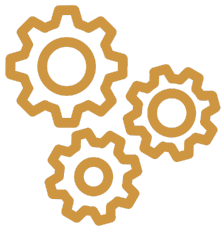
Analysis Plan

Project Name: Identifying Interventions to Increase Affordable Care Act Uptake through a Systematic Review and Meta-analysis

Project Code: 2302

Date Finalized: 2/14/2023



This evaluation is part of the Office of Evaluation Sciences (OES) American Rescue Plan Act of 2021 (ARP) portfolio. The ARP was designed to address immediate needs related to the pandemic, with a specific focus on addressing historically disparate outcomes across race, class, and geography that were further exacerbated by the pandemic. As federal programs are innovating and finding new ways to achieve these goals, the OES portfolio of evaluations will measure whether ARP-funded interventions are working as intended and share lessons learned.

In support of the ARP Equity Learning Agenda, OES is working with agency partners to better understand how to improve awareness, access, and allocation of ARP programs and resources, focusing on ARP programs with equity goals. This set of evaluations will be intentional and strategic in building evidence to understand the role of ARP programs and supported interventions in improving outcomes for historically underserved populations.

Project Description

Over 10 million uninsured individuals are eligible for subsidized coverage through the Affordable Care Act (ACA) marketplaces, and millions more are projected to become eligible for marketplace coverage following the end of the COVID-19 Public Health Emergency. It is, thus, of critical importance that policymakers understand what types of outreach are most effective at increasing ACA take-up among prospective marketplace enrollees. Protecting and strengthening equitable access to high quality and affordable healthcare is also a Department of Health and Human Services (HHS) strategic goal, and the HHS Evidence-Building Plan for fiscal years 2023-2026 includes a priority question related to "increasing choice, affordability, and enrollment in high-quality healthcare coverage."

OES conducted a systematic review and identified 35 published and unpublished randomized, quasi-experimental, and observational evaluations of interventions to increase ACA take-up. The systematic review process included a systematic search of databases and registries, hand-searches of reference lists, and outreach. We met with U.S. federal government organizations like the Council of Economic Advisors, the Centers for Medicare and Medicaid Services, and the U.S. Health and Human Services Office of the Assistant Secretary for Planning and Evaluation; OES was directly involved in the design and implementation of two published

studies[1] and four unpublished studies on strategies to increase enrollment. We also conducted outreach to research- and data-focused organizations like the Abdul Latif Jameel Poverty Action Lab (J-PAL), the National Association of Health Data Organizations (NAHDO), the Robert Wood Johnson Foundation, and the Urban Institute. We placed a general call for relevant studies in NAHDO's monthly newsletter to its members nationwide and reached out directly to State-based marketplaces (SBMs) and All-payer Claims Databases in 6 of the 18 states with state-based marketplaces. Specifically, one project author collected data from a number of rigorous, unpublished studies by Covered California, California's ACA marketplace. Additionally, we discussed the project with two private health insurance brokers and with academic researchers identified through our networks and registry searches.

This analysis plan will outline the next step of this project: synthesizing findings from eligible studies identified by the systematic review using a meta-analysis. A meta-analysis has a number of advantages in this policy setting. It will allow us to identify pooled average effects across interventions or categories of interventions with potentially more precision than is possible in a single study. We are also interested in identifying variation in the efficacy of various interventions, which will give policymakers a sense of the potential impact of various approaches.[2] For that reason, our primary focus is not identifying a simple average treatment effect; instead, we are interested in exploring heterogeneity in these studies to identify which types of interventions work best.

An additional benefit of this approach is its inclusion of unpublished rigorous studies, which were generally led by federal and state agencies.[3] Including unpublished rigorous studies can generate more accurate estimates of effects than a reliance on published studies alone, if interventions with null effects are less likely to result in publications.[4] For this meta-analysis, we do not anticipate publication bias distorting our pooled effect sizes, as most of the unpublished studies identified in our systematic review were rigorous evaluations conducted in a policy context, rather than with the goal of scientific publication.

