UCD GEARY INSTITUTE FOR PUBLIC POLICY
DISCUSSION PAPER SERIES

Should Early Health Investments Work? Evidence from an RCT of a Home Visiting Programme

Deirdre Coy
Geary Institute for Public Policy, University College Dublin

Orla Doyle
Geary Institute for Public Policy, University College Dublin

Geary WP2020/06
July 21, 2020

Should Early Health Investments Work? Evidence from an RCT of a Home Visiting Programme

BY DEIRDRE COY AND ORLA DOYLE*

Evidence for the short-term impact of early intervention on childhood health is weak and inconsistent. Using rigorous methods, careful hypothesis setting, and socioeconomic contextualisation, we examine the impact of an Irish home visiting programme on child health. The treatment provides mentoring visits from pregnancy until school entry to improve child outcomes through positively affecting parenting. In a context where socioeconomic inequalities in health have yet to emerge, modest effects by age four are found, driven by reduced hospital attendance. Conflicting reports in the literature may thus arise from an over-expectation of hypothesized effects and failure to account for social contexts.

Keywords: Home visiting programme, randomised controlled trial, child health, health inequality, human capital formation

JEI Classification: I14 Health and Inequality, I120 Health Behavior, J13 Fertility; Family Planning; Child Care; Children; Youth

* Acknowledgments. This research is funded by the Health Research Board SPHeRE/2013/1. The trial was registered with controlled-trials.com (ISRCTN04631728) and the AEA RCT Registry (AEARCTR-0000066). All study procedures were approved by the UCD Human Research Ethics Committee, the Rotunda Hospital's ethics committee, and the National Maternity Hospital's ethics committee. Coy, University College Dublin, UCD Geary Institute for Public Policy, Belfield, Dublin 4, Ireland (email: deirdre.coy@ucdconnect.ie). Doyle, University College Dublin, B007 UCD Geary Institute for Public Policy, Belfield, Dublin 4, Ireland (email: orla,doyle@ucd.ie).

Consensus on the positive impact of early intervention on life-cycle human capital formation is building, however conflict persists on its effect on child health (Peacock et al. 2013; Avellar and Supplee 2013). Socioeconomic status, parenting, the prenatal and postnatal environment, and birth outcomes are key factors in the formation of health (Case, Lubotsky, and Paxson 2002; Dooley and Stewart 2007; Gennetian et al. 2010; McGovern 2013; Robinson et al. 2008) and ill health during childhood contributes to poorer labour market outcomes through, and independent of, adult health (Case, Fertig, and Paxson 2005; Currie et al. 2010). This may be mitigated by investment in health directly and indirectly, from before conception to the prenatal period and beyond (van den Berg and Pinger 2016; Currie 2011; Timmons et al. 2012). While health development is possible later in life, the return on investment is inversely related to age (Heckman 2006). Intervention programmes beginning early in the lifecycle may reduce inequalities in health through improving maternal health behaviours, the home environment, and parent-child attachment (Nievar, Van Egeren, and Pollard 2010). However, empirical evidence for these approaches does not consistently align with theory, and weak effects on health are often observed (Filene et al. 2013). Using an Irish early intervention programme that is effective in improving child development (Doyle 2020), we examine the programme's impact on health, while correcting for methodological issues that may contribute to the ambiguity in the field (Kader, Sundblom, and Elinder 2015; Avellar and Supplee 2013). Placing the modest treatment effects in context, we argue that an over-expectation of hypothesised effects and a disregard for social context may explain the conflicting reports in the literature.

Comprehensive early intervention programmes based on visits to the child's home during peak periods of malleability may play a role in fostering both short and long-term health capital (Heckman 2012; Conti and Heckman 2013). Home visiting programmes are designed to promote children's health and development by allowing for programme customisation based on home visitor observations of the family environment and providing a gatekeeper to existing health, educational, and social services. There is some evidence that frequent and long-term programme structures, may enhance and sustain positive health behaviours (Avellar and Supplee 2013; Kader, Sundblom, and Elinder 2015; Nievar, Van Egeren, and Pollard 2010).

Cognitive and non-cognitive capabilities operate dynamically and complementarily with health in shaping children's development (Heckman 2012). Thus, fostering an array of skills should improve health formation (Heckman 2006). However evidence on the health effects of early intervention is mixed. Factors including methodology, research design, and hypothesis setting may contribute to this ambiguity (Peacock et al. 2013; Avellar and Supplee 2013; Filene et al. 2013; Kader, Sundblom, and Elinder 2015). The strongest effects should be observed for programme components that directly target health, such as nutrition education (Avellar and Supplee 2013). Home visiting programmes should also have a greater impact where socioeconomic inequalities in health are

evident (Filene et al. 2013; Heckman 2012). However, where the cost of behavioural change is high, such as smoking cessation, effects may be smaller (Sandner et al. 2018). Furthermore, programme components may contradict one another, affecting the magnitude or direction of hypothesised effects. For example, promoting centre-based childcare, which is associated with many positive child outcomes, also increases the burden from infectious disease (Enserink et al. 2015). Due consideration must also be given to measurement regarding the reality and severity of health problems, such as the early diagnosis of asthma (Apouey and Geoffard 2013; Kader, Sundblom, and Elinder 2015). Thus, in this paper we focus on the motivation and choice of health outcomes, and use socioeconomic contextualisation to shape hypotheses setting and interpret treatment effects.

With respect to methodological considerations, randomised controlled trials (RCT), the design for this study, are used in the most robust evaluations. However, issues often arise regarding sample size and the reporting on major but rare adverse health events (Ioannidis, Stanley, and Doucouliagos 2017). Sample selection is also a concern, as inference may be hampered by attrition (Imbens and Wooldridge 2009). Additionally, there is a trade-off in the reporting on the complexity of health through multiple hypotheses against an increased family-wise error rate (Heckman et al. 2010). Thus, we apply rigorous techniques and robustness checks to address these issues.

This paper examines the impact of a five-year mentor-led home visiting programme on health formation, and its associated behaviours and health service use. Childhood diagnoses and maternal health are secondarily examined. We estimate programme impacts from ages one to four. Overall, health service use and poor health in this low SES sample is high. In line with meta-analysis of similar programmes, we find modest effects. We identify treatment effects at ages two and three on hospital use, but few effects for parent-reported measures. By year four there are no treatment effects, as health in both groups generally improves. Using inverse probability weights to account for attrition generates slightly more conservative results. Treatment effects appear to be driven by the severity of conditions rather than health behaviours. We discuss the results using robustness tests examining sample selection, estimation methods, and trial design. While attrition is high, it appears to be random, and the estimation samples remain balanced. In addition, despite the reduced sample size, the trial is sufficiently powered to detect moderate effects. The findings suggest that measurement may be an issue when examining subjective and rare outcomes, particularly when small effect sizes are anticipated and the complexity of health reporting is noted. Comparing the sample to a large national cohort, we find that socioeconomic inequalities in health are not evident under the age of three in Ireland, further indicating that limited effects should be anticipated within the timeframe analysed.

These findings suggest that differences in the literature may arise from an over-expectation of the size of treatment effects, particularly where socioeconomic inequalities in health appear later in childhood. We argue that in settings where social security and affordable healthcare are widespread, modest effects on health should be expected in the early years, alongside large initial gains in cognitive and non-cognitive development. Long-term follow-up, when health inequalities are larger, may be needed to examine the impact of early intervention programmes on life-cycle health formation.

I. Background

A. Home Visiting Programmes

Home visiting programmes typically involve nurse, social worker, mentor, or paraprofessional led home-based interventions that aim to positively affect children through improving parenting practices (Miller and Macdonald 2011). These interventions are motivated by the primary role of the family rather than the school in driving early inequalities (Heckman 2006). Home visiting programmes differ from centre-based interventions through the convenience of a home setting, alleviating transport and childcare barriers, and allowing greater knowledge of the family background (Glenton et al. 2013; Nievar, Van Egeren, and Pollard 2010). The long-term, frequent nature of home visits are particularly useful when targeting behavioural change and continuous learning, such as tackling obesity (Kader, Sundblom, and Elinder 2015; Wagner and Heinrich-weltzien 2017; Salvy et al. 2017). Home visiting may also affect service use and continuity of care through acting as gatekeepers to health, educational, and social services (Elkan et al. 2004). Reviews of home visiting programmes typically find positive effects on child development, maltreatment, parenting, the home environment, maternal outcomes, and sometimes health (Heckman 2008; Peacock et al. 2013; Avellar and Supplee 2013). However programme type, target group, and social insurance setting are factors in determining intervention success (Nievar, Van Egeren, and Pollard 2010; Peacock et al. 2013) and evidence on the causal pathways of treatment effects is not well-established (Filene et al. 2013; Heckman 2012).

A few systematic reviews have collated the results of home visiting programmes. Peacock et al (2013), including 21 studies, focus on health and development until age six. They find evidence of significant treatment effects depending on the structure of the programme, notably a prevention of maltreatment and appropriate birth weight when visits begin prenatally, as well as consistent benefits for cognition and behaviour, and reduced health issues in older children, and appropriate weight gain in younger children. Benefits are mainly limited to more disadvantaged families. Three of the 21 studies are based in countries with broad social security nets, where few significant findings arise and are noted to

be spurious or counter to hypothesis. Study quality is also flagged as restricting the scope of the review.

Avellar and Supplee (2013) examine the impact of home visiting on health among 12 mainly US-based studies, of which two overlap with Peacock et al (2013). Results are most consistent for social-emotional and behavioural outcomes. They find modest benefits for health care use, including immunisations, hospital attendances, and maltreatment. Few effects are found for other health and birth outcomes. They caution that for most studies, more insignificant findings are reported than significant, and most evaluations fail to account for type I error arising from multiple hypothesis testing.

In a meta-analysis of 51 US home visiting programmes, mean effect sizes are significantly different from zero for maternal, parenting, and child cognitive outcomes (Filene et al. 2013). There are no significant differences in birth outcomes, physical health, or maltreatment. However, professional home visitors and teaching behaviour management techniques are predictors of improved physical health. Finally, in a meta-analysis of 60 US programmes, Sweet and Appelbaum (2004) find that home visiting has a small, significant effect on social-emotional development, but they do not report on other health outcomes.

Disparities across studies are evident between countries. Theoretically, these differences may arise from access to healthcare and corresponding shifts in the socioeconomic gradient in health (Propper, Rigg, and Burgess 2007). In Ireland, where the programme being examined is based, there is no evidence of the socioeconomic gradient in health at nine months (Nolan and Layte 2014), however there is a gradient at age nine despite heavily subsidised access to healthcare among those with low incomes. Sandner et al (2018) examine a home visiting programme in Germany, where access to health services is relatively open. They find few effects by age two on health, health behaviours and service use in children or their mothers, except on maternal mental health and child dental health. The authors argue that effects are only observed on outcomes where the health gradient is steep and that knowledge of the detrimental effects of health behaviours is sufficient, but too costly to change. Recently, there have been two adaptions of the US-based Nurse Family Partnership programme in the Netherlands and the UK. While some health effects at birth are evident in the Netherlands, there are no effects on birth outcomes or health at 24 months in the UK trial (Mejdoubi et al. 2014; Robling et al. 2016). Barnes (2016), in a commentary on the Robling et al. paper, contends that these differences may arise from the selection criteria into the programme, differences in parent-reported and administrative recorded data, and the prioritisation of outcomes across settings.

Thus, the evidence on the impact of home visiting on health is unclear. Meta-analyses indicate small effects on social-emotional (d[1] = 0.10) and physical (d =

[1] d is Cohen's d, a measure of effect size.

0.11) health, with a statistically significant effect on the former only (Sweet and Appelbaum 2004; Filene et al. 2013). Some individual studies report positive effects on birth outcomes, health problems in older children, BMI among younger children, service use, immunisations, and maltreatment. Programme design may cause heterogeneous effects. For example, while programmes are more effective in lower SES groups, they may be less effective for the very disadvantaged (Peacock et al. 2013). When many outcomes are studied, there are often more insignificant than significant findings. Study quality is also a concern, with many failing to address methodological issues such as type I and type II error rates arising from multiple hypothesis testing and trial power, short-term follow-up, and poor quality measurement. Programme effectiveness may be context dependent, with socioeconomic inequality and health system characteristics playing a role.

B. Measuring Child Health

A broad range of outcomes are reported in the home visiting literature on health, however there is no consensus on defining its scope or measurement (Elkan et al. 2004). Given the home visiting context, emphasis should be on health problems that arise in disadvantaged groups, such as mental health issues, obesity, respiratory disease, tooth decay, and problems arising from accidents and injuries (Wickham et al. 2016; Allin and Stabile 2012). However, not all problems where socioeconomic differences emerge have long-term effects, such as more frequent minor health shocks (Anderson et al. 2012; Beattie, Gorman, and Walker 2001). It is also useful to examine whether it is the frequency of health shocks or how families react to shocks that is affected by the programme (Condliffe and Link 2008). Attention should also be given to the predictors of long term health and labour market outcomes, such as high health service use including non-urgent emergency department (ED) usage, chronic conditions like obesity and asthma, traumatic injury, mental health problems, health behaviours and maternal health (Currie et al. 2010; Case, Fertig, and Paxson 2005; Apouey and Geoffard 2013; Maslow et al. 2011; Reilly and Kelly 2010; Christakis et al. 2001). Reasonable expectation should also be placed on programme impact and outcomes that are directly targeted by the programme (Avellar and Supplee 2013).

