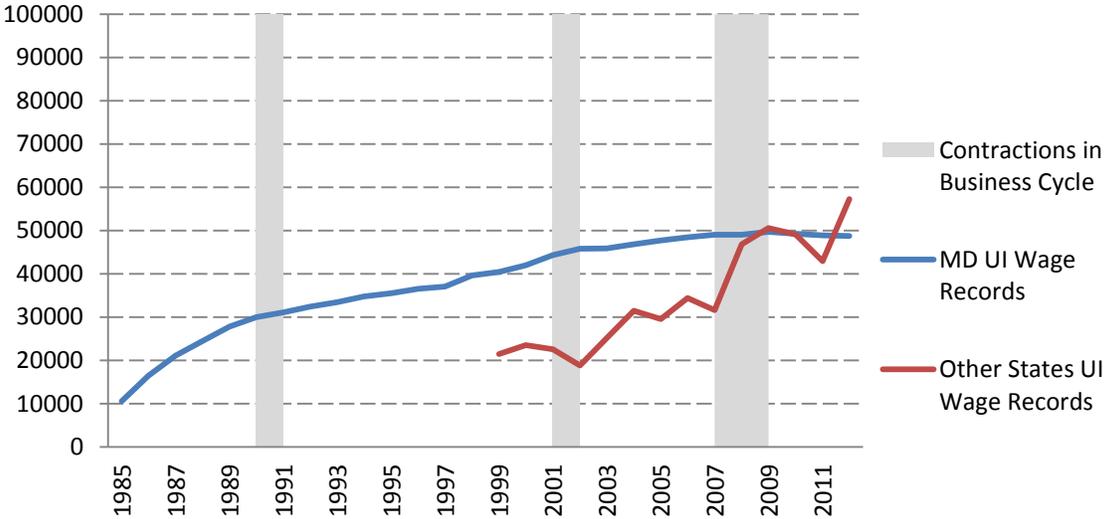
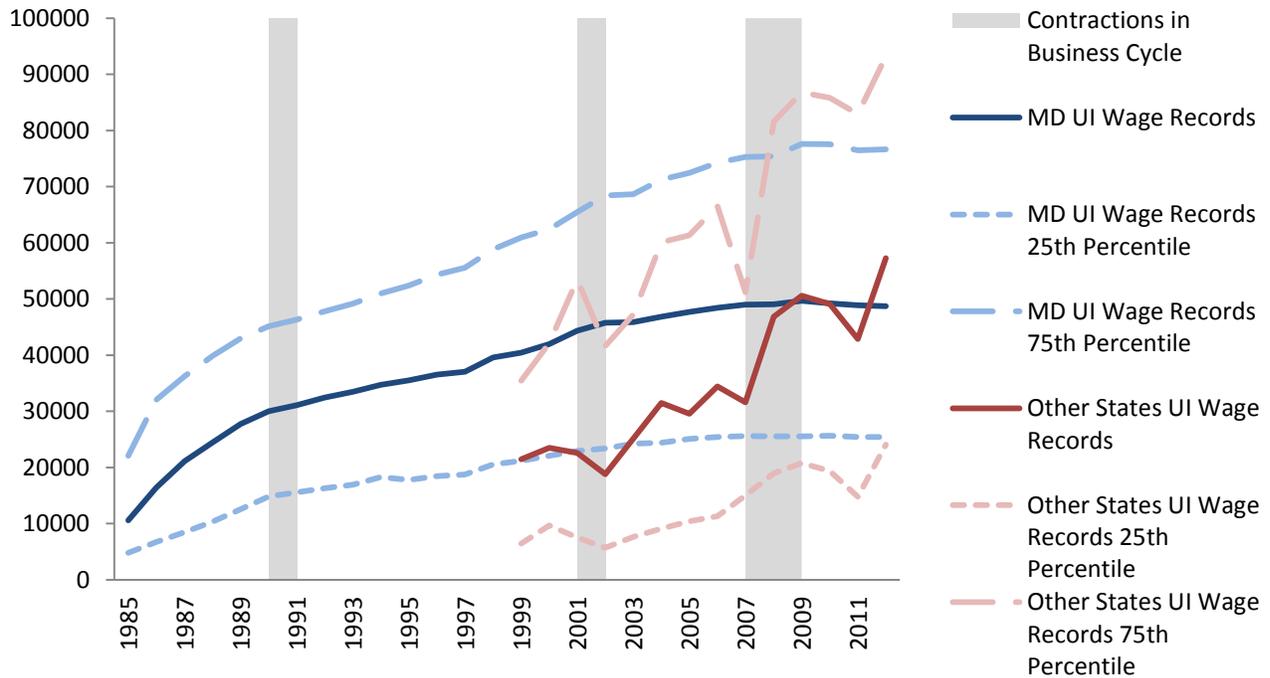


Figure 5, Panel B, adds new insights that further advance the decision relevance of the annual earnings profile information. Median earnings levels are the most common measure of alleged education *outcome*, as this terminology is used in many web-based releases. But the decision making relevance and value of earnings information often lies in other earnings distribution metrics—such as the 25<sup>th</sup> and 75<sup>th</sup> percentile earnings amounts that constitute the boundaries of the middle half of the full distribution; one-fourth of the earners received less than the 25<sup>th</sup> percentile ‘floor’ value and one-fourth of the earners received more than the 75<sup>th</sup> percentile ‘ceiling’ value of this middle half segment of the earnings distribution.