Preregistration Details

This Analysis Plan will be posted on the OES website at oes.gsa.gov before outcome data are analyzed.

---

[1]Goldin, J., Lurie, I.Z., & McCubbin, J. (2021). Health Insurance and Mortality: Experimental Evidence from Taxpayer Outreach. The Quarterly Journal of Economics, 136(1), 1-49. https://doi.org/10.1093/qje/qjaa029. Yokum, D., Hopkins, D., Feher, A., Safran, E., and Peck, J. (2022). Effectiveness of Behaviorally Informed Letters on Health Insurance Marketplace Enrollment: A Randomized Clinical Trial. JAMA Health Forum, 3(3). doi:10.1001/jamahealthforum.2022.0034.

[2] Bryan, C.J., Tipton, E. & Yeager, D.S. Behavioural science is unlikely to change the world without a heterogeneity revolution. Nature Human Behavior 5, 980–989 (2021). https://doi.org/10.1038/s41562-021-01143-3

[3] Of the studies identified in our review, 23 out of 24 unpublished studies were conducted by federal and state agencies.

[4] Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. Science, 345(6203), 1502–1505. https://doi.org/10.1126/science.1255484.

Hypotheses

The systematic review identified 35 studies that fell in two main categories: (1) information-based interventions aimed at addressing learning costs (awareness and understanding of eligibility and access) and (2) enrollment assistance-based interventions aimed at addressing both learning and compliance costs (completing the enrollment process). [5] Of these 35 studies, 26 were aimed at learning costs alone and 9 were aimed at learning and compliance costs. 5 were observational or quasi-experimental studies; 30 were randomized controlled trials.

To build on this, our meta-analysis will synthesize the evidence in each of these categories to estimate average effect sizes. We will also look for heterogeneity in effects within these categories. The goal of the meta-analysis is not a simple test of whether or not interventions in these categories detectably increase ACA take-up; though we will produce estimates of average effects, we are primarily interested in understanding how effects vary.

Data and Data Structure

This section describes variables that will be analyzed, as well as changes that will be made to the raw data with respect to data structure and variables.

Data Source(s):
To summarize the studies identified by the systematic review, we extracted the following data, when possible:

- Citation or publication status
- Study design
- Recruitment or sampling process
- Data collection year
- Open enrollment or not
- Efforts to address missing data
- Unit of analysis
- Effect estimation method
- Intent to treat estimate or not
- Indicator that the analysis included covariates
- Setting (state or nationwide)
- Participant eligibility
- Total sample size
- Multiple treatment groups
- Control group sample size
- Pooled treatment sample size
- Sample sizes for all treatment groups

[5]Herd, P. & Moynihan, D. “How Administrative Burdens Can Harm Health, " Health Affairs Health Policy Brief, October 2, 2020. DOI: 10.1377/hpb20200904.405159.

- Content of treatment
- Delivery method (phone, email, direct mail, etc.)
- Timing of outcome measures
- Cost (based on this OES framework)
- Control group definition or content
- Definitions of all treatments
- Outcome metrics
- Any secondary outcomes
- Method of aggregation
- Timing of outcome measurement
- Control mean
- Pooled treatment mean
- Means for each treatment group
- Pooled treatment effect size
- Effect sizes for each treatment group
- Pooled treatment standard error
- Standard errors for each treatment group
- Any subgroups measured in the study
- Key finding(s)

For the unpublished studies, the summary data largely came directly from the study authors or administrators. For published studies, the data were extracted from papers and appendices. The dataset built through this process is the foundational data for the meta-analysis.

Data Exclusion:
To qualify for inclusion in the meta-analysis, studies identified in the systematic review must meet the following criteria:

1. Compare marketplace enrollment outcomes between a treatment and a true control group[6] (contrast harmonization), excluding studies that compared one version of an intervention to another (such as different ways of wording a reminder about an upcoming deadline);
2. Measure both treatments and outcomes at the household level (measurement harmonization); and
3. Report a sample size, an effect size, and a measure of variance, either directly or by providing enough information that we could calculate these ourselves.