In early childhood, measurement issues arise in selecting outcomes that are reliable as children age. Some diagnoses can be unreliable as true indicators of disease and can be a function of access to healthcare: for example differential effects are expected across SES and treatment groups for chronic diseases such as asthma diagnosis. Asthma is difficult to diagnose in the first five years, therefore early diagnosis is associated with healthcare access and SES. Where a programme increases health literacy, a treatment child may be more likely to be diagnosed with asthma than a control child who has asthma but has not yet been diagnosed (Pedersen et al. 2011). There is also disagreement about the age at which socioeconomic health inequalities begin to emerge, although there is

strong evidence that they are inevitable even in countries with comprehensive health service and welfare structures (Apouey and Geoffard 2013). Thus, our expectations for treatment effects must be considered in line with the anticipated emergence of inequalities. Maternal health and health behaviours, an important factor in determining children's future health, should be considered in periods when fewer inequalities are expected, while acknowledging that it is more difficult to reduce inequalities in health with age (Gall et al. 2019; Allin and Stabile 2012; Perry 2008; Bartlett et al. 2004). Additionally, in considering the socioeconomic gradient, relying on subjective ratings of health may not be sufficient. Evidence on the predictive power of reported poor health in children for adult outcomes is limited, and there is mixed evidence on whether the socioeconomic gradient is weaker for objective measures of health, therefore alternative measures from administrative records and direct measurement may be more reliable (Case, Lubotsky, and Paxson 2002; Nolan and Layte 2014).

Another consideration arises when sample sizes are small. Thus, more commonly arising illnesses and injuries that are predictors of later development should be prioritised over important, but rare outcomes such as major brain trauma (Ioannidis, Stanley, and Doucouliagos 2017; Van As et al. 2016; Sariaslan et al. 2016). Local infections, injuries, and respiratory, gastrointestinal, ophthalmological, and ear conditions are common hospital diagnoses in childhood and are linked to poor general and psychiatric health in adulthood, increased school drop-out, and lower labour market participation (Downing and Rudge 2006; Currie et al. 2010; Bongers et al. 2010; Rahi, Cumberland, and Peckham 2009; Welch and Dawes 2007; Khandaker et al. 2012). Again, age of incidence must be considered: infection and illness typically drive hospital attendance in the first two years, with accident and injury becoming more prevalent as children become active (Downing and Rudge 2006; Olds, Henderson, and Kitzman 1994).

When selecting trial outcomes, careful selection must be made, with consideration that proxies can be misrepresentative, objective reports can contrast with subjective measures, differential treatment effects may distort interpretation, and health issues can emerge at different ages. All the outcomes examined in this paper are detailed in appendix A1, including the data source, purpose of examination, measurement issues and hypothesised programme impact. Evidence on socioeconomic inequalities and life-cycle effects are also recorded. The table also includes the rationale for excluding commonly examined outcomes such as dental visits – primarily due to data availability, selection of a similar but superior alternative, and ambiguity in hypothesis setting.

C. The Intervention - Preparing for Life Programme

Preparing for Life (PFL) is a home-visiting programme for families in a disadvantaged community of North Dublin, Ireland. The programme was

evaluated using a randomised controlled trial and aims to improve school readiness – the social, emotional, cognitive, and physical capabilities needed to benefit from schooling – through early intervention from the prenatal period until the child begins school.

The relative disadvantage status of the PFL cohort can be highlighted using data from Growing Up in Ireland (GUI), a nationally representative longitudinal dataset of Irish children collected during an overlapping period.[2] Relevant data are available at nine months and three years. The PFL mothers are four years younger on average than the national cohort. Fifty-four percentage points fewer PFL mothers are married and they are less likely to have a partner in the home. They leave school one year earlier and are less likely to be employed. Equivalised household income is, on average, 43 percent lower in PFL homes, while receipt of social welfare and housing is more common. Private health insurance is more common in the national sample (57%) than in the PFL cohort (8%), and medical card uptake – for cheaper access to health services - in the PFL cohort is more than twice that of the GUI cohort.[3]

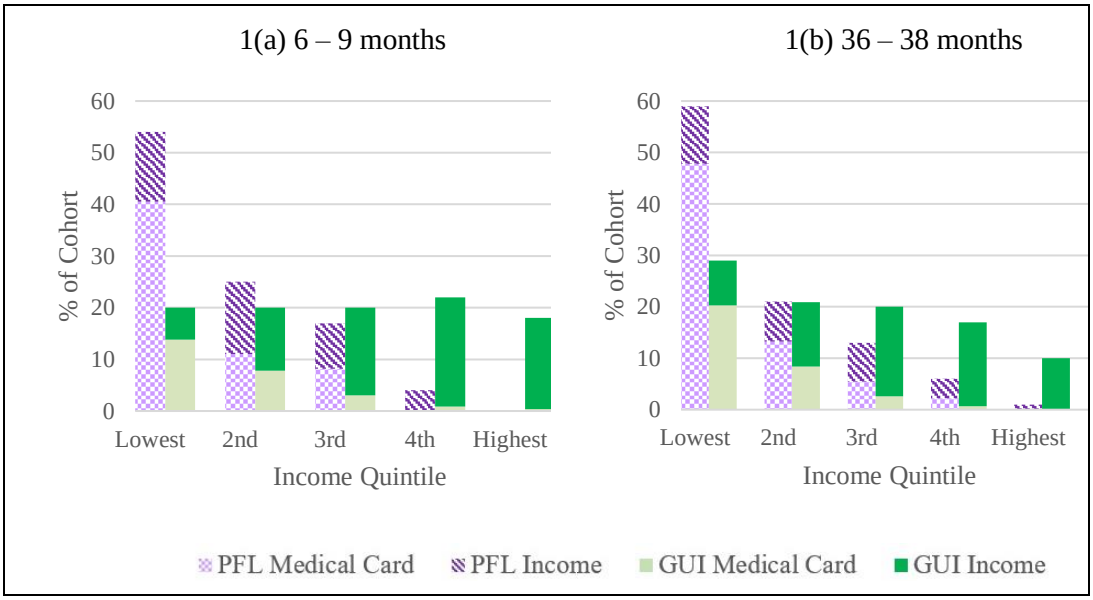
Figure 1 below examines both cohorts based on income. Using the GUI first year income quintiles as a base, 79 percent of the PFL cohort has equivalised household income in the lowest two quintiles. By age three, during the Irish recession, both groups have lower incomes, with 50 percent of the GUI cohort and 80 percent of the PFL cohort in the lowest two quintiles. As a possible mitigating factor to differences in income, medical card uptake is high in poorer households. Medical card possession is higher across all income groups for the PFL cohort. This could arise from better knowledge of social welfare processes in a close-knit, disadvantaged community, or from receipt of medical cards on a

[2] The GUI Infant cohort were born between December 2007 and June 2008. Wave one of data collection took place in 2008-2009 when 11,134 participants – one third of eligible children – were nine months old. Wave two was collected in 2010-2011 when 9,793 participants were 36 months old. PFL participants were born between April 2008 and March 2011. Wave one data was collected between 2008 and 2011, when 172 participants were six to nine months old. In wave two (2011-2014) ages ranged from 36 to 38 months among the 151 participants.

[3] The Irish health system is two-tiered. All residents are entitled to healthcare through the tax-funded public system. Co-payments are charged for some services, for example €100 for an emergency department attendance without referral. Some services, such as primary and dental care are predominantly private. Depending on age, income, illness, and disability status, 30 percent of the population are eligible for reduced or nullified medical fees. These "medical card" holders have subsidised access to private functions of the system and are entitled to free GP, public inpatient and outpatient, dental, aural, optical, maternity and infant care, public health nursing, social work, and some counselling services. Those with income just above the medical card threshold are eligible for free GP visits, however uptake was low over the PFL trial period (Callan et al. 2015). During PFL roll-out, dependents usually had the same entitlement levels as their guardians. Private health insurance is used for hospital, consultant, and elective services. GP, dental, and optical insurance are not common. For an overview on the Irish health system during the PFL recruitment period see McDaid et al (2009).

health rather than income basis.[4] See (Doyle et al. 2017) for more on PFL in the national context.

Figure 1 Cohorts by income quintile and medical card status



Income quintiles based on equivalised household income in the GUI cohort at wave one. Incomes adjusted to December 2008 real euros by year and month of reporting. Sampling weights are applied.

During the PFL trial, intervention and control participants received developmental toy and book packs each year, facilitated access to local services, invitations to social events and public health workshops on stress control and healthy eating, access to enhanced pre-school, and a professionally taken photograph of their child. Intervention parents also received three intensive parenting supports including access to home visits from a trained mentor from pregnancy until age 4/5, baby massage classes in year one, and group parenting classes in year two. Tip sheets were used to inform mentor visits. Eighty-three percent of intervention families received at least one home visit, with 65 percent receiving at least 30 home visits (PFL Evaluation Team 2016).[5]

The programme aimed to improve child health through enhancing parenting practices and the childhood environment. Mentors focused on prenatal care,

[4] Medical card eligibility is typically based on household income and expenditures, such as commuting costs. In 2010 five percent of medical cards were awarded on a health basis, where it was deemed on a case-by-case basis that undue hardship from chronic illness or disability would otherwise arise (McDaid et al. 2009; Callan et al. 2015). We cannot identify recipients who receive the medical card based on health in either dataset.

[5] On average, families received 51 hours of home visits over 50 sessions lasting 59 minutes each. Older, employed mothers with higher cognitive resources received more home visits. Forty-three percent of families engaged with the group parenting classes, where five two-hour-long sessions was offered. Sixty-two percent engaged with the baby massage classes, where five sessions were offered between birth and ten months.

child nutrition, managing common illnesses, providing adequate supervision, parenting practices and behaviours, and creating a safe home environment. Of the 210 tip sheets, 32 focused on the mother and her supports, 60 on social and emotional development, and 105 on physical wellbeing and motor development (PFL Evaluation Team 2016). Through these components, the programme should impact child health when socioeconomic inequalities in these outcomes exist, such as social-emotional development, general health, diet and overweightness, service use for minor and major illnesses, accident occurrence, and the incidence of respiratory and infectious disease.

Previous research on the perinatal outcomes of PFL examined gestational age, birthweight, Apgar scores, maternal outcomes, and mode of delivery. The programme positively affected rates of spontaneous onset of labour and caesarean section (Doyle et al. 2014). Another study examined child health at six, 12, 18, 24, and 36 months based on parent reports of overall health, health issues, hospital stays, accidents, immunisations, chest infections, and wheezing or asthma. Significant effects on wheezing and asthma are reported, with effects for boys on health issues, accidents, and chest infections (Doyle et al. 2015). Over the same period, a study of diet quality finds positive effects on protein consumption at 24 and 36 months and across all food groups at 24 months. There are no effects on consumption of fruits and vegetables, dairy, grains, or sugary and fatty foods (O'Sullivan, Fitzpatrick, and Doyle 2017). No previous studies of the PFL programme examine health outcomes at age four, maternal health at any period, or child hospital records, the focus of this paper.

II. Data

The programme was evaluated using a registered randomised controlled trial. Data was compiled from parent reports, direct measurement, and hospital records, and contains baseline and longitudinal child, maternal, environmental, and implementation data. For more on the trial design see Doyle (2012). Eligible participants were pregnant women living within defined geographic bounds over the recruitment period between 2008 and 2010. Through two maternity hospitals and self-referral, 233 women were recruited.[6] After informed consent was received, participants were randomised into intervention and control arms (N_{INT} = 115, N_{CON} = 118) using computerised, unconditional randomisation. Retention rates over time are available in appendix A2[7]. For the parent surveys, retention rates decreased from 71 to 63 percent over years one to four. Direct measurement data are available for 45 percent of the sample at age 4, and hospital records are available for 47 percent. These datasets are smaller due to

[6] This represents 52 percent of the eligible population based on birth statistics. A survey of 102 eligible non-participants conducted when children were aged four indicate that non-participants are more likely to be older, employed during pregnancy, and better educated that those who joined (PFL Evaluation Team 2016).

[7] Data is available for 40 percent of participants at the four time points we study and child hospital records. A further 28 percent are missing from two or fewer estimation samples.

the research design. For direct measurement, the child needed to be present during the interview and agree to be measured. Consent to collect hospital records was sought after initial consent was granted, therefore hospital data is not available for those who left the programme or refused the additional consent. Balance between the two groups and adjustments to better reflect the randomised sample are discussed during robustness testing.

We examine child health on a yearly basis until age four using hospital records, direct-measurement, and parent-reported data. Maternal health and health behaviours are also examined. We study outcomes during infancy (first year), toddlerhood (second and third years), and the preschool period (fourth year). This allows us to examine the trajectory of health formation in a disadvantaged cohort while accounting for the rapid changes in early skill formation and the priority of health indicators by age group, although variations in normative development cannot be dismissed (Hack 1999; Timmons et al. 2012).

A. Outcomes

Our primary outcomes concern subjective and objective measures of child physical and mental health, health behaviours, severity of health issues, and service use. Using parent reported surveys, binary outcomes indicating whether the child has an adequate diet and is in good health are examined. Using continuous indicators from the parent survey, the number of health problems leading to GP visits and a score for social-emotional development (Ages and Stages: Social-Emotional) are examined. From hospital records, reporting from post-birth until age four, the number and cost of attendances, and the number of urgent and less urgent ED visits are examined. At age four, direct measurements of weight and height are available, from which we generate age and gender dependent BMI z-scores and an indicator of overweightness. To elucidate our findings, we examine maternal reports of health capability, including binary measures of depression risk and health status, and continuous measures of GP visits and weekly alcohol consumption. Additionally, common child ED diagnoses are examined: head and other injuries, and respiratory and gastrointestinal infections.