In Figure 5, Panel B, the color coded dotted lines show the respective 25<sup>th</sup> and 75<sup>th</sup> percentile earnings amounts that band the median annual earnings amounts shown in Figure 5, Panel A, for the each of the two administrative data sources. Both the 25<sup>th</sup> and 75<sup>th</sup> percentile out-of-state earnings values continued an increase that had begun at the end of the 2001-2002 recession, but with more positive slope, while the Maryland upper and lower boundaries of the middle half segment remained basically flat.

Figure 5, Panel B, is one reason why we will be promoting formation of a collaborative mid-Atlantic regional research initiative to pursue a common research design using available state longitudinal data system resources. Mobility within the region is known to be high, but the causes and consequences of this mobility for individuals, the states and the region are not understood. Successful pursuit of this collaborative enterprise can serve as a template for adoption by other regions.

**FIGURE 5. Panel B: 25th and 75th percentile bands of real annual earnings amounts by source**



## 6. Five-year segments of the full time span covered

*Figure 6* is a four-panel presentation of our longitudinal coverage of cohort member earnings separated into five five-year segments plus the most recent two years. This and following subsections speak directly to the web-based Economic Success Metrics information currently available at <http://www.collegemeasures.com>.

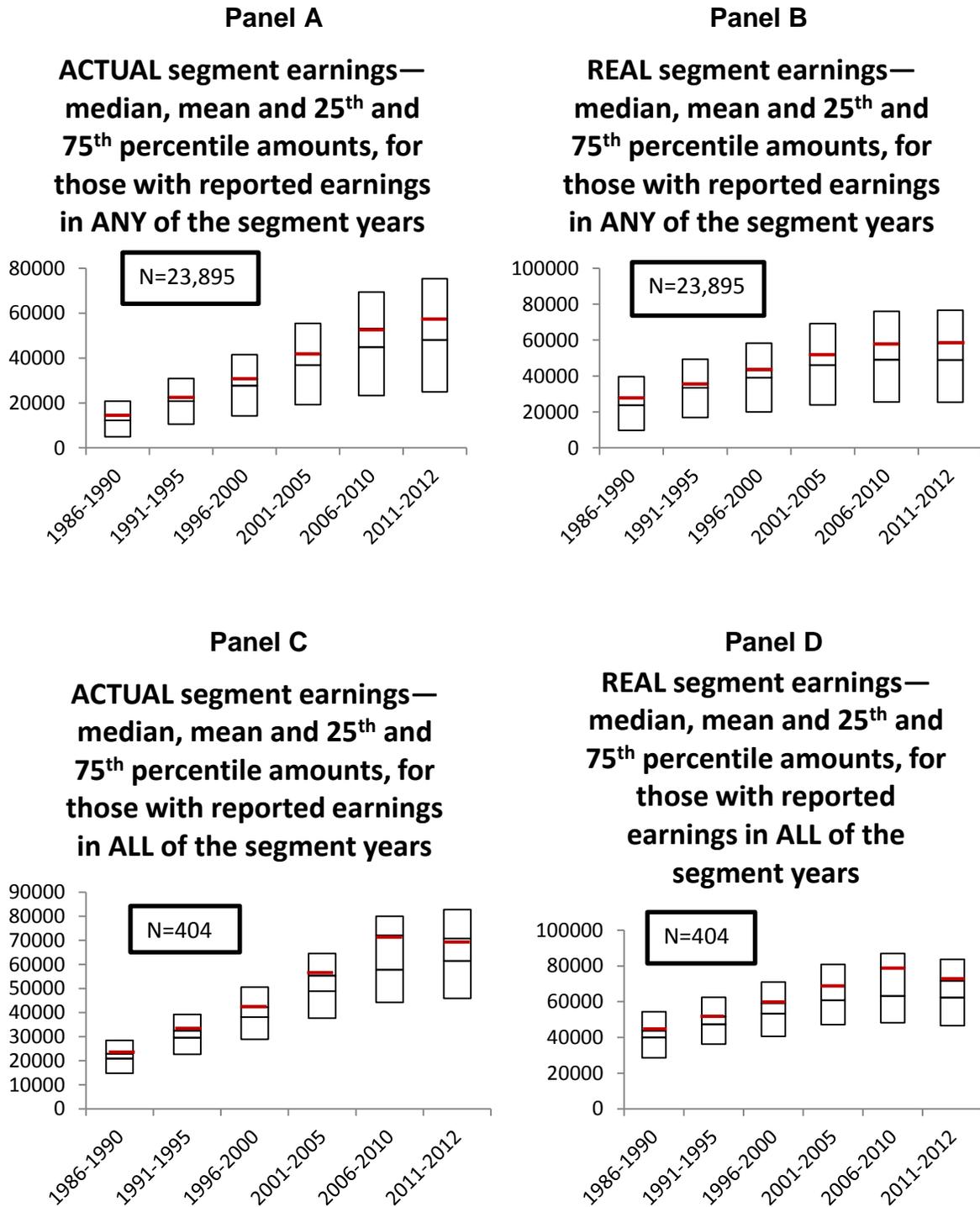
The College Measures initiative represents an important advance in the availability of relevant return-on-investment information long sought by institutional and student constituencies. The new tools being developed by the initiative, and the expansion of time coverage to 10 years after postsecondary graduation that is underway, combine to increase the value of the initiative's products.