This narrowed the list of 35 studies identified in the systematic review to 25.

[6] We did not intentionally exclude quasi-experimental designs with this criterion; however, none of the quasi-experimental or observational studies identified in the systematic review met the other two criteria, so the meta-analysis only includes randomized interventions.

Outcomes to Be Analyzed:
The outcome we will focus on is household-level enrollment in marketplace coverage. Again, we are primarily interested in identifying heterogeneity across interventions aimed at increasing ACA enrollment.

Imported Variables:
We do not plan to import variables from other datasets.

Transformations of Variables:
It is common in meta-analyses to convert effect sizes to standardized mean differences, so we may take this approach using the effect sizes and standard errors we collected for each study, in combination with their sample sizes. However, the R package we plan to use can accommodate any effect size, in combination with a measure of variance, so this may not be necessary except as a robustness check.

We plan to create two additional variables. First, an indicator of whether the study took place before or after the COVID-19 Public Health Emergency was declared on January 31, 2020[7] will be created using the timing of the outcome measure.

Second, we will create a study quality indicator variable based on the quality rating scheme from the Journal of the American Medical Association. This scheme covers a range of methods, rating them from 1 to 5, with "properly powered and conducted randomized controlled trials" rated 1, and "opinion of respected authorities; case reports" rated 5. The majority of the included studies are likely to be rated as a 1, due to our selection criteria which required studies to include a comparison between a treatment and a true control group. We plan to have two team members rate the studies based on the methods captured in our summary data to ensure inter-coder reliability; any disagreement will be reconciled through a closer look at study materials. This indicator will be used in subgroup analyses to identify heterogeneous effects across study quality.

Transformations of Data Structure:
We will transform the data structure to account for multi-arm studies and facilitate robust variance estimation, an approach for dealing with non-independent effect sizes (i.e., estimates of treatment effects from multiple experiments in a single study that rely on the same control group).[8]

---

[7] "Determination That A Public Health Emergency Exists." (2020). https://aspr.hhs.gov/legal/PHE/Pages/2019-nCoV.aspx.

[8] Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*(1), 39–65. https://doi.org/10.1002/jrsm.5. Effect-size dependence based on who is conducting the study is another consideration in this realm. However, given the sample size, we do not plan to use two clustering variables (control group and study ownership). Instead, we may re-run the analysis clustering on study ownership as a robustness check.

Specifically, we will restructure the data to work with the R package we will use for these analyses, robumeta. The data collected from our systematic review is stored in a table with a row for each paper or study; we will instead create rows for each treatment arm, with a column indicating the study to account for dependence between estimates from treatment arms using the same control group in the analysis. Eleven of the studies that met the inclusion criteria for the meta-analysis had multiple treatment groups, for a total of 37 treatment groups; combined with the other 14 studies, this means we will have 51 coefficients in the final meta-analysis based on this transformation.

Treatment of Missing Data:
We do not anticipate missing necessary data among the studies that qualified for inclusion in the meta-analysis. If we have questions about specific variables, we will contact study authors or calculate them using other reported statistics.[9] These calculations will be noted in the record of analysis.

Descriptive Statistics, Tables, & Graphs

The following tables and visualizations will be central to this project:

- A flow diagram outlining the systematic review process;
- A series of descriptive visualizations of the studies identified in the systematic review, eg., bar graphs of where and when the studies took place, a visualization of the different intervention types used, etc.;
- A summary of findings table of the studies in the systematic review; and
- A forest plot displaying effect estimates and confidence intervals for both individual studies and meta-analyses, illustrating heterogeneity in effect sizes and the weights of studies in the meta-analysis.

Statistical Models & Hypothesis Tests

This section describes the statistical models and hypothesis tests that will make up the analysis — including any follow-ups on effects in the main statistical model and any exploratory analyses that can be anticipated prior to analysis.

