Pre-Analysis Plan for Year 1 of the Michigan Contraceptive Access, Research, and Evaluation Study

The Short-Run Impact of Subsidizing Contraceptives: Experimental Evidence from the First Year of the Michigan Contraceptive Access, Research, and Evaluation Study

Martha Bailey, University of Michigan and NBER (PI)

Jennifer Barber, University of Michigan

Vanessa Dalton, University of Michigan and Planned Parenthood of Michigan

Daniel Eisenberg, University of Michigan

Vanessa Wanner Lang, University of Michigan

Julia Kohn, Planned Parenthood Federation of America

Draft: September 29, 2017
Updated: February 12, 2020

1. Introduction

This document describes the analysis plan for the papers that will result from the first year of the Michigan Contraceptive Access, Research, and Evaluation Study (M-CARES). Its purpose is to pre-specify our planned analyses in order to minimize concerns about specification searching and data mining. Because we will learn from our analysis of the data, we may perform additional analyses that are not specified here. We will post additional analysis plans for years 2-5 after processing results for year 1.

M-CARES aims to quantify the effects of increasing the affordability of contraceptives on the outcomes of women and their families. At Planned Parenthood of Michigan (PPMI) clinics, we will recruit 5,000 women ages 18 to 35 years old to participate. Because PPMI patients with family incomes below the poverty line receive free services, only women with incomes above the federal poverty line and without insurance coverage for contraceptives will be included. Half of the study participants will be randomized to receive a voucher to purchase contraceptives of their choice at PPMI within 100 days.

Our analysis will use the randomization of vouchers together with surveys and administrative data (from Census data, tax data, birth records, PPMI clinic records, credit reports) to estimate the short-run and long-run causal effects of subsidizing highly effective contraceptive methods for U.S. women. (This analysis plan, however, focuses only on the outcomes after year 1). Over the course of the study, these rich data allow us to provide novel experimental evidence on outcomes such as contraceptive use, pregnancy, and childbearing. A central contribution of our analysis is that this study allows us to consider a very wide range of additional outcomes, for which there has been no experimental evidence for the U.S. These include physical health and health care use; educational attainment and labor-market outcomes; relationship quality; mental health and well-being; and receipt of public benefits.

Appendix Table A1 summarizes the survey data and administrative data that we plan to use, both sources providing complementary perspectives on women's lives.[1] A great advantage of administrative data is that they contain information from the universe (or a random sample) of the study population, which limits the role of non-response (Ashenfelter & Plant, 1990). In addition, administrative data measure many outcomes better than self-reports on surveys, which ameliorates the potentially large role of this source of measurement error (Bound, Brown, & Mathiowetz, 2001). We supplement administrative data with surveys about dimensions of women's lives not captured in administrative data (for instance, pregnancies, physical and mental health, relationship quality, and overall life satisfaction). We have designed the study to minimize non-response, and our analysis plan lays out our plan to investigate and adjust for non-response.

This analysis plan for year 1 considers two primary outcome domains. First, we examine whether subsidies for contraceptives increase contraceptive efficacy and, especially, the use of long-acting reversible contraceptives (LARCs). Second, we examine whether subsidies for contraceptives reduce unintended pregnancy and childbearing.

We will also conduct supplementary analyses relating to outcomes in the domains of economic self-sufficiency, financial stability, and neighborhood quality. (Analyses of other outcomes and more detailed analyses of long-term outcomes considered in year 1 will be conducted after follow-up surveys 2 and 4 years after enrollment.) We also examine the mechanisms for any effects on primary and supplemental outcome domains. These include analyses of (1) reasons for voucher-nonuse and resulting pregnancy and childbearing outcomes; (2) how vouchers affect out-of-pocket expenditures for reproductive health care and satisfaction with reproductive health care services; (3) more detailed analyses of effects at the extensive and intensive margins, and (4) heterogeneity in causal effects by age, race, income groups, as well as other pre-randomization characteristics.

[1] As more data become available, we plan to add more administrative data including school and education records, criminal justice records, as well as records covering children and the family members of these women.

The year 1 analysis plan is organized as follows. Section II describes the design of M-CARES, data sources, and the definition of outcomes. Section III describes our econometric framework and estimating equations. Section IV describes our planned analyses of the PPMI patient population (as compared to the U.S. national population and Michigan population. Section III then describes the population of PPMI patients who are eligible for M-CARES and those who, conditional upon being eligible, elect to participate in M-CARES. Section V describes our expected first stage outcomes based on other studies and reports the resulting power calculations. Section VI outlines our planned analyses of primary outcomes, and Section VII details our supplemental analyses. Section VIII concludes with a discussion of difficulties in interpretations and caveats.

2. The Design of M-CARES Data Sources

M-CARES was motivated by several commonly cited findings. Nearly half of all pregnancies in the U.S. are unintended, and unintended pregnancies are five times more likely to occur for poor women relative to more affluent women (Finer & Zolna, 2016; Sedgh, Singh, & Hussain, 2014). They are also significantly more common among young and minority women.

Evidence regarding the potential costs of unintended pregnancies for individuals, society, and the economy is less commonly cited. In 2011, 42 percent of unintended pregnancies (excluding miscarriages) ended in abortion (Finer & Zolna, 2016). Roughly two thirds of unplanned births were funded by public insurance programs, primarily Medicaid (Sonnfield & Kost, 2015). Quasi-experimental evidence suggests that unintended pregnancies have a variety of long-term implications for the lives of women and their families. Evidence that exploits changes in legislation permitting young women access to the Pill and the roll-out of the first federally funded family planning clinics in the U.S. suggests that unintended pregnancies in the 1960s and 1970s reduced women's educational attainment, employment, career advancement, and family incomes (Bailey, 2006; Bailey, Hershbein, & Miller, 2012; Goldin & Katz, 2002; Hock, 2008). They may also result in decreased marital stability and increased public assistance expenditures (Bailey, 2013; Bailey, Malkova, & McLaren, 2016). Ultimately, unplanned pregnancy may limit the life opportunities for children, contributing to the cycle of poverty (Ananat & Hungerman, 2012; Bailey et al., 2016).

Behavioral barriers such as inconsistent or inappropriate method use often result in contraceptive failures. Forty-one percent of unintended pregnancies occur among women who are using contraception in the month they become pregnant (Sonfield, Hasstedt, & Gold, 2014), suggesting that methods which do not require adherence, or long-acting reversible contraceptives (LARCs), may be an important tool for addressing high rates of unintended pregnancy. Although lack of provider training and contraceptive counseling are important barriers to increasing adoption of LARCs (Harper et al., 2015), financial barriers remain significant as LARCs have larger up-front costs compared to other reversible methods (Trussell, 2011).

Financial barriers to contraceptive access are likely to become more salient in the near future. President Donald J. Trump and the Republican-controlled Congress have promised to pass sizable funding cuts for family planning care. They plan to cut public funds to organizations like Planned Parenthood (Republican Party, 2016), which includes cutting both Title X funding as well as Planned Parenthood's ability to receive Medicaid funds. Second, proposed cuts to Medicaid (via block granting to states) are likely to cause significant cuts to services provided, including contraceptive care. Both of these changes threaten to eliminate funding for the 8 million women currently served using public funds (Sonfield & Benson Gold, 2012). In addition, changes to the Affordable Care Act (ACA) may increase financial burdens among women with private insurance for receiving family planning services and reduce insurance coverage, thereby increasing the need for publicly-funded services as their funding is diminished.[2]

Observational evidence suggests that the elimination of public funding for family planning services would increase rates of unintended pregnancy in the U.S. by 68 percent (Jennifer J. Frost et al., 2016) and that the public costs of these births would increase by 75 percent (Jennifer J. Frost, Sonnfield, Zolna, & Finer, 2014). Strong inferences about the policy effects of such cuts, however, are tempered by well-known limitations of observational evidence. M-CARES will provide novel experimental evidence regarding these important questions.

2.1 M-CARE Study Design and Inclusion Criteria

The population of interest is women at risk of unintended pregnancy in the U.S. For logistical reasons, M-CARES focuses on women in Michigan. Table 1 shows that Michigan women are very similar in their socio-demographic characteristics to women in the U.S. The one exception to this is that Michigan has significantly fewer Latina residents. Michigan also falls around the national median of many key behaviors related to contraceptive use, such as cohabitation, marriage, age at first birth, nonmarital childbearing, and teenage childbearing (Lesthaeghe & Neidert, 2006).

Table 1. Comparison of Women in Michigan to Women in the U.S. (excluding Michigan)

	U.S. women 18-35 years old		Michigan women 18-35 years old	
	Mean	S.D.	Mean	S.D.
White (non-Hispanic)	0.511	(0.500)	0.662	(0.473)
Black (non-Hispanic)	0.166	(0.372)	0.197	(0.398)
Hispanic	0.234	(0.423)	0.069	(0.253)
Married	0.268	(0.443)	0.246	(0.431)
High school graduate	0.877	(0.328)	0.905	(0.294)
Some college	0.618	(0.486)	0.655	(0.475)
College degree	0.310	(0.462)	0.295	(0.456)
Wage income ($2016)	15,120	(17,220)	13,932	(15,399)
Number of kids	0.800	(1.185)	0.811	(1.202)
Number of kids less than 5 years old	0.339	(0.633)	0.349	(0.652)

Notes: Data is taken from the 2015 American Community Survey, for women ages 18-35 years old with family income less or equal to $80,000 ($2016).

2.1.1 Inclusion Criteria

M-CARES will recruit 5,000 women at Michigan PPMI clinics over at least 12 months who face potentially large out-of-pocket costs for highly effective contraceptives, especially LARC methods. In order to be eligible to participate in the study, the woman needs to be

(1). 18-35 years old,

(2). physically capable (biologically female and fecund) and at risk of having a pregnancy (has sex with men),

[2] The ACA increased access to reliable contraceptives and family planning services for women with insurance, by increasing insurance coverage, requiring health insurance policies to cover family planning services, and reducing copays for these services (Jennifer J. Frost, Frohwirth, & Zolna, 2016).

(3). not pregnant at the time of enrollment and not wishing to become pregnant in the next 12 months, and

(4). face some out-of-pocket costs for contraceptives at PPMI.

M-CARES' assessment of #1-#3 are based on women's answers to the surveyor and a brief set of screening questions. M-CARES' assessment of #4 is based on PPMI's own assessment of a patient's fee scale. When a patient schedules an appointment at PPMI, PPMI uses a brief assessment of the patient's ability-to-pay for services. This includes questions about the patient's sources of income, the number of family members, and her insurance coverage. Based on this information, PPMI assigns each patient a "fee scale" that is used to determine the patient's out-of-pocket costs for services. Patients who are below the federal poverty line (FPL) (fee scale 1 or A) will not be charged costs for contraceptive services by PPMI, regardless of their insurance coverage. Other patients are assigned fee scale of 2/B (101-150% FPL), 3/C (151-200% FPL), 4/D (201-250%), or 5/E (250%+ of FPL). Unless they have insurance to cover their visit, patients will be charged 25%, 50%, 75% or 100% for the services they receive at their PPMI visit.

***UPDATE 02/12/2020

The original study design *implicitly* included another eligibility criterion: only women coming to PPMI for a standard clinician visit are assessed their fee scale and sit in the waiting room, so these were the only women who were recruited. On May 13, 2019 (following a brief pilot in a handful of sites), the study expanded recruitment to women coming to PPMI for *non*-clinician visits. Non-clinician visits are typically very brief and include women picking up their contraceptive supplies, quick follow up visits to pick up test results (e.g., STI/STD testing, pregnancy), and other visits that do not require meeting with a clinician.

****

2.1.2 Screening and Recruitment

We have contracted with NORC at the University of Chicago to hire and train professional surveyors to recruit women in PPMI clinic waiting rooms.

Recruitment proceeds as follows. A NORC surveyor will sit at a desk with information about the M-CARE study. During registration with PPMI, each woman will receive a small card from the PPMI receptionist indicating her fee scale and insurance coverage for her visit (Figure 1).

Figure 1. M-CARES Card Received at Check-in

M-CARES
MICHIGAN CONTRACEPTIVE ACCESS, RESEARCH AND EVALUATION STUDY

Email: m-carestudy@umich.edu Toll-free: 1-844-864-8258
Website: sites.lsa.umich.edu/m-carestudy

	Yes	No
Here for clinician visit?	☐	☐
Has insurance?	☐	☐
Insurance used for today's visit?	☐	☐

Fee scale: ☐ 1/A ☐ 2/B ☐ 3/C ☐ 4/D ☐ 5/E

VID: VD:

Study conducted by UNIVERSITY OF MICHIGAN

After check-in, the NORC surveyor will approach each woman and introduce herself. Surveyor will ask women to complete a 5-minute screening survey, offering a $10 reimbursement for her time. To protect participant's confidential information and privacy, the screening survey will be self-administered on a tablet. Answers to the questions will determine if a woman meets the inclusion criteria. (See Appendix D for oral script and screening survey).

If a woman meets the basic inclusion criteria, the tablet will invite the woman to participate. Participation requires that the patient (1) agrees to be contacted to complete subsequent surveys; (2) consents for M-CARES to use her administrative data and the administrative data of her family.[3]

If the patient chooses to participate, the tablet will walk her through the informed consent process with assistance from a professional survey worker as needed. Enrollment will be conducted on the electronic tablet, which will encode consent responses and personal information such as name, SSN, date of birth, as well as contact information. (See Appendix D for informed consent form).

2.1.3 Voucher Randomization

The tablet will then randomize women to receive PPMI vouchers (which we call "PPMI gift cards") for participating women and, for women randomized to receive a voucher, display the dollar amount of the gift card and a voucher identification number (VIN) on the tablet. The surveyor will write down the VIN and the voucher amount (VD) on the business card and return the card to the woman.

Each patient who meets the inclusion criteria and elects to participate in the study has an equal chance of receiving a voucher. Vouchers will be individually assigned and linked to name, birth date, and date of enrollment to prevent trading or giving away the vouchers. Vouchers can be used for *any* contraceptives and related services at PPMI for up to 100 days and participants may return to the PPMI clinic multiple

[3] On the consent form, "family" is explained to mean children and other individuals who are found on administrative records.

times to redeem them if desired.[4] Initially, t~~T~~he voucher amounts reflect the total out-of-pocket costs for an uninsured woman to have a Liletta IUD inserted after applying the PPMI the sliding scale (see Table 2A). Liletta is the lowest cost IUD which costs $492, including insertion and the medically required pregnancy test, for patients with a fee scale of 5/E. Out-of-pocket costs for women without insurance and fee scale 2/B to 4/E are lower as indicated.

**UPDATE 02/12/2020

After spending around six months in the field, the study team discovered that PPMI *very rarely* inserted the Liletta IUD. Rather than making any contraceptive (including a generic IUD) free, the voucher was reducing the price of the name-brand IUDs, such as Paraguard, Skyla, and Mirena, by 50%. Based on these reports, the study team increased the amount of the voucher to cover the *name-brand* IUDs after applying the sliding scale. This change took place on March 4, 2019. The revised voucher amounts are presented in the Table 2B.

In addition, the withdrawal of PPMI from Title X on November 4, 2019, increased the price of contraception for most women, including fee scale 1/A patients age 22 and older. The study team opted to continue to make any contraceptive up to the cost of a name-brand IUD free by covering 100% of the cost for every group with out-of-pocket costs. The revised voucher amounts are presented in Table 2C. The two notable changes are (1) the large increase in the cost of contraceptive in the control group after November 4, 2019, and (2) the increase in out-of-pocket costs for 1/A women that made them eligible for vouchers.

****

~~Although the voucher amounts are determined as the PPMI price for the Liletta,~~ M-CARES participants can use the voucher to purchase *any* type of contraceptives within 100 days. The voucher could, for instance, be used to select a more expensive IUD or contraceptive method (e.g., an Implant). Participants would pay out of pocket for the cost exceeding the voucher. The voucher could also be used for birth control pills, injections, hormonal patches, or any other kind of birth control (excluding abortion).

In addition, the tablet will present standard information provided by PPMI to inform the woman about the benefits and risks of different contraceptive methods and encourage her to talk to her health care provider about her needs. After this, the M-CARES participant will proceed with her appointment as planned. All contraceptive decisions and discussions with health care providers will take place in the clinic after recruitment.

Following the participant's clinical visit, we will invite her to take a 25-minute baseline survey in the PPMI waiting room. We will offer the respondent $60 cash to take the baseline survey in the clinic or $40 to take it online after she leaves the clinic.

Table 2A. Exact Voucher Amounts by Income Group before March 4, 2019

Women's Income Group	PPMI Sliding Scale: % of Fee Charged	Randomly Assigned Voucher Amounts toward Remaining Out of Pocket Cost	
</= 100% FPL	0%	*No voucher assigned*	
101-150% FPL	25%	$0	$123
151-200% FPL	50%	$0	$246
201-250% FPL	75%	$0	$369
>/= 251% FPL	100%	$0	$492

Note: FPL=Federal Poverty Line.

[4] "Related services" are services medically required for the use of a specific contraceptive device. For an IUD, related services include a pregnancy test and insertion in addition to the device.

Table 2B. Exact Voucher Amounts by Income Group, March 4, 2019, to November 3, 2019

Women's Income Group	PPMI Sliding Scale: % of Fee Charged	Randomly Assigned Voucher Amounts toward Remaining Out of Pocket Cost	
</= 100% FPL	0%	*No voucher assigned*	
101-150% FPL	25%	$0	$223
151-200% FPL	50%	$0	$446
201-250% FPL	75%	$0	$669
>/= 251% FPL	100%	$0	$892

Note: FPL=Federal Poverty Line. Voucher amounts are not exactly 2 times the pre-March 4 amounts, because the cost of the pregnancy test and insertion remained the same.

Table 2C. Exact Voucher Amounts by Income Group on or after November 4, 2019

Women's Income Group	PPMI Sliding Scale: % of Fee Charged	Randomly Assigned Voucher Amounts toward Remaining Out of Pocket Cost			
		Ages 18-21		Ages 22-35	
</= 100% FPL	0%	*No voucher assigned*		$0	$107
101-150% FPL	25%	$0	$267	$0	$411
151-200% FPL	50%	$0	$533	$0	$533
201-250% FPL	75%	$0	$800	$0	$800
>/= 251% FPL	100%	$0	$1,066	$0	$1,066

Note: FPL=Federal Poverty Line.

2.2 Data Sources

A central contribution of our analysis is that a combination of survey data and administrative data provide complementary perspectives. Appendix A describes the data sources used in this study, both for the analyses in year 1 as well as records that we will incorporate in future years. We include a summary of the data available from each data source in Appendix A Table A1, as well as provide a detailed description of the data sources and linking procedures. This section provides a brief overview of these data sources.

2.2.1 Screening Survey

The 5-minute screening survey takes place before recruitment into the study. Answers to its questions determine if a woman meets the inclusion criteria and also gather pre-randomization information. This information allows us to examine differences between the eligible and ineligible population as well as the correlates of study participation. In addition to the eligibility questions, it gathers socio-demographic characteristics (age, race, marital and cohabitation status, number of children, and education), information about the reason and payment for the PPMI visit, childbearing history, the use of contraceptives, and contraceptive method satisfaction.

2.2.2 Baseline Survey

The 25-minute baseline survey asks more detailed questions about how the participant used the voucher or, if she did not, why not; any changes in contraceptive use (or intent to change contraceptive use) since the screening survey; work and income; school enrollment; religion; birth and pregnancy history; birth control and healthcare access; plans for the future; childhood environment; attitudes and beliefs about contraception, relationship quality; and physical and mental health.

2.2.3 Subsequent Surveys

We will also follow M-CARES participants in subsequent surveys currently scheduled for 2 months and 4 years after recruitment. These surveys will be designed to measure the long-term effects of this intervention in many dimensions and include many of the same questions as the baseline survey: attitudes about and use of reproductive health care; pregnancies (including plans and intentions) and their outcomes; physical health and health care use; educational attainment and labor-market outcomes; relationship quality; mental health and well-being; and participation in programs. We do not yet have follow-up surveys drafted, because we are waiting on the year 1 results to inform their construction. We will submit a second pre-analysis plan after we analyze the results from year 1.

2.2.4 Administrative Records

With informed consent, we will link M-CARES participants and their family members to their administrative records as described in Appendix A. In addition to socio-demographic information, these records include, (1) PPMI patient records; (2) credit reports; (3) Tax Data (including IRS 1040s data); (4) birth certificates; (5) Census, American Community Surveys and Current Population Survey (CPS) Annual Social and Economic Supplement (ASEC) from the Michigan Research Data Center (RDC);[5] (6) Education data (K-college and National Student Clearinghouse, NSC); and (7) public program participation, including Supplemental Nutritional Assistance Program (SNAP), Temporary Assistance for Needy Families (TANF), and Medicaid MSIS and Medicaid T-MSIS, (8) criminal justice records. Note that most of these sources provide pre-randomization information data that we can use in our analysis. We expect to link almost all of M-CARES participants to (1)-(4) and the short-form Census (5). Of course, records which are specific to Michigan (or the use of a particular program or service, e.g., (7)) will miss individuals who lived out of state before enrollment or move out of state after enrollment.

PPMI records
PPMI patient data contain medical records with information on services obtained at PPMI, including the type of services, the date of services, the results of any tests or medical diagnoses, and payment information (amount and source). We will use these data to examine the use of contraceptives, pregnancy incidence, and use of reproductive health services, such as abortions.

Credit reports
Participants' credit records contain rich information on their credit history, such as credit accounts' payment status, outstanding balance, credit limit, and payment history (see Avery, Calem;, and Canner (2003) for a detailed discussion of credit data). Based on results of previous work, we expect to link almost all M-CARES participants to credit records.[6] We will use this data to study the impact of voucher assignment on financial well-being.

Tax data
We expect to link every study participant to the tax records. Tax data allow us to characterize tax filing participants, before and after the intervention, in terms of (1) living circumstances (living with parents, single headship, living with married or unmarried partner, etc.), (2) the number of children in household (and when they were born and age at first birth), (3) renter/owner status, and (4) neighborhood quality (an important metric for standard of living). In addition, tax data allow us to assess (5) educational investments (via credits for these expenditures), (6) exact income from wage earnings in the previous year, (7) receipt/eligibility of the Earned Income Tax Credit, and (8) eligibility for other public assistance programs.

[5] Note: surveys such as the ACS and CPS will only contain information on a random sample of M-CARES participants.
[6] The Oregon Medicaid Experiment study was able to link 97 percent of participants to the credit bureau data (Finkelstein et al., 2012).

Birth certificates

Birth certificates are available for respondents and their children from the Michigan Department of Health and Human Services (MDHHS). We expect to find approximately 77 percent of participants and 91 percent of participants' children in these data.[7] Each birth record contains the date of birth, mother's marital status, father's name, source of payment, gestation, number of prenatal visits, and the date of the first prenatal visit. This data allows us to validate mother reports of the number of children and study several types of outcomes: (1) childbearing (number of births, timing/spacing of births); (2) childbearing and partnership stability (non-marital childbearing, no father on the birth certificate); (3) heath of a pregnancy (number of prenatal visits, the timing of the first prenatal visit, gestation length), and (4) receipt of public services (whether the birth was paid for by Medicaid).

Census, ACS and CPS

Census data lets us observe the following characteristics in 2000, 2010, 2020, and 2030: (1) living circumstances (living with parents, single headship, living with married or unmarried partner, etc.), (2) the number of children in household (and when they were born and age at first birth), (3) renter/owner status, (4) incarceration status, (5) neighborhood quality, and (6) address. American Community Survey (ACS) and Current Population Survey (CPS) data contain information for a random sample of M-CARES participants. If the sample overlap is large enough, we will use these records as a supplement to the Census. These surveys allows us to characterize participants in a similar way as the Census, but on a more frequent basis between Census years.

Education data

We will use two types of administrative sources for education data, the National Student Clearinghouse (NSC) and Michigan education data (K-college). The NSC provides information on enrollment, the beginning and ending date that a student is enrolled during each term, whether a student is enrolled full or part-time, whether a student has earned a degree, and the date the degree is earned. For the subset of students participating in the "DegreeVerify" service, we will be able to also obtain college major and degree type (e.g., BA). We expect to link the vast majority of our study participants to NSC data if they have pursued post-secondary education. We will work directly with school districts in Michigan to link our participants to their education records, therefore the final sample of linked respondents depends on district-specific rules for releasing data. In addition, we will not be able to link participants who obtained education outside of the state. Michigan K-college school records include classes taken, achievement test scores, grades, absenteeism, and school delinquency reports. Together, NSC and Michigan K-college records enable us to study education enrollment and attainment of the study participants.

Public program participation

We will use several types of data available to analyze public program participation: Medicaid, Supplemental Nutrition Assistance Program (SNAP), and Temporary Assistance for Needy Families (TANF). These data allow us to observe program enrollment and expenditures for study participants.

Criminal justice records

We expect that criminal records data will be available through the RDC starting in the summer of 2018. These data track an individual on a quarterly basis, collecting information their arrests, prison entries, and incarceration status (these data are not state-specific and include records from all contributing data providers).

2.2.5 Data Sharing

M-CARES surveys collect information on sensitive topics. Given the breadth of individual information reported in these surveys, study participants and Human Subjects Boards may be concerned about the

[7] In ACS2015, among children who reside in Michigan, are under 18 years old, whose mothers are 18-35 years old, and with family income less or equal to $80,000, 91 percent were born in Michigan. Among women 18-35 years old residing in Michigan with family income less or equal to $80,000, 77 percent were born in Michigan.

indirect identification of study participants (even when the data are de-identified). To maximize response rates and limit the disclosure risks to study participants, we, therefore, will limit the sharing of de-identified participant information with the M-CARES team and people who work for organizations that make sure M-CARES research is done safely and properly, such as the University of Michigan and government research offices.

Similarly, administrative data are protected by the private (including PPMI and credit reporting agencies) and government agencies that collect them. These agencies are also concerned about the indirect identification of individuals, and we will not be authorized to disclose individual outcomes from these records. Only aggregated output meeting the strict disclosure guidelines of the relevant agency will be released for publication. These disclosure standards are also appropriate for PPMI (HIPPA) data and we will hold published results from the PPMI data to federally mandated security and disclosure standards.

Our consent form outlines these security and confidentiality measures for study participants. We ask study participants to share their individual data only with the M-CARES team and people who work for organizations that make sure M-CARES research is done safely and properly, such as the University of Michigan and government research offices. The consent form also tells participants that we will publish the results of the study, but we tell them that we will not publish any information that would identify individual participants or family members.

3. Statistical Analyses and Methods

Our study is interested in how increasing the affordability to contraceptives affects outcomes. Let Y_{ij} be one outcome in domain, j, for individual, i. Examples of three domains of interest include contraceptive efficacy; unintended pregnancy and childbirth; economic self-sufficiency; financial security; and neighborhood quality. We normalize outcomes within a domain so that the sign of each outcome represents *more* of the outcome in the relevant domain.

3.1 Main Estimating Equations

We estimate both the reduced-form effects and the two-stage least squares estimates of receiving a voucher for contraceptives. The reduced-form equation is

$$Y_{ij} = \tau Voucher_i + \boldsymbol{X}_i'\boldsymbol{\beta_1} + \gamma_c + \varepsilon_{1i}, \quad (1)$$

where $Voucher_i$ is a binary variable equal to 1 if an individual i is selected to receive a voucher and 0 otherwise; $\boldsymbol{X}_i$, is a set of exogenous covariates (defined below) (see Appendix B for variable description); γ_c is a set of clinic fixed effects which will absorb between-clinic variation in physician recommendations, availability of appointments, and unobserved characteristics of patients; and ε_{1i} is the error term. The main coefficient of interest τ, often called the "intention-to-treat" (ITT) estimate, captures the average difference in means in an outcome between the treatment group (individuals randomly selected to receive a voucher) and the control group (individuals not selected to receive a voucher). The ITT estimate answers the policy question: what is the net, causal effect of reducing out-of-pocket costs for contraceptives (capped at the cost of a generic IUD) to zero for women seeking reproductive health care on outcomes?

Another relevant policy question is: what is the causal effect of increasing the efficacy of contraceptives that women use? The answer to this question differs from the ITT estimate for several reasons. First, not all women offered a voucher will alter the efficacy of their contraceptive methods. Second, even if a woman switches to a more effective method initially, she may not remain on a more effective method for the duration of our study.[8] In addition, some women who switch to more effective methods would have done

[8] One study found that around 7 percent of nulliparious women who selected IUDs had them removed (Brockmeyer, Kishen, & Webb, 2008).

so even without the voucher. Finally, women in our control group may switch to more effective contraceptives without receiving a voucher.

We, therefore, estimate the effect of increasing access to contraceptives within the following two-equation model, where the first-stage equation is

$$\text{(2)} \qquad Contraceptive\ Efficacy_i = \pi_1 Voucher_i + \mathbf{X}_i'\pi_2 + \gamma_c + \varepsilon_{2i},$$

and the second-stage equation is

$$\text{(3)} \qquad Y_i = \delta_1 Contraceptive\ Efficacy_i + \mathbf{X}_i'\delta_2 + \gamma_c + \varepsilon_{3i}.$$

Estimating this model using two-stage least squares (2SLS), the estimate of δ_1 is given by the ratio of the reduced form and first stage coefficients (τ / π_1).

***UPDATE 02/12/2020

To account for unforeseen changes in the study design, one analysis will restrict attention to women at PPMI for clinician visits—a population recruited from the beginning of the study. For this group, we expect the effects of *Voucher* before March 4 (when the voucher covered up to the cost of the Liletta, ~50% of name-brand IUDs) and on or after March 4 (when the voucher covered up to 100% of the cost of name-brand IUDs) to differ, because the full cost of having an IUD inserted was free only on or after March 4. Per section 7.2 of the year 1 pre-analysis plan, we will analyze this change in voucher amount by including an indicator variable in equations (1) and (2) for whether the individual was recruited on or after March 4 as well as the interaction of this indicator variable with *Voucher*. The indicator variable for *Post-March4* in equations (1) and (2) will capture any changes in the control group on or after March 4. The difference between the coefficients for *Voucher* and *Voucher*Post-March4* in equation (1) will illustrate if the change from a voucher amount at 50% of the cost of an IUD to 100% of the cost of an IUD changed outcomes. This interaction in equation (2) will allow the increase in the voucher amount to influence outcomes through its impact on contraceptive efficacy.

We also expect the effect of *Voucher* after November 4 to differ, because costs to the control group of having an IUD inserted increased as shown in Table 2C. In a second analysis, we will analyze women fee-scale 2 and above at the clinic for clinician visits after the November 4th change in voucher amount in equations (1) and (2) by including women recruited on or after March 4 and including an indicator variable for whether the individual was recruited on or after November 4, as well as the interaction of this indicator variable with *Voucher*. In equation (1), the indicator variable will capture any changes in the control group on or after November 4. The difference between the two coefficients for *Voucher* and *Voucher*Post-Nov4* in equation (1) will tell us whether the change in the value of the voucher changed choices directly and, in equation (2), whether the change in the value of the voucher changed choices through increases in contraceptive efficacy.

In addition, we will examine non-clinician visits began (roughly May 13) separately. Using the same framework as outlined above, we examine equation (1) and (2) but add an indicator variable for post-May13 recruitment, an indicator the non-clinician visit indicator, and the interaction between the two. In addition, we will test whether the interaction of the indicator variable for post-May 13 with *Voucher*, the interaction of the indicator for non-clinician visit with *Voucher*, and the triple interaction of on or after May 13, non-clinician visit, and *Voucher*. The idea is to examine the differential treatment effects for clinician visits separately from non-clinician visits to learn about heterogeneity in the treatment effect. This analysis also informs our understanding of the external validity of treatment effects by considering a different population of women.

****

In the absence of an experiment, we expect that contraceptive use and outcomes are correlated with unobservable factors, which render OLS estimates of δ_1 biased and inconsistent (we investigate this directly in section 7.4). Direct comparisons between women who use more effective methods and those who do not may capture a variety of differences between the groups. For instance, more career-interested women with higher expected wage growth and may desire fewer children and also be more likely to use more effective contraceptives. Therefore, comparing the wages of women using more effective contraceptives with those of women using other methods may conflate differences in women's career investments with the effect of more reliable contraceptives.). The advantage of randomizing $Voucher_i$ breaks this endogeneity and provides a valid instrument for $Contraceptive\ Efficacy_i$.

The causal interpretation of the 2SLS estimate turns on two main identifying assumptions: financial barriers (i.e., out-of-pocket costs) are both (1) relevant to women's decisions about which contraceptive method to use and that (2) voucher assignment is exogenous and excludable. This study is premised on assumption (1) that financial barriers matter, which is born out of a variety of studies regarding the determinants of heath care utilization (see Finkelstein et al. (2012) for an overview). Moreover, the randomization in the study ensures that the exogenous assignment of vouchers in (2) is met.

Excludability is more difficult to test and requires that receiving a voucher for contraceptives have an effect on outcomes *only* by increasing the efficacy of contraceptives. This assumption seems plausible as the voucher can only be used for contraceptives at PPMI. Moreover, women in both the treatment and control groups receive cash benefits for completing the screening and baseline surveys, implying that the effects of these cash benefits should be the same in the two groups.

One alternative channel could be that vouchers increase spending on other categories for women already intending to purchase an expensive contraceptive method on the day they enroll in M-CARES. For these women, the voucher would act as a cash transfer, allowing the women to spend money saved for contraceptives on something besides contraceptives in the short term (e.g., her credit card payment or rent). To examine the empirical importance of this channel, we ask women on the screening survey about the reason for their PPMI visit and which method they plan to get that day. We will also examine whether a woman's total debt (as measured on her credit reports) is reduced in the month she enrolls in the intervention for women who had already planned to get an expensive contraceptive. These analyses may suggest adjustments to the analysis if there is a quantitatively important violation of the exclusion restriction.

It is possible that receiving a voucher can have other effects on outcomes (e.g., a voucher can imbue a recipient with a positive or optimistic feeling), but it seems unlikely that this indirect effect would influence outcomes in multiple domains over the many years in the study.

Under these assumptions, we interpret the 2SLS estimate as the local average treatment effect, or LATE (Imbens & Angrist, 1994). The 2SLS estimate, δ_1, identifies the causal effect of contraceptive efficacy among the women who shift the efficacy of their contraceptives after receiving a voucher and who would not have shifted their contraceptive efficacy without the voucher.

3.2 Covariates

Specifying exact definitions of covariates is difficult, because we do not know the distributions of these variables before we collect the data and, therefore, cannot know *ex ante* how best to define them. When we see the distributions of these variables, we may alter categorical, dummy, or continuous variables due to a paucity of individuals in certain categories or skewness if doing so is defensible and enhances what we learn from this study. Without this knowledge, this section outlines our best guess at what these covariates should be. Their inclusion is intended to increase precision by accounting for differences in characteristics between the treatment and control groups within the same clinic location that occur by chance. We describe how we plan to construct all variables in Appendix B.

Covariates, $\boldsymbol{X_i}$ in equations (1) through (3), will consist of two sets: $\boldsymbol{Z_i}$, which includes demographic characteristics available from the pre-randomization screening survey and administrative data; and $\boldsymbol{C_i}$, which includes binary variables for pre-randomization use of different types of contraceptives. Because these are pre-randomization outcomes, the probability of treatment is expected to be uncorrelated with these covariates. We will, however, examine the sensitivity of our results to the inclusion of both subsets of variables. $\boldsymbol{Z_i}$, covariates include

- Age group dummy variables (age groups: 18-19, 20-22, 23-25, 26-29, 30-35)
- Race group dummies (race groups: White, African-American, other or mixed race)
- Fee-scale dummies (four groups previously defined)
- Education group dummies (education groups: less than high school, high school degree, some college, college degree or more)
- Number of previous births dummies (groups: 0, 1, 2, 3+ previous births)
- Married or cohabit with partner
- Time elapsed between the intervention and outcome measurement (linear measure and squared term).[9]

$\boldsymbol{C_i}$, covariates include

- Use of no contraceptive in screening survey
- Use of birth control pills in screening survey
- Use of LARC in screening survey
- Use of condoms or withdrawal in screening survey
- Use of any other method in the screening survey.

3.3 Heterogeneity and Quantile Analyses

The ITT and LATE analyses implicitly assume that the treatment effect does not vary with individual's pre-randomization characteristics. We will explore treatment heterogeneity across pre-treatment characteristics by interacting $Voucher_i$ in either equation (1) or (2) with either binary or continuous covariates. This interaction specification is more powerful that stratifying by group, which seems appropriate given our smaller sample sizes. For some outcomes, we will also estimate quantile treatment effects, because we are interested in the impact of contraceptive efficacy on certain parts of the outcome distribution. These quantile treatment effects describe the treatment-induced differences in the treated and control distributions (rather than the treatment effects for any individual), which is relevant for understanding how increasing contraceptive efficacy may affect the distribution of individuals with large amounts of debt or lower incomes. We define these analyses in more detail in the descriptions of specific analyses below.

3.4 Multiple Outcomes

Many aspects of women's lives may be affected with financial access to more effective contraceptives. Following the analyses of Kling, Liebman, and Katz (2007), we plan to study many potential outcomes in eight domains: (1) contraceptive efficacy, (2) unintended pregnancy and childbearing, (3) economic self-sufficiency, (4) financial security, (5) neighborhood quality, (6) physical health and ability, (7) mental health and well-being, and (8) relationship quality. For the year 1 analyses, we will examine outcomes in the following five domains as standardized indices:[10]

[9] Timing is important because some women will enroll in the study toward the end of the 12-months of recruitment, whereas others will enroll at the beginning. Therefore, women's time in the study may vary by up to one year, which may explain some of the variation in measured outcomes. In balance tables, we will include instead the day that an individual signed up.

[10] Following Kling et al. (2007), we standardize the ITT effects, τ, or LATE, δ_1 , for each outcome from equations (1) and (3) using the mean and standard deviation for the relevant outcome in the control group. In addition, we sign the outcomes within each domain so that increases indicate movement in a common direction. For instance, within the domain of neighborhood quality, an increase in the index would always imply an improvement in that domain.

(1). Contraceptive efficacy: total contraceptive expenditures at PPMI, any contraceptive use, temporal coverage of contraceptive method (e.g., 1 day, 3 months, 6 months, 12 months), use of a high efficacy method, a continuous measure of method efficacy.

(2). Unintended pregnancy and childbearing: number of pregnancies, number of pregnancy tests, number of abortions, number of births, and frequency of use of emergency contraception.

(3). Economic self-sufficiency: wage earnings; enrollment in college or education; employed (any labor income); not incarcerated; income-to-poverty ratio; not single-head of household; not living with parent; reverse coded (-1*) receipt of public dollars (from programs such as TANF, SNAP, Medicaid, EITC, and other programs).

(4). Financial security: reverse coded (-1*) bills sent to collection; reverse coded (-1*) amount owed to collections; reverse coded (-1*) any delinquency where a payment is at least 30 days overdue; reverse coded (-1*) any delinquency where a payment is at least 120 days overdue; total debt.

(5). Neighborhood quality: income-to-poverty ratio in census tract of residence; reverse coded (-1*) teen pregnancy in tract; reverse coded (-1*) share of single-headship in census tract; reverse coded (-1*) share of households receiving public assistance in tract; reverse coded (-1*) share of poor children in tract; share of home ownership in census tract of residence; median house price in census tract of residence; and reverse coded (-1)*number of address changes since enrollment.

Grouping outcomes allows us to examine tests whether there is an *overall* effect of contraceptive efficacy within a single domain, and it also limits the number of statistical tests. However, a shortcoming of this approach is that indices implicitly weight index components equally. Large changes in any one dimension might be averaged together with potentially much smaller or zero effects in other dimensions, implying that the index masks large effects. For this reason, we also report individual outcomes separately in the main analysis, so that individual outcomes of interest can be examined separately. For individual outcomes, we will report multiple-inference adjusted p-values within a given domain based on resampling method of Westfall and Young (1993) with 10,000 iterations.

3.5 Attrition and non-response, price indices, standard error adjustments, and weighting

We expect some attrition in self-reported data. Earlier RCTs involving contraceptive use found attrition of 9 to 13 percent over the course of 12 months (Burke et al., 2010; Harper et al., 2015; Raine et al., 2011). There should be no direct attrition in administrative data—only a failure to link an individual due to an invalid social security number or other incorrect information.

We will investigate whether individual characteristics and voucher receipt are correlated with representation in each sample. We will implement this by estimating a linear regression model using as the dependent variable a variable equal to 1 if the woman was responded to a survey or was linked to an administrative data source. To the extent that attrition or survey non-response is non-random, we will take steps to address issues of non-representativeness using standard propensity-reweighting procedures (DiNardo, Fortin, & Lemieux, 1996; Heckman, Ichimura, Smith, & Todd, 1998). In addition, once we obtain data on selection into study participation (see section 4.1 below), we will examine whether additional reweighting is required so that the characteristics of our participants appear more similar to the national population. These comparisons are presented in Appendix C Table C2.

We will express all monetary variables to common dollars using the Consumer Price Index for Urban Wage Earners and Clerical Workers (CPI-W).

We will estimate heteroskedasticity-robust Huber-White standard errors for all analyses (Huber, 1967; White, 1980). In models that consider the evolution of outcomes across time, we will additionally cluster standard errors at the level of an individual participant.

4. Selection into M-CARES and the First Stage[11]

This section lays out our plan to examine the characteristics of the PPMI population, the PPMI sub-group that takes our screening survey, the PPMI sub-group that elects to participate in our study, and the characteristics of the first stage. This section outlines each of these analyses.

4.1 External validity and enrollment reports

PPMI has agreed to provide M-CARES with aggregated information on patients who come to the recruitment clinics during the recruitment period as well as detailed, individual information for women who elect to participate in the M-CARE study. To understand the external validity of our study, we examine how M-CARES participants compare to (1) the PPMI population, (2) the PPMI sub-group who agrees to take the screening survey, (3) two national samples of women: the NSFG 2013-15 respondents and female respondents from ACS 2015, and (4) the same ACS sample for the state of Michigan. The variables examined across all samples include those outlined in section 3.2 and add, for M-CARES, PPMI patients, and the NSFG samples, variables capturing the use of contraceptives, including birth control pills, LARCS, condoms, withdrawal, no method, and other methods (see Appendix B for description of the variables). These comparisons are presented in Appendix C Table C1.

4.2 Verification of randomization (balance tests)

As described previously, assignment of a voucher is randomly assigned by a computer and given to individual respondents. This prevents surveyors from undoing the randomization and respondents from trading or giving away vouchers. It is possible, however, that the balance of characteristics fails by chance. As is standard in experimental research, we examine randomization by comparing treatment and control differences in pre-randomization characteristics. To do this, we regress a binary dependent variable equal to 1 if an individual was assigned a voucher on the covariates,

$$Voucher_i = \boldsymbol{X}_i'\boldsymbol{\rho} + \varepsilon_{4i}, \quad (4)$$

where $\boldsymbol{X}_i$ is described in section 3.2. Additionally, we use Huber-White standard errors to account for the fact that errors of a linear probability model are heteroskedastic (Huber, 1967; White, 1980). We will also test whether clinic fixed effects explain voucher assignment. If so, we will examine the partial, joint correlation of the $\boldsymbol{X}$ net of these clinic-level differences.

A summary test of covariate balance is a heteroscedasticity-robust Wald test of joint significance of the covariates. Under the null hypothesis that vouchers were randomly assigned, the covariates should not jointly predict the assignment of a voucher and the p-value of the test statistic should be outside the range of conventional statistical significance. The main advantage of this test over more common balance tests for means is that it accounts for the correlations among covariates and the *joint relationship* of the group of covariates with the likelihood of receiving a voucher, aggregating all information in the relevant covariates into a single test statistic. As such, this is a more powerful test than more standard t-tests of differences in the means of individual characteristics. Because it may be easier to reject the null hypothesis of random assignment (and this should only occur by chance), it is important to consider the scientific significance of any resulting imbalance in characteristics (McCloskey, 2005).

A feature of our regression-based test is that it also describes the potential scientific importance of non-representativeness: the magnitudes of the regression coefficients conveniently quantify which characteristics are more or less likely to result in a linked observation after controlling for other record characteristics. Appendix C Table C3A presents an example of this analysis which we will do for each of our linked (survey and administrative) data sources described in section 2.2. If we use samples that combine

[11] Due to updates in the study design, the Tables in Appendix C will be modified accordingly.

administrative data creating new samples, we will also examine whether covariates jointly predict receiving a voucher and being in this sample.

As a complement to these findings, we will also present standard mean comparisons for the same set of covariates. An example of this table appears in Appendix C Table C3B. It includes the means for the control group, and the mean differences in characteristics, or ρ from (4), the standard errors of the estimated coefficient, and the inverse propensity-score reweighted estimates if applicable. We will make control/treatment comparisons for each of our linked (survey and administrative) data sources described in section 2.2 and, if possible, augmenting these tables with lagged outcome variables of interest.

4.2.1 First Stage

Our first stage analysis will examine the relationship between receiving a voucher and contraceptive efficacy represented in equation (2). The simplest framework for thinking about how a voucher might matter comes from Michael and Willis (1976). In their model, the number of children is a random variable, and couples choose a contraceptive strategy to reduce the monthly probability of conception. In addition, contraceptive strategies have a price. Couples optimize by choosing an ex ante distribution of pregnancies with the mean, μ^*, to maximize utility net of the costs of the contraceptive strategy, max $U(\mu)$-$C(\mu)$.

The framework of Michael and Willis provides testable predictions regarding how different contraceptive methods might alter the number of pregnancies. Michael and Willis consider a simple division of costs of attaining a pregnancy distribution, μ, using contraceptive strategy, *j*, into a fixed cost, α_j, and a marginal cost, β_j. These fixed and marginal costs might represent monetary costs, time costs, or psychic costs and may be due to the use of methods *or* adoption of different behaviors. The cost of using strategy *j* to attain an *ex ante* pregnancy distribution, μ, is given by $c_j = \alpha_j + \beta_j (\mu_N - \mu)$, where μ_N indexes the expected distribution of pregnancies in the absence of any contraceptive strategy (including abstinence). The term $\mu_N - \mu$ is, therefore, the expected number of pregnancies averted. The (constant) marginal cost of averting a pregnancy, β_j, might be a behavioral cost (abstinence or withdrawal), the inconvenience or discomfort of birth control use (barrier methods), or the necessity of purchasing supplies (as with condoms or the birth control pill). Fixed costs include the up-front, out-of-pocket costs of getting a new method; the price and time costs of going to the doctor; and potentially the psychic and time costs of learning about the method and adopting different behaviors (e.g., scheduling an appointment every three months to get a Depo shot). The implications of this model for our analysis is that the fixed costs of contraceptives reflect a range of things—not just the out-of-pocket costs for selecting the method.

Importantly for our analysis, high efficacy methods (or methods with low marginal costs of preventing pregnancies) tend to have high fixed costs. For instance, the up-front out-of-pocked costs of getting an IUD at PPMI without the sliding scale is $492, including insertion and a pregnancy test. The IUD, however, has a low marginal cost of preventing pregnancy, as no action or transaction with a partner is required at the time of intimacy.

The total costs of a contraceptive strategy for preventing the desired number of pregnancies determines contraceptive use, the endogenous variable in equation (3). The random assignment of vouchers is intended to reduce the fixed cost, α_j, of highly effective (and low marginal cost) contraceptive methods. Because the fixed costs (especially in some cases the knowledge about methods and the psychic discomfort of getting some of them) will not fall to zero, we do not expect every treated participant to adopt more effective methods.

The main conceptual issues for our study relate to how one might measure the efficacy of a contraceptive strategy. This requires both a time dimension as well as a definition of method efficiency. There is no clear theoretical guidance at the outset. However, to the extent we measure contraceptive efficacy poorly, it provides misleading inferences. We will, therefore, explore alternative measures of contraceptive use so that engaged readers may rescale the reduced-form estimates with their preferred metric.

Our baseline estimate is whether a woman used a high or moderate efficacy (tier 1 or tier 2) method of contraception within the first 100 days of enrollment. In addition, we examine whether she ever used a high or moderate efficacy method of contraception at the end of the first year after enrollment. We will also consider as a dependent variable any contraceptive use, high efficacy method (tier 1), the temporal coverage of contraceptive method (this could be 1 day or up to 3 years), whether the woman switched contraceptive method within 100 days, and a continuous measure of method efficacy. Finally, we will present a standardized index of contraceptive efficacy as defined in section 3.4. Appendix C Table C4 presents an example table.

5. Primary Analyses: Contraceptive Efficacy, Pregnancy, and Childbirth

In the first year, we will measure outcomes to understand how vouchers for contraception affect women's choices about contraceptive method, pregnancy, and childbearing outcomes. This section lays out our primary analyses for the unintended pregnancy and childbirth domain of outcomes.

Appendix Table C Table C5 lays out a table illustrating how we will present our results. The most left hand column describes the outcome of analysis. The control mean is reported in column 1. Columns 2 and 3 report the reduced-form estimate (equation 1) and the 2SLS estimate (equation 3). For individual outcomes we report both the unadjusted and the multiple-test corrected p-value within the domain. Given the complementarity of survey and administrative data, we report outcomes separately for these data sources when measured in both. In the first row, we report the results for the index for the entire domain.

5.1.1 Hypotheses

Vouchers reduce the fixed, out-of-pocket costs for using highly effective contraceptives. We, therefore, hypothesize that:

(H1) Vouchers will increase the utilization of these methods.
(H2) Vouchers will reduce the costs of effective contraceptives and should reduce the number of pregnancies and abortions in the first year.
(H3) Vouchers will reduce the number of children born in the first year, but by less than they reduce pregnancies (because many unintended pregnancies are aborted). The longer term effects are more ambiguous.

Note that H2 and H3 are *not* clearly anticipated by theory. If abortions completely offset unintended pregnancies, then the number of pregnancies and abortions could fall while the number of births remain unchanged. In addition, theoretical work suggests that more reliable contraception could increase *intended* pregnancies and childbearing in the short-term (Michael & Willis, 1976). Our hypotheses about pregnancy, abortion, and childbearing, however, reflect the fact that empirical work tends to show that pregnancies, abortions, and childbearing fall in the short-term.

5.1.2 Contraceptive Efficacy

Our first analysis considers how vouchers effect contraceptive efficacy, where higher values indicate lower failure rate methods. These estimates are the first stage analyses and are reported in Appendix C Table C4. In year 1, measures of contraceptive efficacy will be examined using screening data, baseline survey data, and PPMI administrative data measured up to roughly 1 year after enrollment.

5.1.3 Pregnancy, Abortion, and Childbirth

We then examine the impact of contraceptive efficacy on the outcome domain, "Unintended pregnancy and childbearing." The variables used as part of this domain are number of pregnancies, number of births, number of abortions, and frequency of use of emergency contraception. In year 1, measures of unintended pregnancy and childbearing will be examined as close to one year after enrollment as possible in the PPMI data and at exactly one year in other administrative sources. These estimates are reported in Appendix C Table C5.

6. Supplementary Analyses: Economic Self-Sufficiency, Financial Security, and Neighborhood Quality

We will examine measures of economic self-sufficiency, financial security, and neighborhood quality after the first year. This section lays out these supplementary analyses and uses the same estimating strategy and table format as indicated in section 5.1.3.

6.1.1 Supplementary Hypotheses

If vouchers reduce the fixed, out-of-pocket costs for using highly effective contraceptives, and increased use of highly effective contraceptives reduces pregnancies and unintended childbearing, we expect positive *direct* effects in the domains of economic self-sufficiency, financial security, and neighborhood quality. By direct effects, we mean benefits that accrue directly to women by avoiding an unplanned birth (e.g., more time spent working or in school, lower likelihood of financial hardship, and potentially more control of one's living situation).

We also anticipate *indirect* effects to occur through shifts in women's expectations about future childbearing. For instance, if women cannot afford to use (or continue to use) highly reliable contraceptives, many will rationally expect to experience an unintended pregnancy in the near future. This expectation—*even in the absence of an actual unintended pregnancy or birth*—may lower women's *current* career investments. If a voucher allows women to switch to highly reliable contraceptives such as an implant or an IUD, they should rationally expect to have no unintended pregnancies in the near future. Thus, vouchers should raise *current* career investments, by raising the expected returns to these investments.

These expected effects lead us to the following supplementary hypotheses:

(H4) Vouchers will increase economic self-sufficiency. They will
   a. increase career investment, including educational enrollment and completion;
   b. increase wage earnings, through increases in employment, weeks and hours worked, and returns to investments in careers (positive effects may be offset by reductions in work intensity to invest one's career, H4a);
   c. reduce the likelihood of being a single head of household or living with one's parent;
   d. reduce crime, arrests, and incarceration;
   e. reduce receipt of public dollars for programs such as TANF, SNAP, Medicaid, EITC, as well as other programs.

(H5) Vouchers will increase financial security. They will
   a. reduce the likelihood of having any bills sent to collection
   b. reduce the amount owed to collections;
   c. reduce any delinquency where a payment is at least 30 days overdue
   d. reduce any delinquency where a payment is at least 120 days overdue.

(H6) Vouchers will increase neighborhood quality by increasing women's economic self-sufficiency and financial security, making it possible for them to move to areas with better average socio-economic characteristics of the tract of residence.

Because the effects of effective contraception are cumulative, we do not anticipate large measured effects on these outcomes in year 1. Any effects on self-sufficiency, financial security, and neighborhood quality will likely be small and may not statistically significant in the first year.

6.1.2 Economic Self-Sufficiency, Financial Security, and Neighborhood Quality

As described in section 5.1.3, we will examine how vouchers affect outcomes in each domain using both a reduced-form and 2SLS analysis. Self-sufficiency outcomes and financial security will be examined for the first tax year after enrollment. Example tables are presented as Appendix C Table C6 and C7, respectively. Neighborhood quality will be examined for the residence reported in the tax data as of April 2020, and an example table is presented as Appendix C Table C8.

7. Supplementary Analyses

7.1.1 Plans to Use Vouchers

In order for vouchers to affect a participant's behavior in a meaningful way, they need to be used. Use is documented in the first stage. If implementation is successful, we expect that receiving a voucher is strongly and positively correlated with total expenditures at PPMI for contraceptives (including any voucher dollars paid to PPMI) and negatively correlated with measures of out of-pocket payments at PPMI. However, there are many reasons why vouchers may not have been used and we examine those reasons using responses to the baseline survey. We present basic summary statistics for these main variables in Appendix C Table C10. We will supplement this analysis by using linear regressions to examine the partial correlations of covariates in ***X*** model with these outcomes. The results will help us understand the reasons for voucher use and non-use.

7.1.2 Sensitivity Analyses

The inclusion of covariates should not affect the results beyond contributing to precision. We will analyze this systematically by recreating Appendix C Table C4-8 without covariates and with only the **Z** covariates. We will also examine their robustness to excluding clinic fixed effects.

7.1.3 Heterogeneity by Baseline Characteristics

Understanding how the impact of vouchers varies across subgroups could allow subsidies for contraception to be targeted more effectively. In some cases, we do not have strong hypotheses about which groups may benefit more. For instance, it is unclear whether the vouchers may have larger or smaller effects on outcomes of older or younger women, native-born or foreign-born women, women of different racial groups, women with different levels of education or income, or women with different numbers of pre-randomization births.

But in other cases, we have stronger prior beliefs about how the vouchers could matter. For example, because women already using highly reliable contraceptive methods before the intervention had, by revealed preference, shown their ability to find the resources to pay for these methods, the voucher may not matter as much as it would for women using less effective (and with lower up-front costs) methods. Similarly, vouchers may have been most beneficial for women who indicated in the screening survey that they had delayed reproductive health care in the last year because they could not afford it or for whom reliable transportation to get contraception had been a barrier. (This could be the case if a voucher allows women to get a high out-of-pocket cost method like an implant or IUD rather than a method like Depo-Provera, which requires more frequent returns to PPMI.)

We will explore heterogeneity in the effects of voucher across many covariates defined in section 3.2 we well as in responses on the baseline survey. We will obtain these estimates by interacting $Voucher_i$ with a binary variable for a particular category; a set of dummies for a category (for pay scale and delay in birth control use); or a continuous variable. We will also present the number of non-missing observations for each variable (column 1), the first stage (column 2), as well as the four main second-stage outcome indices (columns 3-6). An example table is presented as Appendix C Table C11.

Appendix C Table C12 presents heterogeneity estimates for attitudes about contraception (as well as a contraceptive attitudes index that standardizes across all of the individual outcomes within the domain). In general, we hypothesize that women with more negative attitudes about contraception or more religious women to be less responsive to the intervention. We have no overall hypothesis about the impact of relationship quality, but our hypotheses about one component are informed by the previous literature. Intimate partner violence is correlated with contraceptive coercion and unintended pregnancy (Miller et al., 2010), we hypothesize that voucher receipt and LARC use may have especially large impacts for women in coercive relationships.

7.2 Heterogeneity by Policy Period

The policy environment surrounding federal and state funding for reproductive health may change during our recruitment period. It is difficult to specify in advance how we might model this change, given that we do not know what the relevant parameters are. We can, however, indicate that if this happens, we will examine whether the ITT effects and LATE we estimate vary across the relevant policy periods. For instance, if Title X funding for Planned Parenthood is eliminated effective July 1, 2018, and PPMI will no longer be able to make contraceptives available according to the fee-scale in Table 2, we will estimate an equation that interacts $Voucher_i$ with a dummy variable equal to 1 for the date the new funding model became effective. We will also account for heterogeneity in the population recruited by interacting this group with *Voucher* and a binary variable *PostPolicy* (=1 for individuals after the policy change) as well as the interaction of *PostPolicy* with *Voucher*.

7.3 Extensive versus intensive margins

The focus of the primary analyses on continuous measures of pregnancy and childbearing do not allow us to understand differences in responses on the extensive and intensive margins (e.g. reduction in the number of unintended pregnancies vs. reduction in the likelihood of having an unintended pregnancy). This is an important difference to characterize because the impact of the *first* unintended birth may be substantially more damaging in terms of career trajectory and reduced future expectations, compared to the second unintended birth. In addition, there may be less measurement error on the extensive margin compared to the extensive margin (e.g. any emergency contraceptive use compared to the frequency of emergency contraceptive use). Therefore, we may have power to detect a response on the extensive margin, even if we do not have power to detect a response on the intensive margin. We will use PPMI patient records to conduct a more extensive analysis of measures of (1) any unintended pregnancy: any pregnancy test, any abortion, any emergency contraception use;[12] (2) measures of the intensive margin effects on unintended pregnancy: number of pregnancy tests if >0, number of abortions if >0, and frequency of use of emergency contraception if >0, all defined only for participants with positive values of the dependent variable. Appendix C Table C13 presents an example table.

7.4 Observational Estimates of Contraceptive Efficacy

We will compare the results using our randomly assigned voucher to results estimating equation (3) by OLS. Appendix C Table C14 presents an example table.

7.5 Moral hazard

One potential unintended consequence of increasing access to high-efficacy contraceptives is that it may increase the incidence of STDs. For example, an IUD reduces the expected costs of having unprotected sex and could, therefore, reduce the cost of having sex without condoms. We will examine the extent of moral hazard as a result of the intervention by studying whether voucher receipt is correlated with measures of STD incidence, using PPMI patient records. PPMI records allow us to measure incidence of HIV,

[12] As we are analyze outcomes in year 1, we do not expect a difference between the number of births and any births during this short time period, therefore we do not examine the extensive margin for childbearing in year 1. However, we will in subsequent analyses for years 2-5.

gonorrhea, chlamydia, syphilis, herpes, and HPV. As with measures of pregnancy constructed from PPMI data, if patients get tested at health centers or facilities other than Planned Parenthood, we will not be able to capture these instances in our data. Therefore, our results may understate the impact of voucher receipt on STD/STI incidence. Appendix C Table C15 presents an example table.

8. Analyses Planned for Later Years

Once we have concrete estimates for our first stage and primary outcomes in section 5, we will be able to develop the best plan for how to design subsequent surveys (whether any tweaks to our baseline survey is needed) and whether these analysis present questions requiring the addition of new variables. Our intention is to register this second pre-analysis plan by July 1, 2020, which is early enough so as to eliminate the possibility of examining any 2 year outcomes for any participants before the plan is posted.[13]

9. Contribution and External Validity

M-CARES is the first experimental study to quantify the effects of increasing the affordability of contraceptives on a wide range of outcomes for women and their families. M-CARES will recruit 5,000 women ages 18 to 35 years old who have no insurance and are above the federal poverty line—all of whom are seeking care at PPMI. After the first year of the study, a rich combination of survey data and administrative data allows us to estimate the short-run causal effects of subsidizing highly effective contraceptive methods for U.S. women.

While providing credible internally valid estimates, important limitations to M-CARES external validity require care in interpreting its results. First, M-CARES results are specific to the population it studies: women who are actively seeking reproductive care and go to Planned Parenthood clinics. This population of women potentially excludes (1) insured women who seek reproductive care at places other than Planned Parenthood, (2) women who have little interest in using contraceptives and, therefore, do not seek care, and (3) uninsured women who may be dissuaded from seeking better contraceptive methods by the expectation of high out of pocket costs.

The exclusion of the first two groups does not bear critically on the central research question of interest: how can increasing the affordability of reliable contraception increase use? This is because the decisions of the first and second groups are not affected by the cost of reliable contraception. However, understanding the population-level effects of any policy similar to M-CARES vouchers should adjust for the relevant population accordingly.

In contrast, we expect that providing vouchers as a national policy could have important effects on contraceptive use of low-income women in group (3). This implies that increasing the affordability or contraceptives could affect more people than respond in this study, however the marginal effects of the policy for this group are not recovered by this study. These marginal effects will be informed by our heterogeneity analyses.

Our results are also specific to reproductive care and economic setting of in Michigan. As we show in Table 1, Michigan is fairly representative of U.S. women of childbearing age in terms of educational attainment, wage income, number of children, and marriage rates. However, Michigan also tends to have higher rates of unplanned pregnancy. 44 percent of pregnancies in Michigan were unplanned in 2010, which is one of the highest numbers in the nation outside of the South (Sonfield & Kost, 2015).[14] Michigan's public support for family planning is also less generous than *many* other states.[15] However, Planned Parenthood's sliding

[13] Note that this is because administrative databases are updated with lags, so 2019 education, tax, vital statistics, and program participation data should not be available until after the calendar year of 2019 ends.

[14] Notably, all but three states (Mississippi, Louisiana, and Georgia) have rates of under 50 percent (Sonfield & Kost, 2015).

[15] Michigan is less generous than 31 other states and more generous than nine other states. In some states, Medicaid or other state programs fund family planning services for women between 100% of the poverty line and higher amounts.

scale is in all other U.S. states, making services available on the sliding scale (and typically free to all women below 100 percent of the federal poverty line, although this varies by state).

The extent to which we can generalize M-CARES results to the U.S. population of women of reproductive age depends on the sizes of these groups relative to the population represented by the study participants. Our analysis will consider these comparisons carefully, as well as study heterogeneity in treatment effects that will inform out-of-sample extrapolations.

10. References

Ananat, E. O., & Hungerman, D. (2012). The Power of the Pill for the Next Generation: Oral Contraception's Effects on Fertility, Abortion, and Maternal and Child Characteristics. *Review of Economics and Statistics, 94*(1), 37-51.

Ashenfelter, O., & Plant, M. W. (1990). Nonparametric Estimates of the Labor-Supply Effects of Negative Income Tax Programs. *Journal of Labor Economics, 8*(1), S396-S415.

Avery, R., Calem;, P., & Canner, G. (2003). An Overview of Consumer Data and Credit Reporting. *Federal Reserve Bulletin, February*.

Bailey, M. J. (2006). More Power to the Pill: The Impact of Contraceptive Freedom on Women's Lifecycle Labor Supply. *Quarterly Journal of Economics, 121*(1), 289-320.

Bailey, M. J. (2013). Fifty Years of U.S. Family Planning: New Evidence on the Long-Run Effects of Increasing Access to Contraception. *Brookings Papers on Economic Activity, Spring*, 341-409.

Bailey, M. J., Hershbein, B. J., & Miller, A. R. (2012). The Opt-In Revolution? Contraception and the Gender Gap in Wages. *American Economic Journal: Applied Economics, 4*(3), 225-254.

Bailey, M. J., Malkova, O., & McLaren, Z. (2016). *Do Family Planning Programs Increase Children's Opportunities? Evidence from the War on Poverty and the Early Years of Title X*. University of Michigan Working Paper. Retrieved from http://www-personal.umich.edu/~baileymj/Bailey_Malkova_McLaren.pdf

Bound, J., Brown, C., & Mathiowetz, N. (Eds.). (2001). *Measurement Error in Survey Data* (Vol. 5). Amsterdam: Elsevier.

Brockmeyer, A., Kishen, M., & Webb, A. (2008). Experience of IUD/IUS insertions and clinical performance in nulliparous women–a pilot study. *The European Journal of Contraception & Reproductive Health Care, 13*(3), 248-254. doi:10.1080/02699200802253706

Burke, A. E., Barnhart, K., Jensen, J. T., Creinin, M. D., Walsh, T. L., Wan, L. S., . . . Wu, H. (2010). A randomized trial of the contraceptive efficacy, acceptability, and safety of C31G and nonoxynol-9 spermicidal gels. *Obstetrics and gynecology, 116*(6), 1265-1273.

DiNardo, J., Fortin, N. M., & Lemieux, T. (1996). Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. *Econometrica, 64*(5), 1001-1044.

Dynarski, S., Hemelt, S. W., & Hyman, J. M. (2013). *The Missing Manual: Using National Student Clearinghouse Data to Track Postsecondary Outcomes*. Retrieved from http://www-personal.umich.edu/~jmhyman/dynarski_hemelt_hyman_missing_manual.pdf

Finer, L. B., & Zolna, M. R. (2016). Declines in unintended pregnancy in the United States, 2008–2011. *New England Journal of Medicine, 374*(9), 843-852.

Finkelstein, A., Taubman, S., Wright, B., Bernstein, M., Gruber, J., Newhouse, J. P., . . . Baicker, K. (2012). The Oregon Health Insurance Experiment: Evidence from the First Year. *The Quarterly Journal of Economics, 127*(3), 1057-1106. doi:10.1093/qje/qjs020

Frost, J. J., Frohwirth, L., & Zolna, M. R. (2016). *Contraceptive Needs and Services, 2014 Update*. Retrieved from https://www.guttmacher.org/sites/default/files/report_pdf/contraceptive-needs-and-services-2014_1.pdf

Frost, J. J., Sonnfield, A., Zolna, M. R., & Finer, L. B. (2014). Return on investment: a fuller assessment of the benefits and cost savings of the US publicly funded family planning program. *Milbank Quarterly*.

Goldin, C., & Katz, L. F. (2002). The Power of the Pill: Oral Contraceptives and Women's Career and Marriage Decisions. *Journal of Political Economy, 110*(4), 730-770.

Harper, C. C., Rocca, C. H., Thompson, K. M., Morfesis, J., Goodman, S., Darney, P. D., . . . Speidel, J. J. (2015). Reductions in Pregnancy Rates in the USA with Long-Acting Reversible Contraception: a Cluster Randomised Trial. *The Lancet, 386*(9993), 562-568.

Heckman, J. J., Ichimura, H., Smith, J., & Todd, P. (1998). Characterizing Selection Bias Using Experimental Data. *Econometrica, 66*(5), 1017-1098.

Hock, H. (2008). *The Pill and the College Attainment of American Women and Men*. Florida State University Department of Economics Working Paper no. 2007-10-1. Retrieved from ftp://econpapers.fsu.edu/RePEc/fsu/wpaper/wp2007_10_01.pdf

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability., 1*, 221–233.

Imbens, G. W., & Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica, 62*(2), 467-475. doi:10.2307/2951620

Kling, J. R., Liebman, J. B., & Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica, 75*(1), 83-119.

Lesthaeghe, R. J., & Neidert, L. (2006). The Second Demographic Transition in the United States: Exception or Textbook Example? . *Population and Development Review, 32*(4), 669-698.

McCloskey, D. (2005). The Trouble with Mathematics and Statistics in Economics. *History of Economic Ideas XIII, 3*, 85-102.

Michael, R. T., & Willis, R. J. (1976). Contraception and Fertility: Household Production under Uncertainty. In *Demographic Behavior of the Household* (pp. 25-98). Cambridge, MA: National Bureau of Economic Research.

Miller, E., Decker, M. R., McCauley, H. L., Tancredi, D. J., Levenson, R. R., Waldman, J., . . . Silverman, J. G. (2010). Pregnancy coercion, intimate partner violence and unintended pregnancy. *Contraception, 81*(4), 316-322.

Raine, T. R., Foster-Rosales, A., Upadhyay, U. D., Boyer, C. B., Brown, B. A., Sokoloff, A., & Harper, C. C. (2011). One-year contraceptive continuation and pregnancy in adolescent girls and women initiating hormonal contraceptives. *Obstetrics and gynecology, 117*(2 Pt 1), 363.

Republican Party. (2016). *Republican Platform 2016*. Republican National Committee. Retrieved from https://prod-static-ngop-pbl.s3.amazonaws.com/media/documents/DRAFT_12_FINAL%5B1%5D-ben_1468872234.pdf

Schoeni, R. F., Stafford, F., McGonagle, K. A., & Andreski, P. (2013). Response Rates in National Panel Surveys. *The Annals of the American Academy of Political and Social Science, 645*(1), 60-87. doi:10.1177/0002716212456363

Sedgh, G., Singh, S., & Hussain, R. (2014). Intended and unintended pregnancies worldwide in 2012 and recent trends. *Studies in Family Planning, 45*(3), 301-314.

Sonfield, A., & Benson Gold, R. (2012). *Public Funding for Family Planning, Sterilization and Abortion Services, FY 1980–2010*. Retrieved from https://www.guttmacher.org/sites/default/files/report_pdf/public-funding-fp-2010.pdf

Sonfield, A., Hasstedt, K., & Gold, R. B. (2014). *Moving Forward: Family Planning in the Era of Health Reform*. Guttmacher Institute. New York. Retrieved from https://www.guttmacher.org/report/moving-forward-family-planning-era-health-reform

Sonfield, A., & Kost, K. (2015). *Public Costs from Unintended Pregnancies and the Role of Public Insurance Programs in Paying for Pregnancy-Related Care National and State Estimates for 2010*. Guttmacher Institute. New York. Retrieved from https://www.guttmacher.org/sites/default/files/report_pdf/public-costs-of-up-2010.pdf

Sonnfield, A., & Kost, K. (2015). *Public Costs from Unintended Pregnancies and the Role of Public Insurance Programs in Paying for Pregnancy-Related Care: National and State Estimates for*

2010. Guttmacher. New York. Retrieved from http://www.guttmacher.org/pubs/public-costs-of-UP-2010.pdf

Trussell, J. (2011). Contraceptive failure in the United States. *Contraception, 83*(5), 397-404.

Westfall, P. H., & Young, S. S. (1993). *Resampling-Based Multiple Testing: Examples and Methods for P-Value Adjustment*. New York: Wiley.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica, 48*, 817–830.

Appendix A. Data Sources

A central contribution of our analysis is using both survey data and administrative data to provide a rich perspective on the evolution of respondents' lives. This appendix describes the data sources used in this study, both for the analyses in year 1 as well as records that we plan to incorporate in future years. Table 1 provides an overview of these sources and the text provides more detail about each source. This section is a snapshot of data sources that we currently intend to use; future editions of this pre-analysis plan will add new data sources to this appendix as needed.

Table 1. Summary of Survey and Administrative Data Sources

	Data	Outcomes
Survey Data	Screener survey	Marital status, living arrangements, educational attainment, contraceptive use, childbearing, PPMI fee scale category.
	Baseline and follow-up surveys	Contraceptive method, employment status, occupation, income category, educational attainment and enrollment, childbearing and pregnancy history, healthcare use, pregnancy intentions, attitudes towards contraception and childbearing, expectations about the future, childhood background, relationship quality, mental health and well-being, physical health.
Administrative Data	PPMI patient records	Physical health information, contraceptive use, childbearing and pregnancy history, abortions, medical diagnoses, insurance and payment information for reproductive care, M-CARES voucher use, STD/STI incidence.
	Credit reports	Measures of indebtedness and financial strain, delinquency, and credit score.
	Tax data (OTA, RDC)	Number of dependents in the household, marital status, homeownership, college enrollment, income and sources, number of employers, receipt/eligibility of Earned Income Tax Credit, eligibility for other public assistance programs, number of address changes.
	Birth certificates	Date of birth, marital status at birth, mother education and occupation, father name, education, occupation, SSN, and DOB if acknowledged, pre-natal care, gestational length, birth weight, parity of birth, Medicaid coverage for birth.
	Census, ACS, CPS	Living circumstances, number of children in the household (and when they were born), renter/owner status, incarceration status, neighborhood quality.
	Education data	Education enrollment, educational attainment, achievement test scores, absenteeism, school delinquency.
	Public program participation	Medicaid receipt, SNAP receipt, TANF receipt, Unemployment Insurance, Disability receipt.

	Michigan criminal justice records	Arrests, prison entries, incarceration status.

1. Screening survey

We use the screening survey for two primary objectives: to screen participants on eligibility and to collect pre-intervention data on participant's characteristics. It has 29 questions excluding questions relating to enrollment and consent. Respondents receive $10 in cash for completing it (they are not allowed to skip any questions), and we estimate that it will take about 5 minutes to complete (see Appendix D for the survey). The screening survey is self-administered prior to enrollment and collects basic demographic characteristics of respondents: income category (PPMI fee scale), age, marital status, living arrangements and cohabitation, educational attainment, contraceptive use, and the number of children. We will have this information for all 5,000 M-CARES participants, and will use it to create enrollment reports, study selection into the M-CARES, and conduct balance tests of covariates between treatment and control groups (see details in section 4.2).

2. Baseline and follow-up surveys

The baseline survey is self-administered either on a tablet at the clinic after the PPMI appointment or in another location using a device with access to the internet. Follow up surveys will be collected 2 and 4 years after enrollment. Surveys collect information such as religion, employment status, income category, educational enrollment, fertility history, healthcare use, pregnancy intentions, attitudes towards contraception, expectations about the future, relationship quality, mental health, physical health, and childhood background (see Appendix D for the survey). In addition, respondents in the treatment arm are asked about voucher use and change in contraceptive use. The baseline survey has 118 questions in total (4 questions verify information collected in the screening survey), as well as skip patterns. We estimate the survey should take 25 minutes to complete. Respondents will receive $60 in cash if they complete the survey at PPMI the day they enroll or $40 if they complete it at a later point. We anticipate a response rate of 80 percent to the baseline survey, owing to the large incentives and the ability to allow individuals to take the baseline on site.[16] M-CARES participants will also be asked to complete 2 follow-up surveys, 2 and 4 years after enrollment. Respondents will receive $50 for each completed survey. The follow-up surveys are not yet designed, but will contain questions which will largely overlap with the questions in the baseline survey, allowing us to measure the relationship between voucher receipt and changes in measured outcomes. Follow-up surveys will also ask respondents about pregnancies, pregnancy intentions, and abortions—important variables that may not be available or complete in administrative data.

3. PPMI patient records

We will provide PPMI with full name, date of birth, address, and SSN for all study participants, and PPMI will use this information to locate participants' patient records. These records contain detailed information on participants' socio-demographic information, physical health information, pregnancy and childbearing history, medical history, their PPMI visits (at all PPMI clinics), including the date of the visit, details on diagnoses and procedures and services provided, and payment method and payment amount. The method of payment for the services on the records allows us to identify the services obtained with the M-CARES voucher. We rely on these data in part to estimate how pregnancies, abortions,

[16] For example, Panel Study of Income Dynamics paid respondents $65 in 2009 for a 75-minute interview, with a response rate greater than 95 percent (Schoeni et al., 2013). It should be noted that PSID is a longitudinal survey, and at least partially the high response rate must stem from the long-term relationship with respondents.

childbearing, and contraceptive use is affected by voucher receipt. We expect to link all participants to their PPMI records (going back to the date of participant's first PPMI visit and up to 2023, the end of the study). The date of participant's enrollment from the screening survey allows us to separate patient history data into pre- and post-intervention periods.

4. Credit reports

We will use either Transunion and/or Equifax for credit report data. We will send Equifax and Transunion study participants' names, social security numbers (SSNs), and addresses per their requirements. Equifax and Transunion will link the data, remove identifying information from the dataset (all name, SSN, address) and send Michigan a de-identified file. (This process is consistent with the Fair Credit Reporting Act). Credit reporting companies create a record from information on an individual's experiences with credit, leases, non-credit-related bills, money-related public records (see Avery, Calem, and Canner (2003) for a detailed discussion of credit data). We will rely on two types of information available from credit reports: (1) collection records, which contain data on accounts in collection (e.g. unpaid utility bills), and (2) credit provided by banks, credit unions, and other institutions, which provide information on outstanding balances, payment history, and credit limits, among others. Although these data provide a nearly complete picture of credit history of the general population, they may be less complete for low-income population, such as the ones involved in this study. It is expected that low-income population may be more likely to rely on informal borrowing channels, which would not be captured in these data.

5. Tax records

We have received approval from both the Census Bureau and the Office of Tax Analysis (OTA) to link our study participants to the universe of all tax filers from 1996 to the present (end date will be updated as time moves forward). We are not yet sure which access route we will use, so describe the process of linking to and disclosing results from both sources here.

Tax data allow us to characterize tax filing participants, before and after the intervention, in terms of (1) living circumstances (living with parents, single headship, living with married or unmarried partner, etc.), (2) the number of children in household (and when they were born and age at first birth), (3) homeownership, and (4) neighborhood quality (an important metric for standard of living). In addition, tax data allow us to assess (5) college enrollment (via tax credits for these expenditures), (6) exact income from wage earnings in the household, (7) receipt/eligibility of the Earned Income Tax Credit, and (8) eligibility for other public assistance programs.

Every M-CARES participant who files taxes can be linked to tax data using a protected identification keys (PIK). To link to Census IRS 1040s, we will provide Census with the full name, address, and SSN of the study participants and their families, which will enable linking to PIKs. Census will then provide access to M-CARES team to these data in the University of Michigan Research Data Center (RDC). OTA will also use PIKs, but OTA will not provide the data to us. Instead, M-CARES team member will conduct analyses with the data. OTA will analyze these data and review them for confidentiality before disclosing to the study team. For OTA data, M-CARES team will not have access to any individual's information. It will only review descriptive statistics and regression output of these data released by OTA. We expect to link almost every study participant to their tax records.

6. Census, American Community Survey (ACS), Current Population Survey (CPS) Annual Social and Economic Supplement (ASEC)

The 2000, 2010, and 2020 (expected) Censuses contain the data compiled from the questions asked of all people and about every housing unit in each census year. This includes a detailed enumeration of everyone in the U.S. population by sex, age, race or Hispanic or Latino origin. In addition, variables indicating relationship to household head and marital status allow us to characterize all children living in the household and sub-family.

Every study participant will be linked to census data using a personal identification key (PIK). Available census variables allow us to characterize *every* person in the study, before and after the intervention, in terms of the following characteristics: (1) living circumstances (living with parents, single headship, living with married or unmarried partner, etc.), (2) the number of children in household (and when they were born and age at first birth), (3) renter/owner status, (4) incarceration status, and (5) neighborhood quality (an important metric for standard of living). Additionally, we can identify these outcomes for race/ethnicity subgroups.

Census data are available in the Research Data Centers (RDCs), one of which is housed at the University of Michigan. The Census Bureau has established a specific application and approval process and has specific disclosure protocols to protect the identity of respondents. We have been able to obtain these data to link to our study participants via a standard application process. In addition, we will need to consent all study participants to link them to Census as part of the enrollment process.

ACS and CPS ASEC data will contain information for a random sample of M-CARES participants. If the sample overlap is large enough, we will use these records as a supplement to the Census. These surveys allows us to characterize participants in a similar way as the Census, but on a more frequent or updated basis.

7. Education records

We will use two types of administrative sources for education data, the National Student Clearinghouse (NSC) and Michigan education data (K-college).

The National Student Clearinghouse (NSC) was founded in 1993. Originally, this data source was tied to the student loan industry and gathered enrollment data from participating colleges. The purpose of the database was to allow the loan industry to confirm that a borrower was enrolled and therefore eligible to defer repayment of student loans. Consistent with this history, the NSC today tracks enrollment (but not credits or major) and whether a student has graduated and degree earned. As of 2012, the NSC started using the Classification of Instructional Programs to record a student's major (http://nces.ed.gov/ipeds/cipcode/).

The main strengths of these data are that it is possible to trace K-12 students through their college careers. There are also some notable limitations as described in Dynarski et al. (2013). First, not every college reports to the NSC because reporting is voluntary. This problem was more pervasive in the 1990s and 2000s but has improved more recently. For example, in 2011 93 percent of post-secondary enrollment was documented in the NSC (Dynarski et al. 2013, p. 9); these rates were higher outside the South and the highest in Virginia and Wyoming at 97.4 percent and 99.6 percent coverage, respectively. This means that college attendance in our study period will be captured for the vast majority of the participants. Second, there is missing information on degree type. Dynarski et al. (2013, p. 15) report that in Michigan, which is a focus of our study, the NSC data record only 81 percent of all degrees earned conditional on enrollment in an NSC-participating institution. However, degree coverage is the highest at public colleges at 82 percent, where our study participants are the most likely to be enrolled if enrolled at all. Third, the NSC cannot disclose student information that students or schools elect to block under Family Educational Rights and Privacy Act (FERPA). The share of student information that is blocked to researchers is not observed, but a NSC publication documents that it is highest at 2-year public colleges. Finally, Dynarski et al. (2013) report that the NSC linking algorithm makes some matching errors. Therefore, our study participants' outcomes may be measured with some error. After accounting for these issues, Dynarski et al. (2013, p.17) estimate that the comprehensive coverage rate in the NSC data is around 86 percent in 2011.

As with Census data, study participants will be linked to the NSC using a unique ID number. Linked individuals would be those enrolled in post-secondary institutions that agreed to participate in the NSC. The NSC provides information on enrollment, the beginning and ending date that a student is enrolled during each term, whether a student is enrolled full or part-time, whether a student has earned a degree, and the date the degree is earned. For the subset of students participating in the "DegreeVerify" service, we will be able to also obtain college major and degree type (e.g., BA). We should be able to link the vast majority of our study participants to NSC data if they have pursued post-secondary education.

To supplement NSC data, we will link our respondents and their children to Michigan public education data. We will work directly with school districts in Michigan to link our participants and their children to their education records, therefore the final sample of linked respondents depends on district-specific rules for releasing data. In addition, we will not be able to link participants who obtained education outside of the state. Michigan public education data contain administrative school records, which report school attendance, promotion to the next grade, grades, and graduation date. Together, NSC and Michigan K-college records enable us to study education enrollment and attainment of the study participants.

8. Birth certificates

We will provide Michigan Department of Health and Human Services (MDHHS) participants' full name, date of birth, address, and SSN, which allows MDHHS to locate birth certificates of participants and participants' children. We expect to find approximately 77 percent of participants and 90 percent of participants' children in these data.[17] Each birth record contains the date of birth, mother's marital status, father's name, source of payment, gestation, number of prenatal visits, and the date of the first prenatal visit. This data allows us to validate mother reports of the number of children and study several types of outcomes: (1) childbearing (number of births, timing/spacing of births); (2) childbearing and partnership stability (non-marital childbearing, no father on the birth certificate); (3) heath of a pregnancy (number of prenatal visits, the timing of the first prenatal visit, gestation length), and (4) receipt of public services (whether the birth was paid for by Medicaid).

9. Public program participation

We will use several types of data available to analyze public program participation: Medicaid, Supplemental Nutrition Assistance Program (SNAP), and Temporary Assistance for Needy Families (TANF). These data allow us to observe program enrollment and expenditures for study participants.

10. Michigan criminal justice records

We expect that criminal records data will be available through the RDC starting in the summer of 2018. These data track an individual on a quarterly basis, collecting information their arrests, prison entries, and incarceration status (these data are not state-specific and include records from all contributing data providers).

[17] In ACS2015, among children who reside in Michigan, are under 18 years old, whose mothers are 18-35 years old, and with family income less or equal to $80,000, 91 percent were born in Michigan. Among women 18-35 years old residing in Michigan with family income less or equal to $80,000, 77 percent were born in Michigan.

Appendix B. Variable Definitions

We expect to use the following variable definitions, but these variable definitions may be altered if the distributions of answers are very different from what is expected. For instance, a pre-specified categorical variable may be undefined if very few or no women report being in a particular category; in this case we will modify variable definitions from what is specified below.

1. Screener survey

We use the screener survey information to measure demographic information. Specifically, we are able to control for the following demographic characteristics:

Age
Q7 asks respondent her age, which is entered as an integer. From the answer, we construct several age indicators: *Age 18-19*, which equals to 1 if respondent is 18 or 19 years old at the time of the screening survey, and 0 otherwise; *Age 20-22*, *Age 23-25*, *Age 26-29*, and *Age 30-35*, which are similarly defined.

Race
Q9 asks respondent for her racial background, allowing for multiple responses. We code these responses as follows: *White* equals to 1 if respondent selects "White" as the only category, and 0 otherwise; *African-American* equals to 1 if respondent selects "Black, African, or African American" as the only category, and 0 otherwise; *Other/mixed race* equals to 1 if respondent selects "American Indian, Alaskan Native or Native Hawaiian", "Hispanic or Latino", "Asian or Pacific Islander", or "Another race/ethnicity" as the only category or selects two or more responses, and 0 otherwise.

Co-residence with partner
Q10 asks respondent if she is currently living with a spouse or a romantic partner. We create a variable *Cohabit with partner* equal to 1 if respondent selects "Yes" as an answer, and equal to 0 otherwise.

Marital Status
Q11 asks respondent to select a response to describe her relationship status. We create the following variables: *Married* equal to 1 if respondent selects "Married", and equal to 0 otherwise*; Committed relationship, Casual relationship, No relationship* are similarly defined.

Educational attainment
Q13 asks respondent to select levels of schooling completed so far, allowing for multiple responses. For brevity, we assign numbers to categories in this way: "GED" - 1, "High school diploma" – 2, "Went to college but did not get a degree" – 3, "Vocational Degree/Certificate" – 4, "Associate's degree" – 5, "Bachelor's degree" – 6, "Graduate degree" – 7, "None of the above" – 0. We construct the following variables: *No degree* equals to 1 if respondent selects category 0, and no other categories, and 0 otherwise; *GED* equals to 1 if respondent selects category 1 and no other categories, and equals to 0 otherwise; *High school* equal to 1 if respondent selects categories 2 or 2 and any of the categories < 2, and equals to 0 otherwise; *Some college* equals to 1 of respondent selects category 3 or 3 and any of the categories < 3, and equals to 0 otherwise; *Vocational degree* equals to 1 if respondent selects category 4 or category 4 and of the categories < 4, and equals to 0 otherwise*; College degree or more* if a respondent selects categories 5-7, and 0 otherwise. We will also construct a continuous measure *Years of education*, by assigning median years of education to each education category.

Number of children
Q19 asks respondent to enter the number of live births she has had. We create the following variables: *One child* equal to 1 if a respondent has had 1 live birth and equal to 0 otherwise, *Two children* equal to 1 if a respondent has had 2 live births and equals to 0 otherwise, *Three+ children* equal to 1 if respondent has had three or more live births, and equal to 0 otherwise. Finally, we create a dummy *No children* equal to 1 if a respondent has had 0 live births, and equal to 0 otherwise.

Income/fee scale category
Q5 on the screener collects information on respondent's payment category, from which we create the following variables: *101-150% FPL* equals to 1 if respondent selects category 2 and equals to 0 otherwise; *151-200% FPL* equals to 1 if respondent selects category 3 and equals to 0 otherwise; *201-250% FPL* equals to 1 if respondent selects category 2 and equals to 0 otherwise; *Above 250% FPL* equals to 1 if respondent selects category 5 and equals to 0 otherwise.

Currently using birth control
Q25 asks whether a respondent is currently using birth control. We create a variable *No method used* equal to 1 if respondent selects "No", and equal to 0 otherwise.

Main contraceptive method
Q26 asks respondent to select main method to prevent or delay pregnancy. We create the following variables: *Birth control pills* equal to 1 if respondent selects "Birth control pills", and equal to 0 otherwise; *Condoms, Withdrawal* are similarly defined. *Injections* equals to 1 if respondent selects "Shot", and 0 otherwise. We create a variable *LARC* equal to 1 if respondent selects "Implant" or "Intrauterine device or IUD", and 0 otherwise. Finally, we create a variable *Other method* equals to 1 if a respondent selects any other answer, and 0 otherwise.

2. Voucher and PPMI Services Use

We will construct the following variables relating to the efficacy of the intervention, including take-up of vouchers and expenditures at PPMI following the intervention.

Used voucher within 100 days
This variable measures the take up of the voucher and the efficacy of the intervention. It will be constructed using PPMI data. PPMI records will provide us with data on the amount and date participants use their voucher. *Used voucher within 100 days* is equal to 1 for all study participants who use a voucher within 100 days allowed by the study. If the intervention is a success, we expect receiving a voucher to predict voucher use. (Finding that this is not the case is evidence of a failure of implementation or that participants do not value the PPMI services one could buy with the voucher.)

Total expenditures at PPMI with 100 days
This variable measures the total value of services purchased and is constructed using PPMI data. As part of each study participant's records, PPMI will provide us with data on the amount paid for services purchased at PPMI using any pay source: self-pay, insurance, or voucher. *Total expenditures at PPMI within 100 days* is equal to total payments made to PPMI within 100 days from any pay source. By construction, total voucher amount used at PPMI is 0 for participants in the control group. We expect total expenditures to be higher in the treatment group.

Out-of-pocket (oop) payment in 100 days
This variable measures the amount of out-of-pocket payments made by each study participant, and is constructed from PPMI patient data. We will measure both *Total oop payments at PPMI within 100 days* as well as construct a binary variable documenting *Any oop payment at PPMI within 100 days.*

Share of voucher used
This variable measures the share of voucher used at PPMI. We will add expenditures paid for by the voucher within 100 days of enrollment from PPMI patient records, and divide by the total voucher amount.

Reasons for nonuse
This set of variables captures various reasons for voucher nonuse, collected in the baseline survey Q9 (they are all equal to 0 for participants who have used the voucher by the time they complete the baseline survey). We create dummy variables capturing each of the following responses: "I did not have any costs", "I did not need any Planned Parenthood services", "I did not have time to use it", "I forgot to use it", "I do not want to use it".

Intend to use voucher
These variables measure whether a respondent who has not used the voucher by the time she completes the baseline survey intends to use the voucher, constructed from answers to Q10 on the baseline survey. By construction, these variables are not defined for respondents who have used the voucher by the time of the baseline survey. We create dummy variables capturing each of the following responses: "Yes, I have already scheduled an appointment and will use it then", "Yes, but I have not scheduled an appointment yet", "No, I do not plan to use the gift card".

Intend to use voucher – contraceptive method
This set of variables captures the types of contraceptives respondent plants to purchase with the voucher, constructed from answers to Q11 on the baseline survey. By construction, these variables are not defined for respondents who have used the voucher by the time they complete the baseline survey. These variables are equal to 0 for respondents who do not intend to use the voucher. We group intended contraceptive methods as follows: *Intend to use on pills* equal to 1 if respondent selects "Birth control pills", and equal to 0 otherwise; *Intend to use on injection* if respondent selects "Shot", and 0 otherwise; *Intend to use on LARCs* if respondent selects "Intrauterine device, or IUD" or "Implant", and 0 otherwise; *Intend to use on other contraceptives* equals to 1 if respondent selects any of the other methods, and 0 otherwise.

Any contraceptive use
This variable measures whether the intervention affects the use of any contraceptives. The variable will be constructed using PPMI data, which contains information on contraceptive services provided at PPMI, and survey data, which provides data on self-reported contraceptive use. We will construct a binary variable *Using any method*=1 if a participant receives any contraceptive method at any of PPMI clinics within 100 days of the enrollment date or reports that she is using a method at baseline.

Temporal coverage of contraceptive method
This variables measures whether the intervention affects the temporal coverage of purchased/used contraceptive method. The variable will be constructed from the survey data and PPMI clinic data, using information on the date of clinic visits, and the services obtained at the visits. Specifically, we will use data on the type and quantity of birth control obtained to construct expected method duration. We will define *Duration of method for X months* = 1 if the participant obtains enough birth control to provide coverage for X months. For example, if a participant purchases 3-month supply of birth control pills on the day she is recruited, we infer that she continues to use birth control pills uninterrupted for 3 months. If she chooses an IUD, we infer that she has coverage with her baseline method for 3 years.

Use of high and/or moderate efficacy method (100 days)
These variables measure whether the intervention affects the use of higher efficacy contraceptives. The variable will be constructed using PPMI data and survey data. We will construct a binary variable *Using any high efficacy method*=1 if a participant receives any Tier 1 method (IUD, implant, or sterilization) at any of PPMI clinics within 100 days of the enrollment date or reports that she is using a Tier 1 method at screening or baseline. We will construct another binary variable *Using any moderate efficacy method*=1 if a participant uses any Tier 1 or Tier 2 method (pill, patch, ring, injectable, diaphragm) at baseline or screening or receives at any of PPMI clinics within 100 days of the enrollment date. We construct a third binary variable *Using any high/moderate efficacy method* equal to 1 if *Using any moderate efficacy method*=1 or *Using any high efficacy method*=1, and equal to 0 otherwise.

Use of high and/or moderate efficacy method (1 year)
These variables measure whether the intervention affects the use of higher efficacy contraceptives in the first year. These variables are constructed in an identical manner as the variables measuring contraceptive use within 100 days (see above); the difference is that we measure use of higher efficacy methods 1 year from enrollment.

Switched contraceptive method within 100 days
This variable measures whether the intervention affects participants' decisions to switch their contraceptive method. The variable will be constructed using PPMI data and survey data. We will construct a binary variable *Switched method*=1 if a participant switches her method from what is reported on the screening survey[18] to the baseline (completed after the clinic visit) or the method she obtains from PPMI within 100 days of the enrollment date.[19] Similarly, we define *Switched to higher tier method*=1 if the participant switches to a higher tier method within 100 days of the enrollment date.

Method efficacy
This variable measures the efficacy of each participant's method use based on CDC reports.[20] For instance, a respondent given an implant will be coded as having a 0.05 percent failure rate, whereas a respondent using birth control pills will be coded as having a 9 percent failure rates in every month for which she purchases a supply. The variable will be constructed from PPMI clinic data, using information on the date of clinic visits, and the services obtained at the visits.

3. Pregnancy and Childbirth

Pregnancy tests
We will use two variables to measure whether the intervention affects the likelihood of a pregnancy test at PPMI using PPMI data. We will construct a binary variable *Any pregnancy test*=1 if a participant receives any pregnancy test within 1 year of enrollment. We will also use *Number of pregnancy tests* that counts the number of pregnancy tests ordered by PPMI for the participant within 1 year of enrollment.

Pregnancies
We will use two variables to measure whether the intervention affects the likelihood of pregnancy using PPMI data. We will construct a binary variable *Any pregnancy*=1 if a participant receives any positive pregnancy test results within 1 year of enrollment. We will also use *Number of pregnancies* that counts the number of positive pregnancy tests received by the respondent (for different pregnancies) by the respondent at PPMI within 1 year of enrollment.

Childbirth
We will measure whether the intervention affects childbirth using Michigan Vital Statistics Natality Data (birth certificates). We will construct *Number of births* variable equal to the number of births for the participant within 1 year of respondent.

Abortions
We will measure whether the intervention affected incidence of abortions. We will construct a binary variable *Abortion*=1 if a participant has an abortion within 1 year of enrollment, using PPMI records data. We will also construct *Number of abortions* which counts the number of abortions observed for the participant in PPMI data within 1 year of enrollment.

Emergency contraceptive use
We will measure whether the intervention affected incidence of emergency contraceptive use. We will construct a binary variable *Any emergency contraceptive use*=1 if a participant obtains emergency contraception within 1 year of enrollment, using PPMI records data. We will also construct a continuous measure of frequency of emergency contraceptive use, *Emergency contraceptive use*, which counts the number of times the participant obtains emergency contraceptives at PPMI within 1 year of enrollment.

[18] The screening survey asks, "What is the one main method to prevent or delay pregnancy you are currently using?"
[19] The baseline survey asks, "Did you get a birth control method at [if today's date=screen date: today's] /[if today's date>screen date: your last] visit to Planned Parenthood?", "If yes, what type of birth control did you get at [if today's date=screen date: today's] /[if today's date>screen date: your last] visit to Planned Parenthood?"
[20] https://www.cdc.gov/reproductivehealth/unintendedpregnancy/pdf/contraceptive_methods_508.pdf (downloaded June 28, 2017).

4. Education, Employment, Income, and Public Program Use

Educational enrollment
This variable measures whether a participant is enrolled in any type of schooling. We will use information from the National Student Clearinghouse and Michigan K-college data for linked respondents as well as tax information to create a variable *Any educational enrollment* =1 if participant is enrolled in any type of school in Michigan within 1 year of enrollment in the study.

Educational enrollment – type of enrollment
This variable measures whether a participant is enrolled in a given type of school. We will use information from the National Student Clearinghouse and Michigan K-college data for linked respondents. We will create a set of indicator variables denoting enrollment in a particular type of school. *Enrollment in GED program* =1 if a participant is enrolled in a GED program in Michigan within 1 year of enrollment in the study, *Enrollment in a 2-year program* =1 if a participant is enrolled in 2-year program in Michigan within 1 year of enrollment in the study, and *Enrollment in a 4-year program* =1 if a participant is enrolled in 4-year program in Michigan within 1 year of enrollment in the study.

Educational enrollment – full-time/part-time
This variable measures whether a participant is enrolled in any type of schooling full-time or part-time. We will use information from the National Student Clearinghouse and Michigan K-college data for linked respondents. We will create a variable *Educational enrollment full-time* =1 if a participant is enrolled full-time in any type of schooling in Michigan within 1 year of enrollment in the study. *Educational enrollment part-time* =1 if a participant is enrolled part-time in any type of schooling in Michigan within 1 year of enrollment in the study. In addition, we can supplement this information with tax records.

Employment
This variable measures whether a participant is employed. We will use information from the Office of Tax Analysis and IRS 1040s from the RDC and code this variable as =1 if a participant has more than $1,000 in taxable earnings in the within 1 year of enrollment (>$1,000 is the standard, OTA internal definition of being employed).

Total Income from Earnings
This variable measures how much a participant earned. We will use information from the Office of Tax Analysis and IRS 1040s from the RDC and code this variable as the total income from earnings within 1 year of enrollment in the study. We will also create a variable *Log total income,* equal to the log of *Total income*. If effect of voucher receipt on income is proportional to income, measuring the average percentage change in income is more informative compared to measuring the average change in income.

SNAP/TANF participation
We have obtained an agreement to use administrative records regarding SNAP and TANF for the state of Michigan from the RDC. We will use this data to construct two measures of program participation: *Received any SNAP/TANF payments*=1 if a participant receives any SNAP/TANF payments within 1 year of enrollment and *Dollars in SNAP/TANF benefits received* = total amount of SNAP/TANF benefits received within 1 year of enrollment.

5. Financial well-being

The variables below are measured since the date of enrollment for each participant to assess how voucher assignment is related to financial well-being of participants.

Any collection
This variable measures the incidence of unpaid bills that have been sent to collection. This variable will not capture the full extent of unpaid bills for several reasons: since not all unpaid bills are sent to collection (larger creditors such as hospitals and utility companies are more likely to send bills to collection); some unpaid bills may be collected by the creditors themselves and not collection agencies; there is a delay

between the time of the unpaid bill and the time it is sent to a collection agency. *Any collection*=1 if a participant has a record of a collection since the date of enrollment, and 0 otherwise.

Amount owed in collection

This variable measures the amount owed in collections at the time of the data extract – it will not capture the amount in collections that has been paid. However, as only 11 percent of reported collections items are paid off (Avery, Calem, and Canner, 2003) we expect that this measure will capture most of the money owed. This variable equals to 0 for participants who have no bills sent to collection.

Any delinquency

Any delinquency = 1 if a participant has had any delinquency on any credit account (defined as a payment that is at least 30 days late) since the date of enrollment, and 0 otherwise. If a participant has no open credit since the notification date, this variable equals to 0 by construction.

Any major delinquency

This variable measures more serious delinquency on a credit account. *Any major delinquency* = 1 if a participant has had any delinquency on any credit account (defined as a payment that is at least 120 days late) since the date of enrollment, and 0 otherwise. If a participant has no open credit since the notification date, this variable equals to 0 by construction.

Total debt

This variable measures total debt incurred by the respondent. It includes amount owed in collection, loans, mortgages, liens, credit card debt, school loans, etc.

6. Neighborhood quality

Address changes since enrollment

Number of address changes is constructed from the tax data and measures the number of different addresses reported in the tax returns.

Income-to-poverty ratio in census tract of residence

This variable is constructed from the Census data, and equals to the average income in the census tract of residence divided by the poverty threshold.

Teen pregnancy in tract

This variable is constructed from the Census data, and equals to the number of women under age 20 in the census tract who have a child living with them, divided by the total number of women under 20 in the census tract.

Share of households receiving public assistance in tract

This variable is constructed from the Census data, and equals to the number of households receiving public assistance divided by the total number of households in the tract of residence.

Share of poor children in tract

This variable is constructed from the Census data, and equals to the number of children (under the age of 18) living in households with family income below the poverty threshold divided by the total number of children living in households in the tract of residence.

Share of home ownership in census tract of residence

This variable is constructed from the Census data, and equals to the number of households who own the housing unit divided by the total number of households in the tract of residence.

Median house price in census tract of residence

This variable is constructed from the Census data and equals to the median house value for all housing units in the tract of residence.