

The design of a randomized controlled trial to evaluate multi-dimensional effects of a section 1115 Medicaid demonstration waiver with community engagement requirements

Kristin A. Linn^{a,b}, Kristen Underhill^{c,d}, Erica L. Dixon^{b,e}, Elizabeth F. Bair^{b,e}, William J. Ferrell^{b,e}, Margrethe E. Montgomery^h, Kevin G. Volpp^{b,e,f,g}, Atheendar S. Venkataramani^{b,e,*}

^a Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, 423 Guardian Dr, Philadelphia, PA 19104, USA

^b Center for Health Incentives and Behavioral Economics, University of Pennsylvania, 423 Guardian Dr, Philadelphia, PA 19104, USA

^c Columbia Law School, New York, NY, USA

^d Department of Population and Family Health, Mailman School of Public Health, Columbia University, New York, NY, USA

^e Department of Medical Ethics and Health Policy, Perelman School of Medicine, University of Pennsylvania, 423 Guardian Dr, Philadelphia, PA 19104, USA

^f Department of Medicine, Perelman School of Medicine, University of Pennsylvania, 423 Guardian Dr, Philadelphia, PA 19104, USA

^g Center for Health Equity Research and Promotion, Crescenz VA Medical Center, Philadelphia, USA

^h National Opinion Research Center at the University of Chicago, Bethesda MD and Chicago, IL, USA

ARTICLE INFO

Keywords:

Randomized controlled trial
Medicaid demonstration waiver
Policy evaluation
Prospective survey

ABSTRACT

Section 1115 demonstration waivers provide a mechanism for states to implement changes to their Medicaid programs. While such waivers are mandated to include evaluations of their impact, randomization – the gold standard for assessing causality – has not typically been a consideration. In a critical departure, the Commonwealth of Kentucky opted to pursue a two-arm randomized controlled trial (RCT) for their controversial 2018 Medicaid Demonstration waiver, which included work requirements as a condition for the subset of beneficiaries deemed able-bodied to maintain eligibility for benefits. Beneficiaries were randomized 9:1 to the new waiver program or a control group who would retain their current benefits as part of the existing Medicaid expansion program. To address potential bias from differential attrition from the Medicaid program that would accrue from solely analyzing administrative data, our team designed a rich, prospective, longitudinal survey to collect primary and secondary outcomes from six categories of interest to policymakers: insurance coverage, health care utilization and quality, health behaviors, socioeconomic measures, personal finances, and health outcomes. At baseline, a subset of survey participants was invited to participate in the collection of biometric samples via in-person follow-up visits, and a cross-section were also invited to participate in qualitative interviews. While the demonstration waiver was terminated before the program began, our study design illustrates that it is possible for other researchers and state agencies seeking to evaluate Medicaid demonstration waivers and other demonstration policies to work together to implement high quality randomized trials – even for controversial policies.

1. Introduction

States are increasingly using Section 1115 waivers to implement changes to their Medicaid programs. Consistent with the experimental imperative of demonstration waivers, the Centers for Medicare and Medicaid Services (CMS) stipulates that states must conduct and report the results of evaluations of their waiver programs. Unfortunately, to date these evaluations have yielded limited understanding of whether a given waiver program has achieved its objectives. A 2018 report by the Government Accountability Office (GAO) found §1115 waiver

evaluation designs have typically lacked rigor and generally fail to provide actionable, policy-relevant information [1]. One major reason for this massive gap in evidence is that the majority of waivers have not employed an experimental strategy that randomizes beneficiaries to the new program or an appropriate control. Instead, universal implementation of waiver programs has forced researchers to rely on descriptive snapshots of Medicaid access and beneficiary health outcomes over time. Such analyses cannot reliably measure the impacts of a program, since any observed changes in beneficiary health outcomes could be attributed to other policy changes in the state or societal

* Corresponding author at: Center for Health Incentives and Behavioral Economics, University of Pennsylvania, 423 Guardian Dr, Philadelphia, PA 19104, USA.
E-mail address: atheenv@pennmedicine.upenn.edu (A.S. Venkataramani).

<https://doi.org/10.1016/j.cct.2020.106173>

Received 12 June 2020; Received in revised form 13 September 2020; Accepted 17 September 2020

