UCD GEARY INSTITUTE FOR PUBLIC POLICY
DISCUSSION PAPER SERIES

Early intervention and child health: Evidence from a Dublin based randomized controlled trial

Orla Doyle
UCD School of Economics
UCD Geary Institute for Public Policy,
University College Dublin

Nick Fitzpatrick
UCD Geary Institute for Public Policy,
University College Dublin

Judy Lovett
UCD Geary Institute for Public Policy,
University College Dublin

Caroline Rawdon
UCD School of Psychology,
University College Dublin

Geary WP2015/05
April, 2015

Early intervention and child health: Evidence from a Dublin-based randomized controlled trial*

Orla Doyle[a], Nick Fitzpatrick[b], Judy Lovett[c], Caroline Rawdon[d]

[a] UCD School of Economics and UCD Geary Institute for Public Policy, University College Dublin, Dublin, Ireland. E-mail: orla.doyle@ucd.ie

[b] UCD Geary Institute for Public Policy, University College Dublin, Dublin, Ireland. E-mail: nick.fitzpatrick@ucd.ie

[c] UCD Geary Institute for Public Policy, University College Dublin, Dublin, Ireland. E-mail: judy.lovett@ucd.ie

[c] UCD School of Psychology, University College Dublin, Dublin, Ireland. E-mail: caroline.rawdon@ucd.ie

Abstract

This article investigates the impact of an early intervention program, which experimentally modifies the parenting and home environment of disadvantaged families, on child health in the first 3 years of life. We recruited and randomized 233 (115 intervention, 118 control) pregnant women from a socioeconomically disadvantaged community in Dublin, Ireland into an intervention or control group. The treatment includes regular home visits commencing antenatally and an additional parenting course commencing at 2 years. Maternal reports of child health are assessed at 6, 12, 18, 24, and 36 months. Treatment effects are estimated using permutation testing to account for small sample size, inverse probability weighting to account for differential attrition, and the stepdown procedure to account for multiple hypothesis testing. Following adjustment for multiple testing and attrition, we observe a positive and statistically significant main treatment effect for wheezing/asthma. The intervention group are 15.5 percentage points (pp) less likely to require medical attention for wheezing/asthma compared to the control group. Statistically significant individual main effects which do not survive multiple testing and IPW-adjustment are found for general health (10.0 pp), hospitalizations (8.2 pp), immunizations (8.6 pp), chest infections (12.2 pp) and the number of health problems (d = 0.34). Subgroup analysis reveals more statistically significant adjusted treatment effects for boys than girls regarding fewer health problems (d = 0.63), accidents (23.9 pp), and chest infections (22.8 – 37.9 pp). Our results suggest that a community-based home visiting program may have favorable impacts on early health conditions. As child ill health is costly to society due to an increased demand on health resources and long-term productivity losses, identifying effective interventions to counteract inequalities in health is important from a policy perspective.

Keywords: Randomized controlled trial, home visiting, child health, early intervention

JEL: C12, C93, J13, I14

* Corresponding author: Dr. Orla Doyle, UCD Geary Institute for Public Policy, University College Dublin, Belfield, Dublin 4, Ireland. Tel: +353 17164637, Fax: +353 17161108. E-mail: orla.doyle@ucd.ie.

Introduction

A steep socioeconomic gradient in adverse health during the early years has been well documented (e.g. Case et al., 2002). Yet there is growing evidence that intervening early in the lifecycle, through targeted home visiting programs for example, is a potential mechanism for reducing this gradient (Avellar and Supplee, 2013). Children facing socioeconomic disadvantage often experience poor health outcomes regarding the prevalence and severity of illness, the incidence of disease, and the likelihood of mortality (Chen et al., 2002). They are also at increased risk of developing a number of preventable diseases later in life such as heart disease, diabetes, respiratory infections, and obesity (Galobardes et al., 2004; Komro et al., 2011; German and Latkin, 2012). Poor health during childhood has also been associated with adverse educational and labor market outcomes (Case et al., 2005; Currie, 2004; Currie and Hyson, 1999).

The child health gradient may be attributed to genetic, psychological, and behavioral factors, as well as the direct effect of parental resources (Anderson and Armstead, 1995; Smith, 1999). Income, as a primary resource, may affect the quality and quantity of health care provided, as parents with higher incomes can purchase or produce inputs such as nutritious meals, frequent doctor visits, and provide a safe and stimulating home environment (Mayer, 2002). Furthermore, mothers who have attained higher education may combine health inputs more efficiently (Grossman, 1972; Rosenzweig and Schultz, 1982), such as engaging in preventative care and changing health behaviors during pregnancy.

Yet identifying the causal pathways through which socioeconomic status is related to child health is limited by endogeneity, whereby family circumstances and child health are driven by some common unobserved factor, or reverse causality, whereby child illness negatively impacts on parental resources. To overcome these issues this study utilizes random assignment, which experimentally modifies the parenting and home environment of disadvantaged families, to investigate a mechanism for ameliorating poor health. Ill health during childhood is costly to society in the short run, in terms of increased demand for health resources, and in the long run, in terms of losses in economic productivity. Thus, identifying effective interventions to reduce inequalities in health by counteracting the socioeconomic risks associated with low family income and education is a key goal for policymakers (Marmot, 2005).

A growing body of evidence demonstrates that early intervention can reduce health inequalities and promote health in adulthood (e.g., see Campbell et al., 2014). Early intervention is considered both biologically and economically efficient as development is more malleable early in life (Halfon et al., 2001), thus investments made in this period are likely to generate larger returns than later remedial interventions (Cunha and Heckman, 2007; Heckman, 2007). In particular, given the importance of the fetal environment and maternal behavior during pregnancy on later childhood health, interventions commencing during pregnancy should yield the highest returns (Doyle et al., 2009).

Home visiting programs (HVPs) are one form of intervention which target disadvantaged families in the first years of their children's lives in order to improve health and development (Sweet and Appelbaum, 2004). In general, HVPs provide parents with information, direct instruction on parenting practices, emotional support, and access to community services (Howard and Brooks-Gunn, 2009). They operate through regular home visits provided by trained workers, either professionals such as nurses or child development specialists, or paraprofessionals, such as mentors. HVPs may improve child health by promoting immunization uptake and appropriate care for illnesses, and reducing preventable injuries. Systematic reviews of the effectiveness of HVPs using experimental designs have identified some positive effects on child health, yet the evidence is mixed. For example, Avellar and Supplee (2013) report that five of twelve HVPs identify favorable and significant effects on health care coverage or use, including well-child visits and dental service use, while five of six programs reduce child maltreatment. Another review by Peacock et al. (2013) finds that two of seven HVPs have statistically significant effects on physical growth, including improved birth weight and catch-up growth, and two of six programs have an impact on hospitalizations, illnesses, and injuries, while one study reporting on immunizations also has a positive effect.

However, the existing literature is somewhat limited by the type of methods used to estimate treatment effects. While some experimental HVP studies are derived from large samples, others are constrained by small sample sizes yet utilize large sample test statistics. In addition, many studies estimate treatment effects across multiple health outcomes while failing to adjust for Type-I errors. Attrition is also a common concern in longitudinal trials, and while some studies test for differential attrition, few adequately

account for its effect on treatment outcomes. This article investigates the impact of *Preparing for Life* (*PFL*), a community-based HVP in Ireland, on children's health within the first 3 years of life utilizing methods which counteract common issues in randomized controlled trials (RCTs) including accounting for small sample inference, differential attrition, and multiple hypothesis testing. Specifically, we investigate program impact on children's general health, number of health problems, hospital stays, accidents, immunizations, wheezing/asthma, and chest infections at multiple time points. As early intervention programs often find differential treatment effects by gender (e.g. Anderson, 2008; Eckenrode et al., 2010; Heckman et al., 2010), we conduct a subgroup analysis for boys and girls separately. We also assess the internal validity of the findings by testing for the presence of contamination and differential misreporting.

2 Methods

2.1 Treatment and Setting[1]

This study is a RCT of the Preparing for Life (*PFL)* program. The study enrolled pregnant women from a community in Dublin, Ireland that had above national average rates of unemployment, early school leaving, lone parent households, and public housing (Doyle, 2013). The inclusion criteria included all pregnant women living in the catchment area, regardless of parity. There were no exclusion criteria. Participation was voluntary and recruitment took place between 2008 and 2010 through two maternity hospitals or self-referral in the community. After informed consent was obtained, a computerized unconditional probability randomization procedure assigned 115 participants to an intervention group and 118 to a control group. No stratification or block techniques were used.

PFL is a community-based home visiting program (HVP) which aims to improve children's health and development by intervening during pregnancy and working with families until the children start school at age 4/5 years. *PFL* prescribes twice monthly home visits, lasting approximately one hour, delivered by mentors from a cross-section of professional backgrounds including education, social care, and youth studies. The average number of visits delivered to the intervention group between program entry

[1] The trial was registered with controlled-trials.com (ISRCTN04631728) and was conducted and reported in conformity with CONSORT guidelines. All study procedures were approved by the university and maternity hospitals' respective ethics committees. All participants gave informed consent before taking part in the randomization process.

and 36 months was 51 (SD = 21), which represents 57.8% of prescribed visits and is consistent with other HVPs (Gomby et al., 1999). Thus the majority of participants receive monthly visits and some fortnightly. Mentors received extensive training prior to program implementation and monthly supervision thereafter to ensure fidelity to the program model.[2] Each family is assigned the same mentor over the course of the intervention where possible. The mentors use role modelling, demonstration, coaching, discussion, encouragement, and feedback to deliver the intervention, as well as direct interactions with the child in the presence of the parent. The aim of the visits are to support and educate the parents on key child rearing issues including the identification of developmental milestones and appropriate parenting practices that promote the children's health, and cognitive and non-cognitive development. Each visit is guided by a set of *PFL*-developed 'Tip Sheets' which are based on pre-existing governmental and local non-governmental organizations' recommendations, and present best-practice information on pregnancy, parenting, and child health and development (see Appendix A for an example of a Tip Sheet).[3] There are three sets of age-specific Tip Sheets - pre-birth-12 months, 1-2 years, and 2-4 years. The mentors can choose when to deliver the Tip Sheets within these specific time periods based on the age of the child and the needs of the family. The Tip Sheets are given to the participants at the end of each visit to keep as an on-going resource.

This study refers to the impact of the intervention on child health between program entry and 3 years and a number of Tip Sheets delivered during this period encouraged awareness of child health and are directly related to the outcomes assessed in this article. For example, a Tip Sheet on *immunizing* gives a full immunization schedule from birth to 13 months, while the Tip Sheet on *childhood illnesses* contains information on common childhood illnesses (e.g. fever, croup, ear infections) and caring for a sick child. There are also Tip Sheets on *keeping baby safe* and *kid safe rooms* including information on making the home secure for a child and a room-by-room checklist to ensure a secure environment. In addition, a Tip Sheet on *passive smoking* alerts participants to the risk of exposure to smoke and how to protect children from passive smoking.

[2] Supervision is based on the model commonly used by social workers and is provided two hours per month. Key areas addressed include areas such as participant work, team work, support, administration, and training/development.

[3] There are approximately 150 *PFL* Tip Sheets over the course of the program.

The intervention group are also invited to participate in an additional parenting course (Triple P Positive Parenting Program; Sanders et al., 2003) when their children are between 2 and 3 years old i.e., after they have completed the 24-month assessment. Triple P promotes healthy parenting practices and positive parent-child attachment. Meta-analysis of Triple P has demonstrated positive effects for parents regarding improved parenting practices and for children regarding improved social, emotional, and behavioral outcomes (Sanders et al., 2014). 62.1% of the intervention participants who completed the 36 month assessment took part in some form of Triple P, with the majority availing of Group Triple P which consists of 5 two-hour group discussion sessions and 3 individual phone calls facilitated by the mentors.

While the HVP is the intervention under investigation, it should be noted that both the intervention and control group receive some common supports including developmental materials and book packs. The developmental packs consist of materials such as a baby gym, food utensils, safety items and an assortment of developmental toys. Both groups are also encouraged by letter, mobile phone text message, and Facebook notices to attend public health workshops on stress management and healthy eating which are already taking place in the community. The control group also has access to a support worker who can help them avail of community services if needed, while this function is provided by the mentors for the intervention group. Note that the control group do not receive the HVP, Tip Sheets, or the additional parenting course.[4] Further information on the program and the design of the evaluation has been published elsewhere (Doyle, 2013).