Our intention has been to build on this foundation of new web-based information. Prospective students and institutional leadership teams need to understand whether short-term 5 and 10 year snapshots are predictive of future working life-cycle earnings prospects.

The four panels of *Figure 6* that follow next are:

- *Panel A*—summarizes actual five-year and two-year segment earnings for cohort members that had reported earnings of any amount in any one or more of the reference years in a segment. Median, mean and 25<sup>th</sup> and 75<sup>th</sup> percentile actual dollar amounts are shown.
- *Panel B*—transforms the *Panel A* earnings amounts from actual earnings to inflation adjusted real earnings. This addresses the decision making relevance of increased purchasing power over time.
- *Panel C*—summarizes actual segment earnings for a different defined cohort subpopulation than is covered in *Panel A*. *Panel A* includes “cohort members that had reported earnings of any amount in any one or more of the reference years in a segment”, while *Panel C* includes only those cohort members that had reported earnings of any amount throughout the twenty-eight years. So, those included in *Panel C* are a subset of those included in *Panel A*.
- *Panel D*—transforms the *Panel C* earnings amounts from actual earning to inflation adjusted real earning; again addressing the decision making relevance of increased purchasing power over time.

**FIGURE 6. Segment earnings--median, mean and 25th and 75th percentile amounts for those with reported earnings in defined segment years**



Note: Inside each box, horizontal lines signal 25<sup>th</sup> percentile, median, mean (red), and 75<sup>th</sup> percentile from bottom to top.

The four panels of *Figure 6* on the previous page constitute a clear message that authorized access to extended time span coverage of longitudinal integrated education and workforce administrative data creates a capacity to add decision relevant value to available web-based short-term earnings metrics. The final section of this report describes how we intend to take advantage of this capacity, while offering those in other states suggestions for how they, too, can replicate our research design and adapt it to satisfy their own decision support priorities.

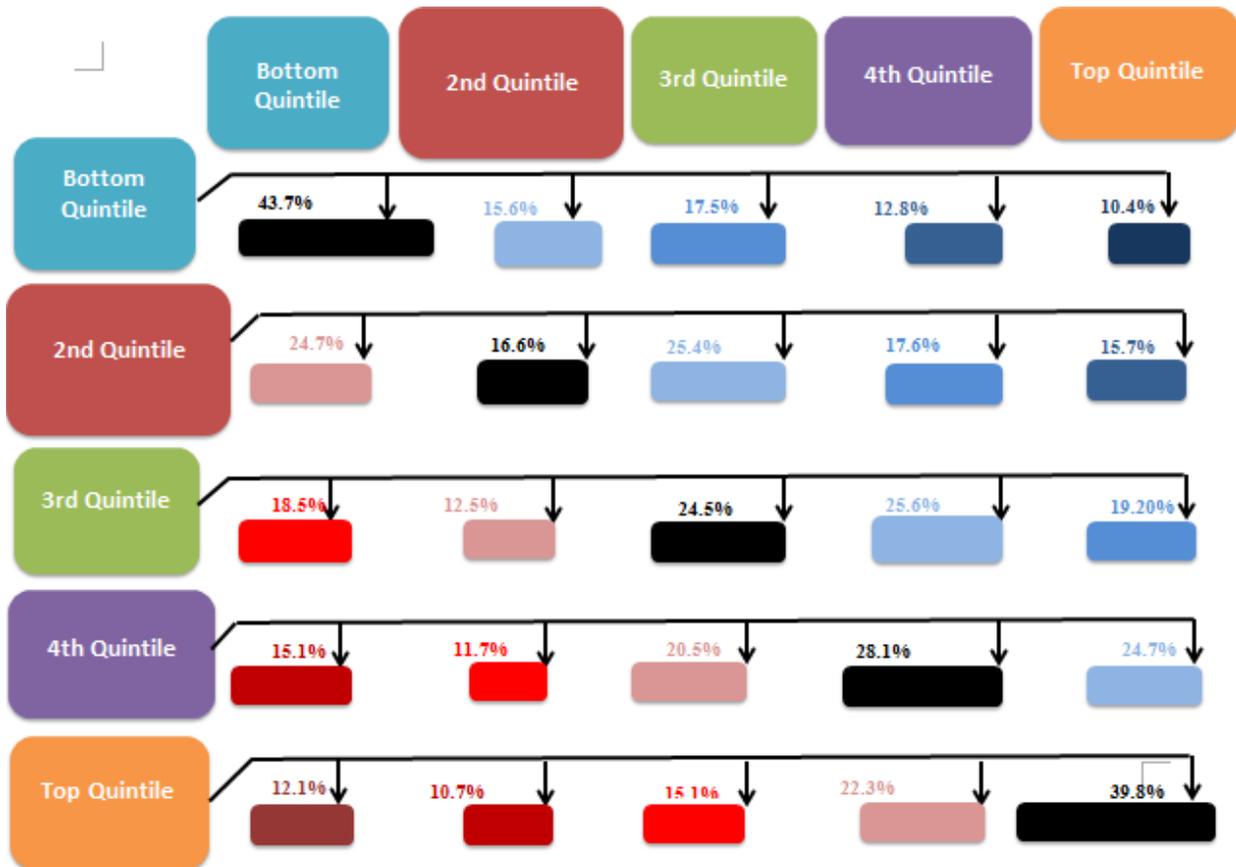
Much of the value added potential of extended time span coverage lies in an improved ability to account for life-cycle sequences of educational endeavor and attainment. This accountability is needed to improve the accuracy of educational *outcome* measures.

