Life Course Centre Working Paper Series

lifecoursecentre

ARC Centre of Excellence for Children and Families over the Life Course

Do Welfare Restrictions Improve Child Health? Estimating the Causal Impact of Income Management in the Northern Territory

Mary-Alice Doyle
School of Economics, The University of Sydney

Stefanie Schurer
School of Economics, The University of Sydney & IZA

Sven Silburn
Menzies School of Health Research, Charles Darwin University

No. 2017-23
December 2017

The University of Queensland Australia

The University of Melbourne

The University of Western Australia

The University of Sydney

Australian Government
Australian Research Council

NON-TECHNICAL SUMMARY

In 2007, the Australian government changed the way that welfare recipients in Indigenous communities in the Northern Territory were paid. Instead of receiving unconditional cash payments, recipients had half of their payments quarantined into a separate account which could only be used for priority goods. One of the stated goals of income management was to improve child outcomes by increasing the share of household income spent on food and other household essentials, and reducing the amount spent on potentially harmful goods such as alcohol and tobacco.

Over the past decade, governments, commentators and journalists have questioned whether income management has achieved this goal. Despite widespread use of such restrictions, and a large literature evaluating the impact of individual transfer programs, there is relatively little empirical evidence that directly compares the impact of restricted with unrestricted transfers. In this paper, we provide a first estimate of the impact of income management on child health using data on birth outcomes.

The findings of our study suggest that income management did not improve one measure of child health outcomes, and, by extension, that income management does not appear to have produced the desired change in household consumption patterns, at least for households with pregnant women. In fact, income management may have had a negative impact on newborn health – lower average birthweights and a higher probability of low birthweight (defined as less than 2500g), over and above what would be expected if a baby was premature.

Other research suggests that policy-driven changes in pregnant women's nutrition have the largest impact on birthweight if introduced in the third trimester. But exposure to income management mattered most for birth outcomes when it occurred in the first or second trimester of the pregnancy, which suggests that channels other than lack of nutrition may have caused the adverse effects. We are unable to identify the channel of the treatment effect, but one possibility could be increasing levels of stress experienced by the mother when income management was implemented. The restriction may have disrupted existing intra-household financial arrangements, by changing the value of resources over which individual household members had discretionary decision-making power. This may have negatively affected consumption patterns or even caused conflict within the household. Another source of stress could have originated from policy implementation problems, such as confusion over how to access funds, the administrative and time burden of allocating funds to various uses, and general dissatisfaction with this policy as part of the NTER policy package (see Cobb-Clark, Kettlewell et al., 2017 who find similar negative effects on school attendance).

Our paper contributes to the literature in three important ways. First, it adds to the emerging international literature on the effectiveness of restricted transfers relative to cash transfers, especially for highly vulnerable and marginalised populations. Second, our research contributes to a wide literature that studies the long-term drivers and consequences of early-life health problems, especially low birthweight. Third, this research contributes directly to the debate on the effectiveness of income management in reducing health disparities.

ABOUT THE AUTHORS

Mary-Alice Doyle is a graduating economics masters student at the University of Sydney, having worked on this paper as part of her thesis. In her professional career, she has held roles as an economist analysing international economies, the Australian labour market, and the retail payments system. Her research interests are in social and health policy, programme evaluation and behavioural insights. Email: mdoy966o@uni.sydney.edu.au

Stefanie Schurer is an Associate Professor and a SOAR Fellow in the School of Economics at The University of Sydney. Her main research interest is the Economics of Human Development. Most of her current research projects explore the evolution of skills, preferences, and health over the life course and the role that parents and the public sector play in determining these skills. From 2014 to 2017 she will be in a research-only position funded by an ARC Discovery Early Career Research Award (DECRA) titled "Exceptional upward mobility against all odds: Noncognitive skills and early-childhood disadvantage". Email: stefanie.schurer@sydney.edu.au

Sven Silburn is a national leader in clinical, epidemiological and evaluative research in child development and education, youth mental health and suicide prevention. Over the past 10 years he has been a chief investigator on research grants totalling over $35m and an author of 75 scientific publications and reports. He has served on several national, state and territory advisory boards and been active in policy advocacy in the areas of the developmental health of Indigenous and non-Indigenous children and suicide prevention. Email: Sven.Silburn@menzies.edu.au

ACKNOWLEDGEMENTS: The authors thank the following people for their valuable feedback on this research: Steven Guthridge, Olga Havnen, Heather d'Antoine, Matthew Gray, Nicholas Biddle, Jim Smith, Liz Moore, Matthew James, David Cooper, Julie Brimblecombe, Gawaian Bodkin-Andrews, Maggie Walter and Dilhan Perera. The authors acknowledge valuable feedback from seminar participants at the University of Chicago, and at Indigenous Affairs Group at the Department of the Prime Minister and Cabinet, Canberra. This study uses data from the Northern Territory (NT) Early Childhood Data Linkage Project, `Improving the developmental outcomes of NT Children: A data linkage study to inform policy and practice across the health, education and family services sectors'', funded through a National Health and Medical Research Council (NHMRC) Partnership Grant (No.1091491), and with Menzies School of Health Research, the NT Government and the Aboriginal Medical Services Alliance Northern Territory (AMSANT) as the project's organisational partners. This study uses administrative data obtained from the NT Departments of Health and Education provided through this NHMRC Partnership Project. The analysis conducted in this project has followed the NHMRC 'Values and Ethics: Guidelines for Ethical Conduct in Aboriginal and Torres Strait Islander Health Research' (2003) and the Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS) 'Guidelines for Ethical Research in Australian Indigenous Studies (2012) (Reciprocity, Respect, Equality, Responsibility, Survival and Protection, Spirit and Integrity'). The researchers are bound by, and the research analysis complies with, all ethical standards outlined in the ethics agreement from the Human Research Ethics Committee of the Northern Territory Department of Health and Menzies School of Health Research (HREC Reference Number: 2016-2611). Mary-Alice Doyle was added as a named investigator to this agreement in January 2017. This research received financial support from an Australian Research Council (ARC) Discovery Early Career Award (DE140100463) and the Australian Research Council Centre of Excellence for Children and Families over the Life Course (project number CE140100027).

lifecoursecentre

(ARC Centre of Excellence for Children and Families over the Life Course)
Institute for Social Science Research, The University of Queensland (administration node)
UQ Long Pocket Precinct, Indooroopilly, Qld 4068, Telephone: +61 7 334 67477
Email: lcc@uq.edu.au, Web: www.lifecoursecentre.org.au

Abstract

In 2007, the Australian government changed the way that welfare recipients in Indigenous communities in the Northern Territory were paid. Instead of receiving unconditional cash payments, recipients had half of their payments quarantined into a separate account which could only be used for priority goods. One of the objectives of the policy was to improve child welfare. Using administrative data on the universe of births in the Northern Territory, this is the first study to evaluate the impact of the policy on children's health outcomes at birth. To identify the causal impact of income management, we exploit its staggered and unsystematic rollout as a source of exogenous variation. Because the policy was intended to create a healthy consumption environment, we expect to see an increase in average birthweight. We find no such effect. Restricting welfare payments reduced birthweight by over 100 grams and increased the probability of low birthweight by around 30 percent. We find no evidence that these results are driven by changes in fertility, the composition of mothers having babies, access to better neonatal health care, or an increase in risky health behaviours (smoking, drinking) during pregnancy. We find that exposure to income management mattered most when it occurred in the first or second trimester of the pregnancy, suggesting that channels other than nutrition may have caused the adverse effects, potentially including increased exposure to stress.

Keywords: welfare restrictions; conditional cash transfers; birth outcomes; Indigenous health; Australia

1 INTRODUCTION

In both developed and developing countries, policymakers and donors routinely place restrictions on how welfare recipients can use their benefits. Restricted welfare sometimes takes the form of in-kind transfers (e.g., food stamps in the US). In other cases, restrictions are placed on the use of cash transfers (e.g., Prospera in Mexico). Despite widespread use of such restrictions, and a large literature evaluating the impact of individual transfer programs, there is relatively little empirical evidence that directly compares the impact of restricted with unrestricted transfers (Gentilini, 2015).[1]

This paper addresses this gap, using a unique natural experiment to test the impact of new restrictions on cash transfers in Australia. Starting in 2007, the Australian Government rolled out 'income management' for welfare payment recipients living in remote Aboriginal communities in the Northern Territory (NT). Recipients previously received payments directly into their bank accounts. Under income management, half of each payment was quarantined in a separate account with restrictions over its use.

This paper focuses on the initial rollout of income management in the NT in 2007, which represented the beginning of a broader rollout of the policy (Figure 1). Income management currently applies to welfare recipients in the NT (orange), and to recipients in certain other parts of Australia (purple and dark green). More recently, a modified version of income management has been trialled in communities in Western Australia and South Australia (light green). In early 2017, the Australian Government announced an intention to expand that program to additional sites, citing it as a policy tool to promote the health and wellbeing of residents in disadvantaged communities.

[1] We use the term 'restricted transfers' to incorporate in-kind transfers, as well as cash transfers that come with restrictions or conditions over how recipients should use their funds. As Das and Do (2005) note, the economics of these policy options is very similar.

Figure 1: Income management in Austrtalia, 2017

Source: Australian Department of Social Services

One of the stated goals of income management was to improve child outcomes by increasing the share of household income spent on food and other household essentials, and reducing the amount spent on potentially harmful goods such as alcohol and tobacco (FaCHSIA 2008). Until now, evaluators had judged that it was not possible to quantify the impact of income management on this goal (AIHW 2010, FaHCSIA 2011); this paper provides a first estimate.[2]

Using the gradual rollout of income management across communities as a source of exogenous variation, we estimate its causal impact on early-life health outcomes. We use administrative data covering the universe of births in the NT from the NT Early Childhood Data Linkage Project. We focus on birthweight; if income management was successful in increasing household food consumption during pregnancy and reducing consumption of excluded goods, such as alcohol and tobacco (which can harm development in utero), we would expect an

[2] The research team are also currently investigating the impact of income management on school attendance rates. Preliminary findings are reported in an unpublished manuscript by Cobb-Clark et al. (2017).

increase in birthweight. We compare outcomes for newborns in communities where income management was introduced before or during the pregnancy, with outcomes for newborns in communities where income management was not yet implemented at birth, or was implemented very late in the pregnancy. This approach is similar to that of Almond, Hoynes et al. (2011) who estimated the causal impact of the US food stamp program. We are able to satisfy the two identification assumptions of this approach: that the rollout schedule appears exogenous, and that trends in 'treatment' and 'control' groups were similar before the rollout.

Our paper contributes to the literature in three important ways. First, it adds to the emerging international literature on the effectiveness of restricted transfers relative to cash transfers, especially for highly vulnerable and marginalised populations. Policy makers often justify the restriction of welfare payments by appealing to social preferences or paternalism, especially when the consumption of certain goods has either negative (e.g. excessive consumption of alcohol and tobacco) or positive (e.g. investment in education and health care) spill-over effects for communities (Currie & Gahvari, 2008).

Our context is different from programs that have been studied previously. Similar to many other in-kind transfers or conditional cash transfers, the goal of income management is to improve social and economic well-being in Aboriginal communities by increasing the consumption of priority goods and decreasing the consumption of goods with potential negative spill-overs. But unlike programs in other countries which typically target only discretionary income or additional payments, at the time of the rollout, Australia's scheme was compulsory for all welfare recipients in the affected communities, limiting their ability to spend core welfare entitlements (see Mendes et al., 2014). Our analysis therefore also contributes to the literature on place-based policy interventions aimed at boosting local development interventions that have been widely implemented in disadvantaged regions of developed countries (Neumark & Simpson, 2015). Income management was intended to be a community-

level intervention, with benefits expected by policymakers to permeate throughout the community through positive spill-over effects.

Second, our research contributes to a wide literature that studies the long-term consequences of early-life health problems, especially low birthweight. Birthweight has been used as key indicator of infant health and welfare and the central focus of infant health policy (Almond et al. 2011). The reason is that low birthweight children have been found to be delayed or impaired in their cognitive development, educational attainment and later-life health and economic outcomes (see Filio, Guryan et al., 2014 for a recent review). Some find that the adverse effects of low birthweight may persist across generations (Black, Devereux et al. 2007, Currie and Moretti, 2007, Victora, Adair et al. 2008). These negative effects imply a large potential cost to society. From a public policy perspective, low birthweight also implies an immediate public cost in terms of increased health care expenditures and pressures on public hospitals.

Third, this research contributes directly to the debate on the effectiveness of income management in reducing health disparities. To date, what is known about the initial rollout of income management in the NT comes from qualitative evidence which can at best be described as mixed. Some sources report widespread dissatisfaction with implementation problems and the prescriptive nature of the scheme (Yu et al., 2008). Evidence based on qualitative data obtained from non-randomly sampled survey participants suggest that some Aboriginal Australians believed that income management had benefits, while others believed that income management had not improved their lives (Central Land Council, 2008; AIHW, 2010).

The findings of our study suggest that income management did not improve birth outcomes for Aboriginal children in the Northern Territory. In fact, the reform appears to have reduced average birthweight of children exposed to income management in utero before the start of the third trimester of the pregnancy. The adverse effects are strongest at the bottom end of the

birthweight distribution, which is consistent with our finding that income management also increased the risk of low birthweight. Our results cannot be explained by changes in the composition of the mothers who were willing to have a baby during the roll out of the policy, or in fertility. The effects are also not driven by better access to neonatal care, that may increase the detection of pregnancy problems, or an increase in the maternal consumption of alcohol or cigarettes. We conducted a series of robustness checks that exclude the possibility that our results are driven by seasonal variations, small communities, and model assumptions.

Other research has found transfer policies aimed at boosting nutrition to have the greatest effect on birthweight in the third trimester of pregnancy. But exposure to income management mattered most for birth outcomes when it occurred in the first or second trimester of the pregnancy, which suggests that channels other than lack of nutrition may have caused the adverse effects. We are unable to identify the channel of the treatment effect, but one possibility could be increasing levels of stress experienced by the mother when income management was implemented. Based on the policy design, we see two potential avenues through which income management could have caused additional stress. The restriction may have disrupted existing intra-household financial arrangements, by changing the value of resources over which individual household members had discretionary decision-making power. This may have negatively affected consumption patterns or even caused conflict within the household. Another source of stress could have originated from policy implementation problems, such as confusion over how to access funds, the administrative and time burden of allocating funds to various uses, and general dissatisfaction with this policy as part of the NTER policy package (see Cobb-Clark, Kettlewell et al., 2017 who find similar negative effects on school attendance).

The remainder of this study proceeds as follows. Section 2 describes the institutional context of the income management policy. In Section 3 we review the relevant literature, including

international evidence on the effectiveness of restricted cash transfers and evidence from the Australian income management experience. In Section 4 we outline the identification strategy and the economic theory underlying the mechanisms through which income management may affect child wellbeing. Section 5 describes the data and empirical framework. Estimation results, and a series of robustness checks, are presented in Section 6. Section 7 discusses the implications of our findings and concludes. An appendix provides supplementary material.

2 POLICY BACKGROUND

2.1 THE NORTHERN TERRITORY

The Northern Territory is a vast geographic area, stretching over 521,000 square miles (about two times the size of Texas) and covering approximately one-sixth of the Australian continent. Around half of its approximately 246,000 residents live in the capital city of Darwin. Aboriginal and Torres Strait Islanders make up 25.5 percent of the Northern Territory's total population despite constituting only 2.8 percent of the Australian population overall. The Northern Territory is governed by its own local government in conjunction with the Australian Federal Government and approximately half of the land in the Northern Territory is Aboriginal-owned as a result of the *Aboriginal Land Rights (Northern Territory) Act of 1976*.

Aboriginal kinship relationships are complex, dynamic and not easily captured by non-Aboriginal notions of family, or based on physical living arrangements (Lohoar et al 2014). In particular, as people see themselves in relation to others in their local communities as well as in other remote areas, it is common for children and adults to move between households. Raising children is a collective responsibility; Aboriginal children are given a great deal of autonomy to develop their skills by exploring their environment under the watchful eyes of the community at large (Lohoar et al 2014, Muir & Bohr 2014).

Education experts and community leaders have struggled to find ways to ensure that Aboriginal children can access “Western cultural capital" while at the same time nurturing their Aboriginality and Aboriginal culture (McTaggart 1991, p. 297). However, levels of school engagement among indigenous children in the NT remain well below the national average (Commonwealth of Australia 2017). Similarly, over the past decade, extensive attention has been given to efforts to ‘close the gap’ in health outcomes between Indigenous and non-Indigenous Australians. Official targets aim to halve the gap in child mortality between 2008 and 2018, and to close the gap in life expectancy by 2031, but these targets are not currently on track to be achieved (Commonwealth of Australia 2017).

2.2 WHAT IS INCOME MANAGEMENT?

In early 2007, the Northern Territory Board of Inquiry into the Protection of Aboriginal Children from Sexual Abuse released its report titled *Ampe Akelyernemane Meke Mekarle “Little Children are Sacred”* (2007). The report called for immediate government action to address relatively high rates of child sexual abuse in remote communities. It emphasised the need to consider child neglect, alcoholism and inadequate education as long-term contributors to abuse.

In mid-2007 in response to the report, the Australian Government announced the Northern Territory Emergency Response (NTER). The NTER included a range of policies, such as alcohol and pornography bans, additional police presence, night patrols, child health checks and housing and land reform, in addition to income management. These policies applied to residents in 73 remote Aboriginal communities and outstations, and 10 town camps.[3] To facilitate the targeted nature of these policies towards Aboriginal Australians, the government suspended the *Racial Discrimination Act 1975*.