Available online 07 October 2020

1551-7144/© 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

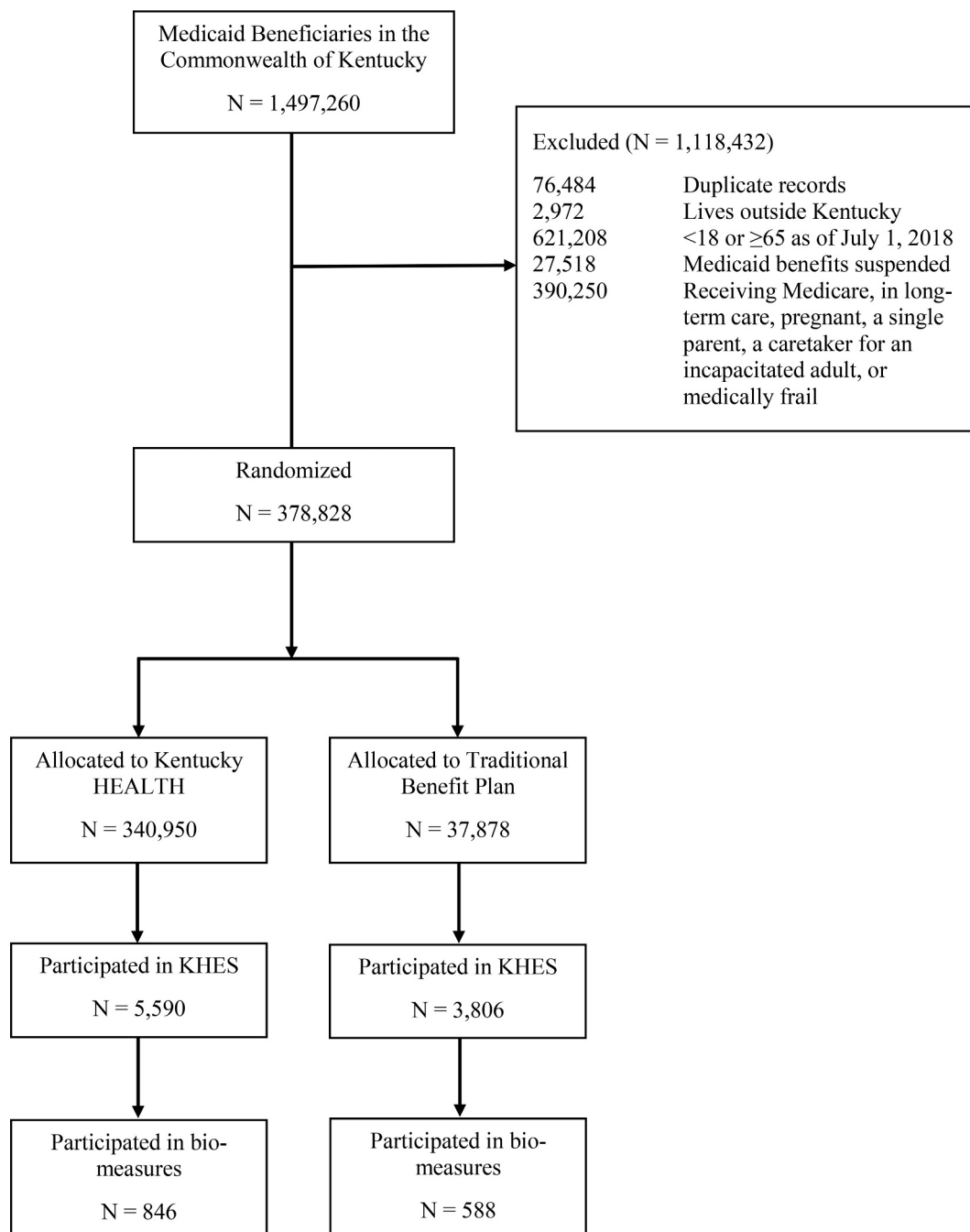


Fig. 1. Flow diagram of eligibility, randomization, and participation in the Kentucky Health Experiment Survey (KHES).

trends in health and economic opportunities [2,3]. This ambiguity underscores the need for evaluation designs to include carefully considered comparison groups [4,5].

Recognizing the critical need for conducting scientifically rigorous waiver evaluations, we worked with state policymakers to design an unprecedented randomized controlled trial (RCT) of the Commonwealth of Kentucky's §1115 demonstration waiver. The waiver program, known as Kentucky HEALTH (“Helping to Engage and Achieve Long-Term Health”), sought a variety of changes, most notably community engagement requirements and cost-sharing as a condition for continued eligibility for adults considered able-bodied. The Commonwealth hypothesized that the program would improve beneficiary health as a result of “able-bodied, working age adults [experiencing] the dignity of a job, of contributing to their own care, and gaining a foothold on the path to independence.” [6] An alternate

possibility, however, was that the program would lead to coverage losses among beneficiaries who did not meet or report their required activities, as well as a reduction in program access among potential future beneficiaries [7,8]. The types of research designs thus far used in evaluations of Medicaid waivers would not allow policymakers to credibly distinguish these two possibilities [1,9].

In this report, we describe the RCT portion of the evaluation that would have been conducted had Kentucky HEALTH been implemented as planned on July 1, 2018. Program implementation was delayed several times due to litigation, and Kentucky HEALTH was ultimately ended by executive order on December 16th, 2019. We discuss design and implementation challenges and general lessons that may be relevant to evaluations of Medicaid waiver programs – or other public programs – in other states. In light of implementation delays before the ultimate removal of the waiver, we also include a description of

beneficiary departure from Medicaid by arm in the time between notice of randomized group assignment and October 2018.

2. Materials and methods

2.1. Intervention – Kentucky HEALTH

The Kentucky HEALTH demonstration waiver aimed to introduce community engagement (work requirements) and cost-sharing requirements, as well as remove vision and dental benefits except as accessed through health behavior incentives via the My Rewards program. The program was intended for beneficiaries aged 18–64 considered able-bodied (i.e., not pregnant, disabled, or medically frail). A more detailed description of all Kentucky HEALTH program components can be found in the original study protocol that was submitted to the Centers for Medicare and Medicaid services (see Supplementary Material).

2.2. Control group

Beneficiaries assigned to the control group would not be subject to any of the requirements of Kentucky HEALTH.

2.3. Eligibility for randomization

Currently enrolled Medicaid beneficiaries between age 18–64 with either a valid address or phone number were eligible for randomization. Exclusion criteria included having an address outside of Kentucky, being suspended from Medicaid, being medically frail or in long term care, Medicare dual-eligible individuals, and pregnant women.

2.4. Randomization

A total of 378,829 Medicaid beneficiaries met the study eligibility criteria as of February 2, 2018. 90% of eligible participants were randomized to the intervention ($N = 340,951$) and 10% to the control group ($N = 37,878$). Assignment of 10% to the control group, instead of the typical 50% (which would maximize statistical power) was done based on the Commonwealth's stated policy preference to involve as many individuals as possible in the intervention. In addition, randomization was applied at the individual, rather than household, level due to administrative challenges in identifying (fluid) households. Randomization was conducted by National Opinion Research Center at the University of Chicago (NORC), details of which are provided in the Supplementary Material. The flow diagram in Fig. 1 contains details about the number of individuals in the population who met exclusion criteria and the number of eligible individuals who were randomized.