2.2 Data collection and Variables

All interviews are conducted on tablet laptops by trained interviewers who are blind to participants' treatment status; although consistent with other non-clinical interventions it is not possible to blind participants. Participants can choose to complete the

[4] Care as usual, which is available to all pregnant women and infants in Ireland, is as follows: Expectant mothers are provided with an initial family doctor (G.P.)/obstetrician appointment at 12 weeks and a further 5 examinations for first time mothers and 6 for subsequent pregnancies. Antenatal classes are provided by local public maternity hospitals free of charge. Following birth, a G.P. examination is carried out for the baby at 2 weeks and mother and baby at 6 weeks. The mother is entitled to free in-patient, out-patient and accident and emergency/casualty services in public hospitals in respect of the pregnancy and the birth and is not liable for any hospital charges. In addition, checks by a public health nurse are generally carried out in the home in the weeks after birth and when the infant is 9, 18, and 24 months, but they are not mandatory. A schedule of immunizations in provided free of charge at birth, 2, 4, 6, 12, and 13 months.

interviews in their home or in a local community centre. Each participant is given a €20 (~$21) shopping voucher on completion of each interview. Child health is assessed at 6, 12, 18, 24, and 36 months.[5] The measures assessed are based on areas of child health which the mentors specifically target as part of the program through attempting to change the preventative health care behavior of the parents. A limitation of this study is the reliance on maternal reports of child health rather than hospital/medical records or a formal diagnosis. However, this approach is consistent with much of the HVP literature (e.g. Culp et al., 2007; Kemp et al., 2011), and a HVP study verifying maternal reports using hospital records found no evidence of misreporting regarding episodes of hospitalization (Koniak-Griffin et al., 2003). In addition, parental reports have been deemed acceptable for research purposes (Pless and Pless, 1995), particularly regarding acute health care use for children under the age of 3 (D'Souza-Vazirani et al., 2005). Below, we test for the possibility of differential misreporting across the intervention and control groups. Yet future research should consider the use of medical records in order to verify parental reports.

General health is assessed using maternal ratings of the child's health on a 5-point scale. Binary measures denoting whether the child had good health (good, very good, excellent) or not (poor, fair) are created.[6] The number of health problems experienced by the child is assessed using maternal reports on whether the child required medical attention for any health problems.[7] A discrete measure denoting the total number of health problems experienced is generated. Binary measures of hospital stays are derived using maternal reports on whether the child spent at least 1 night in hospital. Binary measures of accidents are generated from two questions using maternal reports on whether the child had an accident which required medical attention. Binary measures of immunizations are created based on maternal reports on whether the child received the recommended 4 month, 6 month, and 13 month immunizations respectively. Due to the prevalence of respiratory illnesses in young

[5] The first three interviews take place within a 3 month window around the child's birthday, while the last two interviews take place within a 6 month window. On average, the children were 6.3 months, 12.3 months, 18.3 months, 24.6 months, and 37 months when the 6, 12, 18, 24, and 36 month interviews took place respectively.

[6] The original 5 category variable was also tested using an ordered logit regression and the results were the same.

[7] Mothers were asked to select all the applicable health problems from the following list: chest infections, ear infections, feeding problems, sleeping problems, wheezing/asthma, skin problems, sight or eye problems, failure to gain weight or grow, persistent or severe vomiting, persistent or severe diarrhoea, fits or convulsions, excessive crying, accidents, other health problems. If they selected 'other health problems' the interviewer would ask them to specify. The most prevalent health problems at 36 months are listed in Appendix Table B1.

children, separate binary measures of wheezing/asthma and chest infections are generated using maternal reports on whether the child received medical attention due to wheezing/asthma or chest infections. With the exception of immunizations, each health outcome is assessed in relation to the child's health in the previous 6 months for the first four interviews and the previous 12 months for the last interview.[8]

2.3 Empirical Model and Estimation

This study adopts an intention-to-treat approach, regardless of the number of home visits delivered or Triple P attendance. The standard treatment effect framework describes the observed outcome Y_i of participant $i \in I$ by:

$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0) \qquad i \in I = \{1 \dots N\} \tag{1}$$

where $I = \{1 \dots N\}$ denotes the sample space, D_i denotes the treatment assignment for participant i ($D_i = 1$ for the intention-to-treat sample , $D_i = 0$ otherwise) and ($Y_i(0)$, $Y_i(1)$) are potential outcomes for participant i. We test the null hypothesis of no treatment effect on child health outcomes via:

$$Y_i = \beta_0 + \beta_1 D_i + \epsilon_i \tag{2}$$

Equation 2 is estimated via two methods. Firstly, t-tests from OLS regressions on the continuous child health outcomes and chi-squared tests from logistic regressions on the binary outcomes are estimated. Secondly, permutation-based hypothesis testing is used as an alternative method of assessing the statistical significance of the observed treatment effects. Permutation testing is more suitable than standard bivariate tests, such as t-tests, as it does not depend on distributional assumptions and thus facilitates the estimation of treatment effects in small samples (Ludbrook and Dudley, 1998). A number of simulation studies have found that permutation testing has superior power advantages over parametric t tests, particularly if the data are skewed and the degree of skewness is correlated with the size of the treatment effect (e.g. Hayes, 1996; Mewhort, 2005; Keller, 2012). A permutation test relies on the assumption of exchangeability under the null hypothesis. If the null hypothesis is true, which implies that the program

[8] General child health, number of health problems, and hospital stays are assessed at every point in time. Accidents, wheezing/asthma, and chest infections are assessed at 12, 18, 24, and 36 months, and immunizations is assessed at 6, 12, and 18 months.

has no impact, then taking random permutations of the treatment indicator does not change the distribution of outcomes for the intervention or control group.

Permutation tests calculate the observed test statistic that is generated by comparing the mean outcomes of the intervention and control groups. Next, the data are repeatedly shuffled so that the treatment assignment of some participants is switched between the groups. The *p*-value for the permutation test is computed by examining the proportion of permutations that have a test statistic more extreme than the observed test statistic. In this study, we use permutation tests based on 100,000 replications, to estimate the program's impact on child health. We report *p*-values from two-sided tests in order to test the hypothesis that the program may have either a positive or negative effect on health outcomes. For example, the intervention group may have better health than the control group if their parents engage in preventative health care as encouraged by the mentors. However, it is also possible that the intervention group display poorer health, such as more medical visits for health problems, as the parents are more cognisant of potential health issues. Effect sizes are calculated using Cohen's *d* for continuous variables and marginal effects (ME) for binary variables.[9]

2.4 Robustness Checks

Due to differential attrition, the estimation samples at each time point may not be representative of the original randomized sample. This may bias the estimation of treatment effects if the type of participants who drop out of the study or do not complete a particular assessment differ across the intervention and control groups. An inverse probability weighting (IPW) procedure is applied to deal with this issue (see Doyle et al., 2013, for a full description). This involves estimating logistic regression models predicting the probability of completing an interview at each assessment point by modelling attrition as a function of baseline characteristics. Between 8% and 12% of 50 baseline variables predict attrition from the intervention group at any time point based on bivariate tests and the corresponding figure for the control group ranges from 8% to 22%. Thus, there is more evidence of differential attrition in the control group. In

[9] The Cohen's *d* is used to define the strength of a relationship. A Cohen's *d* ranging from 0.0 to 0.2 is deemed a small effect; values ranging from 0.2 to 0.8 represent a medium effect; and values greater than 0.8 illustrate a large effect (Gravetter and Wallnau, 2004). For example, a Cohen's *d* of 0.5 implies that the observed group difference is equal to half of the pooled standard deviation. The marginal effects are calculated based on the average derivative.

general, participants with poorer baseline characteristics are more likely to drop-out, for example, they tend to be younger, have less education, less likely to work, lower self-esteem, poorer parenting skills, and lower IQ.

Given the sample size and the large number of potential covariates, it is not possible to control for all baseline predictors, thus, the Bayesian Information Criterion (BIC; Schwarz, 1978) is used to determine which covariates are included in the logistic models used to generate the IPW weights. The BIC, which measures goodness of fit, is estimated for different combinations of baseline variables while accounting for the number of variables included in the model. First, 50 baseline variables are included in a model of attrition and the BIC is calculated and stored. Next, one variable is excluded and the BIC is calculated and compared to the stored BIC. If the new BIC is more than 2 points smaller than the stored BIC (i.e. a lower BIC indicates a model with greater prediction), the new BIC is stored and the process continues by testing all possible combinations of variables until the optimal set of baseline predictors has been identified. The set of variables which result in the lowest BIC can be found in the Appendix Table B2. The logistic models are calculated separately for the intervention and control groups, at each time point. A similar method is adopted in Campbell et al. (2014).

The probabilities generated from these logistic models are then applied as weights[10] in the estimation of treatment effects (regression models and permutation testing) so that a larger weight is applied to participants that are underrepresented in the sample due to missing observations.[11]

Analysing the impact of the program on multiple child health measures increases the likelihood of a Type-1 error and studies of RCTs have been criticized for overstating treatment effects due to this 'multiplicity' effect (Pocock et al., 1987). In order to assess the robustness of our results we apply the stepdown procedure described in Romano and Wolf (2005) to our individual permutation tests. The stepdown procedure involves

[10] In terms of the distribution of the IPW weights, the majority of participants receive a weight less than 2 and there are very few outliners. The mean and standard deviations of the weights at each time point is 1.20 (1.02) at 6 months; 1.31 (1.35) at 12 months; 1.33 (0.59) at 18 months; 1.23 (0.34) at 24 months; and 1.32 (0.79) at 36 months. We re-estimated the results by giving the 3 participants who had a weight above 2.5 the average weight, and found that it did not change the results for all but one variable (number of health problems at 24 months).

[11] Two participants who did not complete the baseline assessment yet completed interviews at later time points are assigned an average weight at each time point.

calculating a t-statistic for each null hypothesis in a family of related outcomes and placing them in descending order. Using the permutation testing method, the largest absolute observed t-statistic is compared with the distribution of maximal permuted t-statistics. If the probability of observing this statistic by chance is high ($p \geq 0.1$) we fail to reject the joint null hypothesis that the treatment has no impact on any outcome in the family being tested. If the probability of observing this t-statistic is low ($p < 0.1$) we reject the joint null hypothesis and proceed by excluding the most statistical significant individual hypothesis and test the subset of hypotheses that remain for joint significance. This process of dropping the most significant individual hypothesis continues until only one hypothesis remains. 'Stepping down' through the hypotheses allows us to isolate the hypotheses that lead to a rejection of the null. This method is superior to the Bonferroni adjustment method as it accounts for interdependence across outcomes.

The child health outcomes are placed into 7 families for the individual tests and the stepdown tests. For the stepdown procedure, the outcome measures included in each family should be correlated and represent an underlying construct. In this case the measures included in each family are identical variables measured at different time points.[12] The stepdown tests are only estimated for the families where we identify statistically significant differences in the individual tests.

2.5 Additional Analyses

To test for differential treatment effects by gender, subgroup analysis is conducted separately for girls and boys using the methods described above (i.e., IPW-adjusted chi-squared/t-tests, permutation tests, and stepdown tests). In addition, we test for the presence of contamination and differential misreporting using an IPW-adjusted and unadjusted permutation test. For these analyses, the stepdown procedure is not applied as only one outcome is considered.

3 Results

3.1 Sample Description

[12] The 7 families include: whether or not the child has good health, number of health problems experienced by child, whether or not child stayed in hospital, whether or not child had an accident, child's immunizations, whether or not child suffers from wheezing/asthma, whether or not child suffers from chest infections.