In infancy, when the socioeconomic gradient is not as well established in similar social security contexts, we hypothesise that there will be limited to no differences between the groups, but that effects will emerge from the second year onwards (Apouey and Geoffard 2013). For the secondary outcomes, we expect reductions in injuries in years three and four. Details on outcome motivation, calculation, and hypothesis setting are available in Appendix A1.

III. Methodology

A. Estimation Strategy

As shown in appendix A3, the characteristics of the intervention and control groups are balanced indicating that randomisation holds in our estimation samples, thus mean comparisons between the groups indicate causal effects.[8] We take an intention-to-treat approach to test the null hypothesis of no effect, β, on a given health outcome, *Y*, for a participant, *i*, from our sample N = {1 . . . *i* . . . n}, for the intervention group, *D* = 1, compared to the control group, *D* = 0, such that:

$$\beta = E[Y|D_i = 1] - E[Y|D_i = 0]$$

B. Inference

The statistical significance of the treatment effects are calculated using standard *t*-tests, permutation-based hypothesis testing, and permutation-based hypothesis testing where a stepdown procedure is applied to adjust for multiple hypothesis testing. This approach is taken as linear regression analysis is based on asymptotic assumptions that may not be appropriate given the relatively small sample size used here (Freedman 2008). As such, we use the non-parametric permutation testing which is based on the assumption that if the null hypothesis of no treatment effect is true, all potential outcomes are known exactly (Cohen and Dupas 2010). Therefore, reassigning the intervention indicator, *D*, will have no effect on the outcome. By randomly shuffling exposure to the intervention, 100,000 datasets and test statistics are created. The true test statistic is then ranked among those generated from the shuffled data and the null hypothesis can be rejected with a confidence level of 0.9 if the true test statistic lies in the 0.05% tails of the distribution. As a robustness check we rerun the tests using linear regression analysis.

C. Multiple Hypothesis Testing

A shortcoming of previous home visiting studies is that those examining a wide range of outcomes typically do not make corrections for multiple hypothesis testing (Avellar and Supplee 2013). As the number of outcomes increases, so too does the probability of one or more false rejections of null hypotheses. In this paper, we examine 8 to 9 primary outcomes and 7 to 8 secondary outcomes at each time point. The joint probability of rejecting all the null hypotheses, *j*, is

$$1 - (1 - \alpha)^{\theta} \cong j$$

where α is the significance level for the rejection of the null and θ is the number of tested outcomes. Thus, there is an 86 to 90 percent probability of falsely

[8] Balancing tests for the randomised and estimated samples are available in appendix A3 and are discussed under robustness tests.

rejecting one or more hypotheses at a 10 percent reporting level. Collating the primary and secondary outcomes over four time periods, there is a minimum 99.8 percent probability of a type I error. Thus, to mitigate the family-wise error rate, a stepdown procedure is used as outlined in Romano and Wolf (2005). Blocks of outcomes are formed for joint hypothesis testing based on the estimation sample. The block constitutes a joint test of the null hypothesis of no treatment effects for any of the outcomes in that block. The result from this test is recorded, with the most significant recorded effect in single hypothesis testing. Step-wise ordering is used such that each successive joint test removes one outcome from the joint null hypothesis based on its significance in single testing. Here, the stepgroups are formed based on the estimation sample, for example, all parent reported primary outcomes in the first year are placed in one stepgroup. Thus, for the primary outcomes, eight joint tests and one single test (for overweightness) are conducted over the four time points, reducing the probability of spurious rejections to 61 percent. While reduced, the error rate remains high and will be factored into the discussion of the results. We diverge from Heckman et al. (2010) and other papers on PFL (see, for example, Doyle et al. (2015)) in reporting two-sided *p*-values as the assumed direction of the treatment effects is ambiguous for some outcomes, for example respiratory infections. We focus on the *p*-values associated with multiple hypothesis testing adjusted estimates in our presentation and discussion of results, unless otherwise stated.

D. Inverse Probability Weighting

Inverse probability weighting (IPW) is employed to improve the generalisability of our inferences to the randomised sample (Wooldridge 2007). Here, intervention and control means are weighted by the inverse of the group-specific probability of being in a given estimation sample. The weights are generated from a logistic regression based on selection into the estimation sample using baseline characteristics. Model specification was completed using a combination of stepwise regression, the Bayesian information criteria (BIC), and best subset selection. Given the sensitivity of logistic regression models to the number of parameters, *I*, the BIC was preferred to the Akaike information criterion (Steyerberg et al. 2000; Peduzzi et al. 1996; Schwarz 1978). Forty-six to 49 baseline characteristics were used based on full data availability and an anticipated role in selection.[9] The difference between BIC scores to indicate supersedence were in the range $d = 0.25n$ for $n \in \{4 \dots 12\}$, such that $I \leq (\frac{S}{10})$ and $2 - |d|$ was minimised, where S is the number of stayers in a given estimation sample (Kass and Raftery 1995; Steyerberg et al. 2000; Peduzzi et al. 1996). This approach hinges on the assumption that participation at each period can be predicted from observed baseline characteristics. Where that is accepted,

[9] When forming weights, there must be no missing data for any participant. Thus the number of characteristics varies, with exclusion arising where there is a missing response at a given timepoint. When examining baseline equivalence between the groups, we use all available data, regardless of missing responses – for example, household income.

inferences can be interpreted as reflecting expected treatment effects if the programme was implemented in a similar population (Imbens and Wooldridge 2009).

E. Presentation of Results

In the results section, we examine each outcome over time for the intervention (blue, dotted) and control (red, continuous) groups, plotting means and standard deviations, for continuous outcomes. Means and standard deviations for inverse probability weighted outcomes are presented in lighter sub-plots. Beneath the graph we include the corresponding sample size, and *p*-values from permutation tests, permutation tests adjusted for multiple hypothesis testing, and permutation tests adjusted for multiple hypothesis testing where inverse probability weights have been applied.

To further identify any group differences and quantify the size of the results, effect sizes are calculated using Cohen's *d* for continuous outcomes and odds ratios for binary outcomes. Cohen's *d* (*d*) is the difference in the means of the two groups divided by the pooled standard deviation. Interpretation of Cohen's *d* effect sizes indicates that $d = 0.2$ is small, $d = 0.5$ is medium, and $d = 0.8$ is large (Lenth 2001; Jacob Cohen 1988). Filene et al (2013), in a meta-analysis of home visiting programmes, find a non-significant effect size of 0.11 for physical health outcomes. From a power analysis of this trial, a minimum detectable effect size of 0.4 is expected (for more on this refer to section V on power analysis and sample size under robustness checks). Where an effect size is in a direction indicating worse outcomes, a superscripted c (^{c}) is used. For the binary outcomes, odds ratios (*OR*) are presented as a measure of association, a comparison of the relative odds of the occurrence of a binary outcome given exposure to the intervention.

IV. Results

A. Summary Statistics

Balancing tests and descriptions of the baseline variables are available in appendices A3 and A4. These tests show that both the original and the estimation samples are balanced. Table one shows that the cohort is relatively disadvantaged with mothers leaving school at the mean age of 17, having cognitive ability in the low-average range (Wechsler 1999), and high levels of unemployment and social housing. Intervention mothers know significantly more about infant development and are more likely to have a physical health condition. Children do not differ in their birthdate, but the proportion of girls in the control group is significantly greater.

Table 1 Summary Statistics: Maternal characteristics at recruitment

	N (INT/CON)	M_{INT} (*SD*)	M_{CON} (*SD*)	*P*
Age at baseline	205 (104/101)	25.46 (5.85)	25.30 (5.99)	0.840
Married[1]	205 (104/101)	0.14 (0.35)	0.18 (0.38)	0.960
Age leaving education	187 (96/91)	17.41 (2.53)	17.43 (3.08)	0.960
WASI IQ	205 (104/101)	82.06 (12.32)	80.91 (12.88)	0.519
Unemployed[1]	205 (104/101)	0.43 (0.50)	0.41 (0.49)	0.699
Participant in social housing[1]	204 (103/101)	0.55 (0.50)	0.55 (0.50)	0.985
First child[1]	205 (104/101)	0.54 (0.50)	0.50 (0.50)	0.548
Pregnancy was planned[1]	203 (103/100)	0.29 (0.46)	0.30 (0.46)	0.897
KIDI Knowledge of infant development	205 (104/101)	72.25 (7.60)	69.82 (8.19)	**0.028**
Smokes during pregnancy[1]	205 (104/101)	0.51 (0.50)	0.48 (0.50)	0.610
Has a mental health condition[1]	205 (104/101)	0.28 (0.45)	0.24 (0.43)	0.511
Has a physical health condition[1]	205 (104/101)	0.75 (0.44)	0.62 (0.49)	**0.053**
Health insurance[1]	202 (102/100)	0.09 (0.29)	0.07 (0.26)	0.677
Child's age on 01/01/12	200 (101/99)	2.11 (0.70)	2.12 (0.68)	0.909
Child is female[1]	194 (98/96)	0.46 (0.50)	0.64 (0.48)	**0.013**

N is sample size, $_{INT}$ is the intervention group, $_{CON}$ is the control group, *M* is the mean, and *SD* is the standard deviation. *p* is the *p*-value from a two-tailed permutation test over 100,000 permutations of the mean difference between groups. *p*-values less than 0.1 are highlighted. [1] binary measure

B. Primary Outcomes

Tables two and three display parent reported and hospital outcomes respectively. Means and standard deviations (plotted symmetrically around the mean) from non-weighted and weighted tests are graphed. Results including means and standard deviations are in appendix A5 and, using weights, appendix A6.[10]

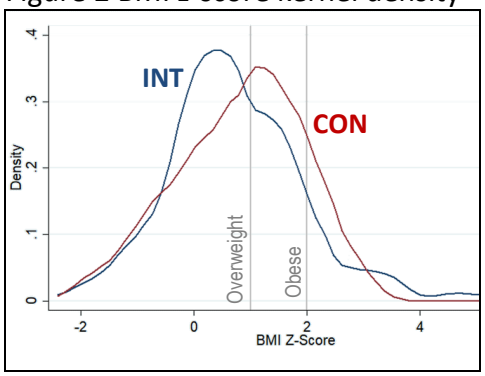
Table two shows the programme's impact on parent-reported child health outcomes. Diet adequacy is examined as a marker of parental health behaviours that directly affects child health. Given the focus on nutrition in the PFL

[10] We do not examine the effects by gender due to the reduction in power this would incur.

curriculum, we expect PFL to positively impact diet, with effects declining with age (Kader, Sundblom, and Elinder 2015). Panel 2a shows that diet quality declines over time: 87 to 88 percent of the groups consumed an adequate diet at age one, declining to 22 to 34 percent for the control and intervention groups respectively by the fourth year. A significantly greater proportion of the intervention group, a 20 percentage point difference, have an adequate diet at age three ($p^1 = 0.034$; $OR = 2.47$). The application of IPW leads to slightly more conservative estimates. This result aligns with the literature, with expected findings in the very early years mitigated by the high level of diet adequacy in both groups.

Given this focus on nutrition, PFL is also expected to reduce the incidence of overweightness. Figure two is a graph of the kernel density of BMI z-scores for both groups at age four,

Figure 2 BMI z-score kernel density



including cut-offs for obesity and overweightness. While 41 percent of the control group are overweight, compared to 26 percent of the intervention group, this difference is not statistically significant (N_{INT}=53, N_{CON}=51, *p* = 0.120, *OR* = 0.51). However, when IPW weights are applied, the intervention group is significantly less likely to be overweight (*p* = 0.075, *OR* = 0.46). Yet, the continuous BMI z-scores indicates no significant differences between the groups when using non-weighted (p^1= 0.653; *d* = 0.09) or weighted (p^1= 0.497; *d* = 0.14) estimates. Thus, these results suggest that the programme impacted the tail but not the whole BMI distribution.

Next, parent reported child health status (panel 2b) and the number of health problems

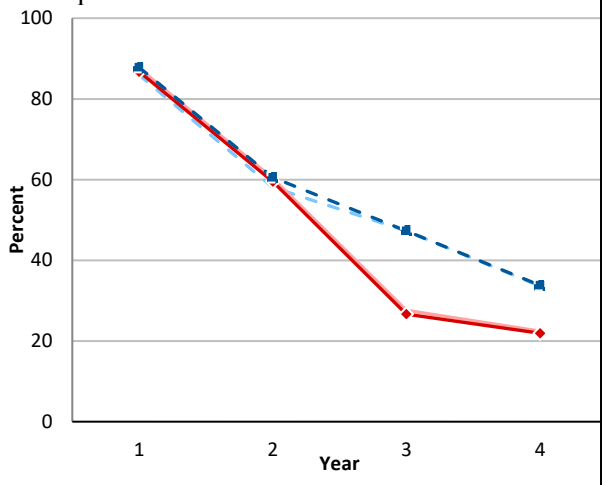
leading to health service use (panel 2c) are examined. Different programme components have been shown to be both positively (teaching behaviour management techniques) and negatively (group parenting classes) associated with child health, and evidence points to larger effects with age (Filene et al. 2013; Peacock et al. 2013). We anticipate any effects on health status and health problems to be modest and to emerge later in childhood, with effects on health status preceding those on health problems (National Research Council 2004). The majority of parents, 64 to 77 percent, report children to be in excellent or very good health at all ages, with no significant differences across groups. The number of health problems peak in years one and two, with no significant differences. Again, the use of weights does not affect the significance. The effect sizes on health problems in years two, three, and four (*d* = 0.10 – 0.15) align with results from the Filene (2010) meta-analysis on physical health (*d* = 0.11), although are more conservative when IPW is applied.

In reviews of health effects from similar programmes, the most consistent, largest effects are found on social-emotional development. Using a scale identifying social-emotional developmental delay and disabilities, we thus anticipate the intervention group will score significantly lower on the scale (higher scores are indicative of more problem behaviours). Panel 2d shows the mean score for both groups over the four years, with a cut-off score for risk of developmental issues highlighted. Contrary to hypothesis, social-emotional development is not significantly different across the groups at any timepoint, with very small or negative effect sizes (d = 0.12[c] - 0.10). Only at age three do programme effects (d = 0.10) align with meta-analysis (d = 0.10) findings (Sweet and Appelbaum 2004).

In table three, treatment effects on hospital recorded markers of health are presented. Service use in both groups is high. Within the first four years, 94 percent of participants attended hospital at least once, with a peak of 69 percent in the second year and a low of 43 percent in the fourth year. We examine the number of hospital attendances per year, incorporating emergency, inpatient, and outpatient services in panel 3a. To reflect the gravity of required service use, we also examine the cost of hospital use (panel 3b). By age four, the average intervention child has attended the hospital 4.3 times, compared to a mean of 7.3 by control participants. This is driven by significant differences and medium effect sizes in the second ($p = 0.031$; $d = 0.47$) and third ($p = 0.015$; $d = 0.53$) years. However, this is only

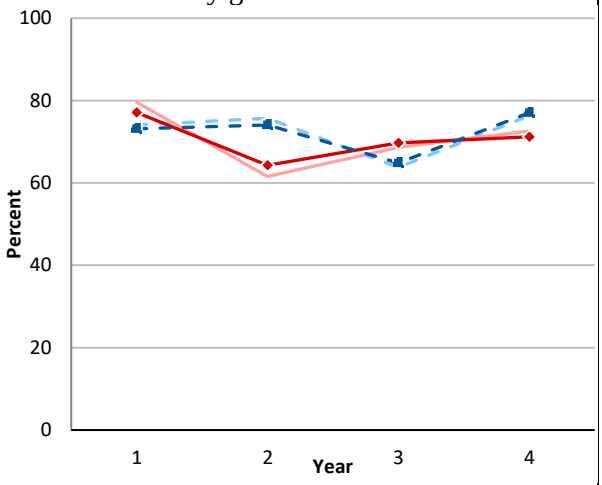
Table 2 Intervention and control group parent-reported outcomes

2a Adequate diet



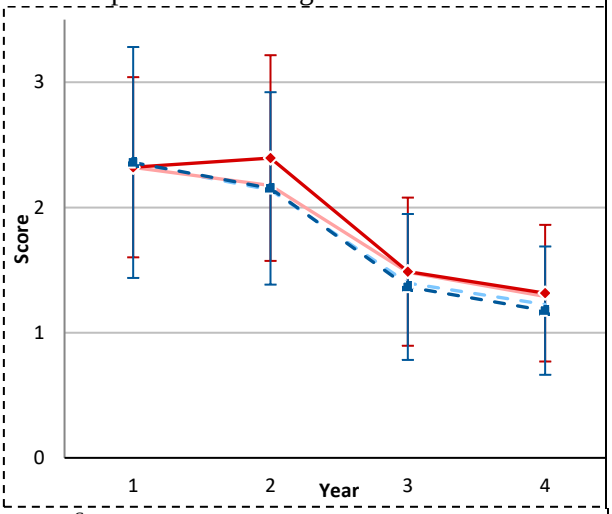
p^S	0.848	0.910	**0.009**	0.107
p^M	0.978	0.910	**0.034**	0.360
OR	1.10	1.04	2.47	1.82
p^W	0.951	0.997	**0.057**	0.451
OR^W	0.87	0.91	2.35	1.75
N	165	165	149	147

2b Excellent or very good health



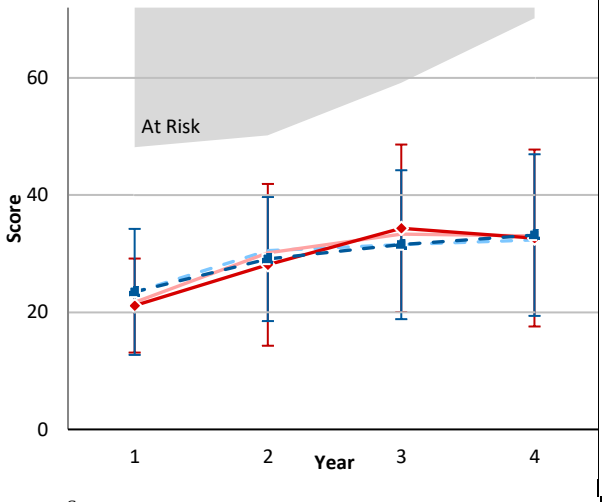
c	0.561	0.181	0.537	0.428
p^M	0.916	0.528	0.537	0.799
OR	0.81	1.59	0.80	1.35
p^W	0.887	0.628	0.897	0.942
OR^W	0.73	1.94	0.81	1.21
N	165	165	150	147

2c Health problems leading to GP & ED visits



p^S	0.887	0.354	0.528	0.428
p^M	0.887	0.735	0.887	0.672
d	0.02[c]	0.15	0.10	0.13
p^W	0.902	0.994	0.880	0.922
d^W	0.02[c]	0.02	0.08	0.06
N	159	150	150	147

2d Social and emotional development



p^S	0.444	0.797	0.535	0.919
p^M	0.890	0.954	0.777	0.919
d	0.12[c]	0.04[c]	0.10	0.02[c]
p^W	0.921	0.953	0.683	0.886
d^W	0.09[c]	0.01[c]	0.07	0.02
N	165	166	150	147

Intervention unweighted mean. Control unweighted mean. Intervention weighted mean. Control weighted mean. Unweighted standard deviation bars, plotted symmetrically around the mean, are indicated and capped at zero. *p*-values (*p*) are from two-tailed permutation tests over 100,000 permutations of the mean difference between groups. Where adjustments are made for multiple hypothesis testing ([1] [2]), outcomes are jointly tested at each time point. p^S is the *p*-value from single tests. p^M is the *p*-value with multiple hypothesis testing adjustments. p^W is the *p*-value with multiple hypothesis testing adjustments and inverse probability weights. *d* and OR are the Cohen's *d* effect size and odds ratio using unweighted data. d^W and OR^W are the Cohen's *d* effect size and odds ratio using inverse probability weighted data. [C] Cohen's d is not in the hypothesised direction.

robust to weighting in the second year, although the effect sizes remain large. Hospital costs[11] (panel 3b) for control participants are three times that of intervention participants in the second (p = 0.059 d = 0.38) and third (p = 0.043 d = 0.46) years, with significant differences that are not robust to weighting, despite relatively consistent effect sizes. By age four, costs for both groups decline and they do not significantly differ. Over the four years, control participants use hospital services valued at €1,359 more than intervention participants.

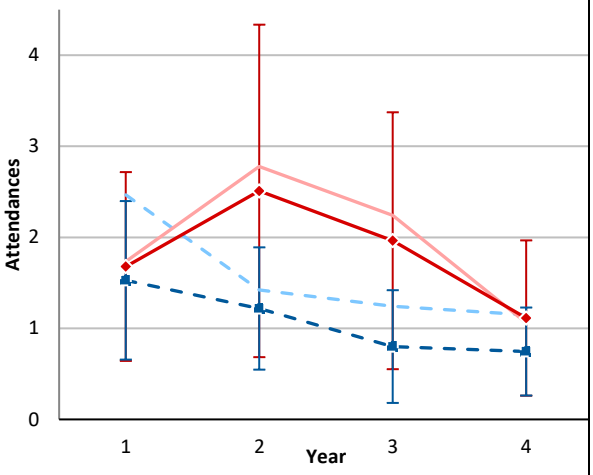
Urgent (panel 3c) and less urgent (panel 3d) ED visits are examined to investigate the programme's impact on health shocks and health behaviours respectively. It is anticipated that PFL will reduce both types of attendance. While there are no differences between the groups for non-urgent ED visits, the intervention group attends for significantly fewer urgent issues in the second year ($p = 0.067$; $d = 0.40$). However, it is no longer statistically significant when weights are applied. A significant difference on urgent attendances in the third year is not robust to multiple hypothesis testing adjustments. As with the other measures of health, we see decreased service use of both urgent and non-urgent cases after the second year. The use of weights leads to a spike in overall and less-urgent attendances in the first year for the intervention group, however these differences are not significant.

In sum, these results show that in the control group, the trajectory of child health is not linear over time, it declines in the second year and improves thereafter. Treatment effects on health are not evident in the first year, when health service use is high for both groups, or in the fourth year when both groups use fewer services and parents generally report good child health. Modest treatment effects are found in the intervening years. At age two, there is a sharp increase in hospital use for the control group and a small decrease in hospital use by the intervention group. Health service use decreases in the control group at age three, however some of the differences between the groups remain significant. By age four, treatment effects on health outcomes have disappeared, as service use declines.

[11] Hospital costs are estimated using publicly available data on the average cost for ED, overnight, and outpatient hospital attendances for children. Further detail and sources are available in appendix A1.

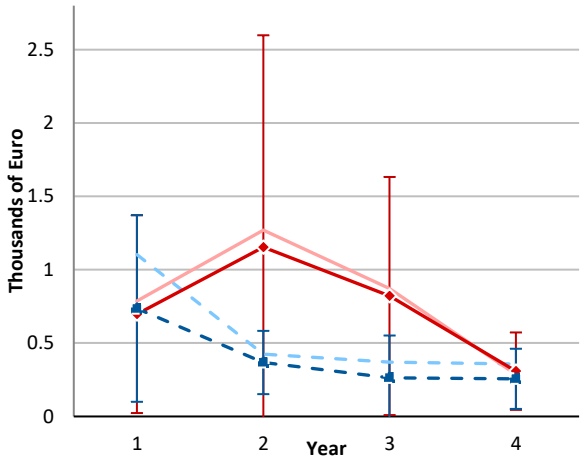
Table 3 Intervention and control group hospital record outcomes

3a Attendances



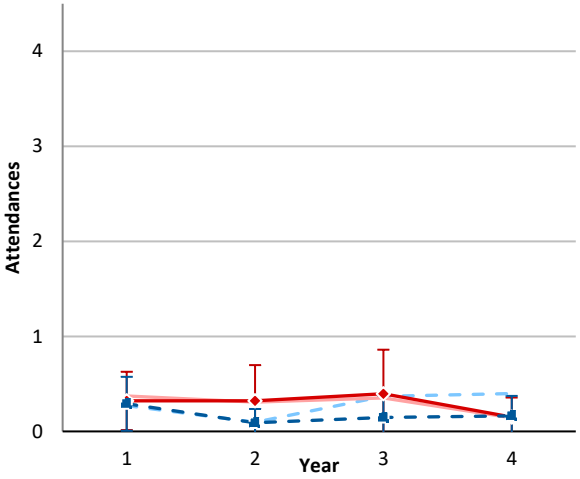
p^S	0.685	**0.014**	**0.006**	0.177
p^M	0.950	**0.031**	**0.015**	0.444
d	0.08	0.47	0.53	0.27
p^W	0.701	**0.092**	0.416	0.866
d^W	0.31[c]	0.46	0.42	0.06[c]
N	108	108	108	108

3b Cost



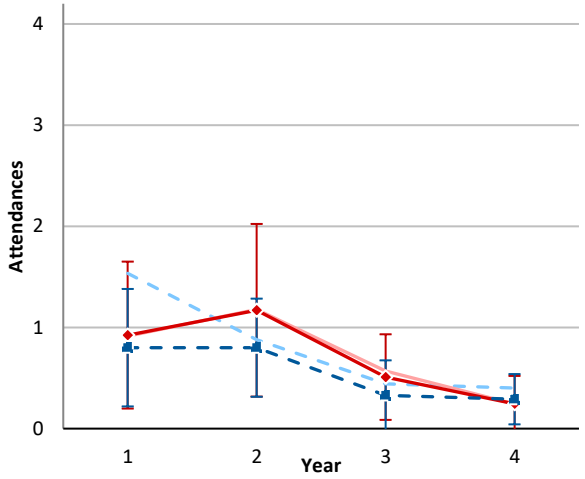
p^S	0.885	**0.022**	**0.013**	0.576
p^M	0.885	**0.059**	**0.043**	0.902
d	0.03[c]	0.38	0.46	0.11
p^W	0.648	0.151	0.413	0.743
d^W	0.21[c]	0.40	0.42	0.15[c]
N	108	108	108	108

3c High triage emergency department attendances



p^S	0.797	**0.033**	**0.074**	0.876
p^M	0.952	**0.067**	0.145	0.876
d	0.05	0.40	0.34	0.03[c]
p^W	0.475	0.139	0.988	0.709
d^W	0.17	0.40	0.02[c]	0.43[c]
N	108	108	108	108

3d Low triage emergency department attendances



p^S	0.632	0.175	0.225	0.666
p^M	0.954	0.175	0.225	0.885
d	0.09	0.27	0.24	0.09[c]
p^W	0.828	0.346	0.915	0.530
d^W	0.35[c]	0.22	0.16	0.30[c]
N	108	108	108	108

Intervention unweighted mean. Control unweighted mean. Intervention weighted mean. Control weighted mean. Unweighted standard deviation bars, plotted symmetrically around the mean, are indicated and capped at zero. *p*-values (*p*) are from two-tailed permutation tests over 100,000 permutations of the mean difference between groups. Where adjustments are made for multiple hypothesis testing ([1 2]), outcomes are jointly tested at each time point. p^S is the *p*-value from single tests. p^M is the *p*-value with multiple hypothesis testing adjustments. p^W is the *p*-value with multiple hypothesis testing adjustments and inverse probability weights. *d* and OR are the Cohen's *d* effect size and odds ratio using unweighted data. d^W and OR^W are the Cohen's *d* effect size and odds ratio using inverse probability weighted data. [C] Cohen's d is not in the hypothesised direction.

C. Secondary Outcomes

In table four we present the effects on child ED diagnoses and maternal health, restricting the reported results to those from multiple hypothesis testing. Full results are available in Appendices A5 and A6. ED diagnoses are examined to investigate the drivers of hospital attendance. Diagnosis is of particular interest in years two and three, where we see significant differences in overall ED attendance and urgent attendances. Forming hypotheses for these secondary outcomes is not clear-cut, with some programme components, such as encouraging increased parent supervision and decreased smoking in the home, expected to have a positive impact on injuries and infections respectively. However, other components, such as encouraging child activity and the use of centre-based childcare, could lead to increased incidents of injury and infection. In line with national attendance patterns for paediatric emergency departments, attendance due to injury is expected to be higher in both groups as children become more mobile, while attendance due to respiratory conditions and other infections is expected to decline with age (Beattie, Gorman, and Walker 2001).

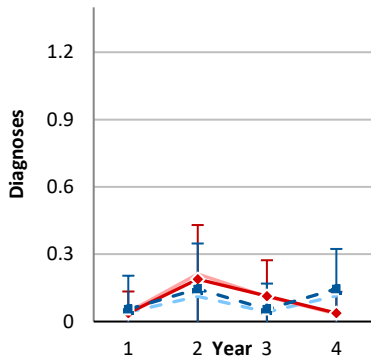
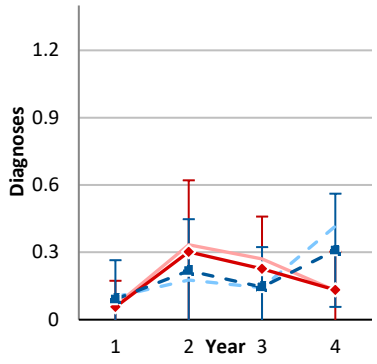
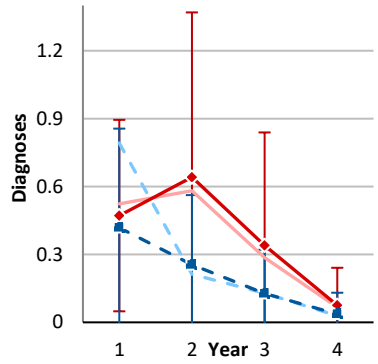
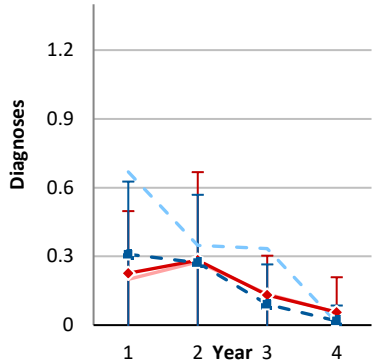
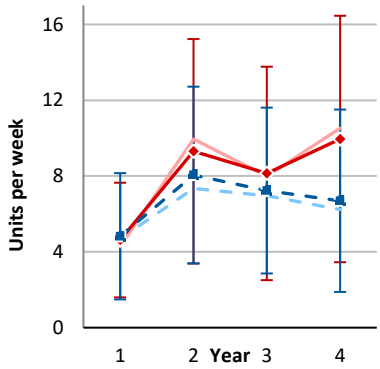
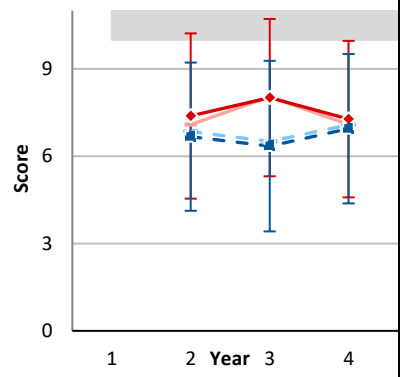
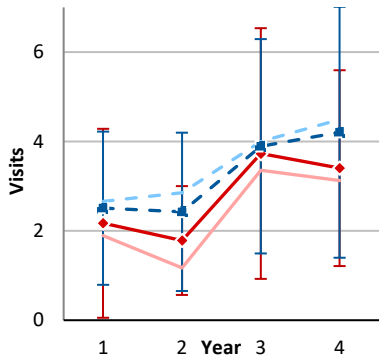
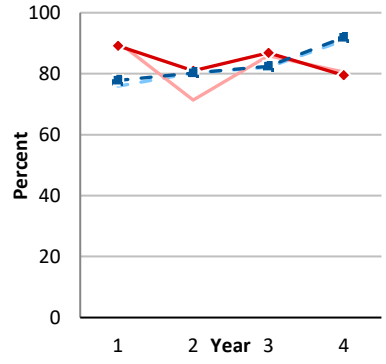
Panels 4a and 4b show the number of attendances for head and other injuries. These are separated, as head injuries are associated with major trauma in children (Van As et al. 2016). Overall, we find no evidence of programme impact when adjustments for multiple hypothesis testing are made. In the second year, ED attendance for injury increases in both groups, in line with the literature (Downing and Rudge 2006). By age four, of the 0.75 hospital visits for intervention participants, 0.15 are for head injuries and 0.22 are for other injuries. At age four, the control group have significantly fewer attendances for non-head related injuries when weights are employed ($p^2 = 0.026$; $d = 0.66$). Panels 4c and 4d show attendance for respiratory conditions and other infections, with respiratory conditions examined to capture issues pertaining to asthma and wheezing. Both types of attendance decline sharply in the third and fourth years in both groups. There are small effects on respiratory conditions in favour of the intervention group in years two and three, however the differences are not significant. When weights are applied, some of the increased intervention attendance in year one is driven by respiratory conditions and other infections, although differences are again not significant.

Panels 4e to 4h shows the impact of PFL on maternal health and health behaviours across the four years. No significant differences are present once adjustments for multiple hypothesis testing are made. Alcohol consumption, displayed in panel 4e, increased in the second year in both groups. Panel 4f shows there is no difference between the groups regarding the risk of

depression, while a significant difference in the third year is not robust to multiple hypothesis testing. Mothers reported two to four GP visits per year (panel 4g), again there were no significant differences across the groups at any time-point. In panel 4f, eight to 22 percent of mothers in both groups report being in fair or poor health over time. A significantly greater proportion of the control group report excellent health in year one, while in the fourth year a significantly higher proportion of intervention mothers are in excellent health. These results are not robust to multiple hypothesis testing when weights are not applied. However, when IPW and the multiple testing procedure are employed, the significantly better health of the control group in the first year remains ($p^2 = 0.048$; $OR = 0.33$).

In sum, these results show that PFL had no impact on the programme's secondary outcomes.

Table 4 Secondary outcomes: Child diagnoses and maternal health

4a Child head injuries  $p^{M.}$ 0.949 0.803 0.626 0.058 *d.* 0.07 [c] 0.10 0.21 0.38 [c]	4b Child other injuries  $p^{M.}$ 0.868 0.779 0.532 0.110 *d.* 0.12 [c] 0.15 0.20 0.41 [c]	4c Child respiratory conditions  $p^{M.}$ 0.746 0.226 0.455 0.457 *d.* 0.06 0.35 0.28 0.15
4d Child other infections  $p^{M.}$ 0.900 0.943 0.535 0.632 *d.* 0.14 [c] 0.01 0.12 0.16	4e Maternal alcohol consumption  $p^{M.}$ 0.855 0.748 0.850 0.281 *d.* 0.03[c] 0.12 0.09 0.28	4f Maternal depression score  $p^{M.}$ - 0.774 0.250 0.706 *d.* - 0.13 0.30 0.06
4g Maternal GP visits  $p^{M.}$ 0.825 0.546 0.858 0.556 *d* 0.09[c] 0.21[c] 0.03[c] 0.16 [c]	4h Maternal good health status  $p^{M.}$ 0.136 0.906 0.841 0.110 *OR* 0.43 0.96 0.71 2.93	

Intervention unweighted mean. Control unweighted mean. Intervention weighted mean. Control weighted mean. Unweighted standard deviation bars, plotted symmetrically around the mean, are indicated and capped at zero. *p*-values (*p*) are from two-tailed permutation tests over 100,000 permutations of the mean difference between groups. Where adjustments are made for multiple hypothesis testing, outcomes are jointly tested at each time point. p^M is the *p*-value with multiple hypothesis testing adjustments. *d* and OR are the Cohen's *d* effect size and odds ratio using unweighted data. [C] Cohen's d is not in the hypothesised direction.

V. Robustness Tests

A. Regression Analysis

To demonstrate the robustness of our results to our chosen estimator, treatment effects are re-estimated using linear and logistic regression analysis for continuous and binary outcomes respectively. Point estimates, standard errors, and *p*-values are available in appendix A7. The results are similar to the main analysis, with some differences in the level of statistical significance where a treatment effect is observed: for example, for hospital attendance in year two and for diet, hospital costs, hospital attendance, and high triage ED visits in year three. Generally, *p*-values from permutation testing are more conservative than those from the regression analysis.

B. Sample Selection

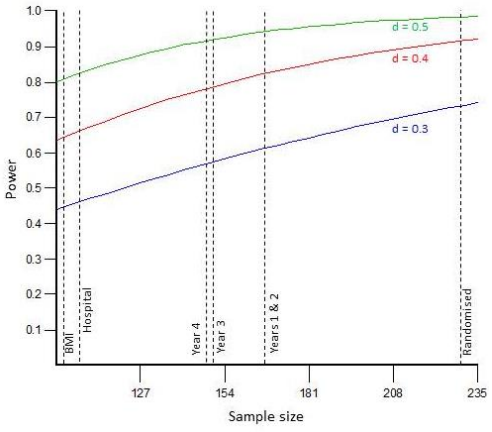
Due to attrition and missing data, we consider sample selection and attrition bias. The sample sizes used for the BMI and hospital record outcomes are particularly low (see appendix A1 for retention rates over time). To examine the impact of sample selection, the primary analysis for each outcome is re-estimated using each estimation sample. For example, 108 observations are available for the hospital record outcomes and between 147 and 166 observations are available for the parent-reported outcomes. Thus, we re-estimate models for all outcomes by restricting the sample only to those participants who are also in the hospital records (n = 100 -106 for the four years). If different results are found to those reported in the main results, any effects (or lack thereof) found could arise from sample selection rather than true differences between the groups. For example, sicker control group participants could be selecting into the hospital sample (or healthier intervention participants, or both). We estimate effects for each estimation sample in appendix A8 and graph the outcomes across hospital record and BMI estimation samples in appendix A9. Generally, the results are robust to sample selection. For all but one outcome, third year hospital costs, where there is a statistically significant difference, significance is consistent across estimation samples. Examining effect sizes, the direct BMI measurement and hospital record samples frequently display the greatest divergence from the reported results. For the hospital sample, five of the nine divergent results lead to increased effect sizes (including one result that is contrary to hypothesis) and four lead to reduced effect sizes. For the direct measurement sample, eight divergent results indicate increased effect sizes. As such we should be cautious that the BMI results may reflect a selected sample with better health in the intervention group.

To examine the impact of attrition, balancing tests between the intervention and control groups for each estimation sample are conducted with and without weights. These results are in appendix A3. There are statistically significant baseline differences between the intervention and control groups for three to 10 percent of variables in each estimation sample. This aligns with expected spurious effects arising from conducting 62 tests. Across the estimation samples, there are some patterns in differences. In the original randomised sample, the groups differed significantly in gender, knowledge of infant development, maternal health conditions, a measure of deprivation, and maternal mastery. Differences in gender remain consistent in nine of the estimation samples and differences in knowledge of infant development remain in five. One difference, the percentage of fathers in paid work, appears in eight of the estimation samples however it was not evident in the original randomised sample. Generally, applying inverse probability weights does not return the sample to a more similar pattern of differences to those in the randomised sample, even though there appears to be some differences between the randomised sample and estimation samples, such as the percentage of fathers in paid work. This could arise as the covariates used to form the weights do not sufficiently predict attrition. Therefore we primarily focus on the estimates calculated without inverse probability weights.

C. Power Analysis

When the PFL trial was designed, it was powered to detect an effect on cognitive development with a sample size of 233 (Côté et al. 2018). A retrospective power analysis has precedence where one fails to reject a null hypothesis and type II error from insufficient power is suspected (Jones, Carley, and Harrison 2003). Here, the estimation samples contain between 104 to 165 observations, thus meaningful but smaller effects may be missed. In meta-analyses of similar programmes, Filene et al (2013) find a small, non-significant effect (Jones, Carley, and Harrison 2003) (*d* = 0.11) on child physical health, while Sweet and Appelbaum (2004) find a small, statistically significant effect on social-emotional development (*d* = 0.096). Based on the estimation sample sizes, an examination of design power for different effect sizes are displayed in figure 3. Power estimation is calculated for minimum detectable effect sizes at a level of 0.1; the probability of a type-I error usually accepted as significant (Duflo, Glennerster, and Kremer 2008). At a power of at least 80 percent, medium effect sizes are minimally detectable (*d* = 0.4 - 0.5), which would exclude the smaller effects found in previous research.

Figure 3 Approximate power for samples, continuous outcomes, $\alpha = 0.1$



D. Socioeconomic Inequality in Health

Propper et al (2007) propose that differences in healthcare access and financing may explain socioeconomic inequalities in health. For home visiting interventions, Sandner et al (2018) suggest that fewer treatment effects may be expected in settings where there is greater access to healthcare or broad social welfare. In Ireland, healthcare for children is broadly financed and often provided by the State, with increased protection for children from disadvantaged families and those with complex medical needs (Staines et al. 2016). In this context, Nolan and Layte (2014) examine socioeconomic inequalities in health for Irish children at nine months and nine years using different cohorts. In line with UK and German studies (Propper, Rigg, and Burgess 2007), they find strong evidence of inequalities at nine years but little evidence at nine months. To test for inequalities within our sample, we compare the PFL cohort to a nationally representative cohort. Where inequality is evident, an effective programme would be expected to mitigate this dispartity through improving parenting and the home environment and acting as a gate-keeper to widely available services (Allin and Stabile 2012). Where inequalities are not evident, we cannot dismiss their existence, however confident hypothesis setting on treatment effects is not possible.

The PFL control group is compared to the Growing Up in Ireland (GUI) Infant cohort at nine months and t three years.[12] As a robustness check, we also examine maternal health outcomes. While evidence on the age at which socioeconomic inequalities in health appear has not reached consensus, it is well established by adulthood (Mackenbach et al. 2008), thus, we expect significant differences between the PFL and GUI groups for maternal outcomes.

Comparing the PFL control group to the GUI cohort, we examine child and maternal health outcomes, *y*, in waves one (6-9 months) and two (36 months) such that,

$$y = \beta_0 + \beta_1 PFL + \beta_n X + \varepsilon$$

Outcomes are chosen that best correspond with our primary and secondary outcomes. No hospital data is available for the national sample. We control for differences in age and year at data collection in all estimations (*X*). Gender is also included as a control due to the high proportion of girls in the PFL control group (64%) compared to the GUI sample (48%). Ordinary least squares and logistic methods are used, with bootstrapped standard errors and sample weights.

A summary of the results are available in table five. Full output is available in appendix A10. In wave one, the PFL control group children are significantly less likely to be in good health compared to the national cohort, however there are no significant differences between the groups with respect to inpatient stays,

[12] Information about GUI and a PFL comparison on SES and healthcare access are presented in the introduction.

birthweight, chest infections, and vomiting and diarrhoea. In wave two, there are no significant differences between the groups for any child outcomes. In contrast, at both timepoints, mothers from the PFL control group are significantly more likely to report poor health. In wave one, there is no significant difference in smoking status, however PFL control group mothers are more likely to be impaired in their daily activities by illness. In wave two, the mothers in the national sample are significantly less likely to smoke.

Table 5 Child & Maternal health inequalities between PFL and GUI participants

Wave	In-patient	Excellent health	Birth Weight	Chest issue	Gastro-intestinal issue	Asthma	Injury	Maternal excellent health	Mother smokes	Mother illness impaired
1	-1.83	**-2.36**	-152.5	0.05	-0.04			**-3.07**	0.80	**1.341**
	(1.12)	**(0.98)**	(130.8)	(0.53)	(0.89)			**(0.53)**	(0.49)	**(0.724)**
2	0.33	-1.01				-3.14	-4.19	**-2.00**	**0.86**	
	(4.71)	(0.91)				(3.58)	(4.25)	**(0.45)**	**(0.45)**	
1	11,179	11,142	11,057	11,184	11,184			11,183	11,182	11,183
2	9,685	9,693				9,694	9,679	9,697	9,609	

Results are from OLS and logistic regressions of PFL control group/GUI national sample status on continuous and binary outcomes respectively, controlling for age at interview, age at GUI collection point, and gender. Bootstrapped standard errors and sampling weights are used. **Bolded** coefficient and standard error: $p< 0.1$

In sum, there is strong evidence of socioeconomic health inequalities for mothers, however, there is little evidence of inequalities for children before the age of three. Thus, the failure to reject many of our null hypotheses on child health may arise as health inequalities have yet to emerge. Although this analysis does not guarantee an absence of inequality, hypothesis setting must be treated with caution. Given that Nolan and Layte (2014) find evidence of health inequalities at age nine, treatment effects on health may not appear until the children are older.

VI. Discussion

In line with the literature, the effects of the PFL home visiting programme on health formation are moderate. Medium sized treatment effects emerge in the second year, mainly driven by severity in hospital use. These effects are also evident into the third year, although they disappear by age four as health service use declines in both groups. Over the four years, this severity is reflected in cost, with the control group generating almost twice the total hospital costs as the intervention participants over the four years, with a mean difference of €1,359 per child. Few treatment effects are found using parent reported measures of child or maternal health. Across both groups, health formation is poor with high hospital use, poor diets, and high parental reporting of ill health.

Our results both align with and differ from the literature. In the US, effects on hospitalisations and emergency department use are found in the Early Intervention Programme by age two (Koniak-Griffin et al. 2003), while the Nurse Family Partnership programme finds effects between the ages of two and four on hospitalisation duration - a proxy for severity - and number of ED visits, but not on ED visits driven by injuries or the number of admissions (Olds, Henderson, and Kitzman 1994). In contrast, during the implementation of the Nurse Family Partnership in the UK there were no effects on hospital or ED attendances by age two (Robling et al. 2016).

This paper contributes to the literature by addressing commonly flagged issues in the home visiting health literature. Outcomes are carefully selected based on life-long importance for health and human capital formation. The ambiguity of hypothesis setting is highlighted, such as the conflicting ways in which home visiting may affect injury and infection and thus overall health and service use. Measuring health and health formation is inherently complex, and this is reflected in the broad range of outcomes studied at different time points. This multiple hypothesis testing is adjusted for, leading to more conservative results. The impact of smaller randomised and estimation sample sizes is also considered. The trial is sufficiently powered to identify medium effect sizes (d = 0.4 – 0.5). The impact of this reduced sample size is particularly notable with regards to the hospital records. Here, in the second and third years, we see five significant differences for effect sizes ranging from d = 0.40 – 0.53, however inverse probability weighting leads to slightly more conservative effect sizes of d = 0.40 – 0.46. While the effect sizes remain economically significant, four of the effects are no longer statistically significant when weights are applied. We also rerun our estimates for each estimation sample to determine if participants with better or worse health select into some parts of the data. While the hospital and direct measurement samples produce the most diverging results, there is no consistent pattern in the intervention group's health, thus there is little evidence that sample selection is affecting the results. A major contribution of this paper is the careful interpretation of our estimates. We focus on the practical and statistical significance of estimates where multiple hypothesis testing has been adjusted for, but without emphasising the impact of inverse probability weights as no systematic pattern in attrition is found. A smaller sample size in some cases may mask treatment effects of interest and is further considered in our interpretation of the results in the fourth year when health issues and service use decline in both groups, and thus only small treatment effects, if any, should be expected.

We show that the ambiguity in hypothesised treatment effects may be a contributory factor to the inconsistent and weak evidence in the home visiting health literature. This arises through two avenues, the measurement of health formation in the pre-school years and the context in which the intervention takes place. Measurement of health formation is difficult. In the early years, the guardian's parenting style and health literacy and the health system are factors in child health formation and its measurement. Commonly studied outcomes such as asthma status can be misleading, where an absence of illness is the target, but prompt diagnosis and treatment is preferred over non-diagnosis or misdiagnosis. A successful intervention aims to both improve child health and parental health literacy, therefore simultaneously increasing and decreasing different types of asthma diagnoses. Therefore, across the home visiting literature, an absence of significant differences should not always be interpreted as a lack of treatment effects. Here, we examine health outcomes where clear treatment effects are expected such as hospital use, BMI and social-emotional health. However, we also examine secondary outcomes, where hypothesised treatment effects are less clear.

Further to this understanding of the complexity of health reporting, the context for the intervention is also considered. The theoretical framework for many home visiting programmes, including PFL, is that the programme impacts health problems arising in disadvantaged environments. However, as with the identification of many health shocks, the age at which these disparities arise should be taken into account. We find no evidence of health inequalities in the PFL cohort compared to their nationally representative peers in early childhood, although there are strong indications of health inequalities among mothers. This could arise from the broad access to healthcare and other welfare and social services in Ireland, or from the inherent difficulty in measuring health formation in the pre-school years. This underscores the need for long-term follow-up when measurement of health formation and the presence of health inequalities are more clear-cut.

A major limitation of this study is sample size, as it precludes us from studying major health shocks, such as traumatic brain injury, that are rare in children but have ramifications throughout the lifecycle. Where even modest effect sizes may be expected or desired, in some samples the study is not sufficiently powered to detect such effects. Given the little evidence of inequalities in health in early childhood, hypothesis of smaller effects may be prudent but not testable. An additional limitation is insufficient detail on GP use to elucidate programme effects on substitution away from the ED arising from increased health literacy.

However, we examine less urgent ED use, which is a substitute for GP use in this particular community, and found no evidence of programme effects. In our study of socioeconomic inequalities, there are limited available outcomes and the ages and frequency of data collection also differ across the two studies. While controls and robustness checks are used to mitigate this, they cannot be disregarded when discussing the apparent absence of health inequalities for the PFL cohort.

In the context of the child health and home visiting literature, our findings are largely expected, although the absence of effects on maternal outcomes and social-emotional development is surprising. Sandner et al (2017) argue that the cost of behaviour change is prohibitive for outcomes such as smoking. We see some evidence of this, with alcohol consumption increasing slightly and child diet quality declining in both groups over time. There is also no evidence that differences in service use are driven by reductions in less urgent use of ED services, another proxy for health behaviour and literacy.

The strongest hypothesised treatment effects in the literature are on social-emotional development, where we find no evidence of effects. A continuous measure is used given evidence that sub-clinical developmental problems in the pre-school years matter for later mental health. The ASQ social-emotional scale was chosen due to its availability at all time points and acceptable rates of internal reliability.[13] In other studies of PFL using different measures of social-emotional development, such as the Child Behavior Checklist, there is some evidence of treatment effects (Doyle 2020). In this paper, however, where we emphasise the trajectory of child health over time, there is no indication of improved social-emotional development using the ASQ measure.

Overweightness in children is associated with both nutrition and activity, therefore diet and ED attendance for accidents and injuries highlight some of the complexity in early childhood health formation. We examine diet and injury-led ED attendances in each year, and BMI at age four. Programme effects on diet, particularly in the earlier years, and BMI were anticipated, with less certain hypothesis on injury. Diet quality decreases in both groups over time, with no differences in the first two years when most children receive adequate nutrition. At age three, although diet quality continues to decline in both groups, the intervention group are significantly more likely to have an adequate diet. At age four, diet quality in both groups declines again and there is no difference

[13] Cronbach's $\alpha = 0.79$ for the Preparing for Life sample at age four where $\alpha > 0.7$ indicates an acceptable level of internal reliability (Doyle and UCD Geary Institute PFL Evaluation Team 2015).

between them, however there is some evidence of programme effects on overweightness at age four. Simultaneously, there is evidence of an increased rate of non-head injuries in the intervention group in year four, a feature of the intervention that encourages increased activity in children. Treatment effects on BMI could therefore be driven by lagged effects on diet, increased activity leading to increased injury, or both. However, without a measure of physical activity we also cannot discount the role of poor supervision, neglect, or another cause in the apparent increase in injury.

The significant effects on the objective hospital records and directly measured outcomes, such as BMI, compared to the subjective parent-reported measures like the ASQ are somewhat surprising. The literature indicates that a flattening of the SES gradient is typically found when using subjective outcomes, arising from differential reporting of health status by SES and/or differences arising from access to healthcare (Reinhold and Jürges 2012). Therefore, we would expect to identify more significant differences for the subjective outcomes. In fact, we find evidence to the contrary. Furthermore, evidence suggests that the reporting of health status in one year should reflect reported service use in the following year (National Research Council 2004). However, we do not find any evidence of this as the direction of the treatment effects on parent reported service use are contrary to those reported for health status in the previous year.

Several theories may explain our time inconsistent results, with effects appearing in the second and third year only. It may reflect the programme's ability to impact health formation only at certain points over the trajectory of development. Programme structure may be a factor in this, particularly in regards the time periods in which pertinent tip sheets are delivered, the building of knowledge and practice over time, or a decrease in programme adherence in some periods. Furthermore, home visiting programmes may simply act as an accelerant in the formation of child health: facilitating knowledge, parenting practices and behaviours, and environmental maintenance that the control participants will engage with of their own accord at a later point. As the intervention took place during the Irish Recession, programme effects may have been mitigated by countervailing forces such as reduced access and increased cost of healthcare among disadvantaged families and the health burden from increased poverty and stress (Reinhard et al. 2018; Case, Lubotsky, and Paxson 2002; Leininger and Levy 2015).[14] Finally, with the low levels of health service use

[14] Between 2005 and 2011 there was a 20 percent increase in healthcare costs in Ireland, with a 10 percent increase in prices such as hospital, outpatient, GP, and dentist fees. Patient burden was further increased through lengthening waiting lists. For medical card holders, a prescription co-payment was introduced in 2010 and increased in 2013 and entitlements to dental services were

in the final year, differences may not be expected at a time when children experience few health shocks. However, the importance of health behaviours such as worsening diets indicates that health shocks may begin to compound as the children age.

The results presented here support the theory that it is not the number of health shocks experienced by children that affects their health formation, but rather their parents' ability to invest in their health. For example, few significant differences emerged regarding general measures of health and illness, however the severity of health issues reflected in high triage ED use and health behaviours such as diet differ between the groups. It is surprising then that we do not observe differences in maternal health behaviours. This may reflect the cost to the parent or the accrual of knowledge – changes in health behaviours for their children's health may result from improved parenting practices and/or the low cost to the parent to implement such changes. However, maternal health behaviours such as reducing alcohol consumption may be too costly for the parent to enact.

Further work in this area should focus on costs, mediation, measurement, heterogeneity, and long-term follow-up. In devising and evaluating home visiting programmes, particularly those that aim to benefit health, it would be useful to study the cost of investment for parents in changing health behaviours, the environment, and parenting practices. This, alongside increased information on the socioeconomic gradient in a given cohort, may be beneficial in unlocking explanations for patterns of skill formation. With respect to the PFL programme, follow-up data in later years could be used to address questions about the apparent fade out of health effects at age four and allow for an examination of chronic and development disease. Hypothesis setting on treatment effects may also be clearer at later points, when socioeconomic inequalities are evident. In addition, detailed knowledge of service use, health behaviours, and rigorously measured health outcomes may help to untangle any lasting impacts. Hospital use among the control children cost €2,981 over four years, almost double the cost for intervention children. While health effects were moderate over the programme cycle, a cost-benefit analysis of the programme would be insightful. Reducing costs while maintaining health through a shift in services may be of interest to policymakers as a significant cost-saving device.

This work is a contribution to the literature on the home visiting effects on health formation by adding to our knowledge of treatment effects on child

cut in 2010. Reduced income limits and stricter eligibility requirements led to the withdrawal of 65,000 medical cards over six months between 2013 and 2014 (Nolan et al. 2014).

health. We find that the home visiting programme has modest effects on health, driven by a reduction in severity in the toddler years. Novel methods are used to mitigate issues that arise in similar studies with respect to multiple hypothesis testing and attrition. We further highlight the importance of careful hypothesis setting in interpreting health formation in complex interventions through considering both measurement and context. The evidence on health effects in home visiting is inconsistent and weak, however with careful interpretation of what health formation entails in the early years, we find that the moderate effects of PFL are, in fact, indicators for a positive trajectory in health formation over the life cycle.

References

Allin, Sarah, and Mark Stabile. 2012. 'Socioeconomic Status and Child Health: What Is the Role of Health Care, Health Conditions, Injuries and Maternal Health?' *Health Economics, Policy and Law* 7 (2): 227–42.

Anderson, Vicki, Celia Godfrey, Jeffrey V. Rosenfeld, and Cathy Catroppa. 2012. 'Predictors of Cognitive Function and Recovery 10 Years After Traumatic Brain Injury in Young Children'. *Pediatrics* 129 (2): e254. https://doi.org/10.1542/peds.2011-0311.

Apouey, Bénédicte, and Pierre-Yves Geoffard. 2013. 'Family Income and Child Health in the UK'. *Journal of Health Economics* 32 (4): 715–27. https://doi.org/10.1016/j.jhealeco.2013.03.006.

Avellar, Sarah A., and Lauren H. Supplee. 2013. 'Effectiveness of Home Visiting in Improving Child Health and Reducing Child Maltreatment'. *Pediatrics* 132 (Supplement 2): S90. https://doi.org/10.1542/peds.2013-1021G.

Barnes, Jacqueline. 2016. 'Short-Term Health and Social Care Benefits of the Family Nurse Partnership Lack Evidence in the UK Context but There Is Promise for Child Developmental Outcomes'. *Evidence Based Medicine* 21 (4): 145. https://doi.org/10.1136/ebmed-2016-110422.

Bartlett, Susan J., Jerry A. Krishnan, Kristin Alice Riekert, A. M. Butz, Floyd J. Malveaux, and Cynthia S. Rand. 2004. 'Maternal Depressive Symptoms and Adherence to Therapy in Inner-City Children with Asthma'. *Pediatrics* 113 (2): 229–37.

Beattie, T. F., D. R. Gorman, and J. J. Walker. 2001. 'The Association between Deprivation Levels, Attendance Rate and Triage Category of Children Attending a Children's Accident and Emergency Department'. *Emergency Medicine Journal* 18 (2): 110. https://doi.org/10.1136/emj.18.2.110.

Berg, Gerard J. van den, and Pia R. Pinger. 2016. 'Transgenerational Effects of Childhood Conditions on Third Generation Health and Education Outcomes'. *Economics & Human Biology* 23 (December): 103–20. https://doi.org/10.1016/j.ehb.2016.07.001.

Bongers, Marloes E. J., Michiel P. van Wijk, Johannes B. Reitsma, and Marc A. Benninga. 2010. 'Long-Term Prognosis for Childhood Constipation: Clinical Outcomes in Adulthood'. *Pediatrics* 126 (1): e156. https://doi.org/10.1542/peds.2009-1009.

Callan, T, B Colgan, C Keane, and J.R. Walsh. 2015. 'Modelling Eligibility for Medical Cards and GP Visit Cards: Methods and Baseline Results'. 515. ESRI Working Papers. https://www.esri.ie/system/files?file=media/file-uploads/2015-11/WP515.pdf.

Case, Anne, Angela Fertig, and Christina Paxson. 2005. 'The Lasting Impact of Childhood Health and Circumstance'. *Journal of Health Economics* 24 (2): 365–89. https://doi.org/10.1016/j.jhealeco.2004.09.008.

Case, Anne, Darren Lubotsky, and Christina Paxson. 2002. 'Economic Status and Health in Childhood: The Origins of the Gradient'. *The American Economic Review* 92 (5): 1308–34.

Christakis, Dimitri A., Chris Feudtner, Catherine Pihoker, and Frederick A. Connell. 2001. 'Continuity and Quality of Care for Children With Diabetes Who Are Covered by Medicaid'. *Ambulatory Pediatrics* 1 (2): 99–103. https://doi.org/10.1367/1539-4409(2001)001<0099:CAQOCF>2.0.CO;2.

Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Erlbaum.

Cohen, Jessica, and Pascaline Dupas. 2010. 'Free Distribution or Cost-Sharing? Evidence from a Randomized Malaria Prevention Experiment*'. *The Quarterly Journal of Economics* 125 (1): 1–45. https://doi.org/10.1162/qjec.2010.125.1.1.

Condliffe, Simon, and Charles R. Link. 2008. 'The Relationship between Economic Status and Child Health: Evidence from the United States'. *American Economic Review* 98 (4): 1605–18. https://doi.org/10.1257/aer.98.4.1605.

Conti, Gabriella, and James J. Heckman. 2013. 'The Developmental Approach to Child and Adult Health'. *Pediatrics* 131 (Supplement 2): S133. https://doi.org/10.1542/peds.2013-0252d.

Côté, Sylvana M, Massimiliano Orri, Richard E Tremblay, and Orla Doyle. 2018. 'A Multicomponent Early Intervention Program and Trajectories of Behavior, Cognition, and Health'. *Pediatrics* 141 (5).

Currie, Janet. 2011. 'Inequality at Birth: Some Causes and Consequences'. *The American Economic Review* 101 (3): 1–22.

Currie, Janet, Mark Stabile, Phongsack Manivong, and Leslie L. Roos. 2010. 'Child Health and Young Adult Outcomes'. *The Journal of Human Resources* 45 (3): 517–48.

Dooley, Martin, and Jennifer Stewart. 2007. 'Family Income, Parenting Styles and Child Behavioual-Emotional Outcomes'. *Health Economics* 16: 145–62. https://doi.org/10.1002/hec.1142.

Downing, A., and G. Rudge. 2006. 'A Study of Childhood Attendance at Emergency Departments in the West Midlands Region'. *Emergency Medicine Journal : EMJ* 23 (5): 391–93. https://doi.org/10.1136/emj.2005.025411.

Doyle, Orla. 2020. 'The First 2000 Days and Children's Skills'. *Journal of Political Economy* 128 (6).

Doyle, Orla, Nick Fitzpatrick, Judy Lovett, and Caroline Rawdon. 2015. 'Early Intervention and Child Physical Health: Evidence from a Dublin-Based Randomized Controlled Trial'. *Economics & Human Biology* 19 (Supplement C): 224–45. https://doi.org/10.1016/j.ehb.2015.09.004.

Doyle, Orla, Colm Harmon, James J. Heckman, Caitriona Logue, and Seong Hyeok Moon. 2017. 'Early Skill Formation and the Efficiency of Parental Investment: A Randomized Controlled Trial of Home Visiting'. *Field Experiments in Labor Economics and Social Policies* 45 (April): 40–58. https://doi.org/10.1016/j.labeco.2016.11.002.

Doyle, Orla, Edel McGlanaghy, Eylin Palamaro-Munsell, and F. M McAuliffe. 2014. 'Home Based Educational Intervention to Improve Perinatal Outcomes for

a Disadvantaged Community: A Randomised Control Trial'. *European Journal of Obstetrics and Gynecology and Reproductive Biology* 180 (September): 162–67. https://doi.org/10.1016/j.ejogrb.2014.06.006.

Doyle, Orla, and UCD Geary Institute PFL Evaluation Team. 2015. 'Preparing for Life Early Childhood Intervention Assessing the Early Impact of Preparing for Life at Forty-Eight Months.' UCD Geary Institute. https://geary.ucd.ie/preparingforlife/wp-content/uploads/2016/08/48MoReportSmallFilePROOF.pdf.

Duflo, Esther, Rachel Glennerster, and Michael Kremer. 2008. 'Using Randomization in Development Economics Research: A Toolkit'. In *Handbook of Development Economics*, edited by P Schultz and J Strauss, 1st ed. Elsevier.

Elkan, Ruth, Denise Kendrick, Michael Hewitt, Jane Robinson, K Tolley, Mitch Blair, Michael Dewey, Debbie Williams, and Kathy Brummell. 2004. 'The Effectiveness of Domiciliary Health Visiting: A Systematic Review of International Studies and a Selective Review of the British Literature'. *Health Technol Assess* 4 (13).

Enserink, R., L. Mughini-Gras, E. Duizer, T. Kortbeek, and W. Van Pelt. 2015. 'Risk Factors for Gastroenteritis in Child Day Care'. *Epidemiology and Infection* 143 (13): 2707–20. https://doi.org/10.1017/S0950268814003367.

Filene, Jill H., Jennifer W. Kaminski, Linda Anne Valle, and Patrice Cachat. 2013. 'Components Associated With Home Visiting Program Outcomes: A Meta-Analysis'. *Pediatrics* 132 (Supplement 2): S100. https://doi.org/10.1542/peds.2013-1021H.

Freedman, David A. 2008. 'On Regression Adjustments to Experimental Data'. *Advances in Applied Mathematics* 40 (2): 180–93. https://doi.org/10.1016/j.aam.2006.12.003.

Gall, Seana L, Natalie Schuz, Benjamin Schuz, Kara Martin, Joan Abbott-Chapman, Nadia Ollington, George C Patton, Terence Dwyer, and Alison J Venn. 2019. 'Childhood Health Motivation and Adult Cardiometabolic Health in the Childhood Determinants of Adult Health (CDAH) Study'. *Health Psychology* 38 (4): 297–305.

Gennetian, Lisa A., Heather D. Hill, Andrew S. London, and Leonard M. Lopoo. 2010. 'Maternal Employment and the Health of Low-Income Young Children'. *Journal of Health Economics* 29 (3): 353–63. https://doi.org/10.1016/j.jhealeco.2010.02.007.

Glenton, Claire, Christopher J Colvin, Benedicte Carlsen, Alison Swartz, Simon Lewin, Jane Noyes, and Arash Rashidian. 2013. 'Barriers and Facilitators to the Implementation of Lay Health Worker Programmes to Improve Access to Maternal and Child Health: Qualitative Evidence Synthesis'. *Cochrane Database of Systematic Reviews*, no. 10. https://doi.org/10.1002/14651858.CD010414.pub2.

Hack, Maureen. 1999. 'Consideration of the Use of Health Status, Functional Outcome, and Quality-of-Life to Monitor Neonatal Intensive Care Practice'. *Pediatrics* 103 (Supplement E1): 319.

Heckman, James J. 2006. ‘Skill Formation and the Economics of Investing in Disadvantaged Children’. *Science* 312: 1900–1902.
———. 2008. ‘Schools, Skills, and Synapses’. *Economic Inquiry* 46 (3): 289–324.
———. 2012. ‘The Developmental Origins of Health’. *Health Economics* 21: 24–29.
Heckman, James J., Seong Hyeok Moon, Rodrigo Pinto, Peter Savelyev, and Adam Yavitz. 2010. ‘Analyzing Social Experiments as Implemented: A Reexamination of the Evidence from the HighScope Perry Preschool Program’. *Quantitative Economics* 1 (1): 1–46.
Imbens, Guido W., and Jeffrey M. Wooldridge. 2009. ‘Recent Developments in the Econometrics of Program Evaluation’. *Journal of Economic Literature* 47 (1): 5–86.
Ioannidis, John P. A., T. D. Stanley, and Hristos Doucouliagos. 2017. ‘The Power of Bias in Economics Research’. *The Economic Journal* 127 (605): F236–F265. https://doi.org/10.1111/ecoj.12461.
Jones, S R, S Carley, and M Harrison. 2003. ‘An Introduction to Power and Sample Size Estimation’. *Emergency Medicine Journal* 20 (5): 453. https://doi.org/10.1136/emj.20.5.453.
Kader, Manzur, Elinor Sundblom, and Liselotte Schäfer Elinder. 2015. ‘Effectiveness of Universal Parental Support Interventions Addressing Children’s Dietary Habits, Physical Activity and Bodyweight: A Systematic Review’. *Preventive Medicine* 77 (August): 52–67. https://doi.org/10.1016/j.ypmed.2015.05.005.
Kass, Robert E., and Adrian E. Raftery. 1995. ‘Bayes Factors’. *Journal of the American Statistical Association* 90 (430): 773–95. https://doi.org/10.1080/01621459.1995.10476572.
Khandaker, Golam M, Jorge Zimbron, Christina Dalman, Glyn Lewis, and Peter B Jones. 2012. ‘Childhood Infection and Adult Schizophrenia: A Meta-Analysis of Population-Based Studies’. *Schizophrenia Research* 139 (1–3): 161–68. https://doi.org/10.1016/j.schres.2012.05.023.
Koniak-Griffin, Deborah, Inese L. Verzemnieks, Nancy L. R. Anderson, Mary-Lynn Brecht, Janna Lesser, Sue Kim, and Carmen Turner-Pluta. 2003. ‘Nurse Visitation for Adolescent Mothers: Two-Year Infant Health and Maternal Outcomes’. *Nursing Research* 52 (2): 127–36.
Leininger, Lindsey, and Helen Levy. 2015. ‘Child Health and Access to Medical Care’. *The Future of Children* 25 (1): 65–90.
Lenth, Russell V. 2001. ‘Some Practical Guidelines for Effective Sample Size Determination’. *The American Statistician* 55 (3): 187–93.
Mackenbach, J P, I Stirbu, A R Roskam, M M Schaap, G Menvielle, M Leinsalu, and A E Kunst. 2008. ‘Socioeconomic Inequalities in Health in 22 European Countries’. *The New England Journal of Medicine* 358 (23): 2468–81.
Maslow, Gary R., Abigail A. Haydon, Carol Ann Ford, and Carolyn Tucker Halpern. 2011. ‘Young Adult Outcomes of Children Growing Up With Chronic Illness: An Analysis of the National Longitudinal Study of Adolescent

HealthChildren Growing Up With Chronic Illness'. *JAMA Pediatrics* 165 (3): 256–61. https://doi.org/10.1001/archpediatrics.2010.287.
McDaid, David, Miriam Wiley, Anna Maresso, and Elias Mossialos. 2009. 'Ireland: Health System Review'. *Health Systems in Transition* 11 (4): 1–268.
McGovern, Mark E. 2013. 'Still Unequal at Birth: Birth Weight, Socio-Economic Status and Outcomes at Age 9'. *The Economic and Social Review* 44 (1): 53–84.
Mejdoubi, Jamila, Silvia C. C. M. van den Heijkant, Frank J. M. van Leerdam, Matty Crone, Alfons Crijnen, and Remy A. HiraSing. 2014. 'Effects of Nurse Home Visitation on Cigarette Smoking, Pregnancy Outcomes and Breastfeeding: A Randomized Controlled Trial.' *Midwifery* 30 (6): 688–95. https://doi.org/10.1016/j.midw.2013.08.006.
Miller, S, Maguire, LK, and G Macdonald. 2011. 'Home-based Child Development Interventions for Preschool Children from Socially Disadvantaged Families'. *Cochrane Database of Systematic Reviews*, no. 12. https://doi.org/10.1002/14651858.CD008131.pub2.
National Research Council. 2004. 'Measuring Children's Health'. In *Children's Health, The Nation's Wealth: Assessing and Improving Child Health*. Washington DC: National Academies Press. https://www.ncbi.nlm.nih.gov/books/NBK92208/#.
Nievar, M. Angela, Laurie A. Van Egeren, and Sara Pollard. 2010. 'A Meta-Analysis of Home Visiting Programs: Moderators of Improvements in Maternal Behavior'. *Infant Mental Health Journal* 31 (5): 499–520. https://doi.org/10.1002/imhj.20269.
Nolan, Anne, Sarah Barry, Sara Burke, and Stephen Thomas. 2014. 'The Impact of the Crisis on the Health System and Health in Ireland'. In *Economic Crisis, Health Systems and Health in Europe: Impact and Implications for Policy.*, 143–78. Maidenhead: Open University Press.
Nolan, Anne, and Richard Layte. 2014. 'Socio-Economic Inequalities in Child Health in Ireland'. *The Economic and Social Review; Vol 45, No 1, Spring (2014): With Policy Papers from 'Future Directions of the Irish Economy' Conference, 10 January 2014*, March. http://www.esr.ie/article/view/107/75.
Olds, David L., Charles R. Henderson, and Harriet Kitzman. 1994. 'Does Prenatal and Infancy Nurse Home Visitation Have Enduring Effects on Qualities of Parental Caregiving and Child Health at 25 to 50 Months of Life?' *Pediatrics* 93 (1): 89.
O'Sullivan, Aifric, Nick Fitzpatrick, and Orla Doyle. 2017. 'Effects of Early Intervention on Dietary Intake and Its Mediating Role on Cognitive Functioning: A Randomised Controlled Trial'. *Public Health Nutrition* 20 (1): 154–64. https://doi.org/10.1017/S1368980016001877.
Peacock, Shelley, Stephanie Konrad, Erin Watson, Darren Nickel, and Nazeem Muhajarine. 2013. 'Effectiveness of Home Visiting Programs on Child Outcomes: A Systematic Review'. *BMC Public Health* 13 (17).

Pedersen, Soren Erik, Suzanne S. Hurd, Robert F. Lemanske Jr., Allan Becker, Heather J. Zar, Peter D. Sly, Manuel Soto-Quiroz, Gary Wong, and Eric D. Bateman. 2011. 'Global Strategy for the Diagnosis and Management of Asthma in Children 5 Years and Younger'. *Pediatric Pulmonology* 46 (1): 1–17. https://doi.org/10.1002/ppul.21321.

Peduzzi, Peter, John Concato, Elizabeth Kemper, Theodore R. Holford, and Alvan R. Feinstein. 1996. 'A Simulation Study of the Number of Events per Variable in Logistic Regression Analysis'. *Journal of Clinical Epidemiology* 49 (12): 1373–79. https://doi.org/10.1016/S0895-4356(96)00236-3.

Perry, Cynthia D. 2008. 'Does Treating Maternal Depression Improve Child Health Management? The Case of Pediatric Asthma'. *Journal of Health Economics* 27 (1): 157–73. https://doi.org/10.1016/j.jhealeco.2007.03.005.

PFL Evaluation Team. 2016. 'Preparing for Life Early Childhood Intervention Final Report: Did Preparing for Life Improve Children's School Readiness'. http://geary.ucd.ie/preparingforlife/wp-content/uploads/2016/09/5654_FP_UCD_Report_Final.pdf.

Propper, Carol, John Rigg, and Simon Burgess. 2007. 'Child Health: Evidence on the Roles of Family Income and Maternal Mental Health from a UK Birth Cohort'. *Health Economics* 16 (11): 1245–69.

Rahi, Jugnoo S., Phillippa M. Cumberland, and Catherine S. Peckham. 2009. 'Visual Function in Working-Age Adults: Early Life Influences and Associations with Health and Social Outcomes'. *Ophthalmology* 116 (10): 1866–71. https://doi.org/10.1016/j.ophtha.2009.03.007.

Reilly, J J, and J Kelly. 2010. 'Long-Term Impact of Overweight and Obesity in Childhood and Adolescence on Morbidity and Premature Mortality in Adulthood: Systematic Review'. *International Journal Of Obesity* 35 (October): 891.

Reinhard, Erica, Richard Layte, Cathal McCrory, Lidia Panico, and Mauricio Avendano. 2018. 'The Great Recession and the Health of Young Children: A Fixed-Effects Analysis in Ireland.' *American Journal of Epidemiology* 187 (7): 1438–48. https://doi.org/10.1093/aje/kwy001.

Reinhold, Steffen, and Hendrik Jürges. 2012. 'Parental Income and Child Health in Germany'. *Health Economics* 21: 562–79. https://doi.org/10.1002/hec.1732.

Robinson, Monique, Wendy Oddy, Garth Kendall, Nicholas de Klerk, Sven Silburn, Stephen Zubrick, John Newnham, Fiona Stanley, and Eugen Mattes. 2008. 'Pre- and Postnatal Influences on Preschool Mental Health: A Large-Scale Cohort Study'. *Journal of Child Psychology & Psychiatry* 49 (10): 1118–28.

Robling, Michael, Marie-Jet Bekkers, Kerry Bell, Christopher C Butler, Rebecca Cannings-John, Sue Channon, Belen Corbacho Martin, et al. 2016. 'Effectiveness of a Nurse-Led Intensive Home-Visitation Programme for First-Time Teenage Mothers (Building Blocks): A Pragmatic Randomised Controlled Trial'. *The Lancet* 387 (10014): 146–55. https://doi.org/10.1016/S0140-6736(15)00392-X.

Salvy, S.-J., K. de la Haye, T. Galama, and M. I. Goran. 2017. 'Home Visitation Programs: An Untapped Opportunity for the Delivery of Early Childhood Obesity Prevention'. *Obesity Reviews* 18 (2): 149–63. https://doi.org/10.1111/obr.12482.

Sandner, Malte, Thomas Cornelissen, Tanja Jungmann, and Peggy Herrmann. 2018. 'Evaluating the Effects of a Targeted Home Visiting Program on Maternal and Child Health Outcomes'. *Journal of Health Economics* 58: 269–83.

Sariaslan, Amir, David J. Sharp, Brian M. D'Onofrio, Henrik Larsson, and Seena Fazel. 2016. 'Long-Term Outcomes Associated with Traumatic Brain Injury in Childhood and Adolescence: A Nationwide Swedish Cohort Study of a Wide Range of Medical and Social Outcomes'. *PLOS Medicine* 13 (8): e1002103. https://doi.org/10.1371/journal.pmed.1002103.

Schwarz, Gideon. 1978. 'Estimating the Dimension of a Model'. *The Annals of Statistics* 6 (2): 461–64.

Staines, Anthony, Kevin P. Balanda, Steve Barron, Yvonne Corcoran, Lorraine Fahy, Louise Gallagher, Tessa Greally, et al. 2016. 'Child Health Care in Ireland'. *Diversity of Child Health Care in Europe: A Study of the European Paediatric Association/Union of National European Paediatric Societies and Associations* 177 (October): S87–106. https://doi.org/10.1016/j.jpeds.2016.04.046.

Steyerberg, E. W., M. J. C. Eijkemans, F. E. Harrell, and J. D. F. Habbema. 2000. 'Prognostic Modelling with Logistic Regression Analysis: A Comparison of Selection and Estimation Methods in Small Data Sets'. *Statistics in Medicine* 19: 1059–79.

Sweet, Monica A., and Mark I. Appelbaum. 2004. 'Is Home Visiting an Effective Strategy? A Meta-Analytic Review of Home Visiting Programs for Families with Young Children'. *Child Development* 75 (5): 1435–56.

Timmons, Brian W., Allana G. LeBlanc, Valerie Carson, Sarah Connor Gorber, Carrie Dillman, Ian Janssen, Michelle E. Kho, John C. Spence, Jodie A. Stearns, and Mark S. Tremblay. 2012. 'Systematic Review of Physical Activity and Health in the Early Years (Aged 0–4 Years)'. *Applied Physiology, Nutrition, and Metabolism* 37 (4): 773–92. https://doi.org/10.1139/h2012-070.

Van As, A. B., D. Schulman, P. Mtambeka, C. Mavengere, and A. Numanoglu. 2016. 'Head Injuries in Young Children; the Forgotten Pandemic'. *Injury Prevention* 22 (Suppl 2): A94. https://doi.org/10.1136/injuryprev-2016-042156.258.

Wagner, Yvonne, and Roswitha Heinrich-weltzien. 2017. 'Evaluation of a Regional German Interdisciplinary Oral Health Programme for Children from Birth to 5 Years of Age'. *Clinical Oral Investigations* 21 (1): 225–35. https://doi.org/10.1007/s00784-016-1781-8.

Wechsler, David. 1999. *Weschsler Abbreviated Scale of Intelligence: WASI Manual*. San Antonio, Texas, USA: The Psychological Association.

Welch, David, and Patrick J D Dawes. 2007. 'Variation in the Normal Hearing Threshold Predicts Childhood IQ, Linguistic, and Behavioral Outcomes'. *Pediatric Research* 61 (June): 737.

Wickham, Sophie, Elspeth Anwar, Ben Barr, Catherine Law, and David Taylor-Robinson. 2016. 'Poverty and Child Health in the UK: Using Evidence for Action'. *Archives of Disease in Childhood* 101 (8): 759. https://doi.org/10.1136/archdischild-2014-306746.

Wooldridge, Jeffrey M. 2007. 'Inverse Probability Weighted Estimation for General Missing Data Problems'. *Journal of Econometrics* 141 (2): 1281–1301. https://doi.org/10.1016/j.jeconom.2007.02.002.

Appendix

For the online appendix for this working paper, visit https://deirdrecoy.com/research