## **7. Comparison of 1986-1990 and 2011-2012 time segment actual earnings quintiles**

We arrive in *Figure 7* at the essence of the business case message, answering the question: What is the predictive power of short-term measured earnings for use in working life-cycle decisions?

*Figure 7* separates two time segment summaries of earnings into quintiles. The question that motivated this matrix display is: Does a cohort member's assignment to an earnings distribution quintile in the 1986-1990 time segment offer predictive insight about where this person will appear in the 2011-2012 earnings distribution quintile?

**FIGURE 7. MD earnings quintile snapshots 1986-1990 vs. 2011-2012**



Note:

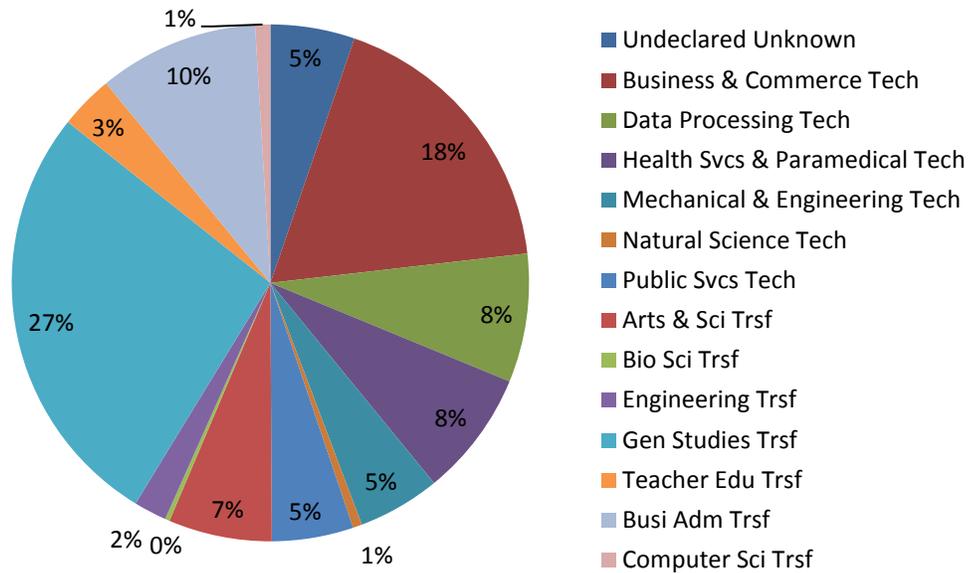
1. All percentages shown in the above figure refer to individuals that appear in both 1986-1990 and 2011-2012 snapshots.

In *Figure 7* the black cells on the diagonal account for the cohort members that appear in the same earnings distribution quintile in both the 1986-1990 and 2011-2012 time segments. Predictive power based on the 1986-1990 quintile assignment is poor--more than half of those assigned to a 1986-1990 quintile appear in a different quintile in the 2011-2012 snapshot.

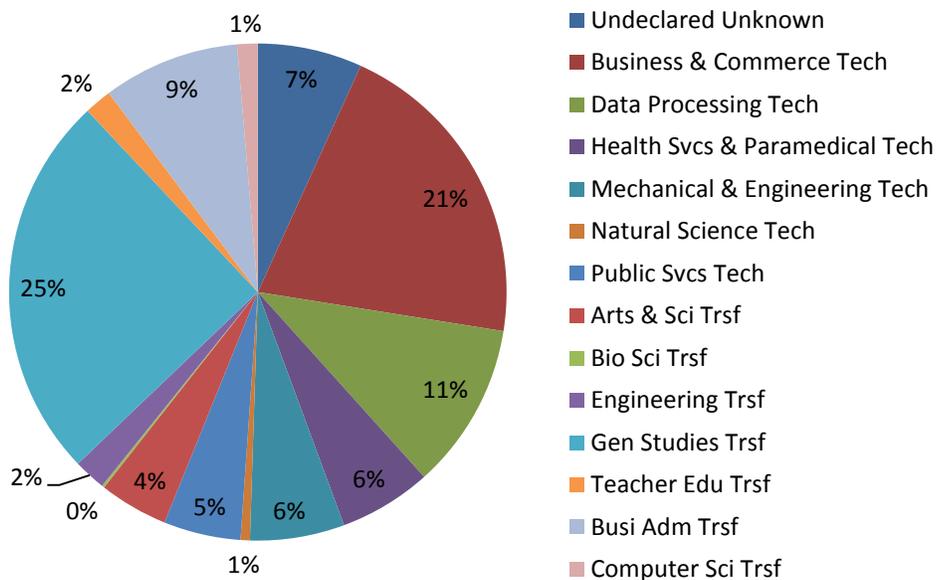
### 8. Cohort member movement up or down in the time segment earnings distributions

The two panels of *Figure 8* show the mix of cohort member community college program assignments represented in the *Figure 7* cohort subpopulations that moved to a higher earnings distribution quintile (Panel A) or to a lower quintile (Panel B).