233 participants were recruited and randomized to the intervention group (n=115) and control group (n=118). Of the participants randomized, 205 completed the baseline interview (intervention = 86%, control = 90%). The Consort diagram in Appendix Figure B1 demonstrates the reasons for this reduction in baseline participation. Appendix Table B3 shows the comparability of the intervention and control groups on all but two of the 21 selected maternal socio-demographic, health, personality, and parenting measures assessed, indicating the equivalence of the groups at baseline.[13] Following baseline, 173 participants completed the 6 month interview (intervention = 72%, control = 76%), 165 the 12 month interview (intervention = 71%, control = 70%), 155 the 18 month interview (intervention = 70%, control = 63%), 165 the 24 month interview (intervention = 70%, control = 71%), and 150 the 36 month interview (intervention = 64%, control = 64%). These attrition rates compare favorably with other HVPs (e.g., Guttentag et al. 2014).[14]

3.2 Treatment Effects[15]

The means (standard deviations) and *p*-values that result from the chi-squared tests of logistic regression coefficients or t-tests of OLS regression coefficients (column 1), as well as the individual permutation test *p*-values (column 2), and effect sizes (column 3) are reported in Table 1. The *p*-values resulting from the traditional tests are very similar in nature to the permutation testing *p*-values, which suggests that the distributional assumptions imposed by the traditional tests are not overly restrictive when applied to the current sample. As the permutation testing procedure may be more reliable in small samples, we focus our interpretation on these results.

[13] In total, the two groups did not differ on 90.5% (114/126) of baseline variables.

[14] 6 participants (intervention = 3, control = 3) completed one of the five waves, 5 participants (intervention = 1, control = 4) completed two of the five waves, 12 participants (intervention = 4, control = 8) completed three of the five waves, 21 participants (intervention = 8, control = 13) completed four of the five waves, and the remaining 133 participants (intervention = 70, control = 63) completed all five waves.

[15] All analysis presented in Tables 1 and 2 are unconditional of any control variables. Given that we use an experimental design, conditioning on covariates should not be strictly necessary, however, including them can improve the precision of estimates (Duflo et al., 2006). The tests were replicated firstly by controlling for baseline variables that may theoretically impact child health including maternal age, parity, medical card status, and maternal physical/mental health conditions. In addition, we also re-estimated the results controlling for a selection of baseline variables that were statistically significantly different between the intervention and control groups including knowledge of infant development, parenting attitudes, self-efficacy, mother's physical health, consideration of future consequences scale, and vulnerable attachment style insecurity score. In both cases, the conditional and unconditional results are very similar. The conditional results are available upon request.

Table 1

Impact of treatment on child health.

	N (*intervention /control)*	$M_{intervention}$ (*SD*)	$M_{control}$ (*SD*)	Chi-squared/t-test p^a	Permutation test p^b	Effect Size ME/d[c]
				(1)	(2)	(3)
Rated Good Health						(ME)
6 months	173 (83/90)	0.93 (0.26)	0.93 (0.25)	0.884	0.835	-0.006
12 months	165 (82/83)	0.94 (0.24)	0.92 (0.28)	0.565	0.586	0.023
18 months	154 (80/74)	0.94 (0.24)	0.84 (0.37)	0.057*	0.049**	0.100
24 months	165 (81/84)	0.95 (0.22)	0.85 (0.36)	0.034**	0.027**	0.105
36 months	150 (74/76)	0.88 (0.33)	0.87 (0.34)	0.855	0.864	0.010
Number of Health Problems						(d)
6 months	173 (83/90)	1.37 (1.62)	1.28 (1.09)	0.647	0.677	0.070
12 months	164 (81/83)	1.31 (1.41)	1.46 (1.25)	0.475	0.481	-0.113
18 months	154 (80/74)	1.34 (1.30)	1.43 (1.28)	0.650	0.650	-0.074
24 months	165 (81/84)	1.20 (1.19)	1.64 (1.42)	0.031**	0.029**	-0.342
36 months	150 (74/76)	1.36 (1.17)	1.49 (1.18)	0.526	0.526	-0.105
Hospital Stay						(ME)
6 months	173 (83/90)	0.10 (0.30)	0.09 (0.29)	0.865	0.817	0.008
12 months	165 (82/83)	0.06 (0.24)	0.06 (0.24)	0.984	0.975	0.001
18 months	154 (80/74)	0.01 (0.11)	0.09 (0.29)	0.051*	0.014**	-0.082
24 months	165 (81/84)	0.02 (0.16)	0.06 (0.24)	0.282	0.305	-0.035
36 months	150 (74/76)	0.05 (0.23)	0.12 (0.33)	0.171	0.161	-0.064
Accident						(ME)
12 months	165 (82/83)	0.05 (0.22)	0.01 (0.11)	0.203	0.194	0.037
18 months	154 (80/74)	0.08 (0.27)	0.05 (0.23)	0.600	0.678	0.021
24 months	165 (81/84)	0.10 (0.30)	0.10 (0.30)	0.939	0.930	0.004

36 months	150 (74/76)	0.14 (0.34)	0.22 (0.42)	0.162	0.157	-0.089
Immunizations						(ME)
4 months	172 (82/90)	0.96 (0.19)	0.88 (0.33)	0.053*	0.045**	0.086
(assessed 6m)						
6 months	165 (82/83)	0.99 (0.11)	0.96 (0.19)	0.340	0.360	0.024
(assessed 12m)						
13 months	154 (80/74)	0.88 (0.33)	0.85 (0.36)	0.670	0.678	0.024
(assessed 18m)						
Wheezing or Asthma						(ME)
12 months	165 (82/83)	0.11 (0.31)	0.13 (0.34)	0.654	0.662	-0.023
18 months	154 (80/74)	0.14 (0.35)	0.19 (0.39)	0.387	0.402	-0.052
24 months	165 (81/84)	0.07 (0.26)	0.21 (0.41)	0.014**	0.009***	-0.140
36 months	150 (74/76)	0.16 (0.37)	0.18 (0.39)	0.722	0.715	-0.022
Chest Infection						(ME)
12 months	165 (82/83)	0.24 (0.43)	0.34 (0.48)	0.188	0.189	-0.093
18 months	154 (80/74)	0.29 (0.46)	0.32 (0.47)	0.620	0.625	-0.037
24 months	165 (81/84)	0.26 (0.44)	0.39 (0.49)	0.096*	0.092*	-0.122
36 months	150 (74/76)	0.28 (0.45)	0.35 (0.48)	0.349	0.348	-0.071

Notes: 'N' indicates the sample size. 'M' indicates the mean. 'SD' indicates the standard deviation. [a] two-tailed *p*-value from a t-test/chi-squared test of the null that the coefficient on treatment assignment from an OLS/logistic regression equals zero. [b] Two-tailed *p*-value from an individual permutation test with 100,000 replications. [c] Effect Size refers to Cohen's *d* for continuous variables and Marginal Effects for binary variables. * p < .10, ** p < .05, *** p < .10.

The intervention and control groups do not statistically significantly differ on ratings of child health at 6, 12, or 36 months. However, the intervention group is statistically significantly more likely to report that their child is in good health at 18 and 24 months. The estimated marginal effect at 18 months implies that a child who is in the intervention group is 10 percentage points more likely to be rated as being in good health by their mother relative to the control group. Overall, a very high proportion of both groups rate their child as being in good health. There are no statistically significant differences between the two groups in terms of number of child health problems reported by mothers at 4 of the 5 time points. Yet at 24 months, the intervention group

report statistically significantly fewer problems than the control group, and the Cohen's *d* statistic implies that the magnitude of the effect is approximately one third of a pooled standard deviation. On average, the intervention group report 0.44 fewer problems per child at this time point. The most prevalent problems at 36 months are chest infections (32%), ear infections (26%), wheezing/asthma (17%), and accidents (16%).

There is a statistically significant difference between the intervention and control groups regarding the proportion of children who spent at least one night in hospital at the 18 month time point only. In regards the substantive effect, the intervention group are 8 percentage points less likely to spend a night in hospital between 12 and 18 months compared to the control group. Figure 1 illustrates that the program may have resulted in the intervention and control groups diverging over the last three time points, although the observed 4 and 6 percentage points differences at 24 and 36 months respectively are not statistically significant. Follow up questions for those who spend at least one night in hospital found that the most common reason for hospitalization throughout the whole period was asthma (14%), accidents (14%), bronchitis (11%), and fever (11%).

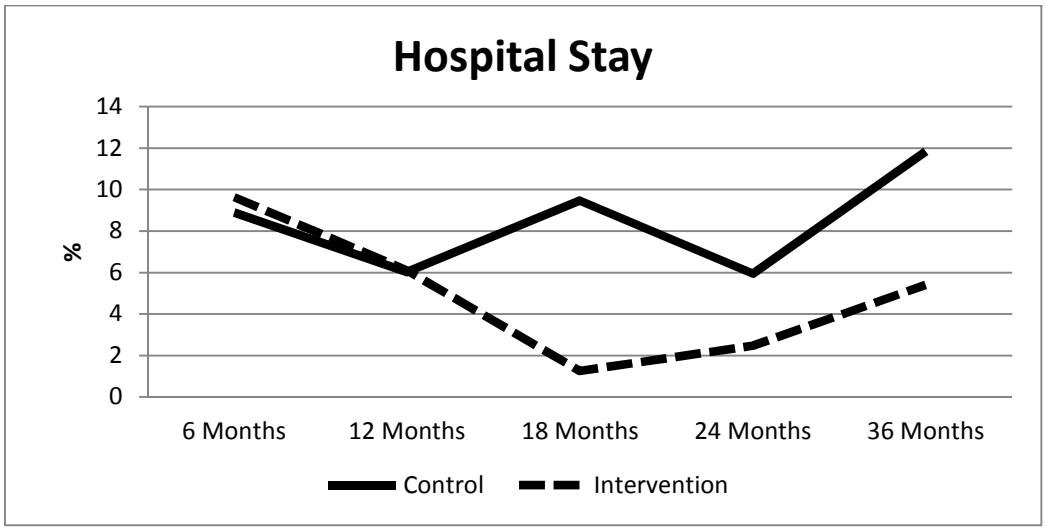

Fig. 1. Rate of hospitalization in the intervention and control group over time.

The results also show that there are no statistically significant differences between the two groups for the proportion of mothers reporting that their child suffered an accident at any time point. As anticipated, the rate of accidents in both

groups rose between 6 and 36 months as the children became more active and independent. However, the rate may also have increased at 36 months due to the 12 month reporting period at 36 months compared to the 6 month reporting period at previous time points. The lack of treatment effects on accidents may be related to the common supports which were provided to both groups and included a number of child-proofing safety items.

A statistically significant treatment effect is found for immunizations by the 6 month interview, such that the intervention group were 9 percentage points more likely to have received their necessary 4 month immunizations relative to control children. However, this effect did not persist for later immunization schedules. A statistically significant treatment effect is also found for asthma/wheezing. This effect implies that at 24 months, a lower proportion of children in the intervention group required medical attention for asthma/wheezing relative to the control group (14 percentage point difference). Similarly, there is a statistical signficiant difference in the proportion of intervention children requiring medical attention for chest infections at 24 months (12 percentage point difference). Diagnosing asthma in early childhood can be difficult, and the majority of childhood asthma onset manifests as wheezing illness in the first 2 to 3 years (Klinnert et al., 2005). The symptoms of wheezing and respiratory infections, such as bronchitis and chest infections, are often very similar to asthma which complicates its formal diagnosis. For example, more than one third of children under two years will wheeze at some point, yet far fewer are given an asthma diagnosis (The Asthma Society of Ireland, 2013). It should be noted that the number of health problems variable incorporates the number of times the child sought medical attention for asthma and wheezing, therefore the significant treatment effect for the number of health problems at 24 months may be driven by the significant treatment effect for asthma/wheezing at 24 months.

In total, the intervention group have more favorable outcomes compared to the control group on 23 of the 30 (77%) measures under investigation, which is statistically significantly different to the 50% we would expect if the program was having no impact, according to a two-sided binomial test ($p = 0.01$).

3.3 IPW-Weighted Treatment Effects[16]

Table 2 reports the adjusted results using the IPW-weights to account for differential attrition and can be interpreted in the same manner as Table 1. The IPW-weighted means (standard deviations) and *p*-values that result from the chi-squared tests of weighted logistic regression coefficients and *t*-tests of weighted OLS regression coefficients (column 1), as well as the weighted individual permutation tests (column 2) are reported.