2.5. Primary outcomes and hypotheses

Primary outcome measures for the study were identified after considering the Commonwealth's objectives in seeking to implement Kentucky HEALTH, guidance from the Centers for Medicare and Medicaid Services (CMS), and the scientific literature on the impacts of Medicaid on health and socioeconomic outcomes. This literature is reviewed in depth in Sommers, Gawande, and Baicker (2017), which broadly focuses on access to coverage; utilization; self-reported health; health outcomes such as glycated hemoglobin, blood pressure, cholesterol levels, depression, diabetes, and cancer; and mortality [10]. Based on these sources, we deemed six core categories of outcome measures to be of greatest policy interest: insurance coverage, health care utilization and quality, health behaviors, socioeconomic measures, personal finances, and health outcomes; Table 1 shows each category with the primary outcome in bold. For each core outcome category, the Commonwealth's primary goal for the Kentucky HEALTH program is shown with the associated two-tailed evaluation hypothesis. All outcomes

were beneficiary-level measures, given that beneficiary health was the primary goal of the waiver. We determined the total sample size for prospective data collection based on the six primary outcomes.

2.6. Secondary outcomes and hypotheses

The non-bolded outcomes in Table 1 comprise our secondary outcomes. Included in the secondary outcomes were biomarker measurements, described further below, which were collected from a subset of high-risk individuals who indicated a diagnosis of diabetes and/or hypertension in the baseline survey.

Built into the evaluation plan was the opportunity to identify, (pre-) specify, and conduct additional analyses in future years. This is critical, as new hypotheses of interest may have emerged from surveys and qualitative interviews or from changes in demonstration waivers or implementation. Information obtained from analysis of data for individuals entering the waiver at the time of first implementation would be useful for structuring hypotheses, data collection efforts, and research designs for future randomizations to examine waiver impacts among individuals who first entered the program after initial implementation, when program features and implementation strategies had stabilized. In this context, the goal of the current document was to balance pre-specification (which minimizes prospects of data mining) and the opportunity to continually learn from the data in a policy relevant manner.

2.7. Data sources

The primary data source would have been a prospective, longitudinal survey of individuals sampled from both the waiver and control arms of the RCT population – the Kentucky HEALTH Experiment Survey (KHES). We opted for primary data collection for two reasons. First, Kentucky – like many other states – does not have an all payer claims database capable of tracking health utilization by individuals across changing sources of insurance. Furthermore, among states that do have all payer claims databases, many are limited to inpatient care. As it is critical to follow beneficiaries who leave the Medicaid program to understand both positive and negative program effects – and to mitigate bias from differential attrition - reliance on Medicaid claims alone could have provided biased estimates of health and utilization effects. Second, administrative data generally have blind spots, including the lack of validated self-reported physical and mental health measures and information on labor force participation. These topics have been reliably interrogated in a number of large-scale surveys.

Survey activities (sampling, fielding the survey, and providing de-identified data to the evaluation team) were contracted out to NORC. The initial KHES assessment, which was planned as a baseline for the original implementation date, attempted to contact 34,191 individuals from the intervention group and 22,556 from the traditional Medicaid control group from April to August 2018, in order to obtain a target sample of 5400 intervention and 3600 control group completed surveys at baseline. In total, NORC obtained 9396 completed baseline surveys, including 5590 from the intervention group and 3806 from the control group (Fig. 1). Although an equal number of samples from the two arms would have been preferred to maximize power for hypothesis testing, NORC anticipated the need to contact a large portion of the 10% randomized to the control arm in order to obtain an adequate number of completed surveys over time. This prompted our team to prescribe a 60/40% composition of intervention/control completed surveys in the survey study design.

Notably, planning an RCT requires careful sample size calculations that assume realistic survey response rates. For the initial KHES survey, the yield rate was 16.7%; using the various response rate formulas compiled by the American Association of Public Opinion Research (AAPOR), the response rate for this study ranges from 29.1% (definition 1) to 48.9% (definition 4) [11,12]. Yield rate is calculated from the total

Table 1
Kentucky HEALTH program goals, primary hypotheses, and primary and secondary outcomes for the RCT.

Kentucky HEALTH program goal	Primary evaluation hypothesis ^a	Primary and secondary outcomes (Primary in bold)
1. To increase transition of current beneficiaries to employer-sponsored health insurance	Beneficiaries moved into Kentucky HEALTH will experience differential changes in insurance coverage (duration, type) compared to traditional Medicaid beneficiaries.	Current insurance status Current Medicaid enrollment Current insurance source Access to employer-sponsored insurance Uptake of employer-sponsored insurance Months on Medicaid/non-Medicaid/uninsured in the past year Frequency and duration of lockouts Timing of changes in insurance
2. To improve beneficiaries' engagement in preventive health care and encourage appropriate use of medical services	Beneficiaries moved into Kentucky HEALTH will have significantly different uptake of preventive health care services, compared to traditional Medicaid beneficiaries.	Annual wellness visit (physical exam) Number of emergency department visits Reasons for emergency department visits Hospitalizations Hospitalizations for ambulatory care sensitive conditions Hospitalizations for SUD Number of primary care visits Number of specialist visits Number of dental care visits Number of vision care visits Usual place of care Type of regular care provider Prescription medication use Foregone medical care Foregone dental care Foregone prescription medications
3. To improve health behaviors among beneficiaries	Beneficiaries moved into Kentucky HEALTH will engage in significantly different health behaviors, compared to traditional Medicaid beneficiaries.	Smoking Substance use
4. To foster socioeconomic advancement among beneficiaries	Beneficiaries moved into Kentucky HEALTH will have significantly different labor force participation and income, compared to traditional Medicaid beneficiaries.	Currently employed Months employed in the past year Hours worked per week Volunteerism (Hours/week) Wages Family income
5. NA ^c	Beneficiaries moved into Kentucky HEALTH will have significantly different amounts of debt and differential banking status, compared to traditional Medicaid beneficiaries.	Consumer and medical debt (composite) Banking status
6. To improve health outcomes among beneficiaries	Beneficiaries moved into Kentucky HEALTH will have significantly different health outcomes, compared to traditional Medicaid beneficiaries.	Self-reported overall health Physical health days Self-reported mental health Mental health days Self-reported dental health Self-reported changes in health status Mortality Biometrics ^b

Abbreviations: HEALTH, Helping to Engage and Achieve Long-Term Health; RCT, Randomized Controlled Trial; SUD, Substance Use Disorder.