**FIGURE 8. Panel A: The 1984 community college program assignments of the cohort members that moved to a higher earnings distribution quintile from 1986-1990 to 2011-2012**



**FIGURE 8. Panel B: The 1984 community college program assignments of the cohort members that moved to a lower earnings distribution quintile from 1986-1990 to 2011-2012**



The two panels of *Figure 8* do not show any obvious relationship between a cohort member's fall 1984 community college program assignment and how each will fare later in earnings improvement or deterioration relative to other cohort members.

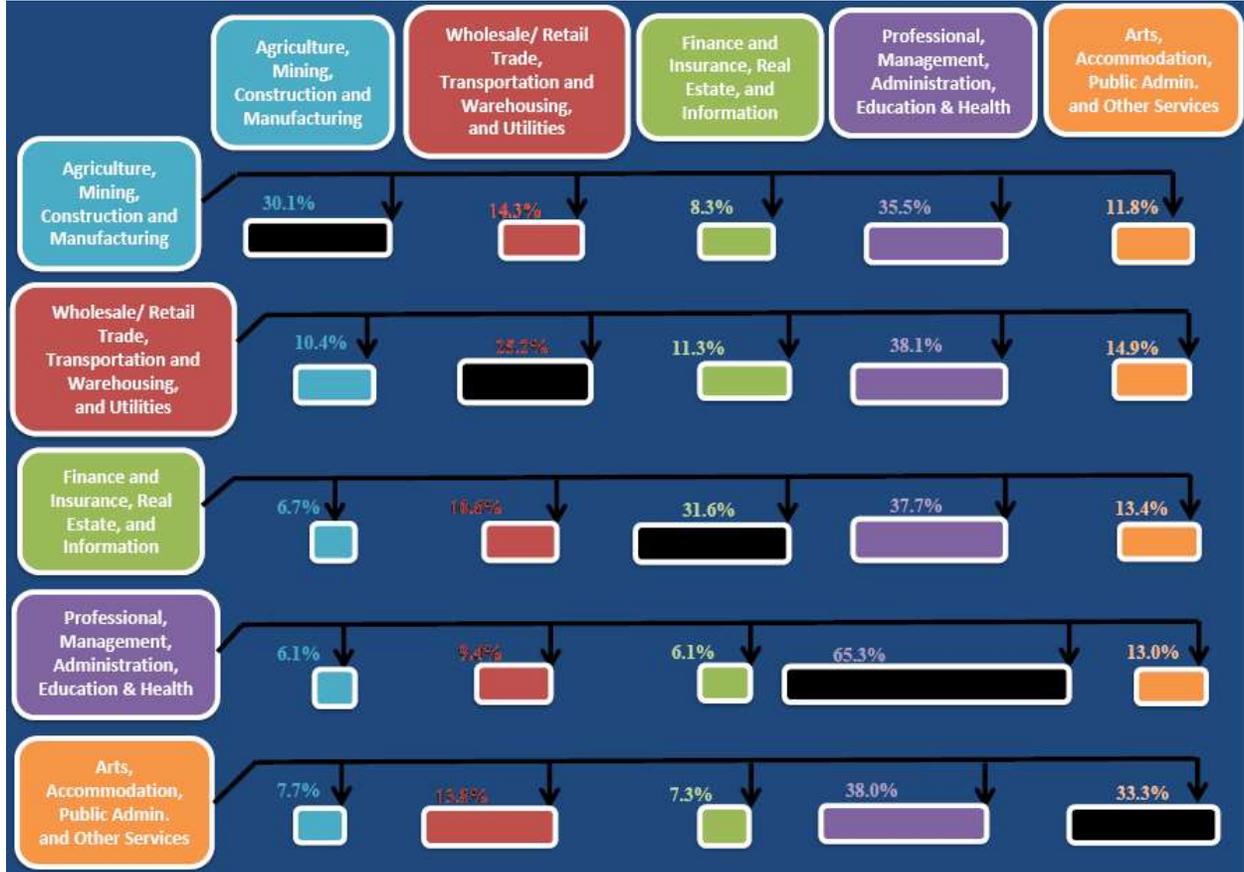
### **9. Comparison of 1985 fourth quarter and 2012 fourth quarter industry affiliations**

Investment in sector-based economic development initiatives is accelerating and increasing. Many assaults on skills shortages are sector-based. An important way in which state longitudinal integrated data systems can contribute to the quality and impact of these initiatives is to advance the capacity of researchers to study the dynamics of worker mobility among industries over an extended time span. Of particular interest in this regard is how the sector mobility dynamics change leading up to, during and after a recession.

The matrix visualization in *Figure 9* is similar conceptually to that achieved in *Figure 7*, but now looking at inter-industry mobility patterns instead of movement within an earnings distribution. Again, the black color-coded cells on the diagonal include the cohort members that were found in the same industry sector in the fourth quarter of 1985 and the fourth quarter of 2012. We did a crossover match from Standard Industrial Classification System (SIC) to North American Industry Classification System (NAICS). We draw attention to the fact that we have consolidated some of the official NAICS taxonomy labels to improve visualization clarity.