As in the unweighted analysis, the traditional tests and the permutation tests result in similar conclusions in most cases. Additionally, adjusting for attrition does not substantially alter the overall pattern of results, yet the number of statistically significant individual differences falls from 7 to 5. For example, the unweighted analysis identifies a statistically significant group difference on maternal ratings of child health at 18 and 24 months, while only the 24 month effect remains significant using the IPW-adjusted analysis. This occurs as the weighting procedure leads to a slight decrease within the intervention group and a slight increase within the control group regarding the proportion of mothers rating their children as having good health at 18 months. In addition, the unweighted analysis identifies a statistically significant group difference for chest infections at 24 months; however this result does not remain significant using the IPW-adjusted analysis.

In the IPW analysis, the intervention group have more favorable outcomes on 20 of the 30 (67%) measures under investigation. Although this percentage is lower than the equivalent figure for the unweighted analysis, it is still statistically significantly different to the 50% we would expect if the program was having no impact, according to a two-sided binomial test (p=0.099).

[16] As an alternative to IPW, multiple imputation (MI) was also used to account for attrition and wave non-response. Missing values were imputed 50 times using the baseline variables which were identified by the BIC process. Analyses were run with the resulting 50 completed data sets and then pooled with Rubin's combination rules (Rubin, 1987). For binary outcomes logistic imputation was used, for the continuous outcomes a multivariate normal model was used. Where possible (i.e. when there was enough variation in the respective outcomes) imputation models were fitted separately for the intervention and control groups. In general, the IPW and MI results are largely equivalent, with one extra statistically significant result (number of health problems at 24 months) found in the MI models. The MI results are available upon request.

Table 2

IPW-adjusted impact of treatment on child health

	N (*intervention /control)*	$M_{intervention}$ (*SD*)	$M_{control}$ (*SD*)	IPW chi-squared/t-test p[a]	IPW Permutation Test p[b]	Effect Size ME/d[c]
				(1)	(2)	(3)
Rated Good Health						(ME)
6 months	173 (83/90)	0.92 (0.27)	0.94 (0.23)	0.580	0.591	-0.021
12 months	165 (82/83)	0.94 (0.24)	0.93 (0.26)	0.833	0.841	0.009
18 months	154 (80/74)	0.92 (0.27)	0.87 (0.34)	0.322	0.302	0.054
24 months	165 (81/84)	0.95 (0.23)	0.84 (0.37)	0.054*	0.037**	0.108
36 months	150 (74/76)	0.88 (0.33)	0.89 (0.32)	0.908	0.910	-0.006
Number of Health Problems						(d)
6 months	173 (83/90)	1.38 (1.73)	1.24 (1.02)	0.414	0.605	0.099
12 months	164 (81/83)	1.37 (1.49)	1.39 (1.12)	0.712	0.926	-0.015
18 months	154 (80/74)	1.42 (1.38)	1.31 (1.24)	0.664	0.651	0.084
24 months	165 (81/84)	1.26 (1.31)	1.71 (1.55)	0.257	0.089*	-0.314
36 months	150 (74/76)	1.42 (1.22)	1.41 (1.24)	0.940	0.942	0.008
Hospital Stay						(ME)
6 months	173 (83/90)	0.09 (0.29)	0.07 (0.26)	0.632	0.637	0.020
12 months	165 (82/83)	0.07 (0.25)	0.07 (0.25)	0.987	0.987	-0.001
18 months	154 (80/74)	0.01 (0.10)	0.09 (0.28)	0.042**	0.027**	-0.077
24 months	165 (81/84)	0.04 (0.20)	0.06 (0.24)	0.596	0.581	-0.022
36 months	150 (74/76)	0.06 (0.25)	0.10 (0.30)	0.425	0.409	-0.039
Accident						(ME)
12 months	165 (82/83)	0.04 (0.21)	0.01 (0.10)	0.171	0.154	0.034
18 months	154 (80/74)	0.08 (0.27)	0.04 (0.21)	0.387	0.405	0.032
24 months	165 (81/84)	0.10 (0.30)	0.10 (0.30)	0.955	0.957	-0.003

36 months	150 (74/76)	0.12 (0.33)	0.21 (0.41)	0.162	0.159	-0.089
Immunizations						(ME)
4 months	172 (82/90)	0.97 (0.18)	0.90 (0.31)	0.071*	0.045**	0.072
(assessed 6m)						
6 months	165 (82/83)	0.98 (0.13)	0.97 (0.18)	0.584	0.597	0.019
(assessed 12m)						
13 months	154 (80/74)	0.89 (0.32)	0.85 (0.36)	0.526	0.536	0.034
(assessed 18m)						
Wheezing or Asthma						(ME)
12 months	165 (82/83)	0.11 (0.32)	0.10 (0.30)	0.812	0.817	0.011
18 months	154 (80/74)	0.17 (0.38)	0.16 (0.37)	0.844	0.846	0.012
24 months	165 (81/84)	0.08 (0.28)	0.24 (0.43)	0.019**	0.013**	-0.155
36 months	150 (74/76)	0.16 (0.37)	0.17 (0.38)	0.939	0.939	-0.005
Chest Infection						(ME)
12 months	165 (82/83)	0.26 (0.44)	0.43 (0.50)	0.102	0.158	-0.177
18 months	154 (80/74)	0.29 (0.46)	0.31 (0.47)	0.746	0.750	-0.026
24 months	165 (81/84)	0.27 (0.45)	0.38 (0.49)	0.134	0.132	-0.115
36 months	150 (74/76)	0.31 (0.47)	0.33 (0.47)	0.833	0.835	-0.018

Note: 'N' indicates the sample size. 'M' indicates the IPW mean. 'SD' indicates the IPW standard deviation. [a] two-tailed *p*-value from a t-test/chi-squared test of the null that the coefficient on treatment assignment from an IPW OLS/logistic regression. [b] Two-tailed *p*-value from an individual IPW permutation test with 100,000 replications. [c] Effect Size refers to Cohen's *d* for continuous variables and Marginal Effects for binary variables. * p < .10, ** p < .05, *** p < .10.

3.4 Adjusting for multiple hypothesis testing

As a final robustness check we conduct stepdown tests to account for multiple testing for all the outcome families where statistically significant individual differences are found. Thus we exclude the accidents family. Both unweighted (column 1) and IPW-adjusted (column 2) stepdown permutation testing *p*-values are presented in Table 3. Three of the six stepdown families survive the stepdown procedure when the unweighted analysis is used, and one of the six stepdown families survive when the IPW-adjustment is made. Statistically significant effects are found in the unweighted and weighted analyses for the asthma/wheezing stepdown family, where the individual

finding of lower reported asthma/wheezing among the intervention group remains statistically significant in the stepdown test. The statistically significant individual findings for the 18 month hospital stays result and the 4-month immunization result survives adjustment for multiple comparisons in the unweighted results only. The effects identified for general health, number of health problems, and chest infections do not survive adjustment for multiple comparisons.

Table 3

Accounting for multiple comparisons.

	Stepdown Permutation Test p^a	IPW Stepdown Permutation Test p^b
	(1)	(2)
Rated Good Health		
6 months	0.835	0.932
12 months	0.922	0.973
18 months	0.156	0.731
24 months	0.128	0.124
36 months	0.971	0.910
Number of Health Problems		
6 months	0.677	0.963
12 months	0.912	0.995
18 months	0.880	0.953
24 months	0.120	0.279
36 months	0.887	0.942
Hospital Stay		
6 months	0.962	0.895
12 months	0.975	0.987
18 months	0.081*	0.186
24 months	0.666	0.924
36 months	0.467	0.889

Immunizations		
4 months (assessed 6m)	0.079*	0.212
6 months (assessed 12m)	0.564	0.597
13 months (assessed 18m)	0.678	0.807
Wheezing or Asthma		
12 months	0.879	0.993
18 months	0.748	0.973
24 months	0.036**	0.035**
36 months	0.715	0.939
Chest Infection		
12 months	0.444	0.215
18 months	0.625	0.935
24 months	0.304	0.375
36 months	0.559	0.835

Notes: [a] two-tailed p-value from a stepdown permutation test with 100,000 replications, [b] two-tailed IPW p-value from a stepdown permutation test with 100,000 replications. * p < .10, ** p < .05, *** p < .10.

3.5 Gender subgroup analysis

A number of studies have investigated the differential impact of early intervention programs by gender (e.g. Anderson, 2008; Eckenrode et al., 2010; Heckman et al., 2010). Many find that such programs are more effective for girls than boys in the long term, particularly in the area of human capital; however, recent work has also found long term effects for men regarding health outcomes (Campbell et al., 2014). To explore the potential for differential treatment effects by gender we conducted a subgroup analysis using the same methodology as above (i.e. IPW-adjusted chi squared/t-tests , individual permutation tests, and stepdown tests). The results reported in Tables 4 and 5 show that we find many more treatment effects in both the individual and stepdown tests for boys, and relatively few effects for girls. The number of findings for boys is considerable given the smaller sample size compared to the main analysis.

In particular, Table 4 shows statistically significant individual treatment effects for boys regarding the number of health problems at 24 months (d = 0.63), hospital stays at 18 months (3 pp), accidents at 36 months (23.9 pp), asthma/wheezing at 24 months (22.5

pp), and chest infections at every time point (22.8 – 37.9 pp). We also find statistically significant stepdown effects for boys in three stepdown families including the number of health problems, accidents, and chest infection. For girls, we only find one positive statistically significant effect in the permutation results (general health at 24 months, 12.6 pp) and one negative treatment effect (accidents at 18 months, 12.9 pp), in addition none of the results survive adjustment for multiple testing.

In total, boys in the intervention group have more favorable outcomes than boys in the control group on 22 of the 30 (73%) measures under investigation, which is statistically significantly different to the 50% we would expect if the program was having no impact, according to a two-sided binomial test (p = 0.016), while the corresponding figure for girls is 19 (63%), which is not significantly different from 50%. Thus, similar to previous research focused on health outcomes later in life, the treatment effects during early childhood are primarily concentrated among boys.

Table 4

IPW-adjusted impact of treatment on boy's health.

	N (*intervention /control)*	$M_{\text{intervention}}$ (*SD*)	M_{control} (*SD*)	IPW chi-squared/t-test p^a	IPW Permutation Test p^b	IPW Stepdown Test p^c	Effect Size ME/d[d]
				(1)	(2)	(3)	(4)
Rated Good Health							(ME)
6 months	75 (42/33)	0.90 (0.30)	0.96 (0.20)	0.343	0.349	0.556	-0.056
12 months	72 (44/28)	0.93 (0.27)	0.98 (0.15)	0.290	0.307	0.604	-0.054
18 months	68 (42/26)	0.92 (0.28)	0.80 (0.41)	0.205	0.226	0.614	0.118
24 months	71 (41/30)	0.91 (0.28)	0.81 (0.40)	0.246	0.251	0.618	0.104
36 months	64 (37/27)	0.82 (0.39)	0.83 (0.38)	0.934	0.934	0.934	0.008
Number of Health Problems							(d)
6 months	75 (42/33)	1.52 (2.20)	1.26 (0.85)	0.527	0.654	0.654	0.156
12 months	72 (44/28)	1.47 (1.74)	1.22 (0.74)	0.454	0.503	0.746	0.187