[3] A town camp is an Aboriginal community situated in a town or city, or close to its boundaries.

Income management was a key part of the NTER. Under income management, recipients faced new restrictions on what they could do with their welfare payments: half of each regular payment was set aside into the recipient's income management account, and could only be directed towards priority needs such as food, housing, bills and clothing. The remaining half was paid into recipients' bank accounts as usual. The full amount of any lump sum payment was also allocated into recipients' income management accounts.

Income managed funds were distributed in consultation with a case officer. Initially, recipients were required to meet with a case officer to create a plan, and funds were then allocated manually in line with that plan in each payment period. For instance, recipients could choose to have part of their income managed funds paid directly to suppliers to cover their bills, rent or debt repayments. They could also have some of their funds credited in their name to a local store to purchase food and household goods, and could leave some funds in their income managed account as savings. Any changes to these allocations were made in consultation with a case officer. Between September and December 2008, towards the end of the rollout period, a debit card system (the Basics Card) was introduced. This was a more flexible system, allowing recipients to load their funds onto the card and use it to purchase items at any participating store.

Since the NTER, income management has been rolled out more broadly. After the reinstatement of the *Racial Discrimination Act* in 2010, a modified version of income management was introduced to cover the whole of the NT, and place-based and voluntary income management was subsequently introduced in certain locations within other states. More recently, in early 2017 the government completed trials of a similar program (renamed the Cashless Debit Card) in a small number of communities in Western Australia and South Australia, and has since announced an intention to expand the policy to additional communities.

2.3 WHO WAS AFFECTED?

Income management applied to the vast majority of welfare recipients living in NTER communities and town camps, and therefore affected most residents in remote communities. While detailed data on welfare payment rates are unavailable, in aggregate around 55 percent of adults in NTER communities were income managed by the end of the rollout period, with around 80 percent having been income managed at some point during the rollout period (AIHW 2010).[4] Though limited information is available on which residents were affected, women and younger adults were more likely to receive welfare payments (AIHW 2010), meaning the rate was likely above-average for pregnant women.

The number of people affected by income management was probably greater than the number of adults receiving welfare payments. Given the large average household sizes in NTER communities (see Table 1 below), many residents who were not themselves recipients were likely living in households with somebody who was. In addition, if income management was successful in reducing consumption of alcohol and reducing financial harassment (or 'humbugging'), it could have had community-wide externalities, potentially contributing to a safer community for all residents.

Recipients did not have the ability to appeal the application of income management, though some exemptions were given, for example, to individuals who resided in an NTER community only temporarily, who had little connection to the community, or who had moved permanently away. By the end of March 2009, 15,125 people were subject to income management, but only 649 exemptions (three percent of all individuals who were, at some point, subject to income management) had been granted (AIHW 2010).

[4] The lower share of residents affected at the end of the period reflects residents moving onto and off of income support payments, for example, due to changes in employment status or eligibility.

3 LITERATURE

3.1 RESTRICTED AND UNRESTRICTED TRANSFERS

There is limited empirical evidence on the relative impact of restricted over unrestricted cash transfers. A small number of recent studies consider this question and report mixed findings. These studies mainly relate to transfer programs in low- and middle-income countries, though some evidence on the US food stamps program is also relevant.

Gentilini (2016) surveys ten studies – from Bangladesh, Cambodia, the Democratic Republic of Congo, Ecuador, Ethiopia, Mexico, Niger, Sri Lanka, Uganda and Yemen – that use either randomised controlled trials or natural experiments to compare transfers of cash to transfers of food. The programs were diverse, with the transfer value ranging from 2½ to 30 percent of households' average expenditure. In some cases, food was given directly to participants, while in others, participants were given food vouchers.

Most studies found no significant difference between the impact of the cash transfer and the restricted transfer. Both increased household food consumption and dietary diversity and reduced the incidence of malnutrition. Contrary to expectations, food consumption was significantly higher under the cash transfer than the restricted transfer in three out of the ten studies, while in others there was no significant difference. Gentilini concludes that there is no clear evidence that either transfer type is more effective. However, based on the limited information on each program's cost, cash or voucher transfers appear to be more cost-effective.

Another recent study from Uruguay estimates the impact of a temporary transfer program on birthweight (Amarante, Manacorda et al. 2016). The program began as a cash transfer, but partway through, 25-50 percent of participants' payments were converted to food vouchers. While the increase in income from the transfer reduced the incidence of low birthweight (by 19-25 percent), the partial conversion of transfers to food vouchers had no additional impact.

However, this program provides only a low-powered test of the impact of restricting transfers on birthweight; the voucher component of the transfer was equivalent to just 6-13 percent of income for the average household, and therefore highly likely to be inframarginal (see Section 4.1).

In the US, the effectiveness of restricting transfers has been studied at length with respect to the food stamps program. Researchers have used a range of methods to identify the impact of food stamps relative to cash transfers, but with no clear consensus. The studies with the clearest identification have been those that examine responses to program introduction or rule changes (Senauer and Young 1986, Hoynes and Schanzenbach 2009, Beatty and Tuttle 2015), or responses to "cash-out" experiments, where participants in certain states had their stamps converted to cash (Wilde and Ranney 1996, Breunig and Dasgupta 2005).

Several studies find that food stamp income is equivalent to cash income (Hoynes and Schanzenbach 2009, Cuffey, Beatty et al. 2016). But some have identified a "cash-out puzzle", finding that food stamps increase food consumption, even among households that already spend more on food than the value of their food stamps. Breunig and Dasgupta (2005) suggest this may be the result of intra-household bargaining dynamics. For instance, in a household with children, the primary carer may have a stronger preference for food expenditure to provide for the children, and may exert more control over the allocation of food stamp income than cash income.

A significant challenge that most of the US studies face is that their identification strategy is based on a program introduction or rule change which involves an increase in income. They must therefore disentangle the effect of welfare restrictions from the effect of higher benefits. This paper builds on the US literature, but uses a context where, like the papers reviewed by Gentilini, the effect of a move from unrestricted to restricted transfers can be observed without

any confounding change in the level of payments. Moreover, Australia's income management policy provides an interesting complement to the existing literature, because the program includes both restricted and unrestricted transfer components. Gentilini (2016) highlights that there is little evidence on these combined program types, and identifies this gap in the literature as an avenue for future research.

3.2 EVIDENCE ON INCOME MANAGEMENT

Within Australia, there have been many attempts to evaluate the impact of income management in its various forms, but with a focus mainly on qualitative data. Two quantitative studies have considered the localised impact of income management on components of household expenditure, but they find differing results. Considering the impact of income management on food expenditure, Brimblecombe, McDonnell et al. (2010) use monthly sales data from ten community stores. They find no evidence that the overall value of food sales changed following the introduction of income management, and no change in the share of sales directed towards fruit, vegetables or tobacco. The authors caution, however, that their results may not be representative of all income managed communities; before the rollout of income management, the stores in this study already provided a voluntary 'Food Card' system to residents, which restricted purchases to nutritious items.[5]

Conversely, Lamb and Young (2011) find evidence that income management may have reduced gambling expenditure. In one gambling venue in each of Alice Springs and Katherine, they find reductions in monthly revenue per electronic poker machine of $450 and $800, respectively, but no impact at other venues. They argue that these two venues are likely to be the ones most frequented by Aboriginal welfare recipients, and interpret their results as

[5] The 'Food Card' program was in use before income management was introduced, and was subsequently provided to welfare recipients as an optional way of accessing income managed funds before the Basics Card was rolled out. The authors also note most of the ten communities had pre-existing alcohol bans.

tentative evidence that income management reduced formal gambling expenditure. However, they note that their findings do not necessarily mean that total gambling expenditure decreased; there may have been a commensurate increase in informal gambling.

Apart from these two studies, government departments have written or commissioned evaluation reports on income management in its various forms. Two reports cover the introduction of income management in the NT, which is also the focus of this paper (Australian Institute of Health and Welfare (AIHW) 2010; Department of Families, Hosing, Community Services and Indigenous Affairs (FaHCSIA 2011)). Other reports cover more recent introductions of the policy, such as 'new income management' in the NT, introduced in 2010 (Bray, Gray et al. 2014), and the 2016-2017 Cashless Debit Card trials (Orima Research 2017).[6]

The two reports covering the initial rollout, which is the subject of this paper, reach broadly positive conclusions. The AIHW report concludes that there is consistent evidence that the policy led to more income being spent on primary needs, and the FaHCSIA report finds that while income management was perceived negatively in its early stages, it is "now seen as beneficial by many people, especially women" (p. 11). No baseline data were collected before income management was rolled out, so benefits found in the evaluation reports relate to survey respondents' perceptions of changes. For instance, 69 percent of surveyed respondents felt that children in their communities were getting more food than before income management was introduced, and 57 percent reported that children were healthier than they had been three years earlier (FaHCSIA 2011).

These two reports cover a broad set of outcomes, but rely mainly on qualitative surveys of community residents and staff involved in administering the program. These surveys were conducted with small, non-random samples so are unlikely to be representative of the treated

[6] See Department of Social Services website for reports on other forms of income management.

population.[7] In addition, they rely on the accuracy of respondents' impressions and recollections, and may be susceptible to response bias, potentially causing respondents to under-report behaviour that is seen as socially undesirable (Buckmaster and Ey 2012).

Importantly, even if the survey data are assumed to be accurate and representative, findings in these reports are unlikely to reflect the impact of income management itself. Many other NTER policies had been implemented by the time of the surveys. Therefore, at best, these surveys would provide information on the impact of the package of NTER policies, but not income management itself.

In terms of quantitative evidence, aggregate data on Indigenous children in remote NT communities show that after income management was introduced, there was a decrease in the share of children diagnosed with anaemia, or who were underweight or stunted before age four (FaHCSIA 2011). But these rates had been trending down over a longer period, so it is not clear whether these changes can be attributed to income management.

This paper builds on the existing evidence, offering a more robust method but narrower focus. Like Brimblecombe, McDonnell et al (2010) and Lamb and Young (2011), we use quantitative data, avoiding the potential for response, recollection or sample-selection bias which may be present in survey data. Our unique dataset allows for two key contributions. First, because we have data on all births in the NT, we can estimate the impact across all affected communities and not a just a subset. Having data on all affected communities also means we can exploit the gradual rollout to estimate a causal effect, moving beyond a 'before and after' analysis. Second, our data allow us to focus directly on health outcomes, rather than changes in spending patterns which may or may not lead to improved health (see Section 4.3).

[7] Respondents were reportedly invited to participate by community brokers or government business managers (AIHW 2010).

3.3 MATERNAL BEHAVIOUR AND BIRTHWEIGHT

There are two potential causes of low birthweight: gestational length and intrauterine growth restriction. The determinants of each are different and complex. The causes of short gestational length (or prematurity) have been found to include the mother's pre-pregnancy weight, history of prematurity, and stress levels during pregnancy. Intrauterine growth restrictions, which lead to below-average weight for normal gestational length, are generally found to be more likely to be affected by maternal nutrition during pregnancy (Kramer 1987). While nutrition is important throughout pregnancy, Almond, Hoynes et al. (2011) find that the effect was largest in the US when food stamps were in place for the whole of the third trimester of pregnancy, with no additional impact if introduced earlier.

There is strong evidence that increased food consumption during pregnancy can increase birthweight through the intrauterine growth channel, and further, that transfer programs can help to increase birthweight through this channel (Bitler and Currie 2005, Barber and Gertler 2008, Almond, Hoynes et al. 2011, Hoynes, Page et al. 2011, Amarante, Manacorda et al. 2016). Therefore, this is the channel through which we expect to see an impact, if any, of income management on birth outcomes.

The magnitude of effects from previous studies are wide-ranging. For instance, the studies described above from the US and Uruguay found that the introduction of food or cash transfer programs reduced the probability of low birthweight by 0.7 to 2.4 percentage points (7 to 25 percent), and increased average birthweight by 13-30 grams (½ to 4½ percent) (Almond, Hoynes et al. 2011, Amarante, Manacorda et al. 2016). But these transfer programs represented a relatively low share of household income. A study of the Oportunidades conditional cash transfer program in Mexico found much larger effects, with a 130 gram increase in birthweight and a 4.6 percentage point decrease in probability of low birthweight (Barber and Gertler 2008).

Outside of the economic literature, nutrition-focused pregnancy and pre-pregnancy interventions have yielded similarly large impacts. For instance, a recent meta-analysis finds that, on average, randomised trials that provide food or fortified food products during pregnancy increase birthweight by 125 grams (Gresham, Byles et al. 2014). Therefore, if there was an impact of income management on nutrition during pregnancy, we may expect an effect of a similar magnitude.

4 CONCEPTUAL FRAMEWORK AND IDENTIFICATION

4.1 ECONOMIC THEORY

Income management is a budgeting tool, aimed at increasing the funds available for purchase of food and other household essentials and preventing recipients from over-consuming addictive substances. If the policy was successful in changing consumption patterns, we would expect to see improved health outcomes. In particular, if the policy led to increased food consumption then nutrition and weight gain during pregnancy should improve, leading to higher average birthweight. It follows that two conditions are required for income management to have this hypothesised effect. First the restriction must cause recipients to consume more household essentials. Second, this change in consumption must affect health outcomes.

Under this first condition, fa restricted transfer must affect household consumption behaviour differently from an equivalent cash transfer if it is to have an impact. A simple model first developed by Southworth (1945) describes the potential for such change.

Households maximise utility by consuming either ‘priority goods’ (X), or other goods (Y). Before income management is introduced, the household has cash income of M + IM.

$$U = U(X, Y)$$

$$s.t.\ P_x X + P_y Y = M + IM$$

After the introduction of income management, the household retains cash income M, but the amount IM is quarantined and must be spent on priority goods. Therefore, the household faces a second restriction:

$$s.t.\ P_y Y \leq M$$

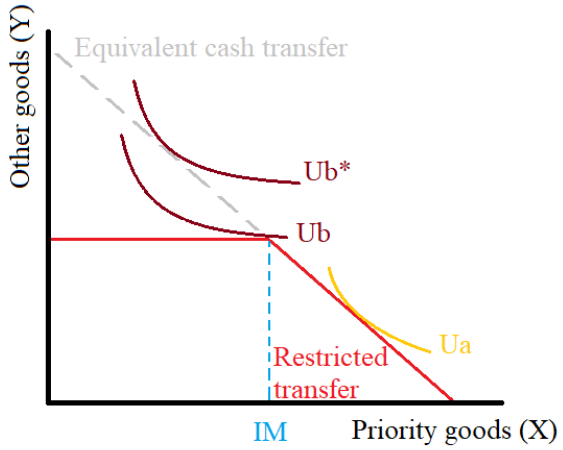
For some households this constraint is already satisfied, because even before the introduction of income management, the household spends more than the minimum amount IM on priority goods ($P_x\ X\ \geq\ IM$). These households are referred to as inframarginal. If a household is inframarginal, the introduction of income management should not impact household expenditure, because the marginal propensities to consume (MPC) priority goods out of IM and M would be equal:

$$MPC_{x,M} = MPC_{x,IM}$$

It is only if the household initially consumes priority goods less than the value of IM that income management would affect consumption; these would be 'extramarginal' households, with $MPC_{x,M} < MPC_{x,IM}$.

Figure 2 demonstrates this model, illustrating two potential responses to the introduction of income management. Household A is inframarginal; it is unaffected by the restriction as it would already optimally spend more than amount IM on priority goods to reach utility indifference curve U_a. However, household B is extramarginal; with the move from cash to restricted transfers, it shifts its consumption towards more priority goods, moving to a lower indifference curve, U_b.

Figure 2: Effect of restricting welfare payments for inframarginal and extramarginal households



Given this simple framework, a first step in our analysis is to test the null hypothesis that birth outcomes are unchanged under income management against the alternative that birth outcomes were affected. The null hypothesis corresponds to the inframarginal case. The alternative hypothesis corresponds to the extramarginal case where the MPC of priority goods (of which the largest component is food)[8] out of income managed funds is greater than the MPC of priority goods out of cash income. If the alternative hypothesis is true – at least for a non-negligible share of households – we would expect income management to improve birth outcomes through increased food consumption.