*Figure 9* shows that a cohort member's industry affiliation in late 1985 is not a reliable predictor of their industry affiliation nearly three decades later. Future diagnostics using this matrix visualization can change the reference base and/or end year to study the timing of particular dynamics. For example, if the end year had been defined as 2006, before the national recession began, would the observed mobility pattern have been different? Or, when additional Degree Information System data are analyzed, will different mobility patterns be found for those that received particular degrees?

**FIGURE 9. Industry Sector Changes 1985 quarter 4 vs. 2012 quarter 4**



Note:

1. All percentages shown in the above figure are based on individuals that appear in both 1985 quarter 4 and 2012 quarter 4 snap shots.

## F. CONCLUSIONS AND NEXT STEPS

Earlier in this report, before presenting our findings, we expressed the following thought:

“We urge reader attention to a relevance criterion that we imposed on ourselves in deciding what to present—what role-based decision maker(s) can be expected to recognize the relevance of this information for their own education investment decisions?”

Then, at the outset of the findings section, we defined two goals to be accomplished in the presentation of findings:

1. Deliver evidence-based advances in the development of a strong business case for continued investment in actionable longitudinal state integrated data systems.
2. Offer states a practical research design that can quickly add power to the business case through expanded geographic and time span coverage.

We applied the combination of a decision relevance criterion and two explicit goals to select nine visualization tools that were then presented and interpreted. We begin this final section with a brief review of each of these nine visualization tools, again asking whether each has made a positive contribution to development of a strong business case for future investment in state longitudinal integrated data systems. Following this summary, we turn to presentation of a series of steps that states can take to replicate the parts of our methodology that align well with their own integrated data system capabilities and decision support priorities. Next, we comment on features of a state longitudinal integrated data system that promote decision making relevance and value. Finally, we identify communication challenges that impede rapid progress.

### 1. The business case relevance of our longitudinal visualization tools

This subsection proceeds through each of the nine visualization tools that were presented in the previous section of this report. The sequence of presentation has been intentional and strategic, advancing the business case step-by-step.

*Figure 1*—number of cohort members by fall 1984 community college program assignment and gender. This is the logical baseline foundation for undertaking a cohort study. Program and gender are two administrative descriptors that set the stage for actionable results. Role-specific action responses would be expected to include postsecondary education leadership teams, currently enrolled and future students, counselors, vendors of career pathways resources, and business human resources administrators.

*Figure 2*—University System of Maryland enrollment and degree counts, 1985-1997. A temporary caution has repeatedly been expressed throughout this report, noting that we expect to expand both the time span and institutional coverage of this case building step. The contribution of this visualization tool, when updated to include the expanded time and institutional coverage, is to show the *longitudinal* progression of cohort members from the common fall 1984 starting point; progression into, through and from subsequent in-state public postsecondary enrollments and credential achievements.<sup>6</sup> This information will advance our ability to improve the clarity of communication about educational *outcomes* that are often attributed wholly to the last institution of record. Acquisition of human capital is a cumulative and dynamic process, including loss of productive capability and related earning potential. Knowing when and how to renew skills and skills relevance remains elusive. Longitudinal information of the type illustrated here can help to inform adult learner decisions about when and how to reengage in postsecondary education. At the same time, this information can prepare governors, legislators and postsecondary leadership teams to make informed investments in institutional capacities to respond to adult learner needs and interests.

*Figure 3*—appearance of cohort members in employment administrative records. Our decision to present *Figure 1*, *Figure 2* and *Figure 3* in that order creates a risk of misinterpretation stemming from a long standing, but now clearly obsolete, tradition of reference to one-way school-to-work transition flows. Therefore, we hasten to correct this potential error of interpretation. Transitions between postsecondary education and employment flow in both from-to directions. The two statuses are often concurrent. Transitions between the two can and do recur over a working life-cycle. Within the postsecondary sector, enrollment in community college and four-year institutions is often concurrent, and sequencing over time can be difficult to disentangle and communicate. The case building value of *Figure 3* is its demonstration that the complex interactions among the statuses documented in *Figures 1-3* can be sorted out and communicated to decision makers. The value of extended integrated data system time span coverage becomes clear when the cumulative contribution of *Figures 1-3* is thought through carefully.

*Figure 4*—frequency of annual appearance in the Maryland UI wage record file, 1985-2012. This was the logical next step to build on the foundation set through *Figure 3*. The career pathways literature often conveys an impression that employment is a stable phenomenon for most workers. *Figure 4* suggests a need to exercise caution in acting on this belief. *Figure 4*, alone, is one of many examples in this report that trigger a flood of new questions about what this new information means and how to realize its potential value for future education investment decisions. *Figure 4* also illustrates a limitation of reliance on administrative data—there are employment circumstances that are not reflected in a state’s UI wage records. Decision makers should be made aware of whether and how omissions should be factored into their use of the available information.

---

<sup>6</sup> For now, we are deferring consideration of whether and how National Student Clearinghouse enrollment and degree information can be incorporated into an analysis of this type.

*Figure 5, Panels A & B*—median, 25<sup>th</sup> and 75<sup>th</sup> percentile real annual earnings amounts. Up to this point in our development of a business case for future investment in state integrated data systems, *Figures 1-4* alignment in time of education and employment statuses should be thought of as intermediate steps along a path to improved, or at least maintained, working life span earnings.<sup>7</sup> The two panels of *Figure 5* introduce in a single visualization tool the actual growth path of cohort member earnings as documented by the in-state and nearby states UI wage records. The 25<sup>th</sup> and 75<sup>th</sup> percentile bands in *Figure 5, Panel B* ease the clarity of communication of this earnings profile to decision makers. Half the cohort members with recorded earnings in any one of the 27 annual calculations between 1986 and 2012 earned amounts that lie between the lower (25<sup>th</sup> percentile) and upper (75<sup>th</sup> percentile) bands. However, one-fourth of the cohort members with reported earnings in an annual calculation earned less than the 25<sup>th</sup> percentile ‘floor’ and another one-fourth earned more than the 75<sup>th</sup> percentile ‘ceiling’ that define the boundaries of the middle half.

*Figure 6, Panels A, B, C & D*—multi-year segments of the full time span covered. Anticipating that many individuals that make institutional and individual education investment decisions are not as conversant as researchers are with statistical concepts and applications, we have begun to deconstruct aggregate cohort earnings visualizations and reassemble the data into more meaningful targeted visualization tools.

*Figure 6 Panel A* and *Panel C* begin the process of identifying subsets of the full cohort that convey useful new information. *Panel B* and *Panel D* cover the same cohort subpopulations as *Panel A* and *Panel C* respectively, the only difference being that the earnings profile is transformed from actual (nominal) earnings to inflation-adjusted (real) earnings amounts. For some investment decision making purposes, particularly by individuals, the relevant measure is prospects for improvement of purchasing power; that is, expected future real earnings. The five-year time segment summaries of earnings shown in *Figure 6* help us to align our findings from extended longitudinal coverage with other web-based research reports that concentrate on shorter time spans after postsecondary graduation. *Panels A-B & C-D* show how important continuity of employment is as an often unseen determinant of reported earnings. *Panels A* and *B* include all cohort members in a multi-year segment that had reported earnings in any one or more of the years defined for the segment. *Panels C* and *D* include only those cohort members that had reported earnings throughout all of the defined segment years.

Together, *Figure 5 Panel B* and the multiple panels of *Figure 6* look beyond the aggregate median profile seen in *Figure 5 Panel A*, which is important because few decision makers, whether institutional or individual, should be comfortable with reliance on aggregate median information alone.

---

<sup>7</sup> We are not asserting that *the* goal of investment in education is enhancement of earnings potential, which serves as a proxy for productive capacity that is needed to achieve economic growth. However, we are saying that the business case for investment in state longitudinal integrated data systems is strengthened if transformation of educational achievement through employment to earnings advance is documented.

*Figure 7*—comparison of 1986-1990 and 2011-2012 time segment earnings quintiles. As we noted in the introduction to *Figure 7* in the previous section of this report, *Figure 7* is “the essence of the business case message, answering the question: What is the predictive power of short-term measured earnings for use in working life-cycle decisions?” The mobility of cohort members among two earnings distribution snapshots taken decades apart is striking. This mobility sends a clear signal to virtually all role-based decision makers that short-term earnings amount recorded soon after transition from an education engagement is not necessarily a useful predictor of future earnings prospects. This point is then reinforced in the next two visualization tools presented in the previous section.

*Figure 8*—cohort member movement up or down in the time segment earnings distributions. Here, we combine the 1984 community college enrollment program assignments data that underlie *Figure 1* with the time segment earnings quintile data presented in *Figure 7* to show the 1984 cohort member program mixes found in the respective up or down movements between earnings quintiles. Future refinements of this visualization tool are expected to be of particular interest and value to institutional program leadership teams that are faced with persistent questions about the return-on-investment of resources devoted to their offerings. Future students, too, will find this information worthwhile as they think about programmatic choices made in the context of peer, counselor and influential person advice and commentary.

*Figure 9*—comparison of 1985 quarter 4 and 2012 quarter 4 industry affiliations. This visualization tool reinforces the message conveyed through *Figure 8*. Inter-industry mobility is frequent. The popularity of industry sector investment strategies and targeting of industry skills shortages should be revisited in the context of accurate understanding about inter-industry mobility dynamics. *Figure 9* is a starting point for design of future research strategies that can be expected to improve the effectiveness and efficiency of sector-based intervention strategies and investments.

We have emphasized the message several times that decision making based on a short-term earnings profile soon after transition from an education engagement is not necessarily a useful predictor of future earnings prospects.

## **2. Steps states can take to replicate or adapt relevant parts of our research design**

We turn next to presentation of a series of steps that states can take to replicate the parts of our methodology that align well with their own integrated data system capabilities and decision support priorities. Optimal selection from this menu of action steps will be state-specific, depending upon a state’s unique build-out design, current status, timing of completion and priorities for use of the system’s expanding capability.

We acknowledge, again, that not all states are pursuing a comprehensive longitudinal P-20W integration capability. Our step-by-step counsel may have little if

any immediate relevance for states that have not placed a high priority on at least public high school and public postsecondary administrative data linkage with employment and earnings administrative data. However, even in such circumstances we think that our series of nine visualization tools should be reviewed to decide whether a higher priority should be placed on integration of these three administrative data components to advance the state's business case for future investment and delivery of increased value to decision makers.