18 months	68 (42/26)	1.43 (1.59)	1.85 (1.34)	0.278	0.304	0.604	-0.286
24 months	71 (41/30)	1.27 (1.51)	2.17 (1.34)	0.019**	0.017**	0.052**	-0.630
36 months	64 (37/27)	1.36 (1.27)	1.78 (1.41)	0.269	0.286	0.664	-0.313
Hospital Stay							(ME)
6 months	75 (42/33)	0.07 (0.26)	0.02 (0.15)	0.312	0.314	0.764	0.048
12 months	72 (44/28)	0.05 (0.23)	0.02 (0.15)	0.475	0.550	0.876	0.031
18 months	68 (42/26)	0.00 (0.00)	0.03 (0.19)	0.321	0.080*	0.736	0.030
24 months	71 (41/30)	0.04 (0.20)	0.08 (0.28)	0.580	0.634	0.801	-0.040
36 months	64 (37/27)	0.10 (0.30)	0.10 (0.31)	0.978	0.969	0.969	-0.002
Accident							(ME)
12 months	72 (44/28)	0.04 (0.20)	0.02 (0.15)	0.671	0.713	0.713	0.016
18 months	68 (42/26)	0.02 (0.14)	0.11 (0.31)	0.125	0.137	0.452	-0.088
24 months	71 (41/30)	0.08 (0.28)	0.11 (0.32)	0.657	0.647	0.885	-0.032
36 months	64 (37/27)	0.09 (0.30)	0.33 (0.48)	0.020**	0.027**	0.052*	-0.239
Immunizations							(ME)
4 months (assessed 6m)	74 (41/33)	0.96 (0.20)	0.85 (0.37)	0.110	0.135	0.427	0.112
6 months (assessed 12m)	72 (44/28)	0.97 (0.17)	0.95 (0.23)	0.635	0.629	0.629	0.025
13 months (assessed 18m)	68 (42/26)	0.91 (0.28)	0.85 (0.37)	0.395	0.437	0.729	0.067
Wheezing or Asthma							(ME)
12 months	72 (44/28)	0.10 (0.30)	0.17 (0.38)	0.427	0.449	0.805	-0.073
18 months	68 (42/26)	0.21 (0.42)	0.24 (0.44)	0.821	0.835	0.969	-0.025
24 months	71 (41/30)	0.10 (0.30)	0.32 (0.48)	0.043**	0.035**	0.130	-0.225
36 months	64 (37/27)	0.17 (0.38)	0.19 (0.40)	0.830	0.833	0.833	-0.022
Chest Infection							(ME)
12 months	72 (44/28)	0.25 (0.44)	0.63 (0.49)	0.030**	0.056*	0.061*	-0.379
18 months	68 (42/26)	0.22 (0.42)	0.47 (0.51)	0.047**	0.049**	0.093*	-0.243
24 months	71 (41/30)	0.32 (0.47)	0.55 (0.51)	0.065*	0.069*	0.069*	-0.228

36 months	64 (37/27)	0.25 (0.44)	0.54 (0.51)	0.027**	0.027**	0.068*	-0.288

Note: 'N' indicates the sample size. 'M' indicates the IPW mean. 'SD' indicates the IPW standard deviation. [a] two-tailed *p*-value from a t-test/chi-squared test of the null that the coefficient on treatment assignment from an IPW OLS/logistic regression. [b] Two-tailed *p*-value from an individual IPW permutation test with 100,000 replications. [c] Two-tailed *p*-value from a stepdown IPW permutation test with 100,000 replications. [d] Effect Size refers to Cohen's *d* for continuous variables and Marginal Effects for binary variables. * p < .10, ** p < .05, *** p < .10.

Table 5

IPW-adjusted impact of treatment on girl's health.

	N (*intervention /control)*	$M_{intervention}$ (*SD*)	$M_{control}$ (*SD*)	IPW chi-squared/t-test p[a]	IPW Permutation Test p[b]	IPW Stepdown Test p[c]	Effect Size ME/d[d]
				(1)	(2)	(3)	(4)
Rated Good Health							(ME)
6 months	98 (41/57)	0.95 (0.23)	0.93 (0.25)	0.820	0.888	0.888	0.012
12 months	93 (38/55)	0.95 (0.21)	0.89 (0.31)	0.299	0.321	0.763	0.060
18 months	86 (38/48)	0.93 (0.26)	0.90 (0.30)	0.667	0.648	0.953	0.030
24 months	94 (40/54)	0.98 (0.15)	0.85 (0.36)	0.068*	0.029**	0.116	0.126
36 months	86 (37/49)	0.94 (0.25)	0.91 (0.29)	0.665	0.679	0.910	0.024
Number of Health Problems							(d)
6 months	98 (41/57)	1.24 (1.05)	1.23 (1.13)	0.954	0.954	0.954	0.009
12 months	93 (38/55)	1.24 (1.10)	1.51 (1.32)	0.284	0.281	0.843	-0.222
18 months	86 (38/48)	1.40 (1.11)	1.07 (1.12)	0.194	0.210	0.706	0.296
24 months	94 (40/54)	1.25 (1.09)	1.46 (1.61)	0.523	0.606	0.782	-0.153
36 months	86 (37/49)	1.48 (1.18)	1.23 (1.12)	0.406	0.482	0.808	0.217
Hospital Stay							(ME)
6 months	98 (41/57)	0.12 (0.33)	0.12 (0.32)	0.946	0.928	0.928	0.004
12 months	93 (38/55)	0.08 (0.28)	0.10 (0.30)	0.812	0.853	0.967	-0.017
18 months	86 (38/48)	0.02 (0.15)	0.11 (0.32)	0.086*	0.104	0.477	0.090

24 months	94 (40/54)	0.04 (0.19)	0.05 (0.22)	0.786	0.761	0.990	-0.013
36 months	86 (37/49)	0.03 (0.17)	0.10 (0.31)	0.222	0.162	0.548	-0.074
Accident							(ME)
12 months	93 (38/55)	0.05 (0.22)	0.00 (0.00)	0.157	0.181	0.470	0.050
18 months	86 (38/48)	0.14 (0.35)	0.01 (0.12)	0.033**	0.042**	0.134	0.129
24 months	94 (40/54)	0.11 (0.32)	0.09 (0.29)	0.756	0.764	0.946	0.020
36 months	86 (37/49)	0.15 (0.37)	0.16 (0.37)	0.968	0.969	0.969	-0.003
Immunizations							(ME)
4 months (assessed 6m)	98 (41/57)	0.98 (0.16)	0.93 (0.25)	0.361	0.321	0.547	0.042
6 months (assessed 12m)	93 (38/55)	1.00 (0.00)	0.98 (0.12)	0.321	0.582	0.745	0.020
13 months (assessed 18m)	86 (38/48)	0.86 (0.35)	0.86 (0.35)	0.995	0.996	0.996	0.000
Wheezing or Asthma							(ME)
12 months	93 (38/55)	0.13 (0.34)	0.05 (0.21)	0.167	0.161	0.442	0.078
18 months	86 (38/48)	0.12 (0.32)	0.12 (0.33)	0.954	0.954	0.998	-0.004
24 months	94 (40/54)	0.07 (0.26)	0.19 (0.40)	0.116	0.126	0.273	-0.124
36 months	86 (37/49)	0.16 (0.37)	0.16 (0.37)	0.995	0.995	0.995	-0.001
Chest Infection							(ME)
12 months	93 (38/55)	0.26 (0.45)	0.29 (0.46)	0.774	0.772	0.772	-0.029
18 months	86 (38/48)	0.36 (0.49)	0.24 (0.43)	0.288	0.302	0.608	0.120
24 months	94 (40/54)	0.22 (0.42)	0.29 (0.46)	0.418	0.414	0.635	-0.079
36 months	86 (37/49)	0.37 (0.49)	0.23 (0.43)	0.195	0.225	0.582	0.142

Note: 'N' indicates the sample size. 'M' indicates the IPW mean. 'SD' indicates the IPW standard deviation. [a] two-tailed *p*-value from a t-test/chi-squared test of the null that the coefficient on treatment assignment from an IPW OLS/logistic regression. [b] Two-tailed *p*-value from an individual IPW permutation test with 100,000 replications. [c] Two-tailed *p*-value from a stepdown IPW permutation test with 100,000 replications. [d] Effect Size refers to Cohen's *d* for continuous variables and Marginal Effects for binary variables. * p < .10, ** p < .05, ***

3.6 Testing for differential misreporting

As discussed above, our measures of child health are based on maternal report which may be subject to measurement error if mothers over or under report their child's health. If mothers in both the intervention and control group under/over report the prevalence of health problems, this will not bias the estimation of treatment effects. However, if differential reporting exists, such that one group under/over reports more than the other, this may bias the results. In this case, one may expect the intervention group to underreport their child's health problems relative to the control group as they may be cognisant of the supports and advice they receive from the mentors about preventive health care measures and appropriate care for their child. Therefore, the treatment effects reported above could be driven by differential misreporting rather than program impact.

Below, we test for the presence of differential misreporting across the intervention and control groups using the Social Desirability Scale-17 (SDS-17; Stöber, 2001) measured at 24 months. This scale uses 16 true or false items to measure behaviors that are socially desirable and infrequent, as well as behaviors that are socially undesirable but frequent, with a higher score indicating an increased tendency to respond to the items in a social desirable manner. Table 6 shows that there is no statistically significant difference between the intervention and control groups regarding the social desirability measure, and the means are comparable to those found in a representative sample (11.29; Stöber, 2001). Overall, this indicates that, although participants may be attempting to answer questions in a way which they believe appears more favorable, there is no difference in the levels to which they are doing this across each group. While this does not necessarily imply that the parents do not misreport, it does increase our confidence that the estimation of treatment effects are not driven by differential misreporting.

Table 6

Testing for differential misreporting.

	N (*intervention /control)*	$M_{\text{intervention}}$ (*SD*)	M_{control} (*SD*)	Permutation Test P^1	IPW-Permutation Test P^2
Social Desirability Scale	165 (81/84)	11.19 (2.77)	11.29 (2.76)	0.814	0.540

Notes: 'N' indicates the sample size. 'M' indicates the unweighted mean. 'SD' indicates the unweighted standard deviation. [1] two-tailed (right-sided) p value from an individual permutation test with 100,000 replications. [2] two-tailed (right-sided) p value from an IPW-weighted individual permutation test with 100,000 replications.

3.7 Testing for contamination

It is also possible that the treatment effects are biased due to contamination. Contamination, also known as spillover effects (Bloom, 2005), may have occurred if participants in the intervention group engage in cross-talk or intentionally or unintentionally share their parenting materials, information, strategies, or advice which they receive from their mentors, with participants in the control group. As the potential for contamination in *PFL* is high given the geographical proximity of the participants and randomization at the individual level, a number of strategies were devised to measure cross-talk and information flows between the two groups (information on these strategies can be found in Doyle and Hickey, 2013).

Here, we test for the presence of contamination using a 'blue-dye' question. At 24 months, participants from the intervention and control group were asked if they have heard of a particular parenting phrase, i.e., *'descriptive praise',* and if they know what this phrase means. The phrase is related to a topic which a greater proportion of participants in the intervention group should be aware of as the mentors discuss and promote this behavior with participants when delivering the program. In addition, there is a Tip Sheet on *'descriptive praise'*. This question may be used to as a proxy for contamination as, if a large proportion of the participants in the control group state that they know what this phrase means and they correctly identify how to engage in this behavior, it is indicative that they may have accessed material or information intended for the intervention group only.

The first row in Table 7 shows that a statistically significantly greater proportion of the intervention group (33%) report knowledge of the phrase compared to the control group (12%) suggesting a lack of contamination. However, in order to provide a more accurate measure of contamination, participants who stated that they had heard of the parenting phrase, yet provided incorrect responses regarding how best to engage in this behavior, were treated as if they reporting not knowing the phase. The test using the proportion of participants who accurately report how to engage in *descriptive praise* is re-estimated and is presented in the second row of Table 7. As before, it shows that a statistically significantly greater proportion of the intervention group (28%) than the control group (6%) report knowledge of the phrase and accurately know how to engage in *descriptive praise*. Again, suggesting that contamination may not be a major issue.

A limitation of this analysis is that is it based on one area of child development only, thus it is still possible that the intervention group may have shared material about child health specifically. However, in the absence of alternative measures, this proxy suggests that contamination may be low in the *PFL* trial at 24 months. Indeed, minimal contamination may be expected as *PFL* is a complex intervention which aims to change the behavior of participants by building relationships between mentors and participants in the intervention group. As it is often difficult to achieve behavioral change, even if contamination between the two groups exists, it may not be enough to significantly affect the results (Howe et al., 2007).

Table 7

Testing for contamination across groups

	N (*intervention /control)*	$M_{intervention}$ (*SD*)	$M_{control}$ (*SD*)	Permutation Test P^1	IPW Permutation Test P^2
Heard the phrase 'descriptive praise'	165 (81/84)	0.33 (0.47)	0.12 (0.33)	0.001	0.001
Heard the phrase 'descriptive praise' & accurately reports how to engage in this behavior	165 (81/84)	0.26 (0.44)	0.06 (0.24)	0.000	0.001

Note: 'N' indicates the sample size. 'M' indicates the unweighted mean. 'SD' indicates the unweighted standard deviation. [1]two-tailed p value from an individual permutation test with 100,000 replications. [2]two-tailed p value from an IPW-weighted individual permutation test with 100,000 replications.

4 Conclusions

Developing policies which seek to reduce socioeconomic inequalities in health is hampered by the predominance of observational studies. This article overcomes this issue by utilizing a RCT design which specifically targets disadvantaged families during a critical period of intergenerational health transmission. The aim of the *PFL* program is to improve children's health and development with the ultimate aim of improving their school readiness skills. The program adopts a holistic view of school readiness in accordance with best practice which considers child health as a significant contributor. The measures used here can accurately assess the program's impact on health as they concentrate on areas which the mentors specifically target as part of the program, such as encouraging immunization, identifying symptoms of illness, as well as the importance of creating a safe child-friendly home environment to ensure that accidents are avoided.

We find that a bundle of parenting interventions provided from pregnancy onwards has some positive and statistically significant effects on child health in the first 3 years of life. As we demonstrate minimal evidence of contamination or differential

misreporting across the intervention and control groups, this indicates a high level of internal validity concerning these results. The strongest main effect, both statistically and substantively, is found for reducing the incidences of wheezing/asthma as this domain remains statistically significant when we account for multiple hypothesis testing and differential attrition. It also has a sizable economic effect representing a 15.5 percentage point reduction. Individual main treatment effects which do not survive adjustment for multiple testing and attrition are found for general health (10.0 pp), hospitalizations (8.2 pp), immunizations (8.6 pp), chest infections (12.2 pp), and the number of health problems (d = 0.34). While the binomial tests show that the probability of observing the number of reported favourable differences by chance is small, it is important to note that this inference is based on the assumption that more reported health problems reflect poorer child health which is considered an unfavorable outcome, rather than heightened parental awareness of health issues, which could be considered a positive outcome.

Similar to previous studies (e.g. Campbell et al., 2014) we find that the impact of early intervention on health outcomes is greater for boys than girls. This is in contrast to studies focusing on non-health outcomes, which typically find greater effects for girls in terms of academic and labor market outcomes (e.g. Anderson, 2008). We detect statistically significant IPW-adjusted stepdown families for the number of health problems (d = 0.63), accidents (23.9 pp) and chest infections (22.8 – 37.9 pp) for boys, and none for girls. The size of these effects are large and represent substantial changed in children's health attributed to the intervention. A possible explanation for the gender differences is the greater vulnerability of boys in the prenatal and infancy periods. In particular, the male foetus has a greater probability of non-survival, premature birth and deformity, and from birth, male children are less mature than girls and more likely to experience a developmental disorder (Kraemer, 2000). Thus, boys may demonstrate a greater need for early intervention, and reap larger benefits in terms of health outcomes. Indeed, a comparison of the girls and boys in the control groups appears to suggest that girls have better health outcomes than boys when only the common supports are received.

The exact mechanisms underlying these main and subgroup treatment effects cannot be determined as the treatment is a bundle of provisions including the HVP and the Triple P program, as well as the additional low level supports also provided to the

control group, including the developmental packs and access to community services. As participants were not randomized to receive different components of the bundle it is not possible to tease out the impact of the different provisions. As the Triple P program is primarily concerned with improving parenting skills (such as engaging in positive parenting techniques), provides no information on health behaviors, and could not affect the earlier findings as it began after the 24 month assessment, it seems likely that the improvements in child health are related to the information disseminated and discussed by the mentors through the Tip Sheets. For example, a Tip Sheet delivered in the first year of life highlights that infants need to be protected from passive smoking, ideally by making the home a smoke-free zone. This may be related to the main finding on wheezing/asthma as cigarette smoke can trigger asthma attacks, chest infections, and other infections (Hofhuis, de Jongste and Merkus, 2003). In addition, other Tip Sheets delivered within the first 2 years also provide information on how minor illnesses can be cared for at home (e.g., fever/high temperature, coughs and colds, vomiting, sticky eyes, thrush) before they develop into more serious complaints which may require hospitalization.

Irrespective of the mechanisms at play, the treatment effects must be attributed to the combined package of supports provided to the intervention group encompassing the home visits and the additional parenting course. On average, the participants received just over 50 visits by the time their child was three years old, with most engaging in one visit per month, and just under two-thirds participated in the Triple P course. The observed results also may be driven by the complementarity of the common supports provided to both the intervention and control groups (e.g. developmental toys) with the parenting treatments. Thus, we cannot be confident that the same effects would have arisen if any of these components were omitted. Similarly, if the common supports were effective, they may have reduced the mean differences between the intervention and control groups.

The impact of the *PFL* program on child health within the first 3 years of life contributes to the HVP literature in terms of the substantive findings and the methods used to generate these findings. An advantage of this study is the use of robust methods to identify the main and subgroup treatment effects. The lack of such investigations in the majority of HVP studies limits the confidence we can place in their conclusions regarding the ability of these interventions to reduce inequalities in health. While some

experimental studies of HVPs identify treatment effects for the management of asthma (e.g. Klinnert et al., 2005), none adjust for the multiplicity effect. This is important as the stepdown procedure used here highlights the implication of failing to address this issue. The majority of the individual treatment effects, including increased immunization uptake, reduced hospitalizations, reduced incidences of chest infection, improved general health, and a reduction in the number of health problems, were not strong enough to survive the stepdown procedure. That the main treatment effect for wheezing/asthma and the subgroup effects for the number of health problems, accidents and chest infections survived increases our confidence in these findings.

Another common concern in longitudinal RCTs is attrition. Our review of HVPs examining child health finds that some studies explicitly test for and find no evidence of differential attrition (e.g. Fergusson et al., 2005). However, the majority do not test or account for non-random dropout which may bias estimates of program impact if differential processes exist in the intervention and control groups. In our study, we find that by 36 months, 36% of the randomized sample do not participate, either due to dropout or wave non-response. While the findings from the unweighted results and the IPW-adjusted results are largely equivalent, we do find two less individual treatment effects and one less stepdown effect in the IPW-adjusted analysis. Thus, given the substantial threat which non-random attrition may pose to internal validity, it is important to test and adjust for its presence.

Observational studies are typically more common than RCTs due to the high costs associated with experimentation. Such costs often account for the small sample size in many trials, yet appropriate statistical methods which acknowledge this issue are often not applied. In this article we compare both traditional and permutation testing methods and find that they produce similar conclusions. This implies that the distributional assumptions imposed by the traditional tests are not overly restrictive in this case. However, it is possible that non-normality may be an issue in other experimental studies with relatively small samples, thus exploring alternative hypothesis testing methods can be informative with respect to correctly identifying program effectiveness.

Another issue which should be noted is the timing of the effects. The main treatment effect is restricted to 24 months, with no statistically significant findings in the main analysis at 36 months. There are a number of potential explanations for this. First, it may be related to measurement error as the recall period for the 36 month health outcomes was one year, while at previous assessment points, parents had to recall the child's health in the previous 6 months. Thus, there may be greater measurement error at 36 months due to recall bias which would introduce more variability into the estimate and impinge the identification of significant effects. Second, early intervention may generate effects at particular time points and, in some cases, those effects may dissipate over time, and in other cases, treatment effects may not be uncovered until later in childhood or adulthood. For example, Campbell et al. (2014) find that the Abecedarian early intervention program has significant treatment effects on the proportion of overweight boys at 24 months, however these effects were not present at 36 and 48 months, yet re-merged at 60 and 96 months. In addition, they found that the treatment group members were less at risk of cardiovascular and metabolic diseases in adulthood, particularly the males. Therefore, it is possible that the main effects found at 24 months may re-emerge later in childhood, or indeed, may have longer term implications for adult health. Third, our subgroup analysis identified a number of treatment effects at 36 months for boys in terms of reduced accidents and chest infections, suggesting a continuity of effects beyond 24 months.

Nonetheless, the main treatment effects for wheezing/asthma at 24 months is relevant from a cost-benefit perspective in terms of immediate savings on health care. A cost-benefit analysis of multiple, primarily US-based, HVPs finds returns ranging from $0.21 to $30.46 per $ invested, with a median return of $1.62 (Washington State Institute for Public Policy, 2014). The proportion of the return generated from health care saving generated by the child and the mother amount to ~10% of these returns, thus demonstrating the potential health related cost savings from intervening early in life through home visiting. The main treatment effect on reducing the incidence of wheezing/asthma is also important as asthma is the most common chronic illness in young children (Currie, 2009) and it is often associated with impaired quality of life throughout childhood (Covaciu et al., 2013). Thus, identifying interventions which improve health is significant from both a clinical and cost-benefit perspective. This

intervention, if one accepts the generalization of the results, may therefore provide a vehicle through which policymakers can reduce the socioeconomic gradient in some important dimensions of child health.

Acknowledgements

We would like to thank the Northside Partnership who provided funding for the *Preparing for Life* evaluation through The Atlantic Philanthropies and the Department of Children and Youth Affairs. We are also grateful to all those who participated in and supported this research, especially the participating families and community organisations, the *PFL* intervention staff, and the Expert Advisory Committee. Thanks also to Prof. James Heckman, Prof. Colm Harmon, Prof. Sylvana Cote, Prof. Cecily Kelleher, Prof. Sharon Ramey, and Prof. Craig Ramey for their guidance and advice throughout the project, and the Early Childhood Research Team at UCD Geary Institute for Public Policy for their contributions to the work.

Funding

The trial was funded by the Northside Partnership through funding provided by The Atlantic Philanthropies and the Department of Children and Youth Affairs. The funding source had no involvement in the study design, collection, analysis, and interpretation of the data, in the writing of the report, or in the decision to submit the paper for publication. The authors are fully independent from the funders.

Conflict of Interests

There are no conflicts of interest.

Contributions

OD designed the study. JL and CR contributed to planning and conducting data collection. NF did the data analysis. OD, CR, and NF wrote the paper. All authors contributed to the interpretation of the data. All authors approved the final version. OD is guarantor.

References

Anderson, N.B., Armstead, C.A. 1995. Toward understanding the association of socioeconomic status and health: A new challenge for the biopsychosocial approach. Psychosomatic Medicine 57, 213–225.

Anderson, M. L. 2008. Multiple inference and gender differences in the effects of early intervention: A re-evaluation of the Abecedarian, Perry Preschool, and Early Training Projects. Journal of the American Statistical Association 103, 1481-1495.

Avellar, S.A., Supplee, L.H. 2013. Effectiveness of home visiting in improving child health and reducing child maltreatment. Pediatrics 132, S90.

Bloom, H.S. 2005. Randomizing groups to evaluate place-based programs. In Bloom HS (ed.) Learning More From Social Experiments: Evolving Analytic Approaches. New York, NY: Russell Sage Foundation, 115–172.

Campbell, F., Conti, G., Heckman, J.J., Moon, S.H., Pinto, R., Pungello, E., et al. 2014. Early childhood investments substantially boost adult health. Science 343, 1478-1485.

Case, A., Fertig, A., Paxson, C. 2005. The lasting impact of childhood health and circumstances. Journal of Health Economics 24, 365-389.

Case, A., Lubotsky, D., Paxson, C. 2002. Economic status and health in childhood: The origins of the gradient. American Economic Review 92, 1308-1334.

Chen, E., Matthews, K.A., Boyce, W.T. 2002. Socioeconomic differences in children's health: How and why do these relationships change with age? Psychological Bulletin 128, 295–329.

Covaciu, C., Bergstrom, A., Lind, T., Svartengren, M., Kull, I. 2013. Childhood allergies affect health-related quality of life. Journal of Asthma 50, 522-528.

Culp, A.M., Culp, R., Anderson, J., Carter, S. 2007. Health and safety intervention with first-time mothers. Health Education Research 22, 285-294.

Cunha, F., Heckman, J.J. 2007. The technology of skill formation. American Economic Review 97, 31-47.

Currie, J. 2004. Viewpoint: Child research comes of age. Canadian Journal of Economics 37, 509-527.Currie, J. 2009. Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development, Journal of Economic Literature, 47, 87-122.

Currie, J., Hyson, R. 1999. Is the impact of health shocks cushioned by socioeconomic status? The case of low birth weight. American Economic Review (Papers and Proceedings) 89, 245-50.

Doyle, O. 2013. Breaking the cycle of deprivation: An experimental evaluation of an early childhood intervention. Journal of the Statistical and Social Inquiry Society of Ireland XLI, 92-111.

Doyle, O., Harmon, C., Heckman, J.J., Logue, C., Moon, S. 2013. Measuring investment in human capital formation: An experimental analysis of early life outcomes. NBER Working Paper, No. 19316.

Doyle, O., Harmon, C., Heckman, J.J., Tremblay, R. 2009. Investing in early human development: timing and economic efficiency. Economics and Human Biology 7, 1-6.

Doyle, O., Hickey C. 2013. The challenges of contamination in evaluations of childhood interventions. Evaluation 19, 180-191.

D'Souza-Vazirani, D., Minkovitz, C.S., Strobino, D.M. 2005. Validity of maternal report of acute health care use for children younger than 3 years. Archives of Pediatric and Adolescent Medicine 159, 167-172.

Duflo, E., Glennerster, R., Kremer, M. 2006. Using Randomization in Development Economics Research: A Toolkit. NBER Technical Working Paper, No. 0333.

Eckenrode, J., Campa, M., Luckey, D. W., Henderson, C. R., Cole, R., Kitzman, H., . . . & Olds, D. 2010. Long term effects of prenatal and infancy nurse home visitation on the life course of youths: 19-year follow-up of a randomized-controlled trial. JAMA Pediatrics 164, 9-15.

Fergusson, D.M., Grant, H., Horwood, L.J., Ridder, E.M. 2005. Randomized trial of the Early Start Program of home visitation. Pediatrics 116, e803-e809.

Galobardes, B., Lynch, J.W., Davey Smith, G. 2004. Childhood socioeconomic circumstances and cause-specific mortality in adulthood: systematic review and interpretation. Epidemiological Reviews 26, 7-21.

German, D., Latkin, C. 2012. Social stability and health: Exploring multidimensional social disadvantage. Journal of Urban Health 89, 19-35.

Gomby, D.S., Culross, P.L., Behrman, R.E. 1999. Home visiting: Recent program evaluations-analysis and recommendations. The Future of Children 9, 4-

26.Gravetter, F.J. Wallnau, L. B. 2004. Statistics for the behavioral sciences (6th ed.). Thompson Wadsworth: CA.

Grossman, M. 1972. On the concept of health capital and the demand for health. Journal of Political Economy 80, 223-255.

Guttentag, C.L., Landry, S.H., Williams, J.M., Baggett, K.M., Noria, C.W., Borkowski, J.G., et al. 2014. "My Baby & Me": effects of an early, comprehensive parenting intervention on at-risk mothers and their children. Developmental Psychology 50, 1482-96.

Halfon, N., Shulman, E., Hochstein, M. 2001. Brain development in early childhood. In: N. Halfon, E. Shulman, M. Hochstein (Eds.), Building Community Systems for Young Children. UCLA Center for Healthier Children, Families and Communities. Hayes, A. 1996. Permutation test is not distribution-free: Testing h0 :) = 0. Psychological Methods 1, 184–198.

Heckman, J. J. 2007. The economics, technology and neuroscience of human capability formation. Proceedings of the National Academy of Sciences 104, 13250-13255.

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P., Yavitz, A. 2010. Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope Perry Preschool Program', *Quantitative Economics,* 1(1), 1-46.

Howard, K.S., Brooks-Gunn, J. 2009. The role of home-visiting programs in preventing child abuse and neglect. The Future of Children 19, 119-46.Hofhuis, W., de Jongste, J.C., Merkus, P.J.F.M. 2003. Community child health, public health, and epidemiology: Adverse health effects of prenatal and postnatal tobacco smoke exposure on children. Archives of Disease in Childhood 88, 1086-1090.

Howe, A., Keogh-Brown, M., Miles, S., Bachmann, M. 2007. Expert consensus on contamination in educational trials elicited by a Delphi exercise. Medical Education 41, 196–204.

Kraemer, S. 2000. The fragile male. British Medical Journal 321, 1609-1312.

Keller, B. 2012. Detecting treatment effects with small samples: The power of some tests under the randomization model. Psychometrika 77, 324-338.

Kemp, L., Harris, E., McMahon, C., Matthey, S., Vimpani, G., Anderson, T. 2011. Child and family outcomes of a long-term nurse home visitation programme: a randomised controlled trial. Archives of Disease in Childhood 96, 533-540.

Klinnert, M.D., Liu, A.H., Pearson, M.R., Ellison, M.C., Budhiraja, N., Robinson, J.L. 2005. Short-term impact of a randomized multifaceted intervention for wheezing infants in low-income families. Archives of Pediatric and Adolescent Medicine 159, 75-82.

Komro, K., Flay, B., Biglan, A. 2011. Creating nurturing environments: A science-based framework for promoting child health and development within high-poverty neighborhoods. Clinical Child and Family Psychology Review 14, 111-134.

Koniak-Griffin, D., Verzemnieks, I.L., Anderson, N.L., Brecht, M.L., Lesser, J., Kim, S., et al. 2003. Nurse Visitation for Adolescent Mothers: Two-Year Infant Health and Maternal Outcomes. Nursing Research 52, 127-136.

Ludbrook, J., Dudley, H. 1998. Why permutation tests are superior to t and F tests in biomedical research. The American Statistician 52, 127-32.

Marmot, M. 2005. Social determinants of health inequalities. The Lancet 365, 1099-1104.

Mayer, S.E. 2002. The influence of parental income on children's outcomes. Knowledge Management Group, Ministry of Social Development; Wellington, New Zealand.Mewhort, D.J.K. 2005. A comparison of the randomization test with the F test when error is skewed. Behavior Research Methods 37, 426–435.

Peacock, S., Konrad, S., Watson, E., Nickel, D., Muhajarine, N. 2013. Effectiveness of home visiting programs on child outcomes: A systematic review. BMC Public Health 13, 17.

Pless, C.E., Pless, I. 1995. How well they remember: The accuracy of parent reports. Archives of Pediatric and Adolescent Medicine 149, 553-558.

Pocock, S. J., Hughes, M. D., Lee, R. J. 1987. Statistical problems in the reporting of clinical trials. A survey of three medical journals. New England Journal of Medicine 317, 426-32.

Romano, J.P., Wolf, M. 2005. Exact and approximate stepdown methods for multiple hypothesis testing. Journal of American Statistical Association 100, 94-108.

Rosenzweig, M.R., Schultz T.P. 1982. The behavior of mothers as inputs to child health: the determinants of birth weight, gestation, and rate of fetal growth. In: V.R. Fuchs, ed., Economic Aspects of Health, University of Chicago Press for the National Bureau of Economic Research, Chicago, pp 53-92.

Rubin, D. B. 1987. Multiple imputation for nonresponse in surveys. New York: Wiley.

Sanders, M.R., Kirby, J.N., Tellegen, C.L., Day, J.J. 2014. The Triple P-Positive Parenting Program: A systematic review and meta-analysis of a multi-level system of parenting support. Clinical Psychology Review 34, 337-357.

Sanders, M.R., Markie-Dadds. C., Turner. K. 2003.Theoretical, scientific and clinical foundations of the Triple P-Positive Parenting Program: A population approach to the promotion of parenting competence. Parenting Research and Practice Monograph 1, 1-21.

Smith, J., 1999. Healthy bodies and thick wallets: The dual relation between health and economic status. Journal of Economic Perspectives 13, 145-166.Stöber, J. 2001. The Social Desirability Scale – 17 (SDS-17): Convergent validity, discriminant validity, and relationship with age. European Journal of Psychological Assessment, *17*, 222-232.

Schwarz, G. 1978. Estimating the dimension of a model. The Annals of Statistics, 6, 461-464.

Sweet, M. A., Appelbaum, M. I. 2004. Is home visiting an effective strategy? A meta-analytic review of home visiting programs for families with young children. Child Development 75, 1435-1456.The Asthma Society of Ireland, 2013. https://www.asthma.ie/get-help/learn-about-asthma/diagnosing-asthma

Washington State Institute for Public Policy, 2014. Benefit-Cost Results – Public Health and Prevention. http://www.wsipp.wa.gov/BenefitCost

Appendix A

Example of *PFL* 'Tip Sheet'

Common illnesses

Babies get ill at some stage, in most cases your baby can be cared for at home as these bouts of illness pass quickly. The following are some tips on what to look out for and what you can do to care for your baby.

Illness	*Things you can do to help your baby*
Fever	• The normal temperature for a baby ranges from 36.5 to 37.2 degrees Celsius. This can be taken under baby's arm. • You should seek medical advice if your baby has a temperature especially if they are looking unwell.
Coughs & Colds	• Keep your baby warm. • Give your baby fluids to drink such as the usual milk feeds. • Seek medical advice if your baby finds it hard to breathe or the cough doesn't go away.
Vomits	• Small vomits are normal after feeds and your baby will grow out of it. • You should seek medical advice if your baby vomits large amounts, forceful or repeatedly.
Sticky Eyes	• Seek medical advice from your doctor or pharmacist in case there is an infection.
Thrush	• On your baby's tongue and mouth, thrush is a white spotted fungus that doesn't brush away when you touch it. • On your baby's bottom, thrush looks like a red rash with white spots. • You should contact your doctor or pharmacist on how to treat thrush.
Tummy Upsets	• If your baby has an upset tummy with vomiting, diarrhoea or both, then they can get dehydrated if they don't drink enough fluids. • Offer your baby small amounts of fluid regularly. • You should seek medical advice for treatment should the problem continue.

Contacting your Doctor

You should always contact a doctor regarding your baby's health if he/she experiences any of the following:

- A purple or red rash that looks unusual.
- A raised or sunken soft spot (fontanelle) on his or her head.
- A fever
- Seems much paler and sleepier than usual and is hard to wake up.
- Has an unusual, non - stop high pitched cry or scream.
- Has a fit or a convulsion.
- Has difficulty breathing.
- Goes blue around the lips or face.
- Is not feeding normally.
- Has unusually dry nappies or less than 3 wet nappies in one day.
- Has diarrhoea at each nappy change.
- Is upset due to a fall or bump to the head.
- Gets an electric shock.
- Is burned or scalded.
- Is bitten by an animal.

If a serious accident/incident happens don't delay getting help, telephone 999 or 112 asking for an Ambulance, Fire Bigade or Gardai.

Appendix B

Figure B1 Consort Diagram

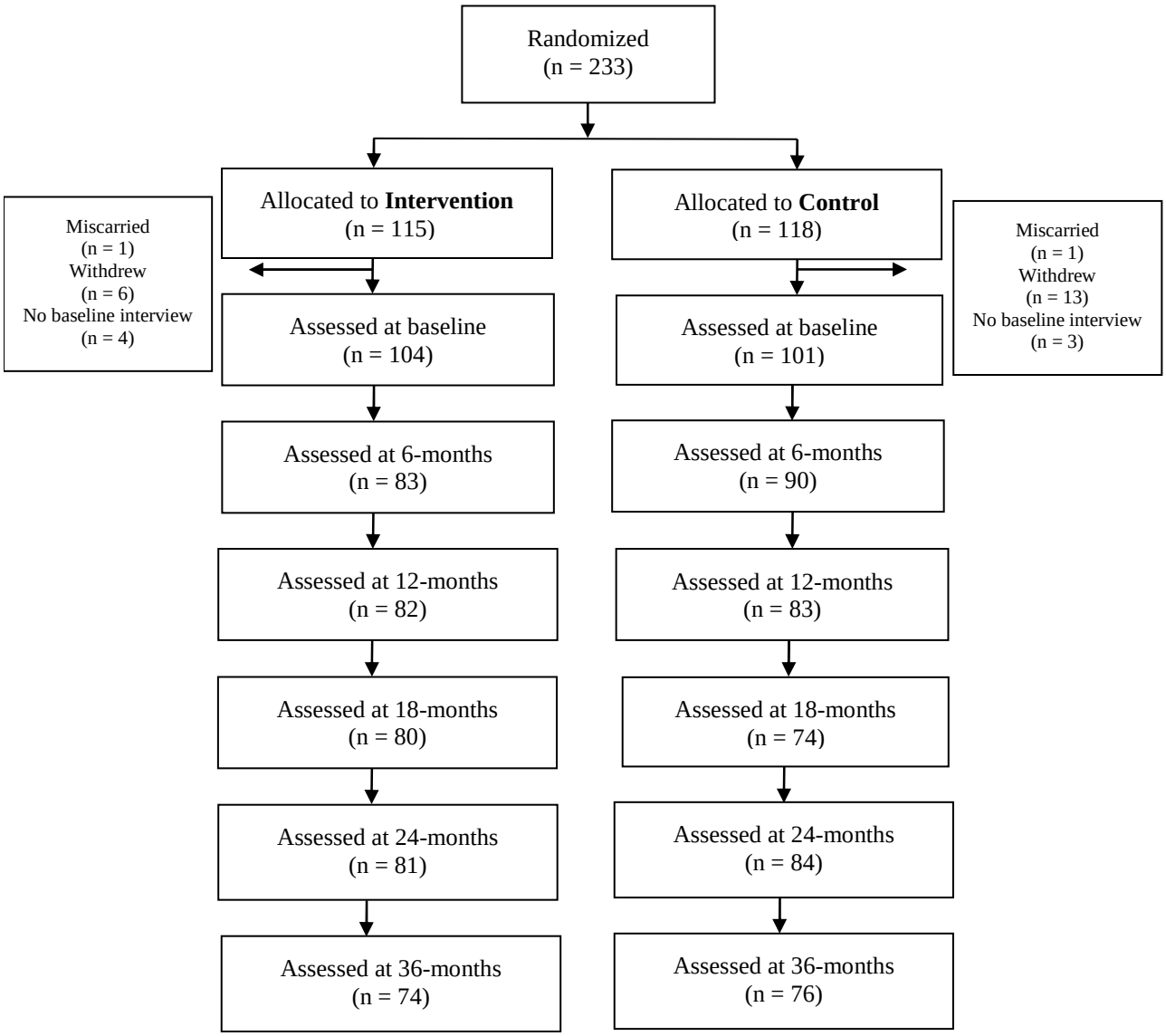

Table B1 Most prevalent health problems at 36 months

	PFL Sample %	National Sample %
Chest Infections	0.32	~
Ear Infections	0.26	~
Wheezing/Asthma	0.17	0.06
Accident	0.16	~
Skin Problems	0.11	0.04
Viral Infection	0.06	~
Sight/Eye Problems	0.05	~
Severe Diarrhoea	0.05	~
Severe Vomiting	0.03	~
Fits/Convulsions	0.02	~
Sleeping Problems	0.01	~
Failure to Gain Weight/Grow	0.01	~
Flu	0.01	~
Constipation	0.01	~
Pneumonia	0.01	~
Kidney Infection	0.01	~

Note that the two measures in the *PFL* and national samples are not directly comparable. For the PFL sample the estimate is based on the proportion of children taken to the GP, health centre or casualty for health problems in the last 12 months. The national sample is based on the Growing up in Ireland (GUI) data which is a representative sample of ~9,000 3 year old children in Ireland. For GUI the estimate is based on the proportion of children diagnosed with a longstanding illness by a doctor at age 3.

Table B2 Predictors of attrition

	Intervention Group	**Control Group**
6 Months	TIPI *emotional stability* score (+), AAPI *parental expectations of children* score (-), AAPI *parental empathy towards children's needs* score (+), has a medical card (+), support from relations (-), married (-), low education (+), employed (-), Irish national (-), has ever taken illegal drugs (-) (10 variables)	WASI *perceptual reasoning* score (+), Pearlin *self-efficacy* score (~), TIPI *conscientiousness* score (-), AAPI *non use of corporal punishment* score (-), AAPI *children's power and independence* score (-), KIDI score (-), low education (+), number of children (-) (8 variables)
12 Months	VASQ *insecurity* score (+), AAPI *parental expectations of children* score (-), AAPI *parental empathy towards children's needs* score (+), drinks alcohol during pregnancy (-) (4 variables)	Rosenberg self-esteem score (+), AAPI *children's power and independence* score (-), KIDI score (-), age (+), number of children (-), low education (+), saves money regularly (-), meets friendly regularly (+), Irish national (-) (9 variables)
18 Months	WASI *perceptual reasoning* score (-), AAPI *parental expectations of children* score (-), AAPI *parental empathy towards children's needs* score (-), support from relatives (-), drinks alcohol during pregnancy (-), uses birth control (+), iron supplements during pregnancy (-), lives with a parent (+) (8 variables)	WASI *verbal ability* score (-), Pearlin *self-efficacy* score (-), Rosenberg self-esteem score (-), VASQ *proximity seeking* score (+), Consideration of Future Consequences Scale score (+), AAPI *parental empathy towards children's needs* score (-), AAPI *children's power and independence* score (-), KIDI score (-), has a medical card (+), exercises regularly (-), has ever taken illegal drugs (+), lives with a parent (-) (12 variables)
24 Months	WASI *perceptual reasoning* score (-), AAPI *parental expectations of children* score (-), AAPI *parental empathy towards children's needs* score (+), support from relatives (-), drinks alcohol during pregnancy (-), knows neighbours (+)(6 variables)	Eats healthily (-), exercises regularly (-), has ever taken illegal drugs (+), satisfaction with neighbourhood (+), Irish national (-) (5 variables)
36 Months	WASI *perceptual reasoning* score (-), AAPI *parental expectations of children* score (-), AAPI *parental empathy towards children's needs* score (-), AAPI *children's power and independence* score (+), support from relatives (-), satisfaction with neighbourhood (-) (6 variables)	WASI *verbal ability* score (-), TIPI *agreeableness* score (-), TIPI *conscientiousness* score (+), TIPI *openness* score (-), AAPI *parental expectations of children* score (-), AAPI *parental empathy towards children's needs* score (-), KIDI score (-), age (-), married (+), experience financial difficulty (+), prior physical health condition (-), exercises regularly (-), has ever used drugs (+), satisfaction with neighbourhood (+)(14 variables)

Note: The table includes the set of variables which resulted in the lowest BIC in models of attrition and are included in the logistic model used to generate the IPW weights. (+) and (-1) indicates a participant with this characteristic has a higher/lower probability of dropping out. Scores on *The Ten Item Personality Inventory* (*TIPI*) range from 1-7 and higher scores are indicative of a greater tendency towards the corresponding personality trait. Positive parenting attitudes was measured using the *Adult Adolescent Parenting Inventory (AAPI)* which measures approaches to parenting and provides an indicator of the endorsement of abuse/neglect. Higher scores indicate a high risk of abuse/neglect. Vulnerable attachment was measured using the *Vulnerable Attachment Style Questionnaire (VASQ)* which assesses respondents' interactions and dependence on other people. Scores above 15 are indicative of depressive disorders. IQ was measured 3 months post-birth using the *Wechsler Abbreviated Scale of Intelligence (WASI)*. Two scores were derived representing verbal ability and perceptual reasoning with higher scores indicating higher ability. Scores on the *Consideration of Future*

Consequences Scale range from 1-15 and higher scores indicate a greater consideration of the future consequences of present behavior. Self-efficacy is measured using the *Pearlin Self-Efficacy* scale. Scores on this scale range from 0-4 with higher scores indicating that the respondent had a stronger feeling of control over her life. The Rosenberg self-esteem scale ranges from zero to 18 with higher scores indicating more maternal self-esteem. The *Knowledge of Infant Development (KIDI)* score represents the percentage of correct responses to questions relating to child development milestones. Higher scores indicate more knowledge of infant development. Low education represents participants who left school after they completed a statewide examination at age 15 to 16 years. Physical health condition indicates whether the mother has ever been diagnosed with any 22 listed physical health conditions. 'Has medical card' is a binary variable indicating the mother is entitled to free medical card in Ireland based on a mean assessment. 'Married' is a binary variable indicating that the mother is married at the time of the baseline assessment. 'Employed' is a binary variable indicating that the mother is in employment (full or part time) at the time of the baseline assessment. 'Has ever taken illegal drugs' is a binary variable indicating whether the mother has ever taken illegal drugs in the past. 'Irish national' is a binary variable indicating whether the mother defines herself as Irish ethically. 'Uses birth control' is a binary variable indicating whether the mother regularly used birth control when she became pregnant. 'Iron supplements during pregnancy' is a binary variable indicating whether the mother is taking iron supplements while pregnant. 'Lives with a parent' is a binary variable indicating whether the mother lives with any of her parents or her partner's parents. 'Age' is the mother's age during pregnancy. 'Number of children' is the total number of children the mother has including the child she was pregnant with at the time of the baseline assessment. 'Saves money regularly' is a binary variable indicating whether the mother saves money on a regular basis. 'Experience financial difficulty' indicates whether the mother reports meeting financial difficulty on a seven point ranging from very easily to with great difficulty. 'Exercises regularly' is a binary variable indicating whether the mother exercises at least 3 times per week for a minimum of 20 minutes. 'Eats healthily' is based on maternal responses to how healthy their eating habits are on a five point scale corresponding to very unhealthy to very healthy. 'Support from relatives' indicates the amount of support mothers felt they received from their relatives on a five point scale ranging from no support to a lot of support. 'Meets friends regularly is based on the frequency of meeting with friends or relatives not living in their household on a five point scale corresponding to on most days to never. 'Satisfaction with neighbourhood' is based on maternal responses to how satisfied they are with their own neighbourhood or area based on a five point scale corresponding to very dissatisfied to very satisfied. 'Knows neighbours' is based on maternal reports how many neighbours they know personally on a five point scale ranging from none to more than 10 people.

Table B3. Baseline maternal characteristics of intervention and control groups

	Intervention (n=104)	**Control (n=101)**	***p*-value**[i]
Weeks pregnant at program entry, mean (SD)	21.59 (7.85)	21.34 (6.95)	.806
Age, mean (SD)	25.46 (5.85)	25.30 (5.99)	.840
Married	14%	18%	.514
Partnered (including married)	78%	84%	.250
Living with parent(s)	57%	47%	.152
First time mother	54%	50%	.548
Low education	34%	40%	.377
Employed	37%	40%	.652
Saves money regularly	47%	51%	.719
Resides in social housing	55%	55%	.985
Prior physical health condition	75%	62%	.053*
Prior mental health condition	28%	24%	.511
Smoking during pregnancy	51%	48%	.610
Drinking during pregnancy%	25%	27%	.761
Drugs ever used	13%	15%	.761
IQ, mean (SD)	82.06 (12.32)	80.91 (12.88)	.519
Vulnerable attachment, mean (SD)	18.24 (3.77)	17.82 (3.98)	.447
Positive parenting attitudes, mean (SD)	5.25 (1.38)	5.12 (1.42)	.499
Self-efficacy, mean (SD)	2.77 (0.63)	2.88 (0.60)	.226
Self-esteem, mean (SD)	12.82 (2.69)	12.78 (2.86)	.930
Knowledge of infant development, mean (SD)	72.25 (7.60)	69.82 (8.19)	.028**

Notes: [i] two-tailed p-value from an individual permutation test with 100,000 replications. * $p < .10$, ** $p < .05$, *** $p < .10$. Low education represents participants who left school after they completed a statewide examination at age 15 to 16 years. Physical/mental health conditions indicate whether the mother has ever been diagnosed with any of the listed conditions. IQ was measured 3 months post-birth using the Wechsler Abbreviated Scale of Intelligence. Vulnerable attachment was measured using the Vulnerable Attachment Style Questionnaire which assesses respondents' interactions and dependence on other people. Scores above 15 are indicative of depressive disorders. Positive parenting attitudes was measured using the Adult Adolescent Parenting Inventory which measures approaches to parenting and provides an indicator of the endorsement of abuse/neglect. Higher scores indicate a high risk of abuse/neglect. The Pearlin Self-Efficacy scale ranges from zero to four with higher scores indicating higher self-efficacy. The Rosenberg self-esteem scale ranges from zero to 18 with higher scores indicating more maternal self-esteem. The Knowledge of Infant Development score represents the percentage of

correct responses to questions relating to child development milestones. Higher scores indicate more knowledge of infant development.