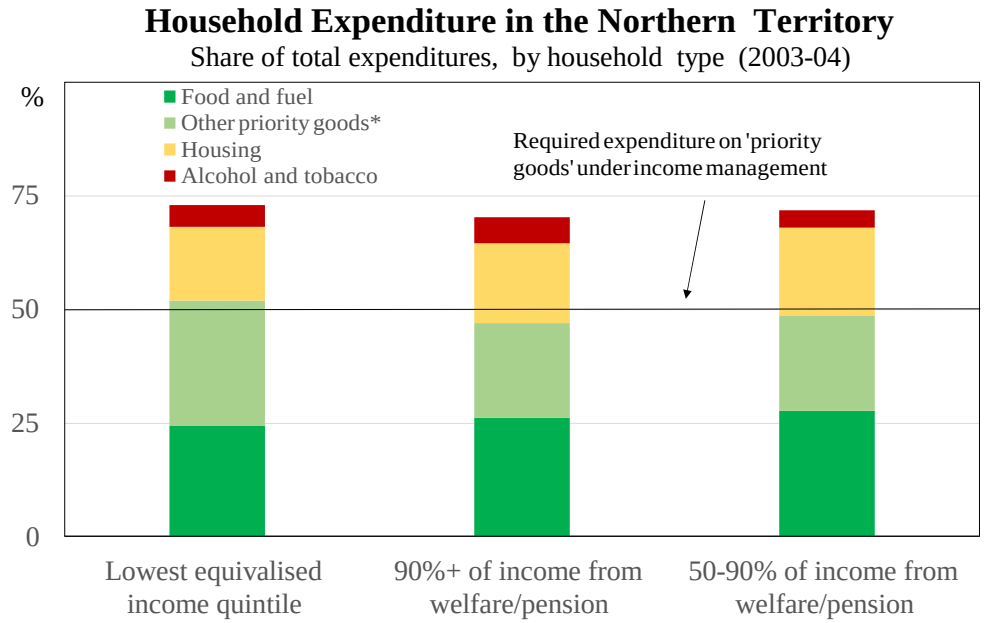
4.2 EVIDENCE ON HOUSEHOLD EXPENDITURE

While we are unable to directly test the impact of income management on household expenditure, aggregate data suggest that the average low-income household in the NT was inframarginal. Community-level data are not available, but pre-rollout ABS data on spending patterns for low-income and welfare-dependent households in the NT provide a proxy.[9] They

[8] Around 65 percent of income managed funds were spent on food during the rollout period (AIHW 2010).
[9] Note that the sample for the ABS Household Expenditure Survey does not cover residents in very remote areas.

show that over 60 percent of average household expenditure was already directed towards priority goods before income management (Figure 3).[10]

Figure 3



* Includes: clothing and footwear; household furnishings and equipment, medical care and health expenses; transport.
Source: ABS Household Expenditure Survey

However, these data do not tell us the distribution of expenditures across households, and are not available for residents in very remote communities. Even if the average household is inframarginal, a significant portion may still have been extramarginal, meaning that we may still observe an impact of the restriction on average outcomes.[11]

4.3 INPUTS AND OUTCOMES

The theoretical framework described above explains the potential for income management to affect household consumption. But such a change is only relevant to policymakers if it translates to outcomes such as improved health, education and wellbeing. As Cunha (2014)

[10] Total household income is likely to be higher than the value of welfare income. The savings rate (i.e., the gap between income and consumption) is unlikely to affect this conclusion as recipients had the option of saving in income managed funds or in cash.

[11] The ABS data may over-estimate housing expenditure for our population. Many residents in remote NTER communities had low or no housing costs.

notes, this is not always the case. When recipients receive food transfers, or are limited to purchasing certain items, they may consume these items instead of close substitutes, with no resulting change in nutrition. For example, in Mexico, Cunha finds that extramarginal transfers of cornflour, cereal and milk powder led to a substitution of consumption away from other grains and sources of protein, but no overall change in food consumption or health outcomes.

A similar dynamic may be relevant in our context, with critics of income management having argued that the requirement to shop at licensed food stores[12] may reduce purchases through less formal channels. For instance, income management may have reduced purchases at local farmers' markets and garage sales. This may even reduce households' purchasing power and consumption of nutritious foods, as products through these less formal channels could be more nutritious or cheaper (AIHW 2010). Alternatively, if the household already spent an adequate amount on food, income management may have increased consumption of non-nutritious foods, which could have no impact or even a negative impact on birth outcomes (Grieger and Clifton 2015). Further, recipients for whom the policy was extramarginal may have circumvented the policy, by trading store cards or items purchased through income management for cash. If prevalent, all of these factors could reduce or prevent any impact of the restriction on health outcomes.

4.4 IDENTIFICATION

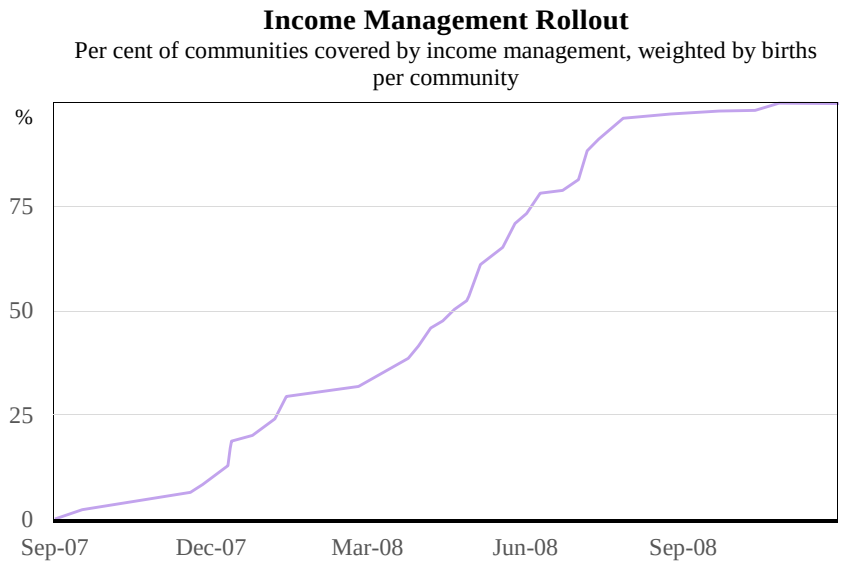
We use the staggered rollout of income management as our identification strategy. Income management was not introduced with evaluation in mind, but using high-frequency data (see Section 5.1 for data description) and the timeline of the rollout, we are able to estimate its

[12] For local stores to accept income management funds, they had to receive a license, indicating that they met certain minimum standards (in terms of their stock of fresh and nutritious foods, for example). In many cases, licenses were provided provisionally at the time of the rollout, with any changes required to meet the minimum standards occurring after the rollout.

impact. Provided that the rollout timeline is exogenous, we can assign a causal interpretation to our results.

Income management was rolled out to 73 communities and outstations and 10 town camps between September 2007 and October 2008 (Figure 4). Importantly for our purposes, it was rolled out on a different timeline from other NTER policies, meaning that our results are unlikely to be confounded by concurrent policy changes. While exact dates of the introduction of other policies are not available, some – such as alcohol bans – were introduced immediately to all communities, while others – such as night patrols and increased police presence – followed a slower rollout and affected only a subset of communities that did not already have these in place before the NTER (FaHCSIA 2011; see Appendix A for details). For instance, just 18 communities had received a new police presence by mid-2008, while 64 communities had received income management (FaHCSIA 2011).

Figure 4



Since income management was rolled out in stages, we can use this time delay to identify an appropriate control group – communities that have not yet received income management but will do so in the future – against which the outcomes of the treatment group can be compared.

We therefore compare outcomes for newborns in communities where income management was introduced before or during the pregnancy (the 'treatment' group), with outcomes for newborns in communities where income management was not yet implemented at birth or was implemented very late in the pregnancy (the 'control' group). A similar approach was used by Almond, Hoynes et al. (2011) to estimate the impact of the US food stamp program.[13] For this identification strategy to work, we must establish that the rollout schedule was exogenous, and that trends in 'treatment' and 'control' groups were similar before the rollout.

The rollout was conducted following a pre-defined timeline, though no information is publicly available explaining the logic of that timeline. In the weeks prior to the scheduled introduction of income management in a given community, there was a consultation period in which Centrelink[14] staff would visit the community, meet with payment recipients, establish relationships with local businesses to allow funds to be allocated to them, and ensure other pre-conditions (such as police presence and support by a Government Business Manager) were met (AIHW 2010).[15] If these conditions were not met, the rollout would be delayed.[16]

The existence of pre-conditions suggest that income management may have been rolled out in larger, better-resourced communities first, as they were more likely to meet the conditions. If true, our treatment effect would be positively biased, as the early-adopting communities (from which our sample has a higher number of 'treated' births) would have better health facilities and access to fresh food. Our identification strategy is strengthened, but does not exclusively rely, on the fact that each community changes status from control to treatment. Indeed, for most

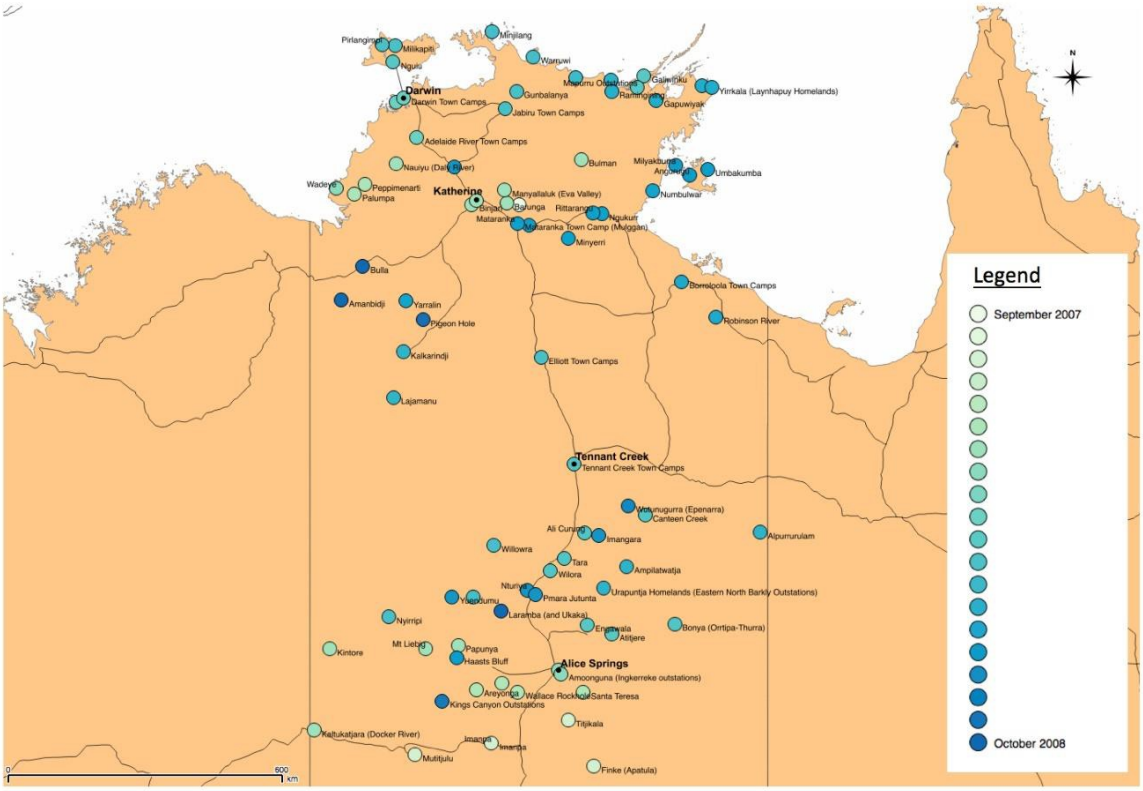
[13] Almond, Hoynes et al. (2011) exploited the staggered rollout over many years and across many counties. Their study focuses on county-level data, whereas we use individual data.
[14] Centrelink is the Australian government agency that distributes welfare payments.
[15] In addition, around 40 communities received money management training prior to the rollout, though as yet we have been unable to find information on which communities these were, how they were chosen, and what the training involved.
[16] The implementation was delayed for 13 of around 90 communities, outstations and town camps. Delays were substantial (that is, more than a few weeks) for four communities (AIHW 2010). Our identification is based on the actual, not planned, rollout dates.

communities, our sample contains both treatment and control observations. But within our sample period, there are 21 communities for which either all births were treated, or all were in the control group. There are two reasons for this. First, one of the communities that received income management on the first day of the rollout, and two of the communities that received it on the last day, had all treatment and all control observations, respectively (9 observations). Second, there are some very small communities for which we observe just few births during the rollout period (39 observations). All these communities still contribute to identification of the causal treatment effect, but we are unable to control for within-community variation for these 48 observations.

If, for instance these communities are extreme outliers, whereby the treatment-only communities displayed very good pre-treatment outcomes, and the control-only communities displayed very poor outcomes, we may over-estimate the treatment effect of income management. But despite these data limitations, we are able to rule out that these fixed-treatment communities are likely to bias our estimation results.

First, the rollout did not follow any clear geographic pattern, suggesting that location-specific characteristics that could affect health and access to health care (e.g., frequency of flooding, access to fresh food and distance to major population centres) were not correlated with the rollout schedule (Figure 5). Income management was rolled out in parallel in two ‘clusters’ (north and south), but with no apparent pattern as to whether very remote communities, larger communities or town camps received treatment first within each cluster.

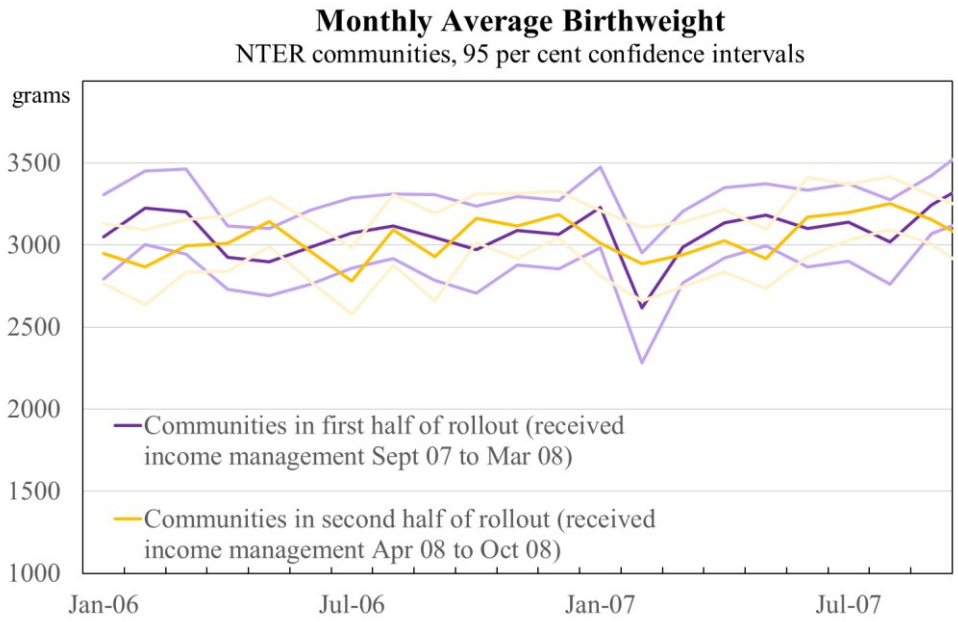
Figure 5: Geographical pattern of income management rollout



Second, in the year prior to the rollout, birth outcomes in communities that received income management early were no different from those that received it later. Based on our administrative data, we find no apparent trend in birthweight in either group prior to the rollout (Figure 6).[17]

[17] The dip in January 2007 represents a seasonal pattern (with generally worse birth outcomes during the wet season). This seasonality is controlled for in our econometric estimation.

Figure 6



Third, earlier and later adopting communities did not differ significantly in terms of average birth outcomes, birth complications, or community characteristics pre-rollout. Table 1 reports mean differences between early- and late-adopting communities[18] before the NTER (data are from our administrative perinatal dataset, and from the 2006 ABS Census). Most birth outcome measures, including obstetric complications, characteristics of the mother and APGAR scores[19] were similar between the two groups before the rollout.[20]

[18] Early and late adopters are defined as communities where income management was implemented between September 2007 and mid-April 2008, and between mid-April 2008 and October 2008, respectively.
[19] APGAR is a test (appearance, pulse, grimace, activity and respiration) given to infants at 1 minute and 5 minutes after birth.
[20] The difference in intrauterine growth restrictions becomes significant at the 10 percent level if we exclude the 21 communities for which there was no variation in within-community treatment status during the rollout period (See

Appendix B: Descriptive statistics).

Table 1 Pre-rollout outcomes and community characteristics

Year prior to NTER (1 July 2006 - 30 June 2007)

Outcome variables	NTER communities: Communities in first half of rollout	NTER communities: Communities in second half of rollout	Difference	Rest of NT
	(1)	(2)	(3)	(4)
Birthweight (grams)	3072.3	3082.2	9.9	3354.3
	(30.13)	(28.71)	(41.67)	(10.98)
Low birthweight (%)	14.4	14.0	-0.5	6.8
	(1.62)	(1.61)	(2.28)	(0.48)
Premature (%)	15.5	15.0	-0.5	7.9
	(1.67)	(1.66)	(2.35)	(0.51)
Obstetric complications				
Due to intrauterine growth restriction (%)	5.1	3.2	-1.9	1.4
	(1.01)	(0.82)	(1.3)	(0.23)
Due to anaemia (%)	9.6	9.9	0.3	2.0
	(1.35)	(1.38)	(1.94)	(0.27)
Due to gestational diabetes (%)	7.2	9.0	1.8	6.8
	(1.19)	(1.33)	(1.78)	(0.48)
Any (%)	43.5	45.9	2.4	24.2
	(2.28)	(2.31)	(3.25)	(0.81)
Other characteristics				
Age of mother	23.9	23.7	-0.1	28.6
	(0.28)	(0.28)	(0.4)	(0.12)
Aboriginal status mother (%)	91.7	94.2	2.5	21.3
	(1.27)	(1.08)	(1.67)	(0.78)
APGAR 1	8.0	8.0	0.0	8.2
	(0.08)	(0.09)	(0.12)	(0.03)
APGAR 5	8.8	8.9	0.1	9.0
	(0.07)	(0.07)	(0.1)	(0.02)
Community characteristics[(a)]				
Community size	388.8	486.4	97.6	na
	(55.96)	(67.09)	(88.77)	
Female share of population (%)	50.8	51.0	0.1	48.5
	(0.6)	(0.75)	(0.98)	
Median age	22.8	22.1	-0.7	31.0
	(0.38)	(0.41)	(0.58)	
Population aged 65+	3.4	3.4	0.0	4.8
	(25.41)	(25.37)	(36.6)	
People per household	5.4	6.5	1.135***	2.9
	(0.23)	(0.22)	(0.32)	
Median personal income	214.6	206.6	-8.0	549.0
	(10.45)	(3.1)	(10.77)	
Median household income	727.4	874.0	146.6**	1192.0
	(47.93)	(42.64)	(65.17)	
Median rent payments	43.9	42.2	-1.7	140.0
	(3.17)	(6.15)	(7.22)	
Labour force share of population (%)[(b)]	39.8	36.4	-3.4	47.3
	(3.24)	(2.81)	(4.36)	

***, ** and * indicate significance at the 1%, 5% and 10% levels, respectively. Standard errors in parentheses.

(a) ABS 2006 Census community profile data; most variables available for 54 NTER communities; rest of NT Census data is average of all of NT.