^a All tests of evaluation hypotheses will consider “beneficiaries” to include all who are beneficiaries in each group at baseline. That is, beneficiaries who transition off the Medicaid program during the 5-year waiver period will be included in analyses, for both the Kentucky HEALTH group and the traditional Medicaid control group.

^b Compared in a high-risk sample that included all individuals who indicated they carried a diagnosis of diabetes and/or hypertension.

^c The evaluation team suggested this category, so there is no associated Kentucky HEALTH Program Goal.

number of sources attempted to be reached, while, the AAPOR response rate formulas vary the denominator by dropping cases for various factors such as changed address with no forwarding, disconnected phone, or ineligibility for survey. In some cases, the number of unreachable cases that would have been ineligible had they been reached is estimated and removed from the denominator in the response rate calculation.

Whether they remained in Medicaid or not, our design specified that baseline survey respondents would be re-surveyed at six months after Kentucky HEALTH implementation to capture immediate waiver effects, one year after implementation, and yearly thereafter for a total of five years. The five-year follow up would allow for an unprecedented long-run examination of the health and socioeconomic trajectories of a low-income population in the Medicaid program and the opportunity to evaluate health outcomes that may require years to develop (e.g., chronic disease severity or mortality) [17–19].

Randomized individuals who were 60 years of age or older as of July 1, 2018 were not eligible for participation in the longitudinal survey. The upper age limit was chosen so that all survey participants had the potential to be exposed to the intervention or control for the full five years of follow up. Recipients of Temporary Assistance for Needy Families (TANF) and the Supplemental Nutrition Assistance Program (SNAP) were randomized but excluded from the longitudinal survey, since these individuals were already subject to work requirements and therefore would not experience the full impact of the waiver's new conditions.

During baseline survey data collection, 2921 individuals were eligible for bio-measure collection based on their responses on the KHES which self-reported presence of diabetes, hypertension, or both. From this high-risk population, NORC sampled and collected 1434 (846 intervention and 588 control) complete bio-measure panels that included blood pressure, pulse, weight, height, waist and hip circumference, and

blood spot data. These data were used to calculate BMI, systolic and diastolic blood pressure, total cholesterol levels, and glycated hemoglobin (A1c, a marker of blood sugar control). Further details about the sampling strategy and power calculations are included in the Supplementary Material.

Drawing on the survey sample, we also purposively recruited a cohort of 127 individuals to complete a one-on-one qualitative interview by phone, which discussed healthcare utilization, health status, experiences with Medicaid, labor force and volunteering, financial circumstances, and perceptions of the Kentucky HEALTH program features. We planned to contact these individuals once per year during the duration of the waiver to discuss experiences with the program and any changes in insurance status. Qualitative data collection was planned for additional groups of beneficiaries, healthcare providers, and program staff in later years.

Administrative data collected and maintained by the Commonwealth was expected to serve as a secondary data source, and it includes basic demographic characteristics (e.g., age, gender, race, level of education, location), family income (continuous percentage as a function of FPL), pre-existing diagnosis codes for medical comorbidities (e.g., diagnoses of chronic diseases), employment history in the year prior to waiver, health service utilization, and assignment to intervention (waiver) or control (traditional Medicaid) group in the RCT. This dataset is updated on a quarterly basis and includes indicators for ongoing Medicaid program participation, reasons for nonparticipation (e.g., enrollment in employer-sponsored insurance premium assistance, program lockout), and current employment status.

Since administrative utilization data is only observed for beneficiaries during periods of Medicaid participation, the KHES would have been critical for capturing utilization outcomes from beneficiaries who transitioned to employer-sponsored insurance, experienced lockouts, or left the Medicaid program for other reasons. Fig. 2 provides a timeline of randomization and planned data collection.

2.8. Sample size

Randomization for the RCT was applied at the population level to all 378,829 Medicaid beneficiaries who met the study eligibility criteria as of February 2, 2018. A portion of randomized beneficiaries from each arm of the RCT were sampled for longitudinal survey data collection. The sample size for the KHES was ultimately decided based on budgetary considerations and response rate projections by the NORC, which led our team to target a 60/40% composition of intervention/control completed surveys. An initial survey sample size of $N = 9000$ was projected by NORC to provide a total sample size of $N = 7250$ completed surveys at the five year follow up time point, assuming approximately 20% attrition across five years.

Given the projection of $N = 7250$ completed surveys at the conclusion of the study, we report here the minimum detectable effect size for a simple comparison of binary outcomes across the two arms (Kentucky HEALTH vs. traditional Medicaid) at a single wave of follow-up using the KHES. Of our six primary outcomes, insurance status, annual wellness visit, current smoking status, and labor force participation are naturally binary. The primary analysis would dichotomize debt as none or more than \$0 USD. Physical and mental health would be dichotomized as 14 or fewer poor health days in the past 30 days for each. In the effect size calculation, we wanted to achieve at least 90% power to reject the null hypothesis using a two-sided test of two independent proportions. Conservatively, we used a Bonferroni adjusted Type I error to control the family wise error (FWE) for six tests, i.e., one for each of the primary outcomes.

Under these specifications, we would have over 90% power to detect a difference in proportions of a primary outcome between arms of approximately 0.05 if the proportion in one group is 0.5 (i.e., the most conservative setting) in year five. Differences in proportions of 0.05 would constitute policy-meaningful effects; for example, a reduction of

the percentage of Medicaid beneficiaries who respond “Every Day” or “Some Days” to the question, “Do you now smoke cigarettes every day, some days, or not at all?” from the baseline rate of approximately 68% to 63% would correspond to 20,000 beneficiaries quitting smoking. The subsequent reductions in heart disease, stroke, and cancer would likely translate into reduced costs and burden on the medical system. Furthermore, we would have greater power for testing outcomes with baseline proportions greater or less than 0.5, as well as greater power in waves prior to year five when the expected level of attrition would be lower than 20%. Effect size calculations for the bio-measure study are provided in the Supplementary Material.