Statistical Models:
To generate pooled estimates of average effects on enrollment, a random effects meta-analysis with inverse variance weighting will be used to identify the minimum variance unbiased estimate.[10] Random effects is the most appropriate way to identify the overall effect of these interventions because of its assumption that there is a distribution of effect sizes, and the studies

---

[9] Algebraic calculation of missing values is a common approach in meta-analyses; see Weir, C.J., Butcher, I., Assi, V. et al. Dealing with missing standard deviation and mean values in meta-analysis of continuous outcomes: a systematic review. BMC Med Res Methodol 18, 25 (2018). https://doi.org/10.1186/s12874-018-0483-0.

[10] Hedges, L. V. (1983). A random effects model for effect sizes. Psychological Bulletin, 93(2):388–395. Hartung, J., Knapp, G., and Sinha, B. K. (2008). Statistical meta-analysis with applications. Wiley series in probability and statistics. Wiley, Hoboken, N.J. OCLC: ocn212627347.

included are estimating different yet related intervention effects; it does not require the assumption that the true effect of intervention is the same across all studies, as a fixed-effect meta-analysis would.[11] In this study, our goal is to understand heterogeneity in effects, not to precisely estimate the mean treatment effect, so a random effects approach allows us to identify variation in effects across studies. Weighting by the inverse variance increases the efficiency of estimates.

To test for heterogeneous treatment effects across the 51 treatment arms in the meta-analysis, we plan to use Cochran's Q test, which tests for differences between three or more matched sets of proportions and is a standard test for heterogeneity in meta-analyses. It tests the null hypothesis that the true treatment effect is the same across studies and variations are simply caused by chance. A test of significance of τ^2, the variance of the underlying effect distribution typically estimated using restricted maximum likelihood (REML), is another test of heterogeneity we may use. The R package metafor facilitates this calculation, testing the null that $\tau^2 = 0$.

To further understand heterogeneity in effects of these interventions, we will conduct meta-regression analyses to identify differences in effect across variables of interest. We intend to use subgroup analyses to explore differences across nominal categorical variables of interest (whether the intervention aimed to address learning costs or learning and compliance costs, the state in which the intervention took place, study-level quality) and meta-regression to explore characteristics of interest that are measured continuously (sample size and year of data collection) or as binary categorical variables (cost of intervention and whether the intervention took place before or after the COVID-19 pandemic was declared a Public Health Emergency). In these analyses, centering continuous variables will make the intercepts more interpretable.

In these models, we will apply four best practices in meta-analyses[12]:

1. Quantifying heterogeneity in effect sizes: We will use subgroup analyses to understand how treatment efficacy differed across types of interventions and study populations.
2. Including all relevant within-study variation in analyses: We will use multilevel modeling approaches to account for the dependence structure of multi-arm studies. This approach involves estimating two sources of heterogeneity – within and between study variation – and is possible using the metafor package in R; the package clubSandwich allows for the estimation of robust standard errors.[13]
3. Adjusting for confounders, including study quality and publication bias: We will include measures of study quality and risk of bias as covariates in our models. We also plan to assess the risk of publication bias in our overall sample.

---

[11] We may also conduct a fixed effect analysis as a robustness check for our findings, though again, the assumption that there is one true effect of intervention in this area seems unlikely.

[12] Tipton, Elizabeth & Bryan, Christopher & Murray, Jared & McDaniel, Mark & Schneider, Barbara & Yeager, David. (2022). Why Meta-Analyses of Growth Mindset and Other Interventions Should Follow Best Practices for Examining Heterogeneity. 10.13140/RG.2.2.34070.01605.

[13] Pustejovsky, J.E. & Tipton, E. (2022). Meta-analysis with robust variance estimation: Expanding the range of working models. *Prevention Science, 23*(3), 425-438.

4. Explaining heterogeneity using moderation analysis: The subgroup analyses and meta-regressions we outline above are driven by theories of which characteristics of interventions and study populations might drive heterogeneity in treatment effects. Using R packages metafor and clubSandwich will allow us to include competing moderators in the same model, which is valuable because these moderators could be correlated.

Confirmatory Analyses:
The average pooled effect will be used to test for significant effects of interventions in these categories. We can also construct confidence intervals around these means using the standard errors.

