



**BOLD
THINKERS
DRIVING
REAL-WORLD
IMPACT**

Addressing Unit Nonresponse with Administrative Data for PACE and HPOG 2.0

David Judkins

Abt Associates Inc.

October 19, 2023

Presentation Overview



- Background on PACE and HPOG 2.0
- Administrative Data Repositories
- Consent
- Approaches and Software
- Challenges
- Scary Stories
- Thoughts for RFP Reviewers



- Major OPRE projects
 - PACE was a collection of 9 workforce training programs informed by the Career Pathways Framework of David Fein – each with its own randomized experiment and follow-up surveys to measure impacts
 - HPOG 2.0 was a collection of 38 local healthcare training programs— also with randomized access and follow-up surveys to measure impacts



- We were concerned that career progress could affect the difficulty of tracing and the willingness of study members to answer the follow-up surveys
 - Might those who fall short of their goals be loath to discuss those shortcomings?
 - Might those who met or exceed their goals be too busy to respond?
 - Response rates were reasonably high but below OMB standards and there were large gaps in response rates between arms, something that the WWC hates to see

Promise of Administrative Data



- Adjustments to the analysis based on administrative data can reduce such biases, but can have unfortunate side effects in the form of increased standard errors or biased estimates of standard errors
- Talk will mention approaches to get bias reduction with minimal side effects of this nature
- No equations -- but should be of use to RFP writers

Administrative Data Repositories



- For PACE and HPOG 2.0, we used administrative data from:
 - National Student Clearinghouse (NSC)
 - National Directory of New Hires (NDNH)
- For PACE, we also collected records from local colleges at some sites



- High quality enrollment from public colleges
- Credential data are not well gradated
- Private colleges are poorly represented
- Training institutions that do not grant degrees are not represented at all
- Also hints of problems at adult training institutions loosely affiliated with community colleges

National Directory of New Hires



- High quality earnings data
 - Both for employees of private and public employers
 - No coverage of self-employment
 - Coverage of gig employment hard to assess
 - Undercoverage of cash wages
- Access is difficult to gain
- But OPRE projects have advantages as a sister agency to OCSS within ACF

Local College Records



- Difficult to negotiate access
- Difficult to obtain updates
- Poor coverage of people denied access to local college program
- Nonetheless, did yield some important insights

Consent



- Institutional Review Board (IRB) will almost certainly require informed consent to access administrative data about the sample in demographic studies
- High hurdle for cross-sectional descriptive studies and general panel studies
- Easier in demonstration studies if consent is a condition of participation



- Use admin data to improve quality of imputations for item nonresponse (income, etc.)
- Use admin data in weighting for unit nonresponse
 - Propensity scoring
 - Plus outcome calibration
 - Plus impact calibration

Weight Calibration



- Generalization of “poststratification” and of “raking”
- General idea is to adjust weights for respondents so that analyses on a smaller sample (such as the respondent sample) agree with analyses on a larger sample (such as the full sample or even the full population in a frame)

Calibrated Weights (on levels)



- Using methods invented by Ralph Folsom and implemented in SUDAAN (RTI software callable from SAS), it is easy to modify weights so that:
 - Weighted distributions of categorical admin variables on the respondent sample agree perfectly with distributions on the full eligible sample
 - Weighted means of continuous admin variables on the respondent sample also agree with counterparts on full sample

Calibrated Weights (on impacts)



- Using method invented by Philip Kott (also of RTI) it is possible (with custom SAS macros + SUDAAN) to further modify weights so that:
 - Weighted regressions on the respondent sample agree with parallel regressions on the full eligible sample

Calibrated Weights – Motivation



- If admin variables sufficiently correlated with survey outcomes, calibration should reduce both bias and variance

Challenges



- Implementation of any of the methods discussed here requires a senior statistician
- Data access is somewhat difficult for NSC and very difficult for NDNH
- NDNH users must pass security checks and use HHS-supplied laptops over slow virtual private networks (VPNs)
- Neither NSC nor NDNH permit archival of their data with survey data at sites such as ICPSR

Challenges (2)



- Variance estimation is more complex when imputation or weight calibration is used
- For imputation, multiple imputation is the standard approach
- For calibration, a replicated approach such as the “jackknife” or the “bootstrap” may be advisable



- The HHS system for supporting NDNH analyses could use some upgrading:
 - The current system suffers from VPN delays and VPN timeouts
 - More critically, a prohibition on the use of C++ compilers mean that STAN cannot be used, a shame because it is the best available Bayesian software
 - Both problems could be solved by moving to “containers”

Scary Stories



- In the PACE evaluation of the HPOG program run by the San Diego Workforce Partnership, we found a very large nonresponse bias on the earnings impact that could only be corrected with the use of NDNH data (OPRE Report 2020-105, Appendix B.3)
- In the PACE evaluation of the HPOG program run the Pima Community College, Abt researchers found very poorly reported enrollment in the NSC (OPRE Report 2020-43, appendix D.3)



- Based on the full sample, the impact of the program at the average of the 12th and 13th quarters after randomization was \$289 and not statistically significant
- **However**, the impact estimated on the three-year follow-up survey, weighted with standard propensity-score weights, was \$738 and statistically significant.
- By using NDNH earnings data, we were able to develop a special set of weights that removed this bias.
- Without the NDNH data, we would have never known that the bias was even present

HPOG program at Pima Community College

- For unknown reasons, NSC reports of FTE months were unreasonably high for students who attend the Center for Training and Development (CTD), a loose affiliate of PCC
- College records showed that the majority of CTD courses resulted in a grade of W (withdrawal)
- These were not properly accounted in the NSC records
- Without the college records, we would have been unaware of the NSC problem

Lessons Learned



- At most PACE sites and for HPOG 2.0, there was little evidence of bias from unit nonresponse
- However, there was a glaring exception
- Until one does the work, it is unknown
- Even then, sometime the administrative data is the source of the problem
- Triangulation from multiple sources best

Thoughts for RFP Reviewers



- Has the offeror discussed various approaches to using administrative data to reduce bias from unit nonresponse with minimal side effects?
- Does the lead statistician have PhD plus ~4 years or Master's plus ~10 years?
- Is there a budget and commitment to ample methods documentation?

For comments or questions about
this presentation, please contact:



**BOLD
THINKERS
DRIVING
REAL-WORLD
IMPACT**

David_Judkins@abtassoc.com

References



Judkins, David, Emily Roessel, and Gabriel Durham. 2022. Career Pathways Long-Term Outcomes Study: Appendices for PACE Six-Year Impact Reports, OPRE Report 2022-69. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. [Career Pathways Long-Term Outcomes Study: Appendices for PACE Six-Year Impact Reports | The Administration for Children and Families \(hhs.gov\)](https://www.hhs.gov/opa/pace/long-term-outcomes-study)

Judkins, David Ross, Sarah Prenovitz, Gabriel Durham, Stas Kolenikov, Emily Roessel, Jacob Alex Klerman, and Robin Koralek. 2023. *Health Profession Opportunity Grants (HPOG 2.0) Intermediate-Term Impact Report Appendix*. OPRE Report 2023-202. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.