2.9. Analysis plan

Here we provide an overview of the statistical analysis plan for the KHES data. The full waiver evaluation plan, which includes analysis specifications for observational administrative data and the qualitative surveys, can be found in the draft evaluation plan that was submitted to CMS (Section 5) that is provided in the Supplementary Material. Our primary approach specified an intent-to-treat (ITT) analysis that would target the effect of being randomized to the waiver group relative to control, regardless of each beneficiary's level of exposure to Kentucky HEALTH. For each outcome, a generalized estimating equation (GEE) would be fit with terms for intervention arm, each follow up survey wave, and interactions between wave and arm. We would also include in the model several pre-specified baseline variables that are listed in the Supplementary Material. We would report robust sandwich variance estimators that account for individual level clustering, and inference would focus on the interactions between wave and intervention arm. Our primary approach for multiple comparison adjustment would be to create a composite outcome based on all survey questions within each of the six outcome domains and perform one hypothesis test for each based on the GEE model. A Bonferroni adjustment would then be used for the six hypothesis tests. As a secondary approach, we would report unadjusted p -values for tests of every primary and secondary outcome as well as adjusted p -values using a method that controls the family-wise error rate at 0.05 for all tests within each outcome domain. Additional details are given in the Supplementary Material.

2.10. Ethical approval and trial registration

The RCT is built into the roll-out of the Section 1115 Medicaid demonstration project, which was being implemented by Kentucky and overseen by the Secretary of HHS. The waiver was considered to be a “demonstration project. .. subject to the approval of department or agency heads. .. that [is] designed to study, evaluate, improve, or otherwise examine public benefit or service programs.” As such, the randomized implementation of the waiver was exempt from internal review board (IRB) approval under 45C.F.R. § 46.104(d)(5), and human subjects concerns are part of review of the waiver by the Secretary [20]. NORC obtained separate IRB approval for all original data collection efforts. Our trial was registered at <https://clinicaltrials.gov> under identifier NCT03602456 on July 26, 2018 [21].

3. Results

3.1. Analysis of beneficiary departures from Medicaid

In any Medicaid population, employment, income, and other life changes can cause a portion of beneficiaries to transition in and out of the program, known as churning. We compared rates of departure from Medicaid by arm of the RCT and with respect to several key subgroups; departure was defined as absence of active enrollment as of August 2019.

As of August 19th, 2019, 26.6% of the randomized study cohort had unenrolled in Medicaid, including 26.7% of the arm assigned to

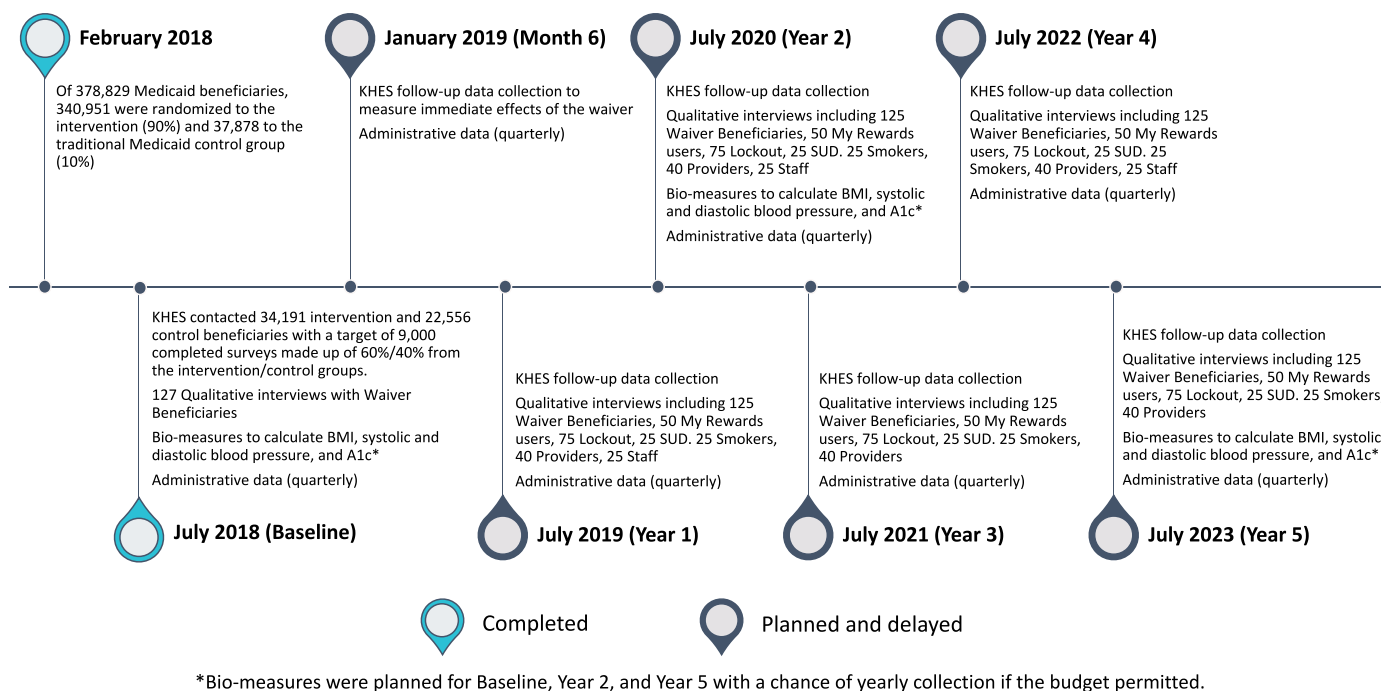


Fig. 2. Timeline of randomization and planned data collection for the Kentucky HEALTH evaluation.

Kentucky HEALTH and 26.1% of the control arm. Of the survey sample, 23.4% had unenrolled overall. The proportion of beneficiaries who had unenrolled was only slightly lower in the control arm at 23.2% versus 23.5% in the intervention arm. Table 2 displays demographic characteristics of the RCT cohort separately by Medicaid enrollment status and treatment arm. Overall, former beneficiaries who were no longer enrolled in Medicaid tended to be younger and more likely to be male and employed. We do not report *p*-values because of the large sample sizes both in the full randomized cohort and survey sample, noting that statistical significance of the differences may not reflect practical or meaningful differences between arms. With the potential for meaningful differences in mind, we designed the KHES to include beneficiary interviews regardless of whether they were enrolled or unenrolled in Medicaid throughout the follow up period to allow for an unbiased comparison of the intervention and control arms.

4. Discussion

This article presents an overview of the protocol for what is arguably the first RCT developed to prospectively evaluate the effects of a Section 1115 Medicaid Demonstration Waiver. The study design addressed significant gaps in research designs used to evaluate demonstration programs to date, particularly around assessing causal effects of Medicaid innovations on beneficiary health. The study design aimed to provide rigorous evidence about the total effects of a multi-component intervention.

Within the broader research design of our evaluation, we embedded original data collection to address gaps in the administrative records typically used in waiver evaluations, natural experiments to isolate the effects of individual different program components, and a large qualitative study (including interviews with current and former Medicaid

Table 2

Administrative demographic characteristics by enrollment and intervention arm for the overall waiver population (N = 378,828).

	Not enrolled in Medicaid (Percent)			Still in Medicaid (Percent)		
	Overall	Intervention	Control	Overall	Intervention	Control
Age, mean (SD)	36.7 (13.0)	36.7 (13.0)	36.8 (13.1)	38.0 (12.5)	38.0 (12.5)	37.9 (12.5)
Female	44.45	44.48	44.20	49.70	49.73	49.39
Race ^a						
White	72.86	72.81	73.28	75.83	75.87	75.50
Black	11.32	11.36	10.99	10.40	10.36	10.68
Asian	1.39	1.38	1.45	1.31	1.31	1.33
Native American	0.18	0.18	0.16	0.15	0.16	0.12
Hawaiian and Pacific Islander	0.11	0.11	0.11	0.07	0.07	0.07
Managed Care Organization ^a						
Aetna Better Health	12.51	12.55	12.14	12.93	12.97	12.55
Anthem Blue Cross Blue Shield	16.96	16.95	17.01	14.14	14.12	14.31
Humana – CareSource	15.96	16.04	15.31	15.48	15.46	15.63
Passport Health Plan	24.27	24.20	24.99	24.50	24.49	24.61
WellCare of Kentucky	30.14	30.11	30.41	32.86	32.86	32.81
SNAP Active	20.95	20.84	22.02	30.69	30.75	30.08
Employed	29.24	29.25	29.15	23.41	23.41	23.49
Has substance use disorder	1.63	1.63	1.60	1.32	1.32	1.38
Number of adults in household, mean	1.5	1.5	1.5	1.5	1.5	1.4

^a Categories may not add up to 100% due to missing values.

beneficiaries, providers, and Medicaid program staff) to better elucidate effect mechanisms. Given our plan to follow a cohort of beneficiaries prospectively, regardless of program enrollment status, our design would have allowed for the addition of secondary outcomes and corresponding survey questions at later follow up times based on information obtained from qualitative interviews or relevant changes in current events. For example, at the 3- and 4-year follow up time points originally planned for July 2020 and 2021, we could have asked KHES participants about the impact of COVID-19 on their ability to access care, self-reported health and mental health, and labor force participation. This flexibility was a strength of our design that would have provided timely information to policymakers looking to propose a renewal, extension, amendment, or expansion of the demonstration.

A number of challenges arise when surveying Medicaid populations, and many methods have been proposed to minimize non-response [13]. Low-income populations tend to have higher rates of illiteracy and lower reading comprehension skills than the general population, which can influence response rates and the quality of the data collected from responders [14,15]. Low income populations also tend to be more mobile than the general population and often live in non-standard housing. In light of these factors, our team partnered with NORC to implement several evidence-based measures [16] that aimed to maximize survey response rates: 1) a pre-survey mailer notified beneficiaries that they would be contacted at a later time and given the opportunity to participate in the KHES; 2) the mailer also explained the value of the data that would be collected for informing future decisions about Kentucky HEALTH and ensured beneficiaries that all individual-level responses would be confidential and not shared with the Commonwealth; 3) the survey was designed to take no longer than 20 min on average; and 4) individuals could participate by phone or by filling out the survey questions online. Despite planning similar measures for future waives of the KHES, loss to follow-up due to out-of-date contact information would have been a concern. NORC planned to use the Accurant locating service to obtain updated contact information for all baseline respondents. Since Accurant relies on credit records as a primary source of contact information, this could have introduced a potential source of bias if contact information was found to be less complete for lower-income survey participants.

Although Kentucky HEALTH was ultimately not implemented, our study design can serve as a resource to other states and research teams planning evaluations of Medicaid waivers. In the following section, we briefly describe several alternative designs our team considered during the planning stage. First, we sought to randomize specific components of Kentucky HEALTH, so as to more reliably isolate the causal consequences of each component. We considered using a factorial design as an efficient approach to estimating main effects and interactions among the waiver components [22]. However, after discussions with the Commonwealth, such a design was ultimately deemed too complicated logistically from an administrative standpoint and likely to increase burden and confusion among beneficiaries who would be randomized to different subsets of the program components. We also contemplated using cluster randomization to assign households as opposed to individuals to the intervention or control. Although this design may have benefitted from higher social acceptability among beneficiaries by avoiding discordant assignment to intervention and control within households, it would have been challenging to correctly identify and track members of dynamic household units. Prior to the Commonwealth's decision to randomize beneficiaries, we also considered a stepped wedge study design where roll out of KY HEALTH would be staggered across areas [23–25]. However, we ultimately concluded that the small number of workforce areas in Kentucky would not provide sufficient power and that the RCT would be more robust to confounding by changes and trends over time in program administration, labor force, and other factors.

In March 2019 - more than a year after we randomized Medicaid beneficiaries in Kentucky - CMS issued a document recommending that

evaluations “be rigorous, incorporate baseline and comparison group assessments, as well as statistical significance testing.” [26] The guidelines call for thoughtful evaluation plans that clearly detail how the proposed design will ensure hypotheses about waiver performance can be tested. The prescribed structure for evaluation plans broadly mimics the typical format of an RCT protocol, requiring specification of the main hypotheses of interest, study design, target population and control group, primary and secondary outcomes including time points at which they will be tested, data sources, and analytic strategy. The guidance recommends a discussion of study limitations and how the chosen design will minimize them. Following the section on limitations, CMS requires that states justify any analysis plan that does not include a comparison group or baseline data analysis. The implication is that less-rigorous, non-randomized study designs for new, untested program modifications should be the exception, not the rule. By encouraging the use of randomization when possible and emphasizing research design best practices, CMS is inviting high quality waiver evaluations that will help policymakers design evidence-based modifications to Medicaid programs in the future. The study design presented in this paper can be used by other researchers as an example of how to meet these new CMS requirements.

4.1. Lessons for policymakers and researchers

During the 12-month planning phase prior to July 2018, our team participated in biweekly calls with state health officials to finalize the design of the waiver evaluation and work out the logistical details of implementation. In addition to biweekly calls, we also traveled as a team to Frankfurt, Kentucky to meet with policymakers in person in September 2017, January 2018, and May 2018. Early on, our team invested time building an understanding of the context and motivation for Kentucky HEALTH, which included the twin challenges of poor health and labor market outcomes among Medicaid beneficiaries and a devastating opioid epidemic. We also spent time in the initial months solidifying our knowledge of various logistical aspects that would shape our approach to the evaluation such as the algorithm that would be used to classify beneficiaries as medically frail and the structure and location of different databases containing individual-level beneficiary information and claims. After learning about the goals and logistics of Kentucky HEALTH, we identified policymakers' main hypotheses about the effects of the waiver and began to suggest possible study designs that would enable us to test those hypotheses. Throughout these early discussions, we conveyed the benefits of a randomized, controlled experiment, emphasizing that it would provide higher-quality evidence about the success or failure of the waiver than a set of observational study designs tailored to the various outcomes and hypotheses. This was particularly important since many other states in the region were engaged in significant concurrent policy interventions, making it likely that ‘difference in differences’ type analyses would be confounded by these concomitant interventions. Importantly, our team benefitted from working with a group of scientifically-minded policymakers in Kentucky who understood the value and rigor that randomization would bring to the evaluation. As a result, state health officials enthusiastically supported our recommendation to do an RCT.

We also obtained buy-in from the state for other design recommendations by iterating frequently with policymakers and involving them in all steps of the decision-making process. As one example, we presented several options for the timing and frequency of follow-up bio-measure collection that would require different levels of investment, allowing the state to choose the option that best balanced their scientific and budgetary priorities. Flexibility to work within some of the state's logistical and budgetary constraints helped build rapport between our respective teams, facilitating trust and progress. For instance, though 1:1 randomization is common in clinical trials, policymakers in Kentucky wanted as many individuals as possible to be in the intervention group and requested a 9:1 allocation ratio between

intervention and control [5]. With only 10% of the Medicaid population assigned to the control group, our survey sampling partner NORC alerted our team that they would not be able to achieve equal group sizes for the KHES when accounting for projected response rates. Given the state's preference for the 9:1 allocation, we modified our plan for the KHES data collection to survey beneficiaries using a 60/40% breakdown of intervention/control participants.

Altogether, our experiences point to several lessons for researchers and policymakers seeking to evaluate Medicaid demonstration waivers using RCTs:

1. Start early with pre-implementation collaboration between researchers and state officials: we utilized a lead time of twelve months to settle on design features with the Commonwealth and plan data collection with NORC.
2. Find well-placed leaders within state government who are supportive of scientific approaches to policy evaluation.
3. Align study design with programmatic goals. For example, while researchers might desire a 50/50 randomization, policy priorities might dictate a higher proportion be part of the intervention group.
4. Collect data on Medicaid beneficiaries even after they exit the program.
5. Consider survey response rate and power: states with smaller Medicaid populations may need to assign more than 10% of the beneficiary population to the control arm, or risk inadequate power to compare study arms.
6. Account for delays: delayed implementation may have important effects on study design and may necessitate additional methods to keep the original cohort engaged, such as postcard mailings or additional follow-up phone calls.

5. Conclusions

Randomized evaluations of Medicaid waiver programs – and other state policies – are more possible to implement than policymakers and researchers may realize. In conjunction with the Commonwealth of Kentucky, we designed an RCT that randomized a cross-sectional cohort of all enrolled Medicaid members in the Commonwealth of Kentucky in February 2018 to continue to receive traditional Medicaid or to receive benefits according to Kentucky HEALTH, a Section 1115 Medicaid waiver demonstration. As part of our research design, we designed a survey to obtain longitudinal data from a subset of the randomized individuals over five years of follow up, as well as qualitative interviews with beneficiaries and providers. Longitudinal follow-up of a representative, randomized sample allows for comparison of long-term health and labor outcomes between individuals assigned to the waiver and those assigned to receive traditional Medicaid benefits. Results from RCTs would provide actionable information to CMS and other states designing Section 1115 Medicaid waivers to promote the health and well-being of the citizens who interact with the Medicaid system.

Our experiences underscore that it is possible for other researchers and state agencies seeking to evaluate Medicaid demonstration waivers and other demonstration policies to work together to implement high quality randomized trials – even for controversial policies.

Funding support

This project was supported by the Commonwealth of Kentucky. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Commonwealth of Kentucky.

Trial registration

Identifier NCT03602456 at <https://clinicaltrials.gov>

Declaration of Competing Interest

None.

Acknowledgements

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. Dr. Volpp has received research support from Humana, Vitality/Discovery, Oscar, and Hawaii Medical Services Association, all unrelated to the subject of this article. He is also a part-owner of the consulting firm VAL Health. Drs. Underhill and Venkataramani have received research support from the Donaghy Foundation and Rx to conduct a separate survey of Kentucky adults about Medicaid policy. The authors would like to thank NORC for implementing the randomization and collecting the survey data. In particular, the authors would like to acknowledge Elizabeth Ormson, MS, and Adrienne Call, MSW, MPP for their contributions.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cct.2020.106173>.

References

- [1] Government Accountability Office, Medicaid Demonstrations: Evaluations Yielded Limited Results, Underscoring Need for Changes to Federal Policies and Procedures, United States Governmental Accountability Office, 2018.
- [2] J. Pearl, Causal inference in statistics: an overview, *Statist. Surv.* 3 (2009) 96–146.
- [3] S. Athey, G.W. Imbens, The state of applied econometrics: causality and policy evaluation, *J. Econ. Perspect.* 31 (2) (2017) 3–32.
- [4] P.W. Holland, Statistics and causal inference, *J. Am. Stat. Assoc.* 81 (396) (1986) 945–960.
- [5] A.S. Venkataramani, K. Underhill, K.G. Volpp, Moving toward evidence-based policy: the value of randomization for program and policy implementation, *JAMA.* 323 (1) (2020) 21–22.
- [6] S. Verma, SPEECH: Remarks by Administrator Seema Verma at the 2018 Medicaid Managed Care Summit, <https://www.cms.gov/newsroom/press-releases/speech-remarks-administrator-seema-verma-2018-medicaid-managed-care-summit>, (2018).
- [7] A.E. Carroll, The problem with work requirements for Medicaid, *JAMA.* 319 (7) (2018) 646–647.
- [8] B.D. Sommers, A.L. Goldman, R.J. Blendon, E.J. Orav, A.M. Epstein, Medicaid work requirements — results from the first year in Arkansas, *N. Engl. J. Med.* 381 (11) (2019) 1073–1082.
- [9] K. Underhill, A. Venkataramani, K.G. Volpp, Fulfilling States' duty to evaluate Medicaid waivers, *N. Engl. J. Med.* 379 (21) (2018) 1985–1988.
- [10] B.D. Sommers, A.A. Gawande, K. Baicker, Health insurance coverage and health — what the recent evidence tells us, *N. Engl. J. Med.* 377 (6) (2017) 586–593.
- [11] T. Smith, A Revised Review of Methods to Estimate the Status of Cases with Unknown Eligibility, https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/FindingE.pdf, (2009).
- [12] A.S. Venkataramani, E.F. Bair, E. Dixon, et al., Assessment of Medicaid beneficiaries included in community engagement requirements in Kentucky, *JAMA Netw. Open* 2 (7) (2019) e197209–e197209.
- [13] D.D. Fredrickson, T.L. Jones, C.A. Molgaard, et al., Optimal design features for surveying low-income populations, *J. Health Care Poor Underserved* 16 (4) (2005) 677–690.
- [14] B.D. Weiss, R.L. Reed, E.W. Kligman, Literacy skills and communication methods of low-income older persons, *Patient Educ. Couns.* 25 (2) (1995) 109–119.
- [15] I. Kirsch, A. Jungeblut, L. Jenkins, A. Kolstad, Adult literacy in America: A first look at the findings of the National Adult Literacy Survey, *Improvement USD0E0oEra*, 1993, p. ed2002.
- [16] P. Edwards, I. Roberts, M. Clarke, et al., Increasing response rates to postal questionnaires: systematic review, *BMJ* 324 (7347) (2002) 1183.
- [17] K. Baicker, S.L. Taubman, H.L. Allen, et al., The Oregon experiment—effects of Medicaid on clinical outcomes, *N. Engl. J. Med.* 368 (18) (2013) 1713–1722.
- [18] B.D. Sommers, E. Arntson, G.M. Kenney, A.M. Epstein, Lessons from early Medicaid expansions under health reform: interviews with Medicaid officials, *Medi. Medicaid Res. Rev.* 3 (4) (2013).
- [19] B.D. Sommers, B. Maylone, R.J. Blendon, E.J. Orav, A.M. Epstein, Three-year impacts of the affordable care act: improved medical care and health among low-income adults, *Health Aff (Millwood)*. 36 (6) (2017) 1119–1128.
- [20] Exemption of Certain Research and Demonstration Projects from Regulations for Protection of Human Research Subjects 48 Fed. Reg. 9266 In:1983, (1983).
- [21] NCT03602456, Evaluation of the Health and Economic Consequences of Kentucky's Section 1115 Demonstration Waiver, <https://ClinicalTrials.gov/show/NCT03602456>, (2018).

- [22] T.B. Baker, S.S. Smith, D.M. Bolt, et al., Implementing clinical research using factorial designs: a primer, *Behav. Ther.* 48 (4) (2017) 567–580.
- [23] C.A. Brown, R.J. Lilford, The stepped wedge trial design: a systematic review, *BMC Med. Res. Methodol.* 6 (1) (2006) 54.
- [24] M.A. Hussey, J.P. Hughes, Design and analysis of stepped wedge cluster randomized trials, *Contemp. Clin. Trials.* 28 (2) (2007) 182–191.
- [25] K. Hemming, T.P. Haines, P.J. Chilton, A.J. Girling, R.J. Lilford, The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting, *BMJ* 350 (2015) h391.
- [26] Centers for Medicare & Medicaid Services, Section 1115 Demonstrations Developing the Evaluation Design, <https://www.medicare.gov/medicaid/downloads/developing-the-evaluation-design.pdf>, (2020).