PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Evaluation Protocol to Assess Maternal and Child Health Outcomes Using Administrative Data: A Community Health Worker Home Visiting Program
AUTHORS	Sabo, Samantha; Butler, Matthew; McCue, Kelly; Wightman, Patrick; Pilling, Vern; Celaya, Martín; Rumann, Sara

VERSION 1 – REVIEW

REVIEWER	Anna Price Centre for Community Child Health, Murdoch Children's Research Institute, Australia
REVIEW RETURNED	20-Jun-2019

GENERAL COMMENTS	Thank you for the opportunity to review this paper. The manuscript describes the protocol for a retrospective analysis that uses propensity score matching to investigate whether Arizona's Health Start Program (a community health worker home visiting program) impacts outcomes related to newborn health, maternal health care and use, and child health and development. Overall, I think the manuscript provides a sufficient level of information and is written clearly. I believe the paper could be strengthened if the following areas were clarified: - Have the authors developed a program logic to guide their choice of outcome measures? A review by Segal and colleagues of nurse home visiting programs designed to reduce child maltreatment found that program logic helped target explicit outcomes and was related to effectiveness, and is a key feature missing from many trials including those targeting developmental and behavioral outcomes. [Please see Segal L, Opie RS, Dalziel K. 2012. Theory! The missing link in understanding the performance of neonate/infant home-visiting programs to prevent child maltreatment: A systematic review. Milbank Q. 90: 47-106.] Tables 1-4 represent aspects of a program logic but there appear to be some inconsistencies across them. A program logic would tie these tables together and refine the logic linking the goals, processes, outputs and outcomes. Without program logic it can be difficult to ascertain why certain outcomes are impacted over others. Some specific examples include: o I can see links between the measurable outcomes in Table 1 and Goals 1-4 of Table 2, but it is not clear where Goal 5 in Table 2 links with Table 1. o I can see links between the Outcomes Indicators in Table 3 and Measurable Outcomes in Table 1, except for "method of delivery" and "inter-pregnancy intervals". o The last paragraph of the Background and Rationale describes safety actions like installing car seats and helping with the safe

	sleep environment. Are there examples of the strategies targeting child development? - The authors describe costs of low birth-weight babies, and also of disseminating findings with Arizona Medicaid. Have the authors considered conducting a cost-evaluation to determine the cost-benefits of the home visiting program? This would help health services (across and between states, and internationally) decide which programs to trial or implement. - “Data collection” and “exclusion criteria” sections indicate that the datasets have already been established. Is this correct and have any of the analyses been performed? - Missing data: As data come from administrative datasets, the authors say that missing data are not anticipated. As a reader who is unfamiliar with the datasets being analysed, does this mean that data are available for every eligible child across all datasets and that none are missing for any child? If so, how do the various services make this possible? - Could the authors please define what the specific social risk factors are relating to “marital status, living situation, race and ethnicity, education level, income, and insurance type, … maternal BMI, and maternal age.” For example, “marital status” on its own is a demographic rather than a risk factor. Is the risk factor “single parent family” or similar? - What are the potential limitations of the study? - Could the authors please provide more information on the following for readers who may not be familiar with these terms: o “universe” of Health Start/births o “honest broker” process o Why do the authors select a HOMVee rating of “moderate” over “high” (i.e. is it because of the study design), and what is the impact of this if any on MIECHV funding and ongoing implementation?

REVIEWER	Dr. Jennifer Enns Manitoba Centre for Health Policy University of Manitoba Canada
REVIEW RETURNED	06-Aug-2019

GENERAL COMMENTS	Overall, the study topic is important and will be of interest to many readers. My impression of the protocol paper is that the science is sound, but the information is not presented in a way that readers will be able to understand very easily. I have therefore suggested the authors revise their paper to improve clarity and readability. Introduction The introduction is a little disorganized - the authors state their aim twice (once in the last sentence of the first paragraph and once in the last sentence of the third paragraph) but then continue providing background information after that. The introduction could also be streamlined somewhat. For example, the last paragraph about HSP could be incorporated into the Methods section where the authors describe the intervention in more detail. A better place to state the aim of the paper is at the end of the introduction or in the objectives section. Methods The language in the methods section is so cluttered with details that I found it difficult to decipher what the authors plan to do. For example, in the Primary Outcome paragraph, the authors state

	their study design, study timeframe and study setting (again) without actually saying what the study outcomes are. The reader is directed to find them in a table instead. Another example – in the stats analysis section, the authors could be more clear and upfront in stating the approach they plan to use and what kind of results this approach will produce, rather than explaining extensively why they are not choosing a different approach and what the specific advantages and disadvantages of their chosen approach are. Throughout the methods section, although much of the information presented is important, it is not presented in a way that most readers would find accessible. The authors should consider using another BMJ Open Protocol paper (or several) as a guide in organizing their manuscript and in assessing how much detail to provide and where. Some of the language used is problematic, for example: - "Data Collection" isn't really an applicable term for what you are describing in this paragraph. I think you mean "Study Cohort Development" or "Identification of Study Population". - You use the term "matching" in the above-mentioned paragraph when what I think you mean is "linking". - One doesn't really "engage a strategy", as the author plan to do in the Dissemination paragraph. Figure 2 – The authors should use a PRISMA flowchart to depict how they identified the study population. http://prisma-statement.org/PRISMAStatement/FlowDiagram.aspx The authors may want to consider adding a section on study limitations. Other: I suggest the authors be more consistent with their abbreviations (e.g. HSP, Health Start, Health Start Program, Arizona Health Start Program). I would be very selective with what you choose to abbreviate as the reader will not be as familiar as you are with abbreviations specific to your study. I suggest the authors have a copyeditor read the manuscript carefully before resubmission. I noted a number of missing sentence articles and instances of misplaced punctuation marks. As a post-script: I have found this resource very helpful in writing papers – Schimel, J. 2012. Writing Science: How to Write Papers That Get Cited and Proposals That Get Funded (1st Edition). Oxford, UK: Oxford University Press. (I have no affiliation whatsoever with this author or his publishing firm)

VERSION 1 – AUTHOR RESPONSE

REVIEWER #1

Thank you for the opportunity to review this paper. The manuscript describes the protocol for a retrospective analysis that uses propensity score matching to investigate whether Arizona's Health

Start Program (a community health worker home visiting program) impacts outcomes related to newborn health, maternal health care and use, and child health and development. Overall, I think the manuscript provides a sufficient level of information and is written clearly. I believe the paper could be strengthened if the following areas were clarified:

1. Have the authors developed a program logic to guide their choice of outcome measures? A review by Segal and colleagues of nurse home visiting programs designed to reduce child maltreatment found that program logic helped target explicit outcomes and was related to effectiveness, and is a key feature missing from many trials including those targeting developmental and behavioral outcomes. [Please see Segal L, Opie RS, Dalziel K. 2012. Theory! The missing link in understanding the performance of neonate/infant home-visiting programs to prevent child maltreatment: A systematic review. Milbank Q. 90: 47-106.]

RESPONSE: We appreciate this comment on a very critical issue. We have carefully read the Segal et al. article and appreciate their contribution to this important topic. Our study is a retrospective evaluation of a long-standing Community Health Worker (CHW) home visiting program. As a program, Health Start is driven by known behavior change theories associated with CHW interventions. To address the issue of a theory driven approach to understanding the performance of CHW home visiting program, we have clarified our use of specific behavior change theories and how they link to CHW core competencies to effectuate change. This is further articulates in a revision to Table 2 (See comment below) which identifies CHW inputs, process and outcomes indicators.

> *Page 6, Par 2.*
> *HSP CHWs identify, screen and enroll pregnant women early in their pregnancy and assists them with obtaining early and consistent prenatal care, provide prenatal and postpartum education, information and referral services, advocacy, and emphasize timely immunizations and developmental assessments for their children (Table 2). HSP CHW home visits are guided by an asset-based approach (34) and two primary theories of behavior change, the Trans Theoretical Model (35) and the Social Cognitive Theory (36). These two behavioral change theories assume, respectively, that behavior modification in individuals is a multistage process in which people move through stages of readiness for change, and that they do so in the context of reciprocal relationships with their environment, behavior and cognition. As trusted members of the community served, sharing both lived experience and cultural knowledge of the population, CHWs are well positioned to support HSP clients. Each CHW home visiting session is structured to promote behavior change through assessment, education, goal planning, referral, advocacy, and follow up activities, coupled with meaningful adult learning models. These interactions are designed to encourage personal agency of adult learners to integrate new knowledge and create a cognitive structure that makes sense of their own surroundings and situations (37). Through behavior change theories and adult learning models, the Health Start Program CHWs privilege the co-construction of knowledge among all participants, assume all are co-learners, and encourage critical thinking about self-sufficiency, empowerment, and personal agency related to the five HSP goals (Table 1).*

2. Tables 1-4 represent aspects of a program logic but there appear to be some inconsistencies across them. A program logic would tie these tables together and refine the logic linking the goals, processes, outputs and outcomes. Without program logic it can be difficult to ascertain why certain outcomes are impacted over others. Some specific examples include:

2a. I can see links between the measurable outcomes in Table 1 and Goals 1-4 of Table 2, but it is not clear where Goal 5 in Table 2 links with Table 1.

RESPONSE: We have carefully reviewed our Tables in light of your suggestions. Page 17, Table 2 has been revised to better link evaluation aims, with CHW inputs, a process indicator and the outcomes indicator.

Table 2. Selected Health Start Program intervention activities (non-exhaustive) performed by Health Start Program CHWs and hypothesized client actions (indicators) evaluated via measurable aims			
Evaluation Aims	*CHW Input*	*Process Indicator*	*Outcomes Indicator*

2b. I can see links between the Outcomes Indicators in Table 3 and Measurable Outcomes in Table 1, except for "method of delivery" and "inter-pregnancy intervals".

RESPONSE: We have attempted to clarify our evaluation aims and better link them to our measurable outcomes (available to measure via administrative data) and articulate the programmatic goals and outcomes linked to "method of delivery" and "inter-pregnancy intervals".

Table 1. Description of Health Start Program goals, evaluation aims and measurable outcomes available through administrative data sources		
Program Goals	*Evaluation Aims*	*Measurable Outcomes*

3. The last paragraph of the Background and Rationale describes safety actions like installing car seats and helping with the safe sleep environment. Are there examples of the strategies targeting child development?

RESPONSE: No these are not relevant to the outcomes of childhood development. This sentence has been removed.

4. The authors describe costs of low birth-weight babies, and also of disseminating findings with Arizona Medicaid. Have the authors considered conducting a cost-evaluation to determine the cost-benefits of the home visiting program? This would help health services (across and between states, and internationally) decide which programs to trial or implement.

RESPONSE: Thank you for this suggestion. We agree that a cost benefit analysis of the Health Start program is much needed and upon completion of the outcomes evaluation we hope to engage with a health economist to complete this important investigation.

5. "Data collection" and "exclusion criteria" sections indicate that the datasets have already been established. Is this correct and have any of the analyses been performed?

RESPONSE: We have shifted the tense of the manuscript to future tense. We have removed 'data collection' as this is a secondary data analysis of administrative data. We have replaced 'exclusion criteria' to Synthetic Comparison Group, which is generated through the propensity score matching process.

Page 7, Par 4

Synthetic Comparison Group
A comparison group of women not exposed to the Health Start Program (non-HSP) will be created using a propensity score matching approach and all other births that occurred in Arizona (derived from VRBD) over the study period to balance representation of subjects in each group.

6. Missing data: As data come from administrative datasets, the authors say that missing data are not anticipated. As a reader who is unfamiliar with the datasets being analysed, does this mean that data are available for every eligible child across all datasets and that none are missing for any child? If so, how do the various services make this possible?

RESPONSE: Thank you for this clarification. This is Administrative data for the use of monitoring public health purposes and as such it encompasses all births within Arizona over the study time frame therefore missing data will be trivial number of instances due to human error in record. During our inspection of the data, we found that the relevant information is missing in only a few instances for both mothers and children. We have clarified this issue in text here:

Page 9, Par 5

Statistical Analysis

Both the HSP enrollment information and VRBD are administrative data sources, established and maintained for public health monitoring purposes. As such, we do not anticipate missing data to be a significant issue. We assume that such instances (as we find them) are very likely to be the result of human error and not any systematic issues with the data collection and/or reporting processes. Where missing-ness does occur in the variables that make up the propensity score model, we will control for these using dummy variables in place of the missing observations. In the case of missing outcome variables, we will restrict the analytic sample to the non-missing observations, and inspect to control variables to verify that there are no systematic differences.

7. Could the authors please define what the specific social risk factors are relating to "marital status, living situation, race and ethnicity, education level, income, and insurance type, … maternal BMI, and maternal age." For example, "marital status" on its own is a demographic rather than a risk factor. Is the risk factor "single parent family" or similar?

RESPONSE: Thank you, we have clarified social and medical risks.

Page 6, Par 1

Health Start Program Intervention

Women are eligible to enroll in HSP if they 1) live in the targeted service area, 2) are pregnant or postpartum with a child under age two, and 3) have one or more social or medical risk factors. Social risks can include but are not limited to: single-parent status, underserved racial or ethnic group, education equal to or less than high school level, income less than $40,000, and Medicaid or no insurance. Medical risks are broad and can include previous preterm birth, low birthweight, chronic disease, high maternal BMI, and substance use. Women can be of any age and there are no income requirements to participate.

8. What are the potential limitations of the study?

RESPONSE: Thank you we appreciate this request and apologize for this major oversight. We have now included a Limitations section, which was somehow deleted from the submitted version.

Page 10, Par 3

Limitations
The primary limitation is the identifying assumption that selection into the HSP is driven by observable characteristics. This is a limitation common to most PSM analyses. Attenuation bias is a possibility, to the extent that HSP mothers are incorrectly identified and linked to state birth certificate data. However, the effect of this would be to underestimate (in absolute value) the magnitude of the resulting coefficients, meaning the true effect is likely to be larger (ceteris parabis). In addition, the analysis may have limited external validity for populations who differ along socioeconomic status, race, and ethnicity.

9. Could the authors please provide more information on the following for readers who may not be familiar with these terms:

RESPONSE: 9a. “universe” of Health Start/births – **we have clarified this to mean ‘all’ Health Start births**

RESPONSE: 9a. “honest broker” process – **we have clarified honest broker.**

Page 8, Par 5

Data management

We established an honest broker process to securely house the four datasets that will be accessed for this study. We designated the Center for Biomedical Informatics and Biostatistics’ Biomedical Informatics Services at the University of Arizona as the honest broker to facilitate the de-identification, transfer, and management of data, as well as maintain protected health information anonymization and HIPAA-compliance. In this role, the honest

broker can identify individuals overlapping between relevant databases, and assign de-identified study codes that would enable cross-linking individuals between the systems.

10. Why do the authors select a HOMVee rating of “moderate” over “high” (i.e. is it because of the study design), and what is the impact of this if any on MIECHV funding and ongoing implementation?

RESPONSE: Thank you for this comment. We employ a matched comparison group design study that meets the published standard for HomVEE’s ‘Moderate’ rating (note: ‘High’ rating reserved for randomized controlled trials). We have clarified this in the Objectives section. MIECHV is a separate entity to HOMVee and has no bearing on the outcomes of the study. Health Start is funded by the state and is not reliant on MIECHV funds to operate. We will not clarify this technical issue in the protocol.

Page 5, Par 3

Objectives

Broadly, the goal for the study is to meet the federal Home Visiting Evidence of Effectiveness (HomVEE) standard for evidence-based effectiveness. We will use a matched comparison group design study that meets the published standard for HomVEE’s ‘Moderate’ rating (note: ‘High’ rating is reserved for randomized controlled trials).

REVIEWER #2

Overall, the study topic is important and will be of interest to many readers. My impression of the protocol paper is that the science is sound, but the information is not presented in a way that readers will be able to understand very easily. I have therefore suggested the authors revise their paper to improve clarity and readability.

RESPONSE: We appreciate your careful review and suggestions for readability and flow. We have attempted to copy edit the manuscript as per your suggestions.

1. Introduction

The introduction is a little disorganized - the authors state their aim twice (once in the last sentence of the first paragraph and once in the last sentence of the third paragraph) but then continue providing background information after that.

RESPONSE: We have revised the introduction for organization and flow.

Background

Over the last decade, the community health worker (CHW) workforce has been recognized by the World Health Organization and several US entities as an evidence-based approach to address health disparities (1-3). In the US, the CHW workforce has gained recognition and visibility, as evidenced by the creation of a US Department of Labor Standard Occupational Classification (21-094) in 2010, to include CHWs as a health profession in the Patient Protection and Affordable Care Act (ACA)(4). According to the American Public Health Association, a CHW is defined as: A frontline public health worker who is a trusted member of and/or has an unusually close understanding of the community served. This trusting relationship enables the worker to serve as a liaison/link/intermediary between health/social services and the community to facilitate access to services and improve the quality and cultural competence of service delivery (5).

Emerging evidence suggests CHWs delivering preventive maternal and child health (MCH) interventions through home visiting improve several important maternal and child outcomes (6, 7). Globally, CHW home visiting interventions are associated with several primary prevention efforts that promote the initiation of any, early, and adequate prenatal care (8, 9), initiation of any and exclusive breastfeeding (7, 10-13), reduction of maternal morbidity and perinatal mortality (14), and the uptake and completion of childhood immunizations (7, 15). In the US, CHW home visiting interventions are associated with decreased incidence of preterm birth (9, 16-18) and low birth weight (9, 16-22), and increases in up-to-date immunizations among newborns and toddlers (23). CHWs share the language, socioeconomic status and life experiences of the community members they serve and are recognized as integral to reducing health inequalities among disenfranchised groups (24). Moreover, CHWs are recognized as integral contributors in collaborative health- and community-based teams and in providing comprehensive care, including attention to the social determinants of health that contribute to health improvements and cost savings (25, 26).

Arizona launched the first iteration of the Health Start Program (HSP) in 1984, when Arizona ranked among the lowest five states for the number of women receiving any or adequate prenatal care (27). HSP is a statewide program that employs CHWs to engage at-risk, low income and racially and ethnically diverse mothers to improve maternal and child outcomes. HSP has been managed by the Arizona Department of Health Services (ADHS), Bureau of Women's and Children's Health since 1992 (28). In 1994, the Arizona State Legislature passed the Arizona Children and Families Stability Act, A.R.S. § 36-697, which formalized and expanded HSP and articulated the purpose, requirements and administration of the program. HSP is a community-based outreach program that identifies, screens and enrolls pregnant women early in their pregnancies and assists them with obtaining early and consistent prenatal care. The program also provides prenatal and postpartum education, information and referral services, client advocacy, and emphasizes timely immunizations and developmental assessments for their children. Since its inception, Arizona Health Start Program's mission has been "to educate, support and advocate for families at risk by promoting optimal use of community-based family health care services and education services through the use of community health workers (CHWs) who live in and reflect the ethnic, cultural and socioeconomic characteristics of the community they serve." (28)

Study Setting

Arizona is the sixth largest state in the nation, with a population of 6.8 million people. The state shares an international border with Mexico and is home to 21 federally recognized American Indian Tribes and Nations, making it uniquely racially and ethnically diverse. Arizona has a higher proportion of Latino (30.9%) and American Indian (5%) residents compared to the nation (17.8% and 1%, respectively) and a comparatively smaller proportion of African American residents (5% compared to 13% nationally) (29).

In 2015, nearly a quarter of the population lived in rural areas, where the poverty rate reached 30%, nearly double that of the national poverty rate (29). Approximately 20% of Arizona families with children live below the federal poverty line, compared to 18% nationally. Poverty disparately effects Arizona's Latino (36%) and American Indian (46%) families and children (29). Arizona ranks as the fifth highest US state for adult female poverty rate in the country, with more than one quarter of Arizona families headed by single-mother households (29). The

initial framework for the HSP was developed in the 1980s and 1990s to address the social determinants associated with the steady decrease in the rate of women receiving early or any prenatal care. In the most recent Arizona Title V Maternal and Child Health Needs Assessment (2017), approximately 74% of pregnant women initiated prenatal care in the first trimester (compared to 61% in 2015 and 81% in 2013), and 7.9% had no prenatal care (29). There were disparities among mothers by race/ethnicity who received prenatal care, notably American Indian mothers having the highest rates of 'inadequate' prenatal care (25%) compared to all women in Arizona (15%) (29).

It is widely recognized that late prenatal care is associated with preterm birth, low birthweight, and infant mortality. In 2014, 9% of babies born in Arizona were premature and 7.2% were low birthweight (29). Historically, low-income mothers have experienced higher rates of premature birth and low birthweight in Arizona (30) and nationally (31). There are also apparent racial disparities for birth outcomes in Arizona. Preterm birth rates are highest among Black (12.2%), American Indian (9.4%), and Latino (9.2%) compared to all preterm births (9.1%) in the state. Preterm births increase the risk of low birthweight; similar trends persist with the highest rates of low birthweight among Black residents (10.32%) compared to White residents (5.36%) and the total Arizona population (7.2%) (29). Preterm and low birthweight baby delivery costs have been shown to be 25 times more than uncomplicated newborn deliveries (32). Although prenatal care and birth outcomes in Arizona have improved over the years, many under-resourced women continue to experience significant challenges and barriers to obtaining health care services.

2. The introduction could also be streamlined somewhat. For example, the last paragraph about HSP could be incorporated into the Methods section where the authors describe the intervention in more detail. A better place to state the aim of the paper is at the end of the introduction or in the objectives section.

RESPONSE: We have revised the introduction for organization and flow. We have incorporated the last paragraph into the Methods section as the reviewer suggests.

3. Methods: The language in the methods section is so cluttered with details that I found it difficult to decipher what the authors plan to do. For example, in the Primary Outcome paragraph, the authors state their study design, study timeframe and study setting (again) without actually saying what the study outcomes are. The reader is directed to find them in a table instead.

RESPONSE: We have revised the Methods Section and specifically the Primary Outcomes section to increase flow and readability.

Page 8, Par 4
Primary Outcomes
HSP is a primary prevention intervention to improve maternal and child health outcomes among at-risk, racially and ethnically diverse, rural and urban mothers and children of Arizona. We will use four Arizona Department of Health Services administrative datasets to evaluate Specific Aims 1-3 including Health Start programmatic data, Vital Records Birth Data, Hospital Discharge Data, and Arizona State Immunization Information System data. Aim 1 (HSP impact on newborn health) will be measured by preterm birth, birthweight, and newborn hospital length of stay and associated charges. Aim 2 (HSP impact on maternal health) will be measured by prenatal care initiation and frequency, method of delivery, maternal morbidities, and inter-pregnancy intervals. Aim 3 (HSP impact on child health) will

be measured by uptake of age-appropriate immunizations, and emergency room and inpatient encounters and charges (Table 3).

4. Another example – in the stats analysis section, the authors could be more clear and upfront in stating the approach they plan to use and what kind of results this approach will produce, rather than explaining extensively why they are not choosing a different approach and what the specific advantages and disadvantages of their chosen approach are.

RESPONSE: We have reviewed and revised our statistical analysis to improve readability and flow.

Page 9, Par 2

Statistical Analysis
The motivation for using PSM to create a synthetic comparison group is to be able to "observe" the "counterfactual" to HSP participation, that is, what would have happened in the absence of the HSP program. This will be done by comparing outcomes between HSP mothers and those "matched" to them by the propensity score. More specifically, the average treatment effect (ATE) generated by PSM will estimate the impact of the program on the population of both HSP mothers and those who "look like" HSP mothers by taking the difference in outcomes between HSP mothers and their matches, and vice-versa.
Our analytic population is of sufficient size to detect meaningful program effects from low-frequency events, including preterm births, low and very low birthweights, maternal morbidity, and differences in immunization and hospitalization rates over a relatively long period of time. This is also true for specific subgroups served by HSP (e.g. Hispanics, Native Americans, economically disadvantaged).
Once we establish proper covariate balance between the intervention and matched-control groups, point estimates of the treatment effects will be estimated by comparing outcomes using Stata version 14 software and specifically the teffects command (45). Following Abadie and Imbens (46, 47), this command considers the fact that propensity scores (i.e. the parameter that determines the comparison population) are estimated when calculating the standard errors and thus generates confidence intervals. The propensity scores will not be used as a covariate in traditional regression analysis because it is less effective in forcing baseline equivalence and assumes the relationship between the score and the outcome is linear (42).

5. Throughout the methods section, although much of the information presented is important, it is not presented in a way that most readers would find accessible. The authors should consider using another BMJ Open Protocol paper (or several) as a guide in organizing their manuscript and in assessing how much detail to provide and where.

RESPONSE: We appreciate this suggestion. We reviewed multiple BMJ Open protocol papers and reorganized our sections for better overall flow, comparable to other protocol papers published by this journal. New organization/headers include: Background, Objectives, Methods (intervention, participants, outcomes), Methods (data management, monitoring, statistical analysis), Discussion, Limitations.

6. Some of the language used is problematic, for example:
    a. "Data Collection" isn't really an applicable term for what you are describing in this paragraph. I think you mean "Study Cohort Development" or "Identification of Study Population".

RESPONSE: 'Data collection' section changed to "Intervention Cohort"

b. You use the term "matching" in the above-mentioned paragraph when what I think you mean is "linking".

RESPONSE: Changed 'matching' to 'linking' in this section describing HSP-VRBD link.

c. One doesn't really "engage a strategy", as the author plan to do in the Dissemination paragraph.

RESPONSE: Agreed. Changed 'engage' to 'initiate'

7. Figure 2 – The authors should use a PRISMA flowchart to depict how they identified the study population. http://prisma-statement.org/PRISMAStatement/FlowDiagram.aspx

RESPONSE: Agreed. Developed Figure 2: Study Flow Chart guided by PRISMA.

8. The authors may want to consider adding a section on study limitations.

RESPONSE: Thank you we appreciate this request and apologize for this major oversight. We have now included a Limitations section, which was somehow deleted from the submitted version.

Page 10, Par 3

Limitations
The primary limitation is the identifying assumption that selection into the HSP is driven by observable characteristics. This is a limitation common to most PSM analyses. Attenuation bias is a possibility, to the extent that HSP mothers are incorrectly identified and linked to state birth certificate data. However, the effect of this would be to underestimate (in absolute value) the magnitude of the resulting coefficients, meaning the true effect is likely to be larger (ceteris parabis). In addition, the analysis may have limited external validity for populations who differ along socioeconomic status, race, and ethnicity.

9. I suggest the authors be more consistent with their abbreviations (e.g. HSP, Health Start, Health Start Program, Arizona Health Start Program). I would be very selective with what you choose to abbreviate as the reader will not be as familiar as you are with abbreviations specific to your study.

RESPONSE: Thank you. We have reviewed for consistency and used Health Start Program or HSP throughout.

10. I suggest the authors have a copyeditor read the manuscript carefully before resubmission. I noted a number of missing sentence articles and instances of misplaced punctuation marks.

RESPONSE: We have engaged a copy editor to review for these grammatical errors.

11. As a post-script: I have found this resource very helpful in writing papers – Schimel, J. 2012. Writing Science: How to Write Papers That Get Cited and Proposals That Get Funded (1st Edition). Oxford, UK: Oxford University Press. (I have no affiliation whatsoever with this author or his publishing firm)

RESPONSE: Thank you for this incredible resource. We have reviewed it and have attempted to apply lessons learned.

VERSION 2 – REVIEW

REVIEWER	Anna Price Centre for Community Child Health, Murdoch Children's Research Institute, The Royal Children's Hospital University of Melbourne.
REVIEW RETURNED	02-Oct-2019

GENERAL COMMENTS	Thanks for the thoughtful and considered revision of this protocol. I have only minor suggestions for revision: 1. In the abstract, what does the term 'unique mothers' mean? Can 'mothers' be used instead? If not, please define the term. (It makes more sense in the context of the methods but less so in the abstract which is often the first thing a reader will read). 2. Sometimes you refer to 'aim' and other times to 'specific aim'. I think using 'aim' is clear as only three are listed as the study objectives. 3. Do you define the HomVEE 'moderate' rating? Apologies if I missed it. If not, it would be a helpful addition. 4. What is an 'asset-based' approach? It would help if this is defined for readers unfamiliar with the term. 5. Regarding CHW competencies, what characteristics does the 'background check' refer to? 6. How were the indicator variables for matching selected? Are the risk factors chosen (e.g. income, education etc) all possible indicators available in the datasets, or the ones the authors considered the best ones? Why is hypertension chosen?

	7. Given the large sample size, it is possible that there will be statistical evidence of differences between groups according to the confidence intervals and p-values. How will you balance these findings with the clinical significance of group differences? And is this a limitation to consider, or something to manage in the analysis? 8. Table 1 and 2 could be combined since the Evaluation Aims and Outcomes are common between them. This would provide a great, informative table. Finally, there are a handful of typos or missing linking words, so a final proof read could help with the final polish. Good luck!

REVIEWER	Jennifer Enns University of Manitoba Canada
REVIEW RETURNED	30-Sep-2019

GENERAL COMMENTS	The paper is much improved in terms of organization and clarity. Just a couple of additional points: Please make it clear early in the abstract that this is a protocol paper for an observational study, not a paper on the observational study itself. Please check the manuscript carefully for spelling errors and typos. I noted several that probably occurred during your editing process, but could easily have been picked up with Microsoft Word spell check before resubmission.

VERSION 2 – AUTHOR RESPONSE

REVIEWER #1

Thanks for the thoughtful and considered revision of this protocol. I have only minor suggestions for revision:

1. In the abstract, what does the term 'unique mothers' mean? Can 'mothers' be used instead? If not, please define the term. (It makes more sense in the context of the methods but less so in the abstract which is often the first thing a reader will read).

   'Unique mothers' was changed to 'mothers' throughout the paper

2. Sometimes you refer to 'aim' and other times to 'specific aim'. I think using 'aim' is clear as only three are listed as the study objectives.

   'Specific aim' was changed to 'aim' throughout the paper

3. Do you define the HomVEE 'moderate' rating? Apologies if I missed it. If not, it would be a helpful addition.

   ***Page 5: Objectives**: Our goal is to describe the research protocol for a retrospective comparative evaluation to assess the impact of Arizona's Health Start Program, a CHW home visiting perinatal support program, on multiple maternal, infant, and child health outcomes. Broadly, the goal for the study is to meet the federal Home Visiting Evidence of Effectiveness (HomVEE) standard for evidence-based effectiveness. We will use a matched comparison group design that meets the published standard for HomVEE's 'Moderate' rating, defined as:*

"1) baseline equivalence established on tested outcomes and demographic characteristics and controls for baseline measures of tested outcomes, if applicable; and 2) no confounding factors; must have at least 2 participants in each study arm and no systematic differences in data collection methods" by HomVEE (note: 'High' rating is reserved for randomized controlled trials) (33).

4. What is an 'asset-based' approach? It would help if this is defined for readers unfamiliar with the term.

   Page 6: Methods; Health Start Program Intervention

   HSP CHW home visits are guided by an asset-based approach (34) and two primary theories of behavior change, the Trans Theoretical Model (35) and the Social Cognitive Theory (36). Identifying assets acknowledges and supports the existing strengths and capabilities of individuals and resources to promote community-driven development and positive change.

5. Regarding CHW competencies, what characteristics does the 'background check' refer to?

   Page 6: Methods; Health Start Program CHW Core Competencies, Roles, & Training

   According to the HSP policy and procedure manual, CHWs must 1) live and work in the service area, 2) reflect the ethnic, cultural and socioeconomic characteristics of the communities they serve, 3) be able to read and write in English, 4) have a high school diploma or General Educational Development, and 5) pass a criminal history background check within the Department of Public Safety records to be eligible to work for the state-funded program.

6. How were the indicator variables for matching selected? Are the risk factors chosen (e.g. income, education etc) all possible indicators available in the datasets, or the ones the authors considered the best ones? Why is hypertension chosen?

   The Indicator variables selected for the matching model were chosen to capture the socio-economic risks and geographic diversity that characterize the population served by Health Start. Because the birth certificates are the single data source for information on non-Health Start mothers, the variable selection was limited to 1) information collected in the certificates, which is focused primarily on the circumstances of the birth and health of the child, and 2) measures that were consistent across the change in birth certificate forms that happened in 2014. Under these constraints, the variables chosen represent what we consider to be the best available for describing the Health Start population. With this in mind, pre-existing hypertension is something of an outlier, given that it is a relatively rare condition and is not predictive of Health Start participation. It is ultimately included in the matching model because we considered it important to include a measure of the mother's health, and it is the only one that is reported and measured consistently between the two birth certificate forms.

7. Given the large sample size, it is possible that there will be statistical evidence of differences between groups according to the confidence intervals and p-values. How will you balance these findings with the clinical significance of group differences? And is this a limitation to consider, or something to manage in the analysis?

   Thank you, we have reviewed your questions and may not completely understand. Therefore, we cannot answer at this time.

8. Table 1 and 2 could be combined since the Evaluation Aims and Outcomes are common between them. This would provide a great, informative table.

   ***Page 16:** Tables 1 & 2 are combined.*

9. Finally, there are a handful of typos or missing linking words, so a final proof read could help with the final polish. Good luck!

Spellcheck completed.

REVIEWER #2

1. Please make it clear early in the abstract that this is a protocol paper for an observational study, not a paper on the observational study itself.

Page 3: Abstract; Methods & Analysis: *This paper outlines our protocol to retrospectively evaluate Health Start Program administrative data for 15,576 mothers enrolled in the program between 2006 and 2015. We will use propensity score matching to generate a statistically similar synthetic control group.*

2. Please check the manuscript carefully for spelling errors and typos. I noted several that probably occurred during your editing process, but could easily have been picked up with Microsoft Word spell check before resubmission.

Spellcheck completed.

VERSION 3 – REVIEW

REVIEWER	Anna Price Murdoch Children's Research Institute, Australia
REVIEW RETURNED	06-Nov-2019

GENERAL COMMENTS	Thanks for the opportunity to re-review your paper. It's looking great. Apologies if I've missed it, but it would be helpful for the manuscript to include an explanation of why only families enrolled antenatally will be eligible for inclusion. All the best with your study.

VERSION 3 – AUTHOR RESPONSE

Dr. Price, thank you very much for reviewing our paper multiple times and providing insightful comments and suggestions. Your feedback has been invaluable.
We would like to address your most recent comment: "…it would be helpful for the manuscript to include an explanation of why only families enrolled antenatally will be eligible for inclusion."
To clarify, our sample includes all women enrolled within 24 months of birth. Eligibility criteria are women who are pregnant and up to their child's second birthday. Please see our modified narrative below; we hope this provides more clarity.
Intervention Cohort Sample Size
Our evaluation intervention group will include all HSP participants enrolled within 24 months of the date of birth of the child during 2006-2015. Of the initial 15,576 records identified through the HSP-to-VRBD data link, 5,911 fall outside of the 24-month (either before or after) HSP enrollment window and will be excluded from all subsequent analyses. The resulting 9,665 HSP-associated births constitute

the basis of this study (Figure 2). Because HSP participants can enroll before or after birth, we will limit the analysis for Aims 1 and 2 (newborn and maternal health outcomes) to those births for which the mother was enrolled during pregnancy. This final criterion results in 6,493 HSP-attributed births for the evaluation of Aims 1 and 2. Aim 3 (child health outcomes) will be evaluated using the larger set of 9,665 HSP-associated births.

Synthetic Comparison Group

A comparison group of women not exposed to the Health Start Program (non-HSP) will be created from all births occurring in Arizona during 2006-2015 (derived from VRBD). After identifying our study population we will use propensity score matching (PSM) to generate a statistically-similar synthetic control group that has, on average, the same observable pre-program characteristics as the HSP mothers (41). The pool of potential comparators will come from all Arizona births that occurred over the study period (2006-2015). This process will be guided by HomVEE standards requiring that the covariates used to balance the treatment and control groups be associated with both treatment status and the outcomes of interest (42). Because the HSP eligibility criteria focus on social and medical risks, we will prioritize these types of measures in the PSM model, in addition to characteristics that have been shown to have strong associations with our outcomes of interest in previous empirical and theoretical work.