(b) Measured as a share of total population.

The only notable differences are observed for some community-level characteristics. Early-adopting communities were smaller on average by 100 community members, and families were smaller by one household member (5.4 versus 6.5). The early-adopting communities were slightly worse off in terms of median household income by A$150, but this appears to be due to smaller household size, since personal incomes were no different. Community composition and median age were not significantly different between early and late adopters, nor were local economic conditions (as proxied by the labour force-to-population ratio).

Table 1 does not fully rule out the possibility that the rollout schedule was intended to target the most in-need communities first, and the least in-need communities last, which would downwardly bias our estimated treatment effect. If true, we would expect the very first communities to have below-average pre-intervention outcomes, and the very last communities to be above-average. But we find that, if anything, the very first-adopting communities had slightly higher pre-intervention birthweight and lower probability of prematurity than other NTER communities (see Appendix B). It does not appear, therefore, that early rollout of income management was targeted towards the communities with the worst pre-intervention outcomes. Furthermore, if we drop all 21 communities with only-treatment or only-control observations, we still find no pre-treatment differences in birth outcomes (see Appendix B).[21]

[21] Similar to Hoynes and Schanzenbach (2009) we also estimated a regression model in which an index of the timing of the reform, indexed to 1 for 16 September 2007, was regressed on pre-treatment community characteristics and levels or changes in birthweight (and other birth outcomes). There is a small and significant negative relationship between the level of pre-treatment birthweight and the timing of the reform, indicating that the reform was implemented later in communities that had, on average, lower birthweight before the intervention. This effect is statistically significant at the 10% level and is of order of magnitude -0.1 days for a one-gram increase in birthweight. We are able to control for these pre-treatment differences in birthweight through community fixed effects in the main regression model. There is also a negative relationship between roll-out date and pre-treatment trends in birthweight (proxied by growth trends between 2000 and 2007); Community growth trend data is highly sensitive to outliers in the small communities for which we do not have data on community-level characteristics from the ABS. Estimating this regression model only on the 49 communities for which we have ABS community characteristics reveals no statistically significant impact of birthweight, birthweight growth trends, and other community characteristics with the roll-out date. Overall, our extended set of control variables explain up to 20 per cent of the variation in the roll-out date, which suggests that most of the variation remains unexplained. This weakness in model fit is a strength for our identification strategy, and the negative coefficient on birthweight operates in the opposite direction from our treatment effects. See Hoynes and Schanzenbach (2009) for similar arguments in the context of the roll out of the Food Stamps program. These results are provided upon

Finally, it is possible that some residents in early-adopting communities moved to late-adopting communities to avoid income management. The scope for residents to move to avoid income management was limited, as eligibility was determined based on place of residents as of 21 July 2007 (one week after the policy was announced). However, even though mobility rates are high in Aboriginal communities, Cobb-Clark, Kettlewell et al (2017) demonstrate that income management did not impact upon mobility.

Overall, our identification method appears robust. The rollout timeline is exogenous to our outcome measures, and there is no evidence of differential pre-rollout conditions or trends in early- and late-adopting communities. In addition, there is scope for policy-driven improvement, as birth outcomes in NTER communities were substantially worse than outcomes in the rest of the NT (Table 1, column 4).

5 DATA AND METHODOLOGY

The above analysis suggests that even though we do not know what the original logic, if any, was behind the rollout schedule, for our purposes it can be treated as though it was randomly assigned. We can therefore apply an analytical approach similar to that used in analysis of cluster-randomised stepped wedge trials; that is, trials where a program is intentionally rolled out across clusters (or communities) on a randomised schedule, to eventually cover the full population (Hemming et al, 2017).[22]

5.1 DATA

The data for this research come from two sources. The main source is administrative hospital records on the universe of infants born in the NT from 1996 onwards, from the NT Early Childhood Data Linkage Project. We identify infants from NTER communities based on their

request.

[22] This is an increasingly common approach in the public health literature.

mother's suburb of residence at the time of birth.[23] We then combine this with data from Appendix A of the AIHW (2010) evaluation report, on the date that income management was introduced in each community, to construct a variable for treatment status.

In our analysis, we use only the subset of births to mothers who resided in a community that received income management, dropping births to mothers who did not live in an NTER community. We limit our sample to infants born during or shortly after the income management rollout period, to include all those who were in their third trimester during the rollout.[24] This gives a total sample of 1,153 births between 17 September 2007 and 31 January 2009. We choose a narrow sample period around the dates of the rollout to reduce the potential for confounding time trends that may affect periods where all observations would be either treatment or control. We also exclude 56 infants born before the beginning of their third trimester.

5.2 Outcome measures

The outcome variables of interest are birthweight and the probability of low birthweight. To separate the two channels that may lead to low birthweight, we also present results on the probability of premature birth. As shown above in Table 1, low birthweight and prematurity are common in remote NT communities – with rates of 14 and 15 percent, respectively in 2006/07, which is around twice the rate of the rest of the NT. This indicates that there is significant capacity for improvement in these outcome measures.

[23] We use a range of sources, including BushTel and *Social Security (Administration) (Declared Relevant Northern Territory Areas – Various) Determination 2010 No. 8* to identify aliases, outstations and alternative spellings for each community.

[24] An alternative is to define the sample to include only infants for whom their third trimester began during the rollout period, though by construction this leads to potential bias of having disproportionately more premature infants in the control group and more overdue infants in the treatment group.

We focus on birthweight for two reasons. First, there is an extensive literature on the impact of maternal disadvantage and behaviour during pregnancy on birthweight (Aizer and Currie, 2014), and in particular the impact of maternal nutrition (Grieger and Clifton 2015). This means that if income management was successful in increasing food consumption, we would expect an increase in birthweight.

Second, birthweight is an important outcome measure in its own right, given its association with later life outcomes and thus the potentially high lifetime costs of low birthweight (Almond, Chay et al. 2005, Almond, Currie et al. 2014). In a survey of the literature, Victora, Adair et al. (2008) report that low birthweight leads to higher risk of chronic disease and of certain types of mental illness. There is also suggestive evidence that lower birthweight is associated with weaker cognitive skills in childhood, and a slight decrease in average years of schooling, reducing lifetime human capital accumulation and income (e.g. Figlio, Guryan et al. 2014)). In sum, poor nutrition in utero has been found to lead to a permanent impairment, with small negative impacts also transmitted to future generations (Almond, Currie et al. 2014).

Evidence of the link between birthweight and chronic disease has also been found in our population of interest. Using data from a health screening program in a remote Australian Aboriginal community, Singh and Hoy (2003) estimate that low birthweight infants faced a 17 percent higher risk of high blood pressure later in life.

This evidence suggests that any policy-driven improvement in birthweight would represent not only the impact, if any, of income management on food expenditure and consumption but, importantly, potential reductions in long-term health risks and enhanced capacity for human capital accumulation.

5.3 ESTIMATION APPROACH

Having shown that the rollout of income management was unsystematic and therefore as good as random, we can exploit this exogenous variation in a simple estimation framework to identify the causal impact of income management on birth outcomes. We estimate the following model with standard errors clustered by community:

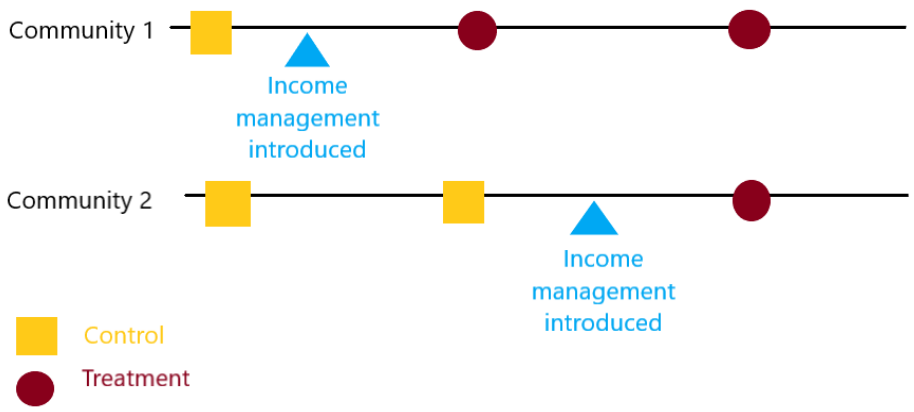
$$Y_i = \alpha_0 + \delta IM_i + \alpha_1 year + \alpha_2 rainfall_{ci} + \eta_c + \epsilon_{ic}, \qquad (1)$$

Where Y_i is the outcome variable for newborn i and IM_i is the 'treatment' indicator, which is equal to 1 for individual i if income management was in place in their community (c) at the beginning of the third trimester of pregnancy, and 0 otherwise.

Figure 7 illustrates the identification strategy. The control group consists both of infants born into community c before income management was introduced, and those born into other communities where income management had not yet been introduced by the beginning of their third trimester.[25] Note that treatment status varies within most communities over time; therefore, infants from the same community may be assigned either to treatment or control status depending on their date of birth.

[25] This definition is based on Almond, Hoynes et al's (2011) findings that the introduction of the food stamp program in the US significantly increased birthweight if it was in place for the full third trimester, but with no additional impact if it was introduced earlier on in pregnancy. Our expectation is therefore that the intended change in behaviour from income management is most likely to be evident if income management was introduced before or at the beginning of the third trimester. In Section 0 we explore whether this holds true in our data. Note that Almond, Hoynes et al (2011) are unable to precisely date the beginning of the third trimester, as they did not have data on gestational age. Our dataset includes gestational age, allowing for a more precise date for the beginning of the third trimester.

Figure 7: Illustration of treatment and control assignment



Our approach is closely related to Almond, Hoynes et al. (2011), who estimated the causal impact of food stamps on birthweight exploiting a staggered rollout across US counties. But importantly, our analysis is conducted at the individual level, whereas Almond, Hoynes et al use a time series of average birth outcomes for each county.[26]

There are clear seasonal patterns in birth outcomes in the NT, with worse outcomes for those born in the wet season (November – April; see Appendix D for an illustration). To control for this seasonality, we include total rainfall (measured in millimetres) for the three months prior to birth the newborn's region.[27] Rainfall data are sourced from the Australian Bureau of Meteorology, based on the total rainfall per month at a weather station in each of four major regions: Darwin, Alice Springs, Katherine, and Gove Airport in the East Arnhem region. This method of controlling for seasonality is preferable to simply using seasonal controls in the models, both because the short time frame of the sample period limits the ability of controls to

[26] Almond, Hoynes et al also use baseline community characteristics and county-time fixed effects. We considered using 2006 ABS Census data for baseline characteristics, but these data are available for just 53 communities. In addition, given our smaller sample size, use of community-time fixed effects leads to over-fitting (though the results are broadly unchanged).

[27] We experiment with different functional forms, for instance, a log transformation of the rainfall data, dummy variables for very high levels of rainfall (likely to represent flooding), and splitting rainfall into quartiles. Results are robust to these different specifications, with our main treatment effect on birthweight ranging between 90-120 grams, depending on the specification.

pick up regular seasonal variation, and because timing of the season can vary from year-to-year.[28]

η_c represents community fixed effects, controlling for unobserved location-specific factors that influence birth outcomes; for instance, differences in the share of the population receiving welfare payments, the size of the community, remoteness of the location, and access to health care facilities. We also include year controls, to capture any time trend; these are essential given that the gradual rollout of the policy introduces a correlation between time and treatment status (Davey, Hargreaves et al. 2015, Hemming, Taljaard et al. 2017).[29] The NTER communities are separated into 88 locations, 83 of which had at least one birth during the rollout period.[30]

Of main interest to our analysis is the sign, size and statistical significance of δ. We test the hypothesis that income management did not affect health outcomes ($\delta = 0$), against a two-sided hypothesis that it did ($\delta \neq 0$). Because we do not observe which families received income management, we consider the estimate of δ as an intent-to-treat (ITT) estimate; the impact on income-managed individuals would likely be larger than this estimate.[31] As noted above, most residents in NTER communities were subject to income management, though we are unable to identify treatment at the individual level. We can interpret δ as causal, if there are no remaining unobserved factors that are explaining birth outcomes and the timing of income management, therefore $cov(IM, \epsilon_{ic}) = 0$. Given that we have shown that the rollout seems to be unrelated to pre-treatment trend and levels of birth outcomes, and that the rollout

[28] Though treatment effects are similar when seasonal controls are instead included in the regression, or when the outcome data are seasonally adjusted pre-analysis using seasonal factors based on the full history (1996-2013) of the data.

[29] Hemming et al (2017) argue that due to the inherent imbalance of treatment and control groups over time, it is essential to control for time, even if coefficients do not appear statistically significant.

[30] The number of communities in the dataset is greater than the number of NTER communities because some outstations were treated on a different timeline to their main community and are therefore listed separately.

[31] To obtain the treatment effect on the treated we would divide the ITT effect by the proportion of mothers in the community who were affected by income management. As of today, we were not able to source this information from Government data.

was not linked to any other factors that explain birth outcomes, we interpret the estimate of δ as causal.

To better understand the mechanism through which income management impacts birthweight, we add an indicator of premature birth in the regressions on birthweight and low birthweight:

$$Y_{i,} = \alpha_0 + \delta IM_i + \alpha_1 year + \alpha_2 rainfall_{ci} + \eta_c + \pi premature_i + \epsilon_{ic} \qquad (2)$$

This allows us to disentangle whether the effect was driven by intrauterine growth or gestational length. If the treatment effect comes through an impact on nutrition during pregnancy, we would expect it to have a larger effect on intrauterine growth than gestational length, and hence for the treatment variable to remain significant when the control for prematurity is introduced. If instead the effect comes through prematurity, the treatment effect should lose significance in this version of the model.

An ordinary least squares model is estimated on average birthweight. For the two binary outcome measures (low birthweight and prematurity), regressions are estimated using both linear probability and probit models.[32] Our main results are estimated using probit models with average marginal effects reported, though some of the sensitivity tests are conducted using linear probability models for ease of estimation, particularly for those that loop through many variations of the model.

Some alternatives to the main model are also presented below. For instance, quantile regressions (on birthweight) help to reveal whether the impact of the policy was different at different parts of the birthweight distribution. As an additional robustness checks, we present propensity score-matched regressions where we limit the control group to those who best resemble the treatment group.

[32] Typically, fixed effects using a probit model encounter an incidental parameter problem, but this does not apply in our case, as we have fixed effects by community with the number of communities fixed.

6 RESULTS

6.1 MAIN RESULTS

In a model without control variables, we find that average birthweight is around 61 grams lower in the treatment group than in the control group (Panel A, Table 2, column 1), and the probability of low birthweight is 2.9 percentage points higher (Panel B, Table 2, column 1). Controlling for season variations and an annual time trend doubles the negative treatment effect to 120 grams, an effect that is statistically significant at the 1% level. Further controlling for community fixed effects increases the treatment effect by about 30% to 164 grams (statistically significant at the 1% level). Similarly, exposure to income management increase the probability of low birthweight by 8 percentage points, or by 50% relative to the mean probability of the control group pre-rollout (Panel B, column 3).

The decrease in birthweight appears to have come through both the intrauterine growth and the gestational length channels. After controlling for prematurity, the treatment effect on birthweight (intrauterine growth channel) declines by 28% to -119 grams (column 4). The treatment effect on low birthweight also declines by almost 50% to 4.8 percentage points. Both treatment effects are statistically significant at the 5% level.

Finally, our estimation results indicate that income management may also have increased the probability of prematurity (Panel C, Table 2), however this treatment effect is statistically significant only at the 10% level.