However, this approach does not account for treatment heterogeneity. Including that source of variance in calculating prediction intervals will yield ranges in which policymakers might expect future effects to fall.[14] With 25 studies, 11 of which have more than one treatment arm (for a total of 51 treatment arms), we are well-positioned to estimate an overall prediction interval; recent studies suggest that prediction intervals with fewer than 20 studies may be too narrow.[15] For this reason, we may not estimate prediction intervals within intervention subgroups, or we may need to take a different approach to estimation, such as drawing bootstrapped samples.

Exploratory Analysis:
We would like to explore heterogeneity in treatment effects across subgroups of consumers (such as across age groups), but it is likely that there are not enough studies with common subgroup analyses to allow this. We may conduct exploratory analyses of those that cover overlapping subgroups. We may also look at additional subgroups of the studies we identified, such as interventions among people who had submitted applications and were found eligible and those among likely uninsured people who had not yet interacted with the marketplace.

Quantifying publication bias in this area is not a priority in this study, but we may decide to explore ways of testing for it using this unique dataset by measuring the skewness of the standardized deviates[16] or an adjusted rank correlation test.[17]

---

[14] For an application of this approach, see Jackson, K.C. and Mackevicius, C. (2021). The Distribution of School Spending Impacts. *National Bureau of Economic Research Working Paper Series. http://www.nber.org/papers/w28517.*

[15] Nagashima, K., Noma, H., and Furukawa, T. A. (2019). Prediction intervals for random-effects meta-analysis: A confidence distribution approach. Statistical Methods in Medical Research, 28(6):1689–1702.
Kontopantelis, E., Springate, D. A., and Reeves, D. (2013). A Re-Analysis of the Cochrane Library Data: The Dangers of Unobserved Heterogeneity in Meta-Analyses. PLoS ONE, 8(7):e69930.

[16] Lin L, Chu H. (2018). Quantifying publication bias in meta-analysis. Biometrics, 74(3):785-794. doi: 10.1111/biom.12817.

[17] Begg CB, Mazumdar M. (1994). Operating characteristics of a rank correlation test for publication bias. Biometrics, 50(4):1088-101. PMID: 7786990.

Finally, we may also want to illustrate that our effects are not driven by any one of our 25 studies. To demonstrate this, we could take bootstrap samples from our data and create a histogram of effects from those samples. This would illustrate the distribution of the estimated pooled effect and downweight the influence of any one study.

Inference Criteria, Including Any Adjustments for Multiple Comparisons:
The key adjustment we plan to make is to use robust variance estimation to account for dependent effect size estimates from studies with multiple treatment arms relying on comparisons to the same control group. In making inferences, we will use 95% confidence intervals and typical p-values ($p < 0.05$, $p < 0.01$, $p < 0.001$).

Limitations:
The key limitations of the systematic review are also relevant considerations for the meta-analysis: in what ways might our sample of studies be biased? Are there ways in which this sample might not adequately represent the universe of interventions and target populations in this area? Our assessment of potential publication bias will be useful in addressing these concerns, and our analyses of correlates of heterogeneity will help us understand how the findings of these studies might generalize to other intervention types, populations, and settings.

Additionally, meta-analyses can be biased toward overestimating effect sizes. Effect sizes tend to be much larger in published studies (a concern we address by incorporating a number of unpublished studies) and studies with small sample sizes (which is less a concern among the studies we include); effect sizes also tend to be much higher in quasi-experiments vs. randomized experiments, and again, we include only experiments with true controls.[18] However, the potential for study-level biases remains, and averaging inaccurate estimates leads to an inaccurate average. Our focus on heterogeneity rather than a precise pooled effect size makes this less of a concern, and our assessments of study-level risk of bias and quality will be useful in understanding this area of potential limitations as well.

[18] Cheung, A. C. K., & Slavin, R. E. (2016). How Methodological Features Affect Effect Sizes in Education. Educational Researcher, 45(5), 283–292. https://doi.org/10.3102/0013189X16656615.