The basic steps to success are:

- Take stock of the legal, regulatory and cultural conditions that together determine what progress in administrative data integration can be accomplished.
- Find out what common identifier, or indirect route to arrival at a common identifier, can be used for data source linkage. Document the impacts identifier availability will have on the accuracy and coverage of linkage.
- Document what historical time coverage will be available at the outset of an integrated data system capability. The expanded federal investment in development of system capabilities began only eight years ago, and initial data loading is still in very early stages in many states. This means that Documentation of state-by-state historical coverage should be broadcast quickly, and then updated, so further development of the business case can focus on what can be learned from the states that have the longest time span coverage of longitudinal integrated data capabilities.

Combine the complementary contributions to success that can be made by future communities of interest and communities of practice. As the portfolio of decision relevant information products grows success on the business case and investment fronts will follow. New investments, in turn, will accelerate and expand the delivery of future decision relevant products.

### **3. Our own future research agenda and priorities**

We turn next to our own future research agenda and priorities. Our highest priority will be incorporation of new Enrollment Information System and Degree Information System data extracts received from the Maryland Higher Education Commission. This will enable us to undertake new connect-the-dots studies of the timing and sequencing of further investments in, and achievements through, further in-state public education over nearly three decades. This mapping exercise, in turn, will open the door to a series of studies of interactive education, employment and earnings dynamics, including refined stratification by educational program, industry affiliation, age group and gender.

Concurrent with our future attention to the research topics described in the previous paragraph, in the current fiscal year we are engaged in skills shortage

research, seeking to improve the clarity of communication about this topic and how practical advances can be made to improve the investment decision process and outcomes.

There are well known gaps in the availability of information that is needed to converge toward a consensus definition of skills shortages and the causes of these shortages. As long as these gaps persist there will be little progress on the clarity of communication front. One of our research goals will be to advance public understanding of the advocacy rationales being made for public investments in alleviation of poorly defined skills shortages.

#### **4. Toward improved clarity of communication**

Our findings support advocacy for the design and management of state longitudinal integrated data systems that include the following features:

- A state integrated data system should have a clearly defined, limited, purpose; it should not be marketed as a way to answer a broad spectrum of hypothetical questions that are difficult to align with high impact decision making. Overly ambitious marketing unnecessarily exposes a state data system to future funding jeopardy. The rigor required for performance-based funding formula use, for example, involves a much higher standard than that associated with basic discretionary consumer use only.
- The individual education level and workforce data field contents, and business rules that define why and when linkage among these is necessary, should be transparent. Here, transparency means that data system advocates and detractors alike should be able to understand what is included and how this content can be used. Parsimony is encouraged—if we cannot explain the positive contribution of a particular facet of overall system design and intended uses we should continue working to achieve an acceptable level of agreement about the value enhancement contributed by this feature, or we should delete it.
- The word *longitudinal* is the keystone needed for development of a strong business case for sustained investment in a state integrated data system. Substantial confusion persists in communities of interest about whether a particular state's SLDS is actually *longitudinal*, meaning that it is capable of linking information about individuals across defined segments and through time.

Public concern about data retention is escalating in response to growing awareness of previously unknown government, social media and business uses of collected information. While there is no evidence that the length of data retention increases identity disclosure risk, public belief appears to have concluded otherwise.

The value derivable from extended time span retention has to be demonstrated. Delivery of such value evidence is the fundamental motivation for this report, which concentrates on linkage among postsecondary education and workforce administrative data sources over nearly three decades.

Use of the word *among* instead of *between* data sources is an important part of our clarity of communication message. Each component includes a diverse portfolio of administrative data sources. A comprehensive state integrated data system design is constructed from multiple early childhood, k-12, postsecondary and workforce data source extracts.

- Realization of value requires more than extended data retention. Retained data must be transformed into decision-relevant information. Decision makers must be able to easily see the relevance of new insights that emerge from analysis of longitudinal integrated data sources for prioritizing the timing and targeting of actions. Examples addressed in this report include:
  - Are recorded average earnings amounts soon after leaving a particular education institution/program an acceptable predictor of future earnings prospects—that is, good enough to be used by future enrollees in this institution/program as a predictor of their own life-cycle earnings prospects?
  - Are cohort-specific employment rates and average earnings amounts recorded soon after graduation acceptable measures of educational *outcome*; that is, acceptable measures of the contribution made by this defined education achievement alone to employment status and earnings amount after graduation?

Our message is a positive one for the SLDS and WDQI stakeholder communities. State longitudinal integrated data systems are on the cusp of something important. These systems have, or can soon achieve, a capacity to deliver new high impact information to those charged with making important education and workforce investment decisions. But the window of opportunity to do so is closing; the flow of federal funds has to be renewed or replaced.

The business case for continuity must be honed and delivered in a matter of months, or a year or two at most. This report is intended to accelerate the timeliness and advance the quality of the business case for continued funding.

Prepared by:

The Jacob  
*france*  
Institute

Merrick School of Business  
The University of Baltimore  
1420 N. Charles Street  
Baltimore, Maryland 21201  
[www.ubalt.edu/jfi](http://www.ubalt.edu/jfi)  
(410) 837-4727