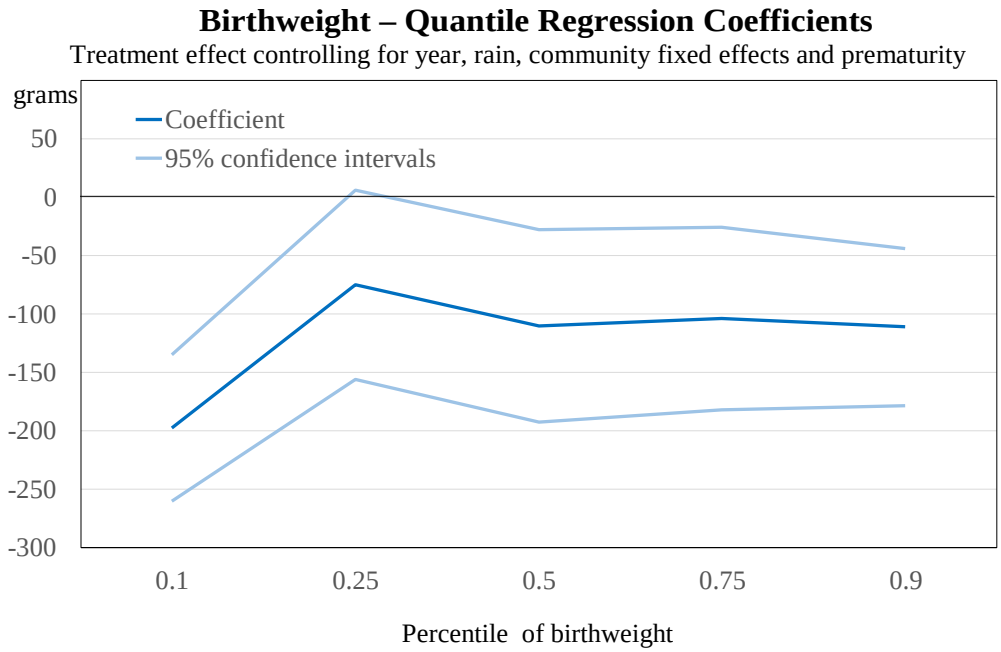
Table 2: Main regression results

	(1)	(2)	(3)	(4)
Panel A: Birthweight (grams, OLS)				
Income management	-60.64*	-119.5***	-163.9***	-118.7**
	(35.95)	(44.77)	(57.58)	(54.76)
Rainfall in 3 months to birth (mm)		-0.180***	-0.101	-0.0873*
		(0.0452)	(0.0620)	(0.0517)
Year (base category = 2007)				
2008		70.94	91.16	53.14
		(52.96)	(56.28)	(52.43)
2009		160.8	169.8*	64.08
		(97.14)	(97.96)	(94.39)
Premature				-932.9***
				(54.42)
Constant	3,161***	3,175***	3,458***	3,465***
	(23.28)	(45.48)	(42.91)	(41.39)
Community fixed effects	No	No	Yes	Yes
Number of communities	83	83	83	83
Observations	1,153	1,153	1,153	1,153
R-squared	0.002	0.011	0.091	0.339
Panel B: Low birthweight (Probit average marginal effects)				
Income management	0.0285*	0.0603***	0.0807***	0.0480**
	(0.0167)	(0.0205)	(0.0275)	(0.0220)
Rainfall in 3 months to birth (mm)		8.90e-05***	8.54e-05***	7.03e-05***
		(2.02e-05)	(2.92e-05)	(1.85e-05)
Year (base category = 2007)				
2008		-0.0385	-0.0538	-0.0207
		(0.0346)	(0.0445)	(0.0385)
2009		-0.0769**	-0.0990**	-0.0275
		(0.0359)	(0.0447)	(0.0431)
Premature				0.314***
				(0.0155)
Community fixed effects			Yes	Yes
Number of communities	83	83	50	50
Observations	1,153	1,153	991	991
Panel C: Premature (probit average marginal effects)				
Income management	0.0275	0.0613***	0.0564*	
	(0.0203)	(0.0235)	(0.0315)	
Rainfall in 3 months to birth (mm)		4.77e-05**	2.45e-05	
		(2.39e-05)	(4.03e-05)	
Year (base category = 2007)				
2008		-0.0438	-0.0523	
		(0.0314)	(0.0385)	
2009		-0.115***	-0.121***	
		(0.0365)	(0.0444)	
Community fixed effects	No	No	Yes	
Number of communities	83	83	50	
Observations	1,153	1,153	1,013	

Community-clustered standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

The finding that average birthweight was significantly affected could represent an improvement in health outcomes, depending on the effect at different parts of the distribution of birthweight. It may be driven by a decrease in birthweight for particularly heavy infants, which may be the result, for instance, of a decline in the incidence of gestational diabetes. But this does not appear to be the case. Quantile regressions results summarised in Figure 8 show that the treatment effect was statistically significant across the distribution, and largest for newborns with very low birthweight.[33]

Figure 8



6.2 ROBUSTNESS CHECKS

Our findings are broadly robust to changes in our estimation model specification, exclusion of potential outliers, and to changes in the sample period. The full set of results from our robustness checks are presented in Appendix C.

[33] This is confirmed in regressions where we include an indicator for whether there were obstetric complications due to gestational diabetes. This variable is highly significant in a regression on birthweight, but its inclusion does not change the magnitude or significance of the treatment effect, either when it is included as an independent control or when interacted with treatment status.

Our main models compare births in treated communities with all control births. Using propensity score matching, we can instead assess whether our results change when we compare treated infants to those in the control group who most closely resemble them. We match treatment to control observations based on the mother's basic demographic characteristics, medical history, hospital of birth, and rainfall prior to birth. Under various matching methods, we find an average reduction in birthweight of 120-130g (significant at the 1 or 5 percent level), an increase in the probability of low birthweight of 5.8-7.9 percentage points (also statistically significant at the 1 or 5 percent level; Table C.1).

Furthermore, changes to the sample period, censoring of outliers and exclusion of partially-treated infants do not meaningfully affect our results (Table C.2). In addition, our negative results persist even if we limit our regression to a 'healthy' birthweight range (2500-5000 grams) for which we obtain a negative treatment effect of 77 grams, although it is not statistically significant (Table C.3). Alternative controls for seasonal effects – for instance using quarter or month controls interaction with the year – yield treatment effects ranging between 70-103 grams (Table C.4).

We also find that the negative treatment effect is robust to dropping communities with less than 10 births in our sample (treatment effect on birthweight: -131 grams, significant at the 5% level; on low birthweight probability: 5.8 percentage points, significant at the 1% level). The negative treatment effect on birthweight is larger for communities in the first half of the roll-out period (-158 grams, significant at the 10% level) than for communities in the second half of the roll-out period (-89 grams, not statistically significant; Table C.5).

Finally, a placebo-style test, in which we re-run our main specification with a 1-6 year lead on treatment timing, on sample periods before income management was introduced (2000-2002; 2001-2003; 2002-2004; 2003-2005; 2004-2006; 2005-2007), reveals no statistically significant

treatment effects on birthweight or the probability of low birthweight in the years prior to the rollout (Table C.6). This suggests that treatment timing does not simply capture unobserved trends in birth outcomes.

6.3 Treatment timing

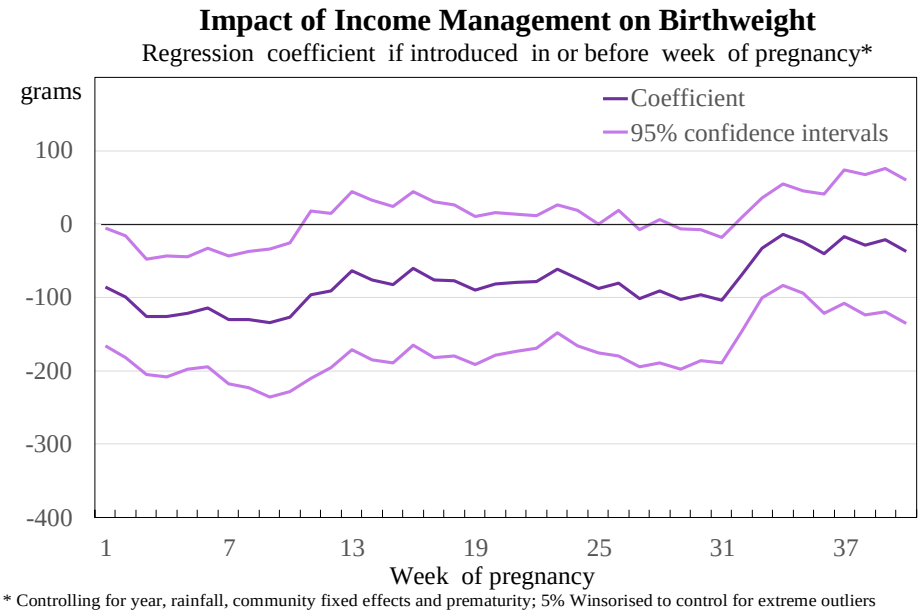
As detailed in Section 5.3, our main definition of treatment status is that income management was in place at or before the beginning of the third trimester of pregnancy – that is, the 28th week of pregnancy. However, there are two potential reasons that this definition could misrepresent the true treatment effect.

First, the impact of nutrition on birthweight is multidimensional (King 2016). It is therefore possible that infants received different 'dosages' depending on when income management was introduced relative to their birth. Although Almond, Hoynes et al (2011) found additional food consumption to be most relevant in the third trimester, nutrition at other times is likely to also affect birthweight. For instance, pre-pregnancy nutrition has been found to have an impact on birthweight over and above the impact of nutrition during pregnancy (Young, Nguyen et al. 2015, King 2016).

Second, the impact of income management on birth outcomes may have come through a channel other than maternal nutrition. For instance, it may be stress due to confusion over the new policy. Stress has been shown to have a negative impact on birthweight because cortisol, which the mother produces during stress, affects the child through the placenta. Previous research has demonstrated the importance of stress on birthweight (Rondo, Ferreira et al. 2003). Some studies identify this impact by the experience of grief during pregnancy (Black et al., 2016; Persson and Rossin-Slater, forthcoming), exposure to hurricanes (Currie and Rossin-Slater, 2013), or exposure to racism (Lauderdale, 2006). The impact of stress on birthweight is not necessarily restricted to the third trimester, but may occur in any trimester of the pregnancy.

To test for these possibilities, we first allow treatment status to occur in any week of the pregnancy. We run regressions defining babies as 'treated' if income management was introduced in or before each of the 40 weeks of pregnancy.[34] Figure 9 shows the estimated treatment effects from this exercise, using the regression specification corresponding to column 4 of Table 2.[35] We find that the impact of income management on birthweight is largest if introduced in or before the first trimester of pregnancy. The negative treatment effect is about 115 grams and statistically significant at the 5% level. Children who were affected by income management only in the final two to three months of the pregnancy did not experience adverse birth outcomes.

Figure 9

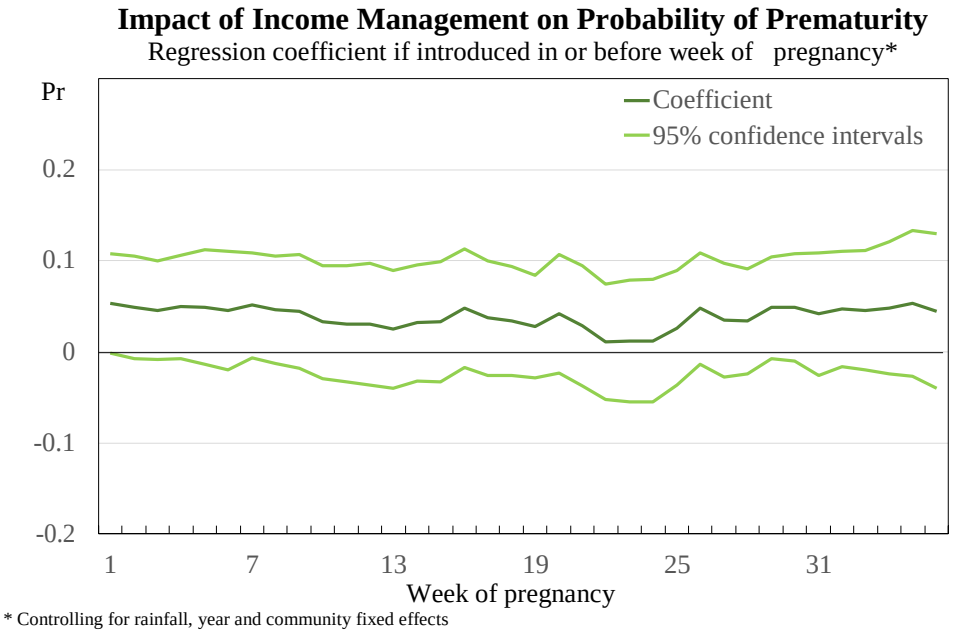


[34] Infants born before the relevant week are excluded from the sample. E.g., those born at week 38 are not included in the regressions for treatment introduction at week 39.

[35] In this graph, birthweight is 5% Winsorised to limit the impact of outliers on any single coefficient, and the sample period is defined on a rolling basis. This makes the coefficient estimate for 28 weeks (our main model) slightly different from that shown in our main results table.

The larger impact in the first trimester appears to reflect intrauterine growth, because we find no corresponding heterogeneity in the treatment effect across the gestation length distribution for prematurity (Figure 9).[36]

Figure 9



We obtain similar results when allowing treatment status to vary by trimester of the pregnancy. We redefine our treatment variable into three categories, indicating whether income management was introduced during or before the first trimester, during the second trimester, or during the third trimester. Introducing three treatment categories, rather than one, reduces the statistical power of our estimation. Nevertheless, our results suggest that the treatment effect is larger the earlier income management was introduced relative to the full-term birthdate (Table 3). Newborns for whom income management was introduced before or during the first trimester were significantly lighter than those who were not exposed to income management at all.

[36] But as noted above, these results vary with model specification; in some specifications there does appear to be a larger and highly significant impact on prematurity during the first trimester.

The large negative effect of income management if introduced in or before the first trimester does not appear to reflect conditions specific to early-adopting communities (which, by construction, had a higher share of births during the sample period for which income management was introduced before pregnancy). The effect persists when we restrict the analysis to the second half of the rollout period (not shown; results available on request).

Table 3: Impact of treatment timing on outcome measures

Controls	(1) None	(2) Year controls and rainfall	(3) Year, rainfall + community FE	(4) Year, rainfall, community FE and premature
Outcome variable: Birthweight (grams, OLS)				
Timing of IM introduction (Omitted category = born before income management introduced in community)				
Before or during first trimester	-34.11	-88.85	-183.6*	-160.9*
	(58.09)	(76.52)	(95.13)	(88.50)
During second trimester	-30.67	-79.47	-95.29	-80.31
	(43.70)	(58.00)	(66.69)	(56.11)
During third trimester	119.4*	85.22	66.23	16.66
	(60.53)	(66.25)	(76.37)	(66.16)
Outcome variable: Low birthweight (linear probability model)				
Timing of IM introduction (Omitted category = born before income management introduced in community)				
Before or during first trimester	0.0117	0.0393	0.0609	0.0470
	(0.0265)	(0.0328)	(0.0454)	(0.0354)
During second trimester	0.0222	0.0478*	0.0435	0.0344
	(0.0216)	(0.0277)	(0.0342)	(0.0265)
During third trimester	-0.0520*	-0.0340	-0.0319	-0.00164
	(0.0285)	(0.0320)	(0.0363)	(0.0309)
Outcome variable: Premature (liner probability model)				
Timing of IM introduction (Omitted category = born before income management introduced in community)				
Before or during first trimester	0.00540	0.0390	0.0244	
	(0.0251)	(0.0311)	(0.0462)	
During second trimester	0.0246	0.0433	0.0161	
	(0.0273)	(0.0327)	(0.0408)	
During third trimester	-0.058**	-0.0457	-0.0532	
	(0.0282)	(0.0336)	(0.0363)	

Community-clustered standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

6.4 Characteristics of the Mother

Income management may have changed fertility decisions and therefore the composition of the pool of women who were willing to have children during this period. For instance, women who were more likely to plan their pregnancy may also pay particular attention to their nutrition during pregnancy and, as a result, have higher birthweight babies; these women may have chosen to postpone having children during the NTER. This does not appear to be the case.

First, we see no impact of the policy on community-level fertility rates. The number of births per resident woman declined slightly following treatment, but this decline was not statistically significant, either as a raw difference, or after controlling for season, year and community fixed effects (Panel A, Table 4).

We also do not see a significant difference in the medical history of women who gave birth after income management was introduced, either in terms of their previous pregnancies (Panel B), or their history of medical complications (Panel C). Hence, it seems that the composition of women who fell pregnant did not change as a consequence of income management.

Table 4: Fertility and mother's medical history

	Treatment	Control	Difference (treatment - control)		Observations
			No controls	With controls[(a)]	
Panel A: Fertility rate (community-level)					
Births per 1000 residents per quarter	8.90	10.10	-1.19	-0.23	337
Births per 1000 women per quarter	15.50	15.60	-0.26	-0.51	297
Panel B: Previous pregnancies					
First pregnancy (%)	25.43	26.68	-1.25	1.14	1153
Number of pregnancies carried to >20 weeks	1.70	1.60	0.10	0.00	1153
Total number of previous pregnancies	2.97	2.86	0.11	-0.05	1153
Panel C: Mother's history of medical complications					
Anaemia (%)	6.88	6.41	0.35	-0.76	1057
Cardiac disease (%)	8.13	6.76	1.17	-3.14	1057
Epilepsy (%)	1.25	1.39	-0.14	0.61	1057
Pre-existing hypertension (%)	1.25	1.04	0.18	0.89	1057
Pre-existing diabetes (%)	4.58	3.12	1.30	1.67	1057
Gestational diabetes (%)	3.13	3.47	-0.35	-0.37	1057
Renal disease (%)	3.96	3.12	0.73	1.05	1057
Syphilis (%)	5.21	3.81	1.07	2.59*	1057
Urinary tract infection (%)	6.04	3.29	2.47**	3.70*	1057
Any complication (%)	57.29	54.59	1.70	-2.17	1057
Number of previous complications	0.81	0.72	0.07	0.03	1057
Medical history unknown (%)	8.92	7.83	1.09	-0.39	1153

OLS or LPM regressions. Controls include rainfall, year and community fixed effects. *** p<0.01, ** p<0.05, * p<0.1

6.5 MATERNAL HEALTH BEHAVIOURS

Alternatively, a negative treatment effect of income management on birth outcomes may be the result of a change in maternal health behaviours. Income management intended to create a healthier consumption environment, however, the stress associated with the rollout may have led to an increase in drinking and smoking by these mothers. We do not have consumption data available, but our perinatal data includes information on the prevalence of drinking or smoking at the time of the first antenatal visit. Treatment is defined to take the value 1 if the first antenatal visit occurred after income management was introduced into the mother's community, and 0 otherwise.

Table 5 presents the estimation results. There is no evidence that income management changed smoking (Panel A) or drinking (Panel B) behaviour of mothers. The treatment effects are precise zeros in both cases (column 3 with full controls), and are not statistically significant. It is therefore not surprising that controlling for smoking and drinking behaviour (and other characteristics of the mother) directly in the regression model does not change the treatment effect of income management on birthweight (see Table E.1 in Appendix E).

Table 5. Impact of income management introduction on smoking and drinking behaviour at first antenatal visit

	No controls	Year rainfall controls	Year, rainfall controls, community FE
	(1)	(2)	(3)
Panel A: Smoking at first antenatal visit	-0.0146	0.0110	0.00705
	(0.0393)	(0.0471)	(0.0470)
Observations	905	905	867
Missing observations[(a)]	280	280	318
Number of communities	81	81	61
Panel B: Drinking at first antenatal visit	0.0181	0.0335	-0.000519
	(0.0198)	(0.0271)	(0.0337)
Observations	991	991	779
Missing observations[(a)]	194	194	406
Number of communities	81	81	52

The probit model drops communities (and the observations contained within those communities) if there is no variation in the outcome measure within that community. In addition, smoking and drinking data are not available for all observations; probability of missing data is not significantly different between treatment and control groups. Community-clustered standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

6.6 Measurement and quality of care

A reduction in birthweight or an increase in the probability of low birthweight does not necessarily indicate a worsening in birth outcomes. The introduction of income management could have increased pregnant women's likelihood of receiving earlier or better quality antenatal care, which may, in some cases, lead to worse measured outcomes. For instance, pregnant women on income management may have had more contact with Centrelink staff or a Government Business Manager in their community. These staff may have recommended regular antenatal visits, or helped to ensure better access to health care. If this was the case, it could affect the interpretation of our treatment effect in two ways.

First, an improvement in access to antenatal care could lead to a higher rate of 'detection' of prematurity, a variable on which we condition our analysis. Indeed, while most indicators of antenatal care were no different between the treatment and control groups, the treatment group had their first ultrasound significantly earlier (by one week on average) than the control group (Table 5). Having an earlier ultrasound may have led to a better estimate of the foetus' gestational age, potentially boosting the measured incidence of prematurity, even if the actual rate did not change.

Second, better access to health care could mean that any serious complications are more likely to be detected and addressed. This may lead to an increase in the frequency of emergency C-sections, which could increase rates of prematurity and decrease birthweight, but would nevertheless be a preferred health outcome. While the treatment group had only a slightly higher rate of emergency deliveries, they were 6.8 percentage points more likely to be born in a major hospital, which may indicate an improvement in quality of care around the time of birth and could affect the timing of delivery.

Table 5: Indicators of behaviour, antenatal care and hospital care

	Level		Difference (T – C)	
	Treatment	Control	Outcome	Pr missing data
Antenatal care				
Number of antenatal visits	8.66	9.05	-0.39	-0.84
	(0.21)	(0.21)	(0.03)	(0.65)
Gestational age at first visit (weeks)	16.18	16.44	-0.26	-1.04
	(0.38)	(0.35)	(0.51)	(0.84)
Gestational age at first ultrasound (weeks)	19.56	20.63	-1.07**	-0.12
	(0.37)	(0.33)	(0.5)	(1.33)
Had a dating ultrasound (%)[(a)]	22.96	19.17	3.79	Na
	(1.83)	(1.57)	(2.4)	Na
Hospital care				
Born in major hospital (%)[(b)]	72.11	65.34	6.77*	Na
	(1.95)	(1.9)	(2.74)	Na
Emergency delivery (%)	20.30	18.85	1.45	Na
	(1.75)	(1.56)	(2.34)	Na
Length of stay in hospital (days)	6.83	6.29	0.54	1.29*
	(0.51)	(0.39)	(0.63)	(0.69)
Admitted to special care nursery (%)	25.43	21.12	4.31*	-0.16
	(1.9)	(1.63)	(2.49)	(0.17)

(a) Date of first ultrasound was during weeks 6-13
(a) Royal Darwin Hospital, Darwin Private or Alice Springs Hospital
Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

We can test both possibilities in a regression framework. To test the measurement hypothesis, we augment our main model with a control for whether the mother had an early ultrasound, which could indicate a more precise estimate of gestational age. We then add a control for birthweight. If measurement of prematurity did improve, we would expect a positive treatment effect on prematurity, even after controlling for birthweight. We find that controlling for early ultrasound does not change the treatment effect, and when we add birthweight to the model, the treatment effect loses significance and the magnitude falls to zero (Table 6). Therefore, it does not appear that the increase in prematurity can be explained by improved data quality.

To test the quality of care hypothesis, we add to our model an indicator for whether the birth was in a major hospital (Darwin or Alice Springs), as well as indicators for the delivery method, to control for potential increases in the identification of emergencies. Our results suggest that these factors may have played some marginal role in prematurity, yet the magnitude of the

treatment effect remains almost the same, at around 5 percentage points (significant at the 10% level). The treatment effect on birthweight remains highly significant and is slightly larger in magnitude (-162 grams).

These findings suggest that our main treatment effect reflects a true worsening of birth outcomes and not a change in measurement or detection of complications.

Table 6 : Regression results controlling for measurement and quality of care

	Hypothesis 1: Measurement		Hypothesis 2: Quality of care		
	(1)	(2)	(3)	(4)	(5)
Outcome	Premature	Premature	Premature	Premature	Birthweight
Income management	0.0572*	-0.0128	0.0562*	0.0520*	-162.4***
	(0.0317)	(0.0310)	(0.0315)	(0.0312)	(55.38)
Early ultrasound	-0.00922	0.0234			
	(0.0282)	(0.0269)			
Birthweight		-0.000278***			
		(1.68e-05)			
Major hospital			0.0797*	0.0646	-11.90
			(0.0469)	(0.0529)	(82.80)
Spontaneous delivery				0.0245	-142.6***
				(0.0358)	(49.83)
Emergency delivery				0.113**	-196.4***
				(0.0540)	(68.83)
Community fixed effects	Yes	Yes	Yes	Yes	Yes
Year and rainfall controls	Yes	Yes	Yes	Yes	Yes
Observations	1,013	1,013	1,013	1,013	1,153

Includes controls for year and rainfall. Community-clustered standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

7 CONCLUSION

7.1 DISCUSSION

Income management was intended to improve child welfare. Over the past decade, governments, commentators and journalists have questioned whether it achieved this goal. Our findings provide a partial answer, suggesting that the policy did not improve one measure of child health outcomes, and, by extension, that income management does not appear to have produced the desired change in household consumption patterns, at least for households with

pregnant women. In fact, income management may have had a net negative impact on newborn health – lower birthweights and a higher probability of low birthweight, over and above what would be expected if a baby was premature.

The treatment effect of a reduction in birthweight of 100g or more is sizable, however it lies within the range of what is possible in the Northern Territory and what has been found in the literature. Although average birthweights have hardly improved in Aboriginal communities in the Northern Territory over the years, seasonal variation within any given year is very high. For instance, babies born in March in the Alice Springs region are up to 120 grams lighter than babies born in the same region in September; such differences are even larger in the Katherine and Barkly regions (see Appendix D). Therefore, our negative treatment effect of income management is equivalent to the impact of normal seasonal variations within any given year.

Randomised trials that provide food or fortified food products during pregnancy have been shown to increase birthweight by 125 grams on average (Gresham, Byles et al. 2014). Almond, Hoynes et al. (2011) found a treatment effect of food stamps – a program that provided a relatively low dose of additional resources for families – increase birthweights in the magnitude of up to 42 grams for black babies in the United States. Conditional cash transfers provided as part of the Oportunidades program (now renamed to Prospera) in Mexico were found to increase birthweight by 130 grams and to decrease low birthweight probabilities by 4.6 percentage points (Barber and Gertler 2008). These treatment effects are very much in line with our findings, although in opposite directions. Separate from the literature on transfer programs, some studies identify the causal impact of nutrition through exposure to Ramadan or famines. Although the evidence is mixed on whether reduced nutrition during Ramadan affects the health of the baby, some studies find birthweight penalties of Ramadan observance in the magnitude of 270 grams (Savitri, Yadegari et al., 2014); others find a treatment effect of exposure to the Dutch famine of around 150 grams (Stein and Susser, 1975).

While we are uncertain of the reason for the negative treatment effect, we can eliminate four potential mechanisms. Income management did not worsen birth outcomes because it reduced birthweights at the very high end of the birthweight distribution, which could have been interpreted as an improvement in health outcomes. It also did not change the composition of the pool of mothers who were willing to have a baby during the reform or affect overall fertility rates. It also did not change risky health behaviours (smoking, drinking) of mothers during the early stages of pregnancy. Finally, we do observe a small increase in access to more specialised neonatal care. But controlling for such access, if anything, increases the magnitude of our negative treatment effects.

We can only speculate on the reasons for our findings, but we can think of at least four potential channels as plausible. First, it is possible that the policy itself created an unhealthy food consumption environment for pregnant women. This would be contrary to intuition and economic theory, which suggests that restricting transfers should either increase consumption of household essentials, or have no impact relative to cash transfers. It is difficult to see how the policy itself could have a negative impact, though one potential channel is that the restriction to spend at least half of welfare income on household essentials acted as a low mental 'anchor' for spending of this type, leading to lower consumption of household essentials than before. We would need data on household consumption patterns to explore this possibility further.

Second, our findings may reflect poor administration of income management that led to a disruption of household consumption. In effect, our results provide a combined estimate of the impact of the policy itself on birth outcomes, and of the impact of the way the policy was implemented. It is possible then that the measured negative treatment effect reflects issues with the process and administration of income management, and with community members' attitudes towards this process, rather than an impact of the restriction itself on behaviour.

This interpretation would be consistent with anecdotal reports of the disruption and ill-will caused by income management and the NTER more generally. The introduction of income management is widely reported to have led to a sense of a loss of freedom, disempowerment, and reduced community control. In the official evaluation reports, some survey respondents stated they perceived the program as 'patronising and dehumanising', with many highlighting the suspension of the *Racial Discrimination Act* a contributing factor to this perception (AIHW 2010). Importantly, while our estimation method allows us to disentangle the impact of income management from other NTER policies, general negative attitudes towards the NTER may have contributed to reduced willingness to comply with the restrictions imposed.

A third possibility is that there was some logic behind the rollout schedule, information for which is not publicly available or reflected in observable community characteristics, that downwardly biases our results. All data that we have to hand suggests that this was not the case; the first communities to receive income management did not have significantly worse outcomes pre-intervention, and the negative treatment effect does not appear to be driven by births in any particular time period or region. In our regression models, community fixed effects also help to control for any time-invariant community characteristics. But to rule out this possibility, future work could explore qualitative differences between communities that are not evident in our dataset. For instance, the impact of income management at the community-level may depend on the level of effective competition between food stores that are licensed to accept income managed funds and any non-licensed food providers. It is possible that the requirement to shop at licensed stores increased those stores' pricing power, leading to general price increases. Although aggregate food price data do not show any notable increase in prices around this time (NT Department of Health 2014), an evaluation report of New Income Management in the NT suggested that prices may have increased (Bray, Gray et al., 2014). Access to community-level (store) data would help to rule this possibility out.

If our results do reflect, at least partly, the process of the rollout rather than the policy itself, we can only conclude that any impact of the welfare restriction on extramarginal households, if one existed, was not large enough to offset the negative effect of the process. We are therefore unable to predict the cost-effectiveness of income management. But given the program's high administration costs, our findings suggest that it is highly unlikely that income management is a cost-effective means of improving newborn health.[37]

A fourth channel through which income management may have had adverse impacts on children's birth outcomes is that it may have changed the way households make consumption decisions and how resources were being allocated. It may even have reduced the bargaining power of women in the household. Thus, income management may have increased the level of stress experienced by mothers because they were no longer in full control of all of their available financial resources. Stress experienced by pregnant women can affect the development of the foetus, as cortisol is passed on to the child through the placenta. Some studies have shown for instance that grief (the death of a grandparent or a relative) adversely impacts upon birth outcomes. Although such treatment effects on birthweight are relatively small, at 11-35 grams (Black et al., 2016; Rossin-Slater and Persson, forthcoming), they demonstrate the important consequences of in-utero exposure to stress. Other studies have shown that maternal exposure to racism or hurricanes may increase the probability of low birthweight (Lauderdale, 2006; Currie and Rossin-Slater, 2014). This suggests that if pregnant women were exposed to more stressful situations as a consequence of income management – for instance interpersonal violence which has been described elsewhere as an important channel

[37] The recent Cashless Debit Card trial costed around $9,000 per recipient per year (including setup costs, not including payments), based on total costs of $18.9 million (DSS 2017), and over 2,000 participants (Orima 2017).

through which economic disadvantage affects birth outcomes (Aizer and Currie, 2014) – then this could help to explain our adverse treatment effects.

7.2 Directions for future work

While our findings provide a partial answer to questions over the effects and efficiency of income management, they also raise new questions. Our findings focus on a single aspect of child health within a short time period. They are therefore not necessarily representative of the long-term impact of income management on children's health and wellbeing. We do not know whether our findings are an anomaly or are representative of a more general negative reaction to income management. Cobb-Clark, Kettlewell et al (2017) find that income management led to a temporary reduction in children's school attendance, which supports our findings and the interpretation of a disruption effect of the policy. Further work focusing on other outcome measures – for instance consumption poverty – would help to answer this question.

Future research could also explore the potential mechanism of the measured negative impact in more detail. When available, linking our perinatal dataset with administrative welfare records would allow identification of treated individuals and their payment rates. Qualitative research suggests that some clients were highly receptive to income management, while others were not (Hand, Katz et al. 2016). Linked data that provides more information about the household would allow for exploration of the sources of this potential heterogeneity.

In the nearer future, our findings could be linked to the child immunisation data which is already available through the NT Early Childhood Data Linkage Project. This would yield further information on whether income management affected mobility, and the impact of this mobility on preventative health treatment. Cobb-Clark, Kettlewell et al (2017) in fact do not find an impact of the policy on mobility, however their results are based on children's school

mobility (when the children were six years or older). In addition, analysis of the Longitudinal Survey of Indigenous Children, which was conducted annually from 2008 onwards and collects extensive information on child health and on whether the child's carers were subject to income management, may shed more light on the channel of any negative impact of the policy.

REFERENCES

Aizer, A., J.M. Currie (2014). “The intergenerational transmission of inequality: Maternal disadvantage and health at birth.” Science **344**(6186): 856–861.

Almond, D., K. Y. Chay and D. S. Lee (2005). "The costs of low birthweight." The Quarterly Journal of Economics **120**(3): 1031-1083.

Almond, D., J. M. Currie and K. Meckel (2014). Fetal origins of lifetime health. Encyclopedia of Health Economics. San Diego, Elsevier**:** 309-314.

Almond, D., H. W. Hoynes and D. W. Schanzenbach (2011). "Inside the war on poverty: The impact of Food Stamps on birth outcomes." Review of Economics and Statistics **93**(2): 387-403.

Amarante, V., M. Manacorda, E. Miguel and A. Vigorito (2016). "Do cash transfers improve birth Outcomes? Evidence from Matched Vital Statistics, Program, and Social Security Data." American Economic Journal: Economic Policy **8**(2): 1-43.

Australian Institute of Health and Welfare (AIHW) (2010). Evaluation of income management in the Northern Territory. Canberra, Commonwealth of Australia.

Barber, S. L. and P. J. Gertler (2008). "The impact of Mexico’s conditional cash transfer programme, Oportunidades, on birthweight." Tropical Medicine & International Health **13**(11): 1405-1414.

Beatty, T. K. M. and C. J. Tuttle (2015). "Expenditure Response to Increases in In-Kind Transfers: Evidence from the Supplemental Nutrition Assistance Program." American Journal of Agricultural Economics **97**(2): 390-404.

Bitler, M. P. and J. Currie (2005). "Does WIC work? The effects of WIC on pregnancy and birth outcomes." Journal of Policy Analysis and Management **24**(1): 73-91.

Black, S. E., P. J. Devereux and K. G. Salvanes (2007). "From the Cradle to the Labor Market? The Effect of Birthweight on Adult Outcomes*." The Quarterly Journal of Economics **122**(1): 409-439.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes. 2016. "Does Grief Transfer across Generations? Bereavements during Pregnancy and Child Outcomes." American Economic Journal: Applied Economics **8**(1): 193-223.

Bray, J. R., M. Gray, K. Hand and I. Katz (2014). Evaluating New Income Management in the Northern Territory: Final Evaluation Report. Sydney, Social Policy Research Centre, UNSW Australia. SPRC Report 25/2014.

Breunig, R. and I. Dasgupta (2005). "Do Intra-Household Effects Generate the Food Stamp Cash-Out Puzzle?" American Journal of Agricultural Economics **87**(3): 552-568.

Brimblecombe, J. K., J. McDonnell, A. Barnes, J. G. Dhurrkay, D. P. Thomas and R. S. Bailie (2010). "Impact of inome management on store sales in the Northern Territory." Medical Journal of Australia **192**(10): 594-554.

Buckmaster, L. and C. Ey (2012). Is income management working? . P. Library. Background note.

Central Land Council. (2008). Reviewing the Northern Territory Emergency Response: Perspectives from six communities. Central Land Council.

Currie, J., & Gahvari, F. (2008). "Transfers in cash and in-kind: Theory meets the data." Journal of Economic Literature **4**: 333-383.
Currie, J. and E. Moretti (2007). "Biology as destiny? short- and long-run determinants of intergenerational transmission of birthweight." Journal of Labor Economics **25**(2): 231–264.

Currie, J. M., M. Rossin-Slater (2013). "Weathering the storm: Hurricanes and birth outcomes." Journal of Health Economics **32**: 487– 503

Cobb-Clark, D.A., N. Kettlewell, S. Schurer and S. Silburn (2017). "The Effect of Quarantining Welfare on School Attendance in Indigenous Communities." Unpublished manuscript. The University Sydney.

Commonwealth of Ausralia, Department of the Prime Minister and Cabinet (2017). *Closing the Gap Prime Minister's Report*, Canberra.

Cuffey, J., T. K. M. Beatty and L. Harnack (2016). "The potential impact of Supplemental Nutrition Assistance Program (SNAP) restrictions on expenditures: a systematic review." Public Health Nutrition **19**(17): 3216-3231.

Cunha, J. M. (2014). "Testing Paternalism: Cash versus In-Kind Transfers." American Economic Journal: Applied Economics **6**(2): 195-230.

Das, J. and Q.-T. Do (2005). "Reassessing Conditional Cash Transfer Programs." World Bank Research Observer **20**(1): 57-80.

Davey, C., J. Hargreaves, J. A. Thompson, A. J. Copas, E. Beard, J. J. Lewis and K. L. Fielding (2015). "Analysis and reporting of stepped wedge randomised controlled trials: synthesis and critical appraisal of published studies, 2010 to 2014." Trials **16**: 358.

Department of Families, H., Community Services and Indigenous Affairs (FaHCSIA), (2011). Northern Territory emergency response evaluation report 2011. Canberra, Australian Government.

Department of Families, H., Community Services and Indigenous Affairs, (2008). Northern Teritory Emergency Response: One Year On. Canberra, Commonwealth of Australia.

Department of Social Services (DSS) (2017). FOI 16/17-123: Costs of the Cashless Debit Card Trials in Ceduna and Kununnurra.

Figlio, D., J. Guryan, K. Karbownik and J. Roth (2014). "The Effects of Poor Neonatal Health on Children's Cognitive Development." American Economic Review **104**(12): 3921-3955.

Fogarty, W., Lovell, M., & Dodson, M. (2015). "Indigenous education in Australia: Place, pedagogy and epistemic assumptions." UNESCO Observatory Refereed e-Journal **4**, 1-21.

Gentilini, U. (2016). "The Revival of the “Cash versus Food” Debate: New Evidence for an Old Puzzle?" The World Bank Research Observer **31**(1): 135-167.

Gresham, E., J. E. Byles, A. Bisquera and A. J. Hure (2014). "Effects of dietary interventions on neonatal and infant outcomes: a systemic review and meta=analysis." American Journal of Clinical Nutrition **100**(5): 1298-1321.

Grieger, J. and V. Clifton (2015). "A Review of the Impact of Dietary Intakes in Human Pregnancy on Infant Birthweight." Nutrients **7**(1): 153.

Hand, K., I. Katz, M. Gray and J. R. Bray (2016). "Welfare Conditionality as a Child Protection Tool." Family Matters(97): 16-29.

Hemming, K., M. Taljaard and A. Forbes (2017). "Analysis of cluster randomised stepped wedge trials with repeated cross-sectional samples." Trials **18**(1): 101.

Hoynes, H., M. Page and A. H. Stevens (2011). "Can targeted transfers improve birth outcomes?: Evidence from the introduction of the WIC program." Journal of Public Economics **95**(7–8): 813-827.

Hoynes, H. W. and D. W. Schanzenbach (2009). "Consumption Responses to In-Kind Transfers: Evidence from the Introduction of the Food Stamp Program." American Economic Journal: Applied Economics **1**(4): 109-139.

King, J. C. (2016). "A Summary of Pathways or Mechanisms Linking Preconception Maternal Nutrition with Birth Outcomes." The Journal of Nutrition **146**(7): 1437S-1444S.

Kramer, M. S. (1987). "Determinants of low birthweight: methodological assessment and meta-analysis." Bulletin of the World Health Organization **65**(5): 663.

Lohoar, S., Butera, N., & Kennedy, E. (2014). "Strengths of Australian Aboriginal cultural practices in family life and child rearing." Child Family Community Australia Paper No. 25 2014.

Lamb, D. and M. Young (2011). "‘Pushing buttons‘: an evaluation of the effect of Aboriginal income management on commercial gambling expenditure." Australian Journal of Social Issues **46**(2): 119-140.

Lauderdale, D.S. (2006). "Birth Outcomes for Arabic-Named Women in California before and after September 11." Demography **43**(1): 185-201.McTaggart, R. (1991). "Western institutional impediments to Australian Aboriginal education." Journal of Curriculum Studies **23**, 297-325.

Mendes, P., Waugh, J., & Flynn, C. (2014). "Income management in Australia: A critical examination of the evidence." International Journal of Social Welfare **23**: 362-372

Muir, N., & Bohr, Y. (2014). "Contemporary practice of traditional Aboriginal child rearing: A review." First Peoples Child and Family Review **9**: 66-79.

Northern Territory Board of Inquiry into the Protection of Aboriginal Children from Sexual Abuse (2007). Ampe Akelyernemane Meke Mekarle "Little Children Are Sacred". Darwin, Northern Territory Government.

Northern Territory Department of Health (2014). Market Basket Survey 2014. Darwin.

Orima Research (2017). Cashless Debit Card Trial Evaluation: Final Evaluation Report.

P. Persson, M. Rossin-Slater (forthcoming). "Family ruptures, stress, and the mental health of the next generation. American Economic Review. Forthcoming.

Rondo, P. H. C., R. F. Ferreira, F. Nogueira, M. C. N. Ribeiro, H. Lobert and R. Artes (2003). "Maternal psychological stress and distress as predictors of low birthweight, prematurity and intrauterine growth retardation." European Journal of Clinical Nutrition **57**(2): 266-272.

Savitri, A., Yadegari, N., Bakker, J., Van Ewijk, R., Grobbee, D., Painter, R., . . . Roseboom, T. (2014). "Ramadan fasting and newborn's birthweight in pregnant Muslim women in The Netherlands." British Journal of Nutrition **112**(9), 1503-1509.

Senauer, B. and N. Young (1986). "The Impact of Food Stamps on Food Expenditures: Rejection of the Traditional Model." American Journal of Agricultural Economics **68**(1): 37-43.

Singh, G. R. and W. E. Hoy (2003). "The association between birthweight and current blood pressure: a cross-sectional study in an Australian Aboriginal community." Medical Journal of Australia **179**(10): 532-535.

Southworth, H. M. (1945). "The Economics of Public Measures to Subsidize Food Consumption." American Journal of Agricultural Economics **27**(1): 38-66.

Stein, Z., M. Susser (1975). "The Dutch Famine, 1944-1945, and the Reproductive Process. I. Effects or, Six Indices at Birth." Pediatric Research **9**: 70-76.

Victora, C. G., L. Adair, C. Fall, P. C. Hallal, R. Martorell, L. Richter and H. S. Sachdev (2008). "Maternal and child undernutrition: consequences for adult health and human capital." The Lancet **371**(9609): 340-357.

Wilde, P. and C. Ranney (1996). "The Distinct Impact of Food Stamps on Food Spending." Journal of Agricultural and Resource Economics **21**(1): 174-185.

Young, M. F., P. H. Nguyen, O. Y. Addo, W. Hao, H. Nguyen, H. Pham, R. Martorell and U. Ramakrishnan (2015). "The relative influence of maternal nutritional status before and during pregnancy on birth outcomes in Vietnam." European Journal of Obstetrics & Gynecology and Reproductive Biology **194**(Supplement C): 223-227.

Yu, P., Duncan, M. E., & Gray, B. (2008). Northern Territory Emergency Response: Report of the review board. Commonwealth of Australia.

APPENDIX A: OTHER NTER POLICY MEASURES

Table A.1 Summary of all NTER measures, their likely impacts, and an assessment of their implications for our identification strategy.

Measure	Possible outcome	Implementation timing	Assessment
Expanding Health Services Delivery Initative	Better prenatal care and child health outcomes	July 2008 onwards	Minimal overlap with IM
Alcohol restrictions	Increase in food spending (as alcohol spending no longer possible), and decrease in maternal health complications	Immediately following NTER Act (August 2007)	No overlap with IM
Night patrols	Community safety	Gradual, but on a slower timeline to than IM (31 operating by 30/6/08, 42 on consultation. Rollout complete by Dec-09). Night patrols already existed in 23 communiteis pre-NTER.	Different timeline from IM
School nutrition program		Rollout by July 2008. But very low takeup rates.	Different rollout timeline, and unlikely to affect birth outcomes.
Child health checks	All children under 15 eligible.	Began in 16 central communities in late July 2007, rolled out to other communities by early 2008.	Different rollout timeline, unclear if prenatal health check were included.
Store licensing	Increased supply of nutritional food. Precondition for IM rollout.	92 licensed from 1 July 2007 to 31 December 2010.	Much slower rollout timeline than IM. Some stores were given provisional licenses (so probably didn't see immediate change in nutritious food), and some communities would have already had better supply of nutritious food (eg town camps).
Change in Centrelink claim rate	When each IM participant met with a Centrelink case officer, some found out that they were eligible for higher payments, others found out they were not eligible for their current payments and had them cancelled.	This was a direct result of the IM rollout, so same rollout timing. People in some communities (which had long consultation periods) may have met with case managers early and been moved to higher payments before IM rollout.	Affected only around 6 per cent of clients who were income managed at some point during the rollout period.
Employment measures	More jobs, households have more money available to spend on childrens' health/education.	From July 2007 to December 2010, 4,100 job placements.	Different rollout schedule.
Increased police presence	Safer communities, lower stress for pregnant women.	New police presence in 18 communities by 30/06/08. Some communities already had a police presence.	Not affecting all communities, different rollout timeline.
Money management trainning	May lead to increased household food consumption due to better budgeting.	40 communities received some training before IM rollout.	Trainning was only received by around half of the NTER communities. We do not know which communities received the training and what the training involved.

Sources: AIHW 2010; FaHCSIA 2011

APPENDIX B: DESCRIPTIVE STATISTICS

Table B.1 Pre-rollout outcomes and community characteristics

	Year prior to NTER (1 July 2006 - 30 June 2007)			
	NTER communities			Rest of NTER
Outcome variables	First 10 communities	Last 10 communities	Difference	
Birthweight (grams)	3193.5	3162.8	-30.70	3056.8
	(71.38)	(60.51)	(95.17)	(23.06)
Low birthweight (%)	14.6	9.5	-5.2	14.6
	(3.91)	(3.4)	(5.26)	(1.26)
Premature (%)	14.6	6.8	-7.9	16.1
	(3.91)	(2.92)	(4.99)	(1.32)
Obstetric complications				
Due to intrauterine growth restriction (%)	4.9	2.7	-2.2	4.2
	(2.38)	(1.89)	(3.1)	(0.72)
Due to anaemia (%)	2.4	4.1	1.6	11.0
	(1.7)	(2.29)	(2.84)	(1.12)
Due to gestational diabetes (%)	12.2	8.1	-4.1	7.7
	(3.62)	(3.17)	(4.88)	(0.95)
Any (%)	50.0	54.1	4.1	43.3
	(5.52)	(5.8)	(8.06)	(1.77)
Other characteristics				
Age of mother	22.6	23.6	1.0	24.0
	(0.6)	(0.64)	(0.88)	(0.22)
Aboriginal status mother (%)	93.9	94.6	0.0	92.7
	(2.64)	(2.63)	(3.76)	(0.93)
APGAR 1	8.2	8.1	-0.1	8.0
	(0.15)	(0.2)	(0.26)	(0.07)
APGAR 5	9.0	8.8	-0.2	8.8
	(0.12)	(0.19)	(0.22)	(0.05)
Community characteristics[(a)]				
Community size	318.8	277.3	-41.47	485.5
	(68.85)	(58.75)	(96.73)	(55.65)
Female share of population (%)	49.5	52.8	3.3	51.0
	(1.11)	(1.29)	(2)	(0.55)
Median age	22.0	22.8	0.8	22.5
	(0.73)	(0.82)	(1.27)	(0.33)
Population aged 65+	3.3	3.8	0.4	3.3
	(56.83)	(76.37)	(104.8)	(19.23)
People per household	5.4	6.0	0.528	6.1
	(0.39)	(0.22)	(0.63)	(0.21)
Median personal income	209.1	211.0	1.9	210.6
	(2.71)	(7.27)	(7.26)	(6.62)
Median household income	838.9	742.5	-96.40	803.3
	(55.81)	(88.59)	(110.2)	(40.05)
Median rent payments	41.8	32.5	-9.3	44.2
	(6.75)	(8.39)	(12.11)	(4.2)
Labour force share of population (%)	39.9	22.8	-17.11**	39.2
	(4.35)	(4.76)	(7.51)	(2.46)

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1, (a) ABS 2006 Census community profile data; most variables available for 54 NTER communities.

Table B.2 Pre-rollout outcomes and community characteristics

Including only communities with at least one treatment and one control observation during rollout period

Year prior to NTER (1 July 2006 - 30 June 2007)

	NTER communities			Rest of NT
Outcome variables	Communities in first half of rollout	Communities in second half of rollout	Difference	
Birthweight (grams)	3071.8	3087.7	15.8	3354.3
	(30.98)	(30.18)	(43.33)	(10.98)
Low birthweight (%)	14.1	14.2	0.1	6.8
	(1.65)	(1.69)	(2.36)	(0.48)
Premature (%)	14.8	15.9	1.1	7.9
	(1.68)	(1.76)	(2.44)	(0.51)
Obstetric complications				
Due to intrauterine growth restriction (%)	5.4	3.0	2.34*	1.4
	(1.07)	(0.83)	(1.36)	(0.23)
Due to anaemia (%)	9.8	9.6	-0.3	2.0
	(1.41)	(1.42)	(2)	(0.27)
Due to gestational diabetes (%)	6.9	9.3	2.4	6.8
	(1.2)	(1.4)	(1.84)	(0.48)
Any (%)	44.1	47.1	3.0	24.2
	(2.35)	(2.41)	(3.37)	(0.81)
Other characteristics				
Age of mother	23.8	23.6	-0.2	28.6
	(0.28)	(0.29)	(0.41)	(0.12)
Aboriginal status mother (%)	91.7	94.6	2.92*	21.3
	(1.3)	(1.09)	(1.71)	(0.78)
APGAR 1	8.0	8.0	0.0	8.2
	(0.09)	(0.09)	(0.13)	(0.03)
APGAR 5	8.8	8.9	0.1	9.0
	(0.07)	(0.07)	(0.1)	(0.02)
Community characteristics (a)				
Community size	396.6	566.1	169.6*	na
	(58.14)	(78.42)	(98.39)	
Female share of population (%)	51.3	50.8	-0.5	48.5
	(0.58)	(0.86)	(1.06)	
Median age	22.7	22.1	-0.6	31.0
	(0.41)	(0.44)	(0.61)	
Population aged 65+	3.4	3.5	0.0	4.8
	(26.8)	(26.25)	(38.3)	
People per household	5.6	6.6	0.987***	2.9
	(0.22)	(0.24)	(0.33)	
Median personal income	204.3	207.0	2.6	549.0
	(3.74)	(3.33)	(5.09)	
Median household income	692.7	862.8	170.1**	1192.0
	(44.52)	(44.97)	(64.7)	
Median rent payments	43.6	37.8	-5.8	140.0
	(3.43)	(5.09)	(6.36)	
Labour force share of population (%)	38.8	35.8	-2.9	47.3
	(3.42)	(3)	(4.62)	

ABS 2006 Census community profile data; most variables available for 54 NTER communities; rest of NT Census data is average of all of NT. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

APPENDIX C: ROBUSTNESS TESTS

Table C.1 Propensity score matching – treatment on treated effect[(a)]

		Before controlling for premature		
	Sample size	Birthweight	Low birthweight	Premature
Matching on: Mother aged under 20, Aboriginal status mother, whether first pregnancy, whether mother has complications in medical history, whether medical history is unknown, year of birth, hospital of birth, sex of baby, regional rainfall in 3 months to birth				
Matching method				
Nearest neighbour	652	-122.0**	0.0785***	0.0706**
		(52.80)	(0.0247)	(0.0313)
Radius of 0.1	830	-129.6***	0.0636**	0.0652***
		(46.54)	(0.0261)	(0.0248)
Kernel	830	-119.2***	0.0576**	0.0593**
		(44.43)	(0.0232)	(0.0249)
Stratified	830	-119.7**	0.0621***	0.0644***
		(48.10)	(0.0234)	(0.0246)

(a) Where treatment is defined as being in utero in a community covered by income management. We are unable to identify welfare recipient status, Bootstrapped standard errors; *** p<0.01, ** p<0.05, * p<0.1.

Table C.2 Sensitivity of results to sample period, partial treatment and outliers

	Sample size	Before controlling for premature			Controlling for premature	
		Birthweight	Low birthweight	Premature	Birthweight	Low birthweight
Main sample (17 Sept 2007 to 31 Jan 2009)	1,153	-163.9***	0.0807***	0.0564*	-118.7**	0.0480**
		(55.48)	(0.0275)	(0.0315)	(52.77)	(0.0220)
Full sample (1996-2009)	11,200	-96.895**	0.037**	0.042*	-59.429	0.014
		(43.565)	(0.019)	(0.022)	(41.127)	(0.016)
Earlier sample (17 Aug 2007 to 31 Dec 2008)	1,139	-159.75***	0.085***	0.062**	-113.473**	0.051***
		(54.784)	(0.025)	(0.031)	(53.009)	(0.019)
Later sample (17 Oct 2007 to 28 Feb 2009)	1,192	-144.61**	0.074***	0.054*	-101.829*	0.042**
		(55.480)	(0.026)	(0.031)	(54.104)	(0.021)
Excluding partial treatment (main sample)	1,006	-105.926	0.038	0.025	-85.943	0.016
		(67.613)	(0.042)	(0.046)	(57.621)	(0.031)
5% Winsorised (main sample)	1,153	-137.85***			-103.480**	
		(49.981)			(47.718)	

Model is estimated using ordinary least squares. Community-clustered standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Table C.3 Main regression, limited to healthy birthweight range (2500g to 5000g)

Outcome:	(1)	(2)	(3)	(4)
Birthweight in grams				
Income management	-17.29	-47.16	-82.91	-76.67
	(33.95)	(41.83)	(57.65)	(58.47)
Rainfall in 3 months to birth		-0.105**	-0.0273	-0.0273
		(0.0439)	(0.0550)	(0.0505)
Year (omitted category = 2007)				
2008		21.77	21.72	17.78
		(39.19)	(40.88)	(42.03)
2009		98.17	80.93	60.49
		(85.27)	(93.48)	(91.36)
Premature				-395.0***
				(53.52)
Constant	3,303***	3,322***	3,691***	3,689***
	(22.15)	(35.60)	(44.00)	(45.14)
Community fixed effects	No	No	Yes	Yes
Observations	996	996	996	996
R-squared	0.000	0.006	0.102	0.139

Model is estimated using ordinary least squares. Community-clustered standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Table C.4 Sensitivity of results to different methods of controlling for seasonal and time trends

	Before controlling for premature		**Controlling for premature**	
	Birthweight	Low birthweight	Birthweight	Low birthweight
Main regression (controls for rainfall, year FE)	-163.860***	0.080***	-118.667**	0.047**
	(55.482)	(0.027)	(52.766)	(0.021)
No seasonal controls (year FE only)	-126.200**	0.044*	-86.093*	0.015
	(48.549)	(0.023)	(46.075)	(0.018)
Dummy for wet season, year FE	-129.533**	0.055**	-89.544*	0.029
	(52.031)	(0.024)	(49.705)	(0.019)
Quarter & year interacted	-117.574**	0.045*	-70.061	0.011
	(54.155)	(0.025)	(49.797)	(0.019)
Month & year interacted	-235.809***	0.134***	-102.729	0.043
	(79.174)	(0.042)	(68.821)	(0.034)
Steps between rollout dates	-233.673***	0.128***	-95.960	0.029
	(85.712)	(0.047)	(77.445)	(0.038)

Model is estimated using ordinary least squares. Community-clustered standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Table C.5 Treatment effect for subsets of sample period

	Sample size	Birth weight	Low birthweight
Main regression	1,153	-118.7**	0.0480**
		(52.77)	(0.0220)
Drop communities with <10 births	976	-131.0**	0.0582***
		(56.91)	(0.0225)
First half of rollout	578	-157.6*	0.0404
		(81.43)	(0.0256)
Second half of rollout	575	-88.78	0.0493
		(78.70)	(0.0340)

Model is estimated using ordinary least squares. Community-clustered standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Table C.6 Placebo test, main sample period

Date range			Years lead	Birth weight	Low Birthweight[(a)]
17-Sep-00	to	31-Jan-02	6	4.357	0.0150
17-Sep-01	to	31-Jan-03	5	54.07	0.00577
17-Sep-02	to	31-Jan-04	4	-18.43	0.0306
17-Sep-03	to	31-Jan-05	3	-60.19	0.0184
17-Sep-04	to	31-Jan-06	2	-6.786	-0.0138
17-Sep-05	to	31-Jan-07	1	-35.56	0.0200
17-Sep-07	**to**	**31-Jan-09**	**Actual**	**-118.7****	**0.0398****

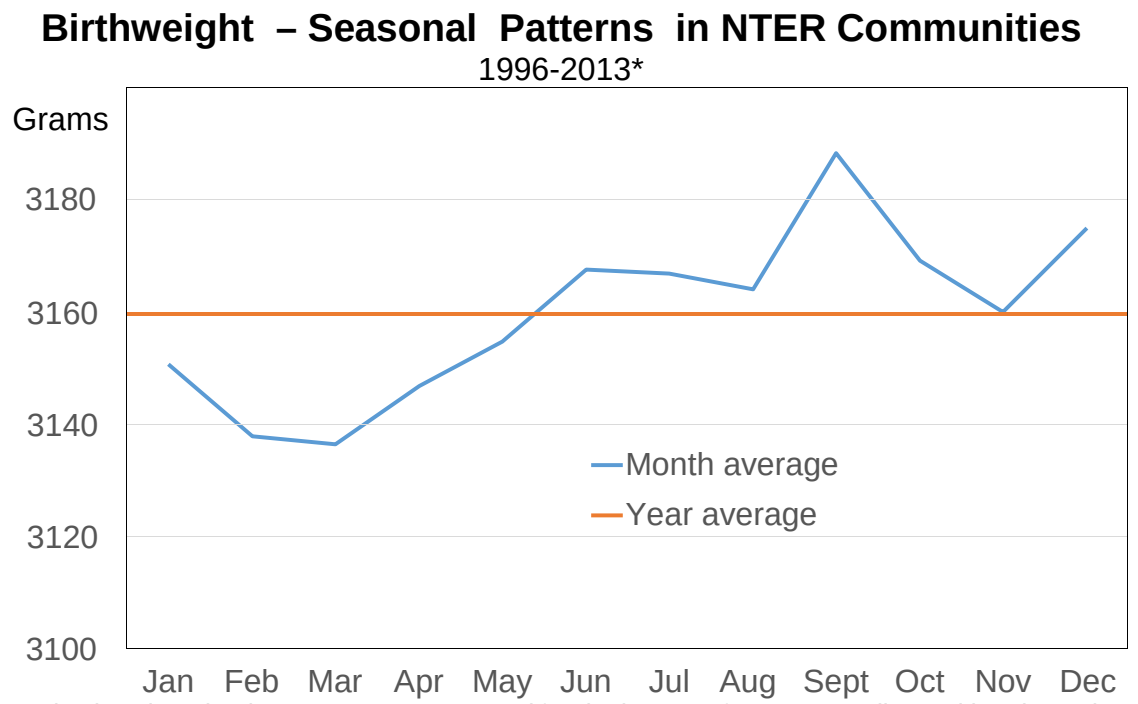
[(a)] Model is estimated using ordinary least squares, and hence results differs from main results table (which uses a probit model); each model controls for year, rainfall, community fixed effects, and prematurity. *** p<0.01, ** p<0.05, * p<0.1.

Table C.7 Treatment effect by region

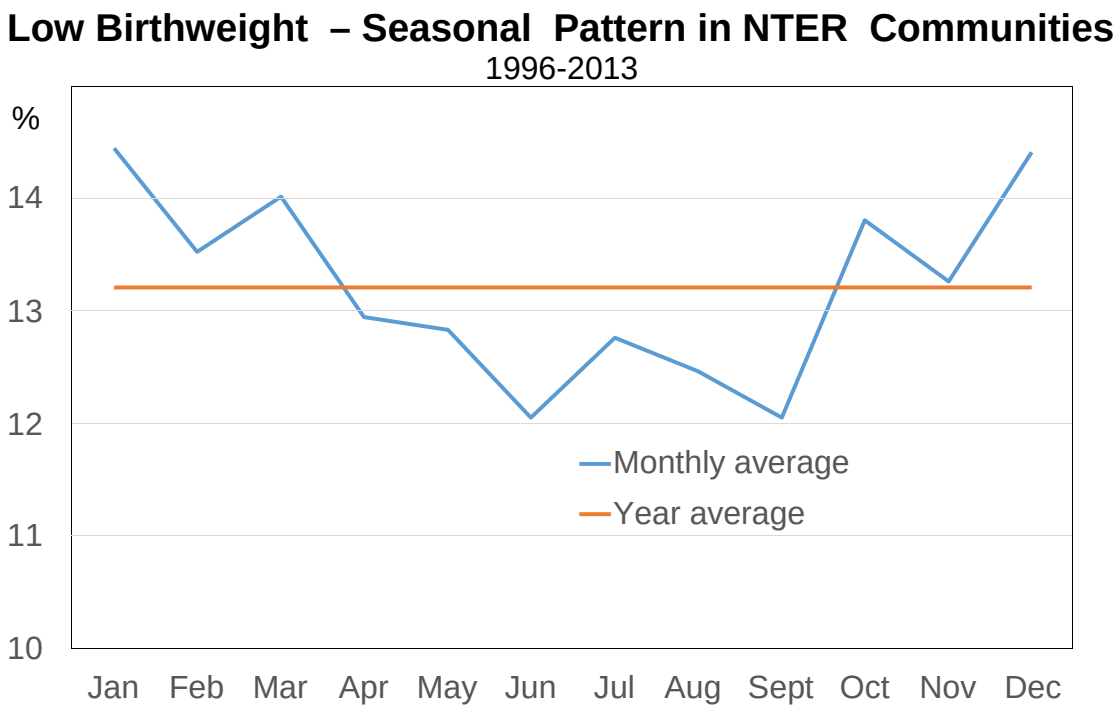
	Sample size	Birth weight	Low birth weight
Darwin region	329	-301.5***	0.103***
		(72.89)	(0.0354)
Central (Katherine and Barkly)	269	-12.53	-0.0550
		(108.3)	(0.0467)
Arnhem region	224	-185.3*	0.0753**
		(97.82)	(0.0332)
Alice region	331	-85.69	0.0840*
		(98.66)	(0.0452)
Model type		OLS	Probit

Each model controls for year, rainfall, community fixed effects, and prematurity. Community-clustered standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

APPENDIX D: SEASONAL PATTERNS

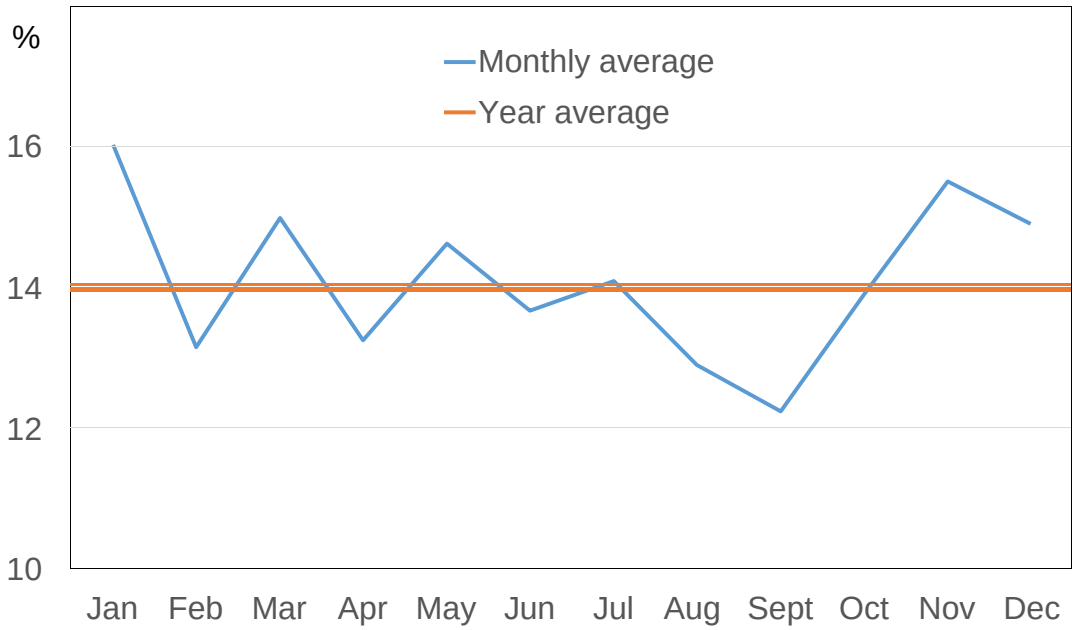


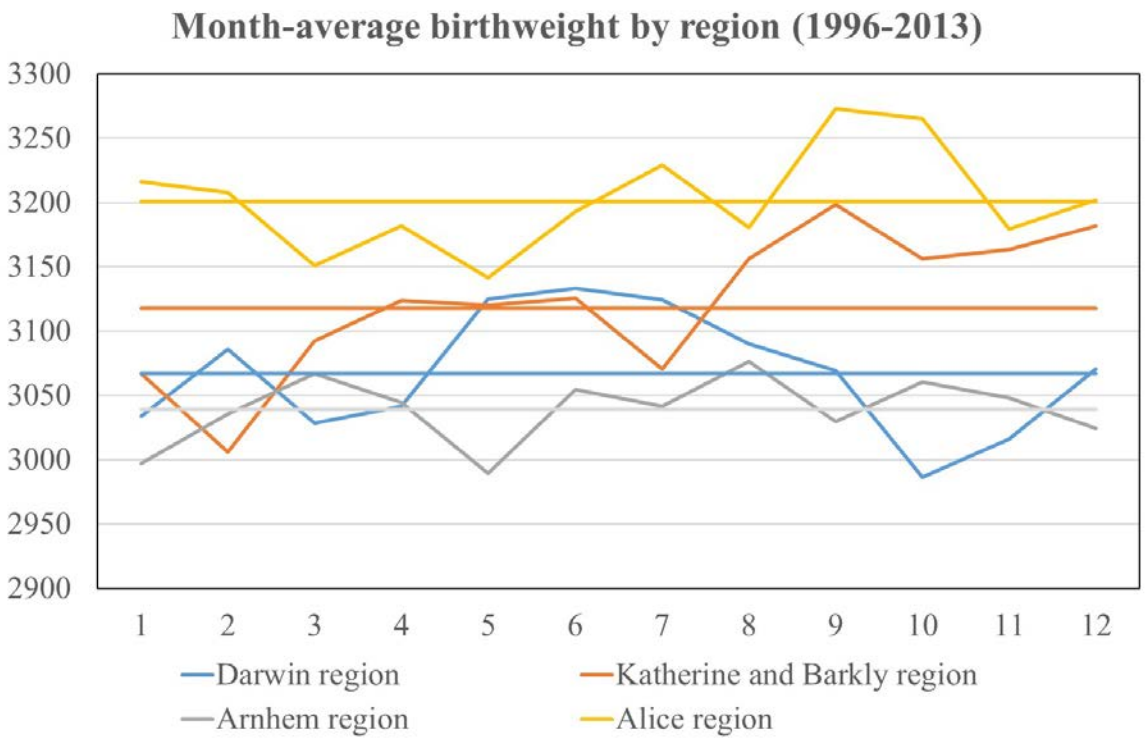
* Series is Winsorised at 10 per cent to control for the impact of extreme outliers, with only a minor effect on the resulting seasonal factors.



Premature Births – Seasonal Pattern in NTER Communities

1996-2013




Month-average birthweight by region (1996-2013)
3300
3250
3200
3150
3100
3050
3000
2950
2900
1
2
3
4
5
6
7
8
9
10
11
12
Darwin region
Katherine and Barkly region
Arnhem region
Alice region

APPENDIX E: TREATMENT EFFECT ON ALCOHOL AND TOBACCO CONSUMPTION

Table E.1 Regression results controlling for maternal characteristics and behaviour

Birthweight	(1)	(2)	(3)	(4)	(5)	(6)
	Before controlling for prematurity			Controlling for prematurity		
Income management	-153.02***	-94.83	-169.45***	-107.37**	-50.30	-123.25**
	(51.99)	(68.03)	(52.73)	(48.28)	(61.93)	(48.27)
Mother's age	11.00***	14.24***	11.01***	9.42***	11.84***	9.35***
	(3.54)	(3.52)	(3.39)	(3.02)	(3.27)	(2.78)
First pregnancy	-70.32	-67.98	-88.16**	-98.47***	-107.71***	-115.20***
	(43.59)	(48.15)	(42.19)	(34.85)	(34.58)	(33.70)
Aboriginal status mother	-326.03***	-231.52***	-223.54***	-268.94***	-198.30**	-180.89***
	(59.04)	(78.65)	(62.47)	(67.21)	(87.95)	(63.32)
Female baby	-95.65**	-146.98**	-93.38**	-122.88***	-155.01***	-118.15***
	(42.28)	(55.88)	(42.09)	(31.95)	(46.36)	(31.12)
Smoking at first antenatal visit (omitted category = no)						
Yes		175.85	-34.65		22.83	-9.19
		(123.21)	(85.98)		(128.88)	(79.09)
Missing			185.31*			246.25***
			(107.91)			(79.23)
Alcohol at first antenatal visit (omitted category = no)						
Yes		-250.56**	-164.89**		-164.08	-153.38*
		(118.06)	(73.55)		(129.76)	(88.17)
Missing			-27.04			-113.23*
			(94.46)			(65.46)
Alcohol at 36 weeks (omitted category = no)						
Yes		128.67	16.08		69.00	29.16
		(151.31)	(111.18)		(149.54)	(101.55)
Missing			-150.40			-17.68
			(91.10)			(61.05)
Smoking at 36 weeks (omitted category = no)						
Yes		-323.85**	-120.06		-165.40	-127.61
		(129.58)	(94.31)		(135.29)	(92.96)
Missing			-257.90***			-268.28***
			(92.27)			(64.59)
Premature				-933.56***	-886.86***	-910.49***
				(53.74)	(81.05)	(53.95)
Constant	3,255.55***	3,273.85***	3,309.65***	3,400.83***	3,406.26***	3,431.75***
	(115.83)	(125.59)	(110.56)	(110.88)	(116.80)	(98.97)
Year, rainfall, FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,153	775	1,153	1,153	775	1,153
R-squared	0.05	0.10	0.09	0.32	0.30	0.34

Community-clustered standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.