WHY DID AUSTRALIA'S MAJOR WELFARE REFORM LEAD TO WORSE BIRTH OUTCOMES IN ABORIGINAL COMMUNITIES?

Mary-Alice Doyle[1], Stefanie Schurer[1,2], and Sven Silburn[3]

[1]The University of Sydney
[2]Institute for the Study of Labor (IZA)
[3]Menzies School of Health Research, Charles Darwin University

11 September 2019

Do not cite or disseminate without permission from the authors

Australia's controversial 'income management' policy restricts Aboriginal welfare recipients to spend 50% of all their government transfers on essential goods (e.g. food, housing, fuel). We evaluate the causal impact of this policy on birth outcomes by exploiting its staggered rollout across communities. The policy aims to improve child outcomes by shifting parents' consumption patterns, but we find no evidence of a positive impact. Instead, it reduced average birthweight by 100 grams and increased the probability of low birthweight by 30 percent, with larger impacts when introduced early in the pregnancy. We explore in depth the mechanisms that explain the unintended consequence.

Keywords: Welfare restrictions, income management, conditional cash transfers, birth outcomes, Aboriginal health, Northern Territory, administrative data.

JEL: D04; I14; I38.

Acknowledgment

The authors thank the following people for their valuable feedback on elements of this research program: Heather d'Antoine, Victoria Baranov, Nicholas Biddle, Gawaian Bodkin-Andrews, Robert Breunig, Julie Brimblecombe, David Cooper, Steven Durlauf, Denzil Fiebig, Marco Francesconi, Matthew Gray, Olga Havnen, James J. Heckman, Matthew James, Liz Moore, Patrick Nolan, Dilhan Perera and Jim Smith; participants of seminars at the University of Chicago, University of Essex, University of Melbourne, Australian National University, University of Western Australia, Institute for Fiscal Studies; and participants of the European Society of Population Economics (Bath, June 2019), of an organised session at the Inaugural Conference of the Asian and Australasian Society of Labour Economists (Canberra, Dec 2017), the Southern Economics Conference (Washington, DC, Nov 2018) and of the 9th Australasian Workshop on Econometrics and Health Economics. 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", which is funded through a Partnership Project between the National Health and Medical Research Council (NHMRC) and the NT Government. This study uses administrative data obtained from the NT Department of Education through this NHMRC Partnership Project. The analysis 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, the ethical standards outlined in the ethics agreement HREC Reference Number: 2016-2611 Project Title: Improving the developmental outcomes of Northern Territory children: A data linkage study to inform policy and practice in health, family services and education (Human Research Ethics Committee of the Northern Territory Department of Health and Menzies School of Health Research). The authors acknowledge funding from an Australian Research Council (ARC) Discovery Early Career Research Award DE140100463 and a University of Sydney SOAR Fellowship (2017-2018).

In both developed and developing countries, policymakers and donors routinely place restrictions on how welfare recipients can use their transfer payments. Policy makers justify such paternalistic welfare policies by appealing to social preferences, especially when transfer payments are used for the consumption of goods that have either negative (e.g. excessive consumption of alcohol and tobacco) or positive (e.g. investment in education and health care) externalities for communities (Currie and Gahvari 2008). Restricted welfare payments sometimes take the form of in-kind transfers or restricted-use vouchers (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 (see Gentilini 2016 for a comprehensive review), there is surprisingly little empirical evidence that directly compares the impact of restricted with unrestricted transfers.[1]

In this paper, we provide first evidence on the effectiveness of restricting welfare payments in the context of a major welfare system reform of a highly-developed OECD country. In 2007, the Australian Government embarked on its most comprehensive and discriminatory welfare reform in recent history. Referred to as ‘income management’ or ‘income quarantining’, this new policy targeted transfer recipients in all Aboriginal[2] communities located in the Northern Territory, one of Australia’s most remote and disadvantaged jurisdictions with the highest share of Aboriginal populations in the country. Recipients previously received payments as a regular unconditional cash transfer. Under

[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, Do and Ozler (2005) note, the economics of these policy options is very similar.

[2] We will refer to people of Aboriginal or Torres Strait Islander descent as simply `Aboriginal' since the vast majority of Indigenous individuals in the Northern Territory identify as Aboriginal singularly or as both Torres Strait Islander and Aboriginal. According to the official definition, “an Aboriginal or Torres Strait Islander is a person of Aboriginal or Torres Strait Islander descent, who identifies as being of Aboriginal or Torres Strait Islander origin and who is accepted as such by the community with which the person associates” (Australian Institute of Aboriginal and Torres Strait Islander Studies 2012).

income management, half of each regular payment was quarantined into a separate account with restrictions over its use. The quarantined amount constituted at the time about AUD150 per week. It could only be spent on household priority items or child-centred goods.

Income management was the cornerstone policy of the Northern Territory National Emergency Response (NTER), which entrusted a taskforce to oversee a range of highly-controversial measures. The NTER reflected the federal government's response to the *Little Children are Sacred* report, a highly publicized inquiry into child sexual abuse and family violence within remote Aboriginal communities (Wild and Anderson 2007). The income management policy however went far beyond the implications of this report. It attempted to rectify what was perceived as the deep-rooted behavioural causes of health and education inequalities in remote Aboriginal communities. The policy's core objective was to improve the welfare of Aboriginal children in general by increasing the share of household income spent on child-centred goods, and by reducing the amount spent on goods viewed as potentially harmful, such as alcohol, tobacco, porn and gambling services.[3] Income management was designed as a community-level intervention. Policymakers expected the benefits of restricting choice to permeate throughout the community by creating positive spill-over effects. It was hoped that income quarantining would make communities safer and improve the bargaining power of women.[4]

[3] The *Welfare Payment Reform Act 2007* stated explicitly that it aims to "promote socially responsible behavior, particularly in relation to the care and education of children" (*Welfare Payment Reform Act 2007* No. 130, 2007 123TB Objects, Section (a)). The government minister responsible for the policy change stated the aim was to "stem the flow of cash going towards substance abuse and gambling and ensure that funds meant to be for children's welfare are used for that purpose" (Brough 2007).

[4] By reducing the amount of physical cash women could access, it was anticipated that women would be protected from excessive demands for money from close and distant family members, a practice known as 'humbugging'. Then Prime Minister John Howard mentioned this aim in a public speech at the Sydney Institute in June 2007.

We are the first to evaluate the impact of income management on child welfare, as measured by early-life health outcomes.[5] To identify its causal impact, we exploit the gradual rollout of the income-management policy across communities as a source of exogenous variation. Our analysis draws on administrative records on the universe of all births from the Northern Territory Data Linkage Study, one of the most comprehensive linked administrative data repositories in the world to study human development of children from very disadvantaged backgrounds (Silburn et al. 2018). We focus specifically on birthweight and the probability of low birthweight. Although the policy affected only those community members who received a government transfer (Australian Institute of Health and Welfare 2010), it affected almost all women with children or women who expected a baby. Our hypothesis is that if income management improved nutrition and reduced risky health behaviors of pregnant benefit recipients, then we would expect to observe an increase in birthweight as a result of the policy.

Our identification strategy builds on a comparison of 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 implemented after birth, or was implemented very late in the pregnancy.[6] We demonstrate that the roll out of the policy was as good as random. The policy was implemented shortly after its announcement, was compulsory, and allowed for almost no exemptions for welfare recipients, meaning there was no capacity for affected individuals to self-select into whether or when they would receive the intervention. The rollout timing did not coincide with pre-

[5] Recently, we also completed an evaluation of the impact of the income-management policy on children's school participation(Cobb-Clark et al. 2018).
[6] Our approach is similar to Almond, Hoynes, and Schanzenbach (2011) who estimated the causal impact of the US food stamp program using its gradual rollout across counties.

existing birthweight trends and was not associated with the degree of economic or health disadvantage of a community.

We find that income management did not improve birth outcomes for Aboriginal children in the Northern Territory. Despite its aims to improve child welfare, the policy reduced average birthweights of children who were exposed to income management in utero by 100 grams. The adverse effects are observed across the birthweight distribution, however they are strongest at the bottom end. Income management increased the probability of low birthweight by almost 30 percent. Exposure to income management mattered most for birth outcomes when it occurred in the first or second trimester of the pregnancy. We interpret our finding as evidence that exposure to the policy had cumulatively harmful effects that required time to materialize, while shorter exposure was less harmful.

We explore in depth the likely mechanisms that explain this unexpected consequence. The negative treatment effect of the policy cannot be explained by changes in fertility or in the composition of mothers who were willing to have a baby. It is also not explained by an increase in maternal consumption of alcohol or cigarettes during pregnancy or by improved access to neonatal care that may have increased the detection of pregnancy problems (potentially leading to premature birth but improved overall health). There is also no evidence that the number of stillbirths decreased, which may have reduced average birthweights because high-risk babies were more likely to survive.

This leaves us with a more qualitative assessment of why the income management policy worsened birth outcomes. First, our findings are consistent with our initial work on the policy's negative impact on children's school participation in the short-run (Cobb-Clark et al. 2018). In this previous work, we ruled out a series of channels including in- and out-migration from communities, which could have explained increased stress levels among the mothers in our sample. Analysing survey data from the Longitudinal Study of Australian Children, we provided tentative evidence that survey participants who moved from standard welfare payments onto income management were more likely to experience an increase in excessive money demands from other family members (humbugging) and in arguments that affected children in the household. It seemed that the policy rollout may have increased financial stress and reduced family functioning (Cobb-Clark et al. 2018). Another important source of stress could have originated from policy implementation problems that led to short-term income insecurity. To test this hypothesis would require access to government transfer data on flow and interruption, which is currently not available for research.

Our research findings are of paramount significance to policymakers. We are the first to provide evidence on the unintended consequence of Australia's major welfare reform, which opted to restrict individual choice to mitigate deep and persistent disadvantage. Unlike programs in other countries, which typically target a relatively small share of household income (Gentilini 2016), Australia's scheme was compulsory and limited the ability of welfare recipients to spend what was for most households their main source of income.[7] A large international literature has studied the negative impacts of poor health behaviours, stress and environmental shocks in utero on birth outcomes (see Aizer and Currie (2014) for an overview), but little evidence exists on the unintended consequences of well-intentioned welfare policies on early-life health. Our study demonstrates that policies that are not carefully designed and tested may unintentionally escalate already poor health outcomes of highly fragile populations.

[7] While we do not have data on the value of household income and welfare payments, AIHW (2010) reports that around one-third of adults in NTER communities were employed, suggesting that the remaining two-thirds are likely to rely on transfer payments as a main source of income.

Birth outcomes are used as early indicator of the health disparities between Aboriginal and non-Aboriginal populations both worldwide and in Australia. In the Northern Territory, the probability of low birthweight for Aboriginal infants has remained steadily high at 14 percent since our measurement began in 1994, compared with 7 percent for the non-Aboriginal Territorians (Su et al. 2018). The early-life health shock experienced by the cohort of Aboriginal children affected by the rollout of the income-management policy is likely to cause further problems down the track. Low birthweight children in general have been found to be delayed or impaired in their cognitive development, educational attainment and later-life health and economic outcomes (Aizer and Currie 2014; Almond, Currie, and Meckel 2014; Figlio et al. 2014; Almond, Chay, and Lee 2005; Fletcher 2011; Currie and Almond 2011). The adverse effects of low birthweight may persist across generations (Black, Devereux, and Salvanes 2016; Currie and Moretti 2007; Victora et al. 2008). Therefore, the income-management policy may have adverse intergenerational impacts for Aboriginal communities, possibly of similar proportions as other historic well-intentioned, paternalistic policies such as the 'forced removal' policy (Silburn et al. 2006).

The remainder of this paper proceeds as follows. Section 1 describes the institutional context of the income management policy. In Section 2 we review the relevant literature, including international evidence on the effectiveness of restricted cash transfers and other evidence from the Australian income management experience. In Section 3 we present a conceptual framework underlying the mechanisms through which income management may affect birth outcomes. Section 4 describes the Northern Territory Data Linkage Study and the variables used for the analysis. In Section 5 we present our empirical framework, including the identification strategy and statistical models. Estimation results and a series of robustness checks are presented in Section 6. In Section 7 we explore the mechanisms that may explain the negative treatment effect of income management on birth outcomes. Section 8 discusses the implications of our findings and concludes. An appendix provides supplementary material.

1 THE INCOME MANAGEMENT POLICY

1.1 Institutional background

The Northern Territory (NT) is a vast geographic area, stretching over 521,000 square miles, covering approximately one-sixth of the Australian continent. It is almost twice the size of Texas, the second largest state in the United States, and ten times the size of the United Kingdom. Around half of its approximately 246,000 residents live in the capital city of Darwin. Aboriginal and Torres Strait Islanders (from here onward referred to as Aboriginal) make up 25.5 percent of the NT's total population, and 2.8 percent of the Australian population overall. The NT is governed by its own local government in conjunction with the Australian Federal Government and approximately half of the land in the NT is Aboriginal-owned as a result of the *Aboriginal Land Rights (Northern Territory) Act of 1976*.

The NT has received heightened attention from both policymakers and clinicians over decades for its high levels of poverty, especially among Aboriginal populations. This attention culminated in early 2007, when the NT Board of Inquiry into the Protection of Aboriginal Children from Sexual Abuse released its report titled *Ampe Akelyernemane Meke Mekarle "Little Children are Sacred"*(Wild and Anderson 2007). The report called for immediate government action to address high rates of child sexual abuse in remote communities. It emphasized the need to consider child neglect, alcoholism and inadequate education and housing 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 (See Appendix A for full list). These policies applied to residents in 73 remote Aboriginal communities and outstations, and 10 town camps, and did not apply to non-Aboriginal towns or communities in the Territory.[8] To facilitate the targeted nature of these policies towards Aboriginal Australians, the government suspended Part II of the *Racial Discrimination Act 1975*, which proscribes equality before the law regardless of race.[9]

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. Towards the end of the rollout period, a debit card ('Basics

[8] A town camp is an Aboriginal community situated in a town or city, or close to its boundaries.
[9] By suspending (excluding) the operation of Part II of the RDA, the members of the communities affected by the NTER legislation were effectively denied the protections afforded by the RDA to every other citizen to challenge legislation that they consider to be in breach of the RDA (Australian Human Rights Commission 2011).

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 Part II of the *Racial Discrimination Act* in 2010, a modified version of income management was introduced to cover the whole of the NT. 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.

1.2 Who was affected?

Income management applied to all 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, we know that around three-quarters of the adults in affected communities were subject to income management at some point during the rollout period, with 55 percent being income managed at a point-in-time after the rollout was complete (Australian Institute of Health and Welfare 2010).[10] Though limited information is available on which residents were affected, women and younger adults were more likely to receive welfare payments (Australian Institute of Health and Welfare 2010). The rate was likely to be almost 100 percent for pregnant women because of their entitlement to a baby bonus lump-sum payment that was fully quarantined.

[10] 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.

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 ('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. Only 649 exemptions, or three percent of all individuals who were, at some point, subject to income management, had been granted (Australian Institute of Health and Welfare 2010).

2 LITERATURE

2.1 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 2.5 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 (that is, achieving similar or better outcomes with lower administrative costs).

Another recent study from Uruguay estimates the impact of a temporary transfer program on birthweight (Amarante 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 6-13 percent of income for the average household. It is highly likely to be inframarginal (see Section 3.1).

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, and Harnack 2016). But others have identified a “cash-out puzzle”, finding that food stamps increase food consumption, even though the vast majority of recipient households 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; that is, in a multi-adult household, different household members might have different preferences over food consumption. When receiving food stamps, they cannot bargain over whether to spend food stamps on food. But when food stamps are cashed out, relative preferences and bargaining between household members will play a more important role in determining household food consumption. In support of this hypothesis, they find that the ‘cash-out puzzle’ is present in multi-adult households, but not in single-adult households.

In summary, there is limited evidence on the impact of restricting transfers, and the evidence that exists does not point towards any consistent finding. A significant challenge that many studies face (e.g., Amarante et al. 2016; Hoynes and Schanzenbach 2009) 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 income. But even among the experimental studies that do not face this challenge, there is no clear finding on the impact of welfare restrictions. Whether and how restrictions affect household behaviour appears to depend on the specifics of how the policy is designed, and how its administration interacts with existing household dynamics (such as the potential intrahousehold bargaining dynamics highlighted by Breunig and Dasgupta).

This paper builds on the existing literature, using a context where, like the experimental studies, 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 programs studied in previous research, because income management payments account for a larger share of household income than the US food stamps program or the programs studied by Gentilini and Amarante. In addition, the program is administered differently; as described above, allocations are more likely to be tailored to each household's needs, as income management recipients were meant to consult with a case worker to jointly determine how restricted funds would be allocated.

2.2 Evidence on income management

Within Australia, there have been 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, and they find differing results. Considering the impact of income management on food expenditure, Brimblecombe 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.[11]

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 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 initial introduction of income management in the NT, which is 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 et al. 2014), and the 2016-2017 Cashless Debit Card trials (Orima Research 2017).[12]

The two reports covering the initial rollout 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,

[11] 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.
[12] See the Department of Social Services website for reports on other forms of income management.

especially women" (p. 11). This finding is replicated also in more recent qualitative report which suggests that some (but not all) clients were highly receptive to income management (Hand et al. 2016).

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 population.[13] 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).

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 said they 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).

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 these 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

[13] Respondents were reportedly invited to participate by community brokers or government business managers (Australian Institute of Health and Welfare 2010).

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 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 the existing 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 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.

2.3 Linking income management with birthweight: nutrition

There are two potential reasons for 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 restriction, which leads to below-average weight for normal gestational length, may be affected by some of these same factors, as well as by by maternal nutrition during pregnancy (Kramer 1987).

Income management is expected to affect birth outcomes by increasing food consumption and improving nutrition during pregnancy. The policy sets aside funds that can only be spent on priority goods, like food (see Section 3 for more details). 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, and Schanzenbach 2011; Hoynes, Page, and Stevens 2011; Amarante 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, and Schanzenbach 2011; Amarante et al. 2016). But these transfer programs represented a relatively low share of household income. A study of the Oportunidades (now named Prospera) 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 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.

3 CONCEPTUAL FRAMEWORK

3.1 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 goods viewed as harmful or addictive. 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, a 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 ($P_x X \geq IM$). For these 'inframarginal' households, 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 1 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. Household B is extramarginal; with the move from cash to income management, it shifts its consumption towards more priority goods, moving to a lower indifference curve, U_b.

[Insert Figure 1 here]

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)[14] 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. Framing the problem in terms of infra- and extramarginal cases implies that we would test against a one-sided alternative hypothesis; this would give us greater statistical power to detect a positive impact, if one exists. While a one-sided test would increase our power to detect an effect (if one exists), we instead test against a two-sided alternative hypothesis; given the negative press around the policy, we wanted to allow for the possibility of a negative impact.

3.2 Are households extramarginal?

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. Figure 2 shows that over 60 percent of average household expenditure was already directed towards priority goods before income management.[15] The reported numbers relate to total household spending, which may be larger than the value of welfare income.

[Insert Figure 2 here]

[14] Around 65 percent of income managed funds were spent on food during the rollout period (Australian Institute of Health and Welfare 2010).

[15] 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.

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.[16]

3.3 Inputs and outcomes

The theoretical framework described above explains the potential for income management to affect household consumption. However, changes in household consumption are only relevant to policymakers if they translate to better outcomes such as improved health, education and wellbeing. As Cunha (2014) 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 (2014) 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[17] may reduce purchases through less formal channels. For instance, income management may have reduced purchases at local 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

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

[17] 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 (see Appendix A).

more nutritious or cheaper (Australian Institute of Health and Welfare 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 DATA

The analysis is conducted with data from the NT Data Linkage Study (*NT-DLS*), which is funded through a Partnership Project between the Australian National Health and Medical Research Council (NHMRC) and the NT Governments (Silburn et al. 2018). The data linkage is managed by SA NT DataLink. So far, the NT-DLS has linked 11 sets of administrative records of children born in the Northern Territory since 1994. For the purpose of this study, we have linked rainfall data from the Australian Bureau of Meteorology (BOM) to obtain information about weather conditions for each community considered in this study, and community characteristics information from the 2006 Census community profile data collected and made available by the Australian Bureau of Statistics (ABS).

We extract from the *NT-DLS* the NT Perinatal Trends files (custodian: NT Chief Health Officer), which include demographic variables, and information on maternal health, the pregnancy, labour, birth and perinatal outcomes. These files contain information on 74,425 children who were born in the Northern Territory between 1994 and 2013. About 40 percent of these children are recorded as of Aboriginal or Torres Strait Islander descent.

4.1 Definition of treatment status

Income management was rolled out in stages in all 73 NTER communities (and associated outstations), and 10 town camps, which we refer to in short as NTER communities. The NTER communities are therefore separated into 88 locations, 83 of which had at least one birth during the rollout period.[18]

To identify children who were living in NTER communities, we use information on the mother's suburb of residence at the time of birth, as recorded in the Perinatal Trends files.[19] We identify suburbs that are located in NTER communities, and link these observations to the date income management was introduced in the community (the schedule is available in Appendix A of the AIHW (2010) evaluation report).

We define a child as being treated if income management was introduced in his/her community before the start of the third trimester in utero. As we have information available on gestational age, we are able to precisely date the beginning of the third trimester. Our definition of treatment status is based on Almond, Hoynes, and Schanzenbach (2011) who find that the introduction of the Food Stamps 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.[20] Our expectation is therefore that the impact of income management is most likely to be evident in infants for whom income management was introduced before or at the beginning of the third trimester.

---

[18] 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 closest large community and are therefore listed separately.

[19] 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.

[20] Note, Almond, Hoynes, and Schanzenbach (2011) are unable to precisely date the beginning of the third trimester, as they did not have data available on gestational age.

4.2 Sample restrictions

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 babies born during or shortly after the income management rollout period, to include all those who were in their third trimester during the rollout.[21] 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 19 births that occurred before the beginning of their third trimester, of which 7 are still births.[22]

4.3 Outcome variables

The outcome variables of interest are birthweight and the probability of low birthweight, which we derive from the NT Perinatal Trend files. Low birthweight is common in remote NT communities – with around 14 percent of infants born with low birthweight in the year before income management was introduced – around twice the rate of the rest of the Northern Territory. This indicates that there is significant capacity for improvement in these outcome measures. As we hypothesize the policy affects birthweight through changes consumption of food and nutrition, we include a control for premature birth. This allows us to isolate the impact of the policy change in intrauterine growth, which (as noted above) we expect to be more responsive to nutrition.

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

[22] However, Table 7 below demonstrates that that our results are not sensitive to these 19 exclusions.

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, Currie, and Meckel 2014; Almond, Chay, and Lee 2005). In a survey of the literature, Victora 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 (Figlio 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 individuals with low birthweight 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 EMPIRICAL FRAMEWORK

5.1 Identification strategy

To identify the causal impact of income management, we exploit its staggered rollout. 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 17 September 2007 and 27 October 2008. As shown in Figure 3, the policy was introduce in roughly 3-5 communities at a time, with 3-4 weeks intervals between introductions. By April 21, 2008, 50 percent of all communities had implemented income management.

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).

[Insert Figure 3 here]

A similar approach was used by Almond, Hoynes, and Schanzenbach (2011) to estimate the impact of the US food stamp program.[23] This approach is also similar to cluster-randomised stepped wedge trials (Hemming, Taljaard, and Forbes 2017), which intentionally roll out a program across clusters (or communities) on a randomised schedule, to eventually cover the full population.[24] 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.

[23] However, a key difference is that Almond, Hoynes, and Schanzenbach (2011) exploited the staggered rollout over many years and across many counties. To analyse this large dataset, their study focuses on county-level data, whereas we use individual data.
[24] This is an increasingly common approach in the public health and program evaluation literature.

The rollout was conducted following a pre-defined timeline.[25] While no information is publicly available explaining the logic of that timeline. demonstrate that the rollout schedule can be considered as good as random, allowing us to isolate the causal impact of income management.

First, income management was rolled out on a different timeline from the other NTER policies, meaning that our results are not confounded by concurrent policy changes (see also Appendix A for details). Second, the rollout did not follow any clear geographic pattern. Figure 4 suggests 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. 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.

[Insert Figure 4 here]

Third, 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. Constructed from our

[25] Note there was a slight difference between actual and planned rollout dates. In the weeks prior to the scheduled introduction of income management in a given community, there was a consultation period in which Centrelink staff (the Australian government agency that distributes welfare payments), 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). In addition, around 40 communities received money management training prior to the rollout, though we have been unable to find information on which communities these were, how they were chosen, and what the training involved. If these conditions were not met, the rollout would be delayed. The implementation was delayed for 13 communities, outstations and town camps. Delays were substantial (that is, more than a few weeks) for only four communities (Australian Institute of Health and Welfare 2010). Our identification is based on the actual, not planned, rollout dates.

administrative data, Figure 5 suggests no apparent trend in birthweight in either group prior to the rollout.[26]

[Insert Figure 5 here]

Fourth, 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[27] before the NTER, using data from the NT Perinatal files and the 2006 ABS Census. Most birth outcome measures, including obstetric complications, characteristics of the mother and APGAR scores[28] were similar between the two groups before the rollout.

[Insert Table 1 here]

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

[26] 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.
[27] 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.
[28] APGAR is a test (appearance, pulse, grimace, activity and respiration) given to newborns at 1 minute and 5 minutes after birth.

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).[29]

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. Yet, we find that at worst, the very first-adopting communities had slightly higher pre-intervention birthweight and similar probability of low birthweight to other NTER communities (see Appendix Table B.1). It does not appear, therefore, that early rollout of income management was targeted towards the communities with the worst pre-intervention outcomes. These conclusions do not change even when dropping the 21 communities with only-treatment or only-control observations (see Appendix Table B.2).

Finally, it could be argued that some residents in early-adopting communities moved to late-adopting communities to avoid income management. Yet, the scope for residents to move to avoid income management was limited, as eligibility was determined based on place of residence as of 21 July 2007, one week after the policy was announced. Even though mobility rates are high in Aboriginal communities, we have demonstrated in our own previous work that income management did not impact mobility (Cobb-Clark et al. 2018). We conclude that the rollout timeline can be considered exogenous to our outcome measure.

[29] 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 17 September 2007, was regressed on pre-treatment community characteristics, levels in birthweight, and rainfall. We find no significant association between any of the variables and the timing of the reform, except for a significant coefficient on household size. Overall, our extended set of control variables in this regression explain up to 12 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 can be provided upon request.

5.2 Econometric model

We exploit the exogenous variation in the rollout schedule in a simple estimation framework to identify the causal impact of income management on birth outcomes. Specifically, we estimate the following model:

$$Y_{it} = \alpha + \delta IM_{it} + \gamma X_{ct} + \eta_c + \theta_t + \delta_{ct} + \epsilon_{ic}, \qquad (1)$$

where Y_{it} is the outcome variable for newborn *i* at time *t*. IM_{it} is the 'treatment' indicator, which is equal to 1 if income management was in place in community (*c*) at the beginning of the third trimester of pregnancy, and 0 otherwise. The control group consists both of newborns born into community *c* before income management was introduced (at time *t-k*), and those born into other NTER communities where income management had not yet been introduced by the beginning of their third trimester, either in time period *t* or *t+k*.[30] Treatment status varies within most communities over time. Therefore, newborns from the same community may be assigned either to treatment or control status depending on their date of birth.[31]

We control for prematurity (X_{ct}), as defined being born before 37 weeks of gestation. This control allows us to identify intrauterine growth factors that influence birth weight. As birth outcomes may differ across communities (see Appendix Figure B.1), we control flexibly for community fixed effects η_c. For instance, birthweights are consistently higher communities located in the Alice Springs region, while they are lowest in the Arnhem Land

---

[30] Our approach is closely related to (Almond, Hoynes, and Schanzenbach 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, and Schanzenbach (2011) use a time series of average birth outcomes for each county.

[31] Almond, Hoynes, and Schanzenbach (2011) 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).

region. These community fixed effects capture unobserved, location-specific factors that influence birth outcomes, but which do not change over time. For instance, they may capture 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 furthermore include controls to capture time trends in birth outcomes (see Appendix Figure B.2). For instance, the probability of low birthweight is highest (14 percent) in the wet season (October to March), and lowest (12 percent) in the dry season (April to September). Controlling for variation over time is essential given that the gradual rollout of the policy introduces a correlation between time and treatment status (Davey et al. 2015; Hemming, Taljaard, and Forbes 2017).[32] More importantly, the time trends may differ by communities, so optimally, we would like to control for time-varying community fixed effects δ_{ct}. Unfortunately, we ask too much of the data, given that the role-out schedule of Income Management stretched only over 13 months.

We deal with this problem with two set of controls. We use year time fixed effects (θ_t), dummy variables for year 2008 (Jan-Dec) and 2009 (Jan), relative to months Sep-Dec in 2007, to capture annual variations in birth outcomes (θ_t). To proxy community-specific time trends, we use variation in weather conditions leading up to the birth of the child. Weather conditions, especially rainfall, can be considered as the key determinant of time variation in birth outcomes in our NTER communities. Total rainfall is measured in millimetres for the three months prior to birth in the newborn's region.[33] Rainfall data are sourced from the Australian Bureau of Meteorology (BOM), based on the total rainfall per

---

[32] (Hemming, Taljaard, and Forbes 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.

[33] 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.

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, both because the short time frame of the sample period limits the ability of controls to pick up regular seasonal variation, and because timing of the wet season can vary from year-to-year.[34] We will conduct a series of robustness checks using alternative time-trend specifications.

5.3 Intention-to-treat (ITT) estimator

Of main interest to our analysis is the sign, size and statistical significance of δ, the impact of income management. 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 households were receiving Government transfers, and thus were affected by 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.[35] As noted above, most residents in NTER communities were subject to income management, because Government transfers include any family payments, including unconditional lump-sum transfers. 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 was not linked to any other factors that explain birth outcomes, we interpret the estimate of δ as causal.

[34] 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 birthweight data.

[35] 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.

We use ordinary least squares to estimate the impact of income management on birthweight, and a binary choice probit model to estimate its impact on the probability of low birth weight. For the latter, we report marginal probability effects evaluated at the mean of all control variables. We furthermore use quantile regressions to demonstrate whether the policy had heterogeneous impacts across the birthweight distribution. As an additional robustness check, we present standard propensity score-matching estimators, where we limit the control group to those who best resemble the treatment group.

The standard errors of our estimated parameters are clustered on the community level. To deal with small sample sizes in some communities, we conduct several robustness checks in which we drop communities with less than 10 births, or drop communities in which there is no variation in treatment status.

6 ESTIMATION RESULTS

We begin by reporting the ITT-estimates from our benchmark model (Section 6.1.). We then move on to report the outcomes of a series of robustness checks (Section 6.2) and a heterogeneity analysis of the treatment effect with respect to the distribution of birthweight and the intensity of exposure to income management (Section 6.3).

6.1 Benchmark model

Table 2 reports the main treatment effects of interest. In a model without control variables (column 1), we find that average birthweight is around 61 grams lower in the treatment group than in the control group (Panel A) with a standard error (S.E.) of 36 grams. The probability of low birthweight is 2.8 percentage points higher (Panel B), with a S.E. of 1.7.

Controlling for rainfall and annual time trends (column 2) doubles the negative treatment effect for birthweight to 120 grams (S.E. 45) and for the probability of low birthweight to 6 percentage points (S.E. 2.1). Further controlling for community fixed effects (column 3) increases the treatment effect by about 30% to 164 grams (S.E. 58). Similarly, exposure to income management increases the probability of low birthweight by 8 percentage points (S.E. 2.8), or by 50% relative to the mean probability of the control group pre-rollout.

The decrease in birthweight appears to have come through both the intrauterine growth and the gestational length channels. After controlling for prematurity (column 4), the treatment effect on birthweight (intrauterine growth channel) declines by 28 percent to minus 119 grams (S.E. 55). The treatment effect on low birthweight also declines by almost 50 percent to 4.8 percentage points (S.E. 2.2). The treatment effects are statistically significant at the 5 percent level.

[Insert Table 2 here]

6.2 Robustness checks

We conduct a series of robustness checks with respect to our estimation model (matching methods), the sample size (dropping observations) and the controls for seasonality (allowing for community-level rainfall data, or months or quarter fixed effects). Our findings are robust to these changes. We will conclude this section with evidence on a Placebo test, which demonstrates that our treatment effects are not driven by unobserved seasonal variations that cause low birthweight and that coincided with the roll out schedule.

Figure 6, depicts the estimated treatment effect (bullet) and its 95 percent confidence interval (horizontal line). 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 newborns to those in the control group who most closely resemble them in terms of observable characteristics. 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 method approaches (red-colored horizontal lines), we find an average reduction in birthweight of 118-150g (significant at the 1 or 5 percent level). The treatment effect on the probability of low birthweight varies between 5.5 and 8.3 percentage points (significant at the 1 or 5 percent level, see Appendix Table C.1).

[Insert Figure 6 here]

Furthermore, changes to the sample period, censoring of outliers and exclusion of partially-treated newborns do not affect our conclusions. The treatment effects remain between minus 77 grams and minus 177 grams. For instance, if we shift the sample period forward by one month (Aug 2007 to Dec 2008) or back by one month (Oct 2007 to Feb 2009), the treatment effect remains around minus 100 grams. Our negative treatment effects 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 no longer statistically significant because of smaller number of births).

When dropping communities with less than 10 births in our sample, the treatment effect of income management on birthweight increases in absolute magnitude to minus 135 grams and is statistically significant at the 1 percent level. The impact on the probability of low birthweight increases to 5.7 percentage points, which is also statistically significant at the 1 percent level. The negative treatment effect on birthweight is larger for communities in the first half of the roll-out period (minus 178 grams) than for communities in the second half of the roll-out period (minus 110 grams), although they do not differ across the two groups in a statistical sense.

Alternative controls for seasonal effects – for instance using quarter or month controls interacted with the year – yield treatment effects ranging between minus 70 and minus 103 grams, each of which are estimated less efficiently. Estimating a version of the model using controls for rainfall at the weather station closest to each community yields a treatment effect of minus 110 grams (note, we had to drop 130 observations for which data were unavailable).

Finally, we conducted a placebo test, in which we re-run our main specification with a one- to six- year lead on treatment timing, using sample periods before income management was introduced (2000-2002; 2001-2003; 2002-2004; 2003-2005; 2004-2006; 2005-2007). We would expect to see no 'treatment' effect in these regressions. Table 3 reports the placebo test results. It reveals no statistically significant treatment effects on birthweight or the probability of low birthweight in the years prior to the rollout. This suggests that treatment timing does not simply capture unobserved trends in birth outcomes that were present before the NTER.

[Insert Table 3 here]

6.3 Heterogeneity in the treatment effect

6.3.1. Where in the birthweight distribution did the treatment effect occur?

Our finding that income management significantly reduced average birthweights 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 newborns, which may be the result, for instance, of a decline in the incidence of gestational diabetes. This does not appear to be the case. Quantile regressions results, summarised in Figure 7, show that the treatment effect was statistically significant across the distribution with a treatment effect of minus 100 grams between the 25th and 90th percentile, and largest for newborns with very low birthweight (below 25th percentile) of roughly minus 200 grams.[36]

[Insert Figure 7 here]

6.3.2. Does the treatment effect differ by length of exposure to income management?

The impact of nutrition on birthweight is multidimensional (King 2016). It is therefore possible that newborns received different 'dosages' depending on when income management was introduced relative to their birth. To test for these possibilities, we vary our definition of 'treated' from our main model (where a newborn is considered treated if income management was introduced at or before the beginning of the third trimester of pregnancy). We run regressions defining babies as 'treated' if income management was introduced in or before each of the 40 weeks of pregnancy.[37] Figure 8 shows the estimated treatment effects by week of in utero exposure. We find that the impact of income management on birthweight is largest if introduced in the first trimester of pregnancy. The treatment effect is minus 115 grams and statistically significant at the 5 percent level. As may be expected, infants who were affected by income management only in the final two to three months of the pregnancy did not experience significantly adverse birth outcomes. These results suggest that higher 'dosage' of the policy increased the negative effect of income management.

[Insert Figure 8 here]

[36] 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.
[37] Newborns born before the relevant week are excluded from the sample, which means newborns born at week 38 are not included in the regressions for treatment introduction at week 39.

7 WHY DID INCOME MANAGEMENT LEAD TO LOWER BIRTHWEIGHTS?

Having identified robust causal evidence of a negative treatment effect of income management on birth outcomes, we are left with the question of why the policy worsened outcomes. We explore four different channels that may explain this effect, including changes in fertility and maternal characteristics (Section 7.1.), maternal risky health behaviours (Section 7.2.), or better access to quality care and therefore higher survival probabilities for at-risk babies (Section 7.3.).

7.1 Fertility and Maternal Characteristics

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. It is possible that these women would have chosen to postpone having children during the NTER. However, there is no evidence supporting this hypothesis in our data.

First, we see no impact of the income-management policy on community-level fertility rates. Table 4, Panel A shows that 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 rainfall, year and community fixed effects. 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. Not surprisingly, controlling for maternal characteristics in our benchmark model (Eq. (1)) does not change the treatment effect of income management on birthweight (Appendix Table C.7).

[Insert Table 4 here]

7.2.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, it may have in fact increased maternal risky health behaviours. On the one hand, mothers may have had more resources available for themselves, if the income management policy reduced 'humbugging'. On the other hand, income management could created a new mental anchor on what should be spent on priority goods, with mothers realising that they could spend up to 50 percent of their government transfers on non-priority goods. If so, income management may have led to an increase in drinking and smoking during pregnancy. 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. We have these data available for about half of our sample.[38]

Table 5 shows that the answer to this question is not straightforward. Women in the treatment group were 4.7 percentage points more likely to smoke at the first antenatal visit, but the S.E. is so large that we have no certainty that this effect is not due to random variation (Panel A). They also seem to be slightly less likely to be drinking alcohol by their first antenatal visit, although this effect is very small with a very large S.E. (Panel B). We thus conclude that the evidence in favor of this hypothesis is not strong. It is therefore not surprising that controlling for smoking and drinking behaviour (and other characteristics of

[38] Treatment in this mechanism analysis is defined as to take the value 1 if the first antenatal visit occurred after income management was introduced into the mother's community, and 0 otherwise. Thee sample of pregnancies covered is different from our main analysis sample.

the mother) directly in the regression model does not change the treatment effect of income management on birthweight (see Appendix Table C.8).

[Insert Table 5 here]

7.3. Access to quality 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 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 government agency staff in their community. These staff may have been able to connect pregnant women with health workers, which may have led to more regular antenatal visits, or helped to ensure better access to health care. This could lead to better monitoring of foetal and maternal health, meaning earlier detection of any serious complications. Higher rates of detection of complications may lead to more frequent emergency C-sections, which would mechanically decrease birthweight but would nevertheless be a preferred health outcome. Ultimately, this could have led to a higher survival probabilities of high-risk, poor-health babies that would have died otherwise.

Table 6 demonstrates that treatment and control group babies do not differ in the probability of receiving antenatal care, however, the treatment group received their first ultrasound significantly earlier (by one week on average) than the control group, and were 9.5 percentage points, or 50 percent, more likely to obtain a dating ultrasound. Yet, the treatment group did not have a significantly higher probability of being born in a major hospital or by emergency C-section, which could have affected the timing of delivery and therefore lead to lower birthweights. The treatment group was however more likely to end up in a special care nursery, which provides extra care to babies with low birthweight. Consistent with this finding, the treatment effect as estimated in our benchmark specification (Eq. (1)) remains robust to controlling for measures of antenatal care and care during birth (see Appendix Table C.9).

Furthermore, income management does not appear to have reduced the probability of stillbirth, which could have explained an increase in low birthweight (Table 7). There are 15 babies in our sample period that were stillborn. In fact, babies in the control group were slightly more likely to be stillborn than babies in the control group, although this treatment effect is not statistically significant (Panel A). In line with this finding, the treatment effect, as defined in our benchmark model (Eq. (1)), is also not sensitive to including infants born very prematurely (Panel B), or to changing our definition of 'low birthweight' (less than 2500 grams) to 'very low birthweight' (less than 1500 grams, Panel C).

[Insert Table 7 here]

8. Discussion and Conclusion

Australia has embarked in the past ten years on a dramatic reform of its welfare system. As one of the very few countries in the OECD, it has opted to restrict individual choice on how welfare recipients in disadvantaged communities can spend their governmental transfers. The initial rollout of the paternalistic 'income management' policy aimed to improve Aboriginal children's welfare. Although the welfare of children took centre stage in the debate over the income management policy and the government's strong commitment to trial such policies all over Australia, little empirical evidence has been brought to bear on the impact of restricting welfare payments on children's outcomes. This study is one of the first attempts to provide sound empirical evidence on the policy's impact on Aboriginal children's welfare. Our findings suggest that the policy did not improve birth outcomes, a key policy parameter in Aboriginal communities. This means that the income management policy either did not produce the desired changes in household consumption patterns, or it entailed unintended consequences which offset any beneficial outcomes.

The treatment effect of a reduction in birthweight of 100 grams is sizable. Yet, it lies within the range of variation observed the Northern Territory and in the international literature. Although average birthweights have hardly improved in Aboriginal communities in the Northern Territory over the years, seasonal variation within any given year is 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 – but it occurs over and above this seasonal variation.

Randomised trials that provide food or fortified food products during pregnancy have been shown to increase birthweight by 125 grams on average (Gresham et al. 2014). Almond, Hoynes, and Schanzenbach (2011) found that food stamps – a program that provided a relatively low dose of additional resources for families – increased 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 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 several channels through which income management may have led to worse birth outcome measurements. First, income management did not reduce average birthweight by bringing heavier infants down to a healthy birthweight - instead, the impact was largest at the lower end of the distribution, making low birthweight babies even lighter. Second, it 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. Third, it also did not change risky health behaviours (smoking, drinking) of mothers during the early stages of pregnancy. Finally, it did not change access to health care services in a way that could have led to earlier detection of health problems, or greater survival probabilities of very unhealthy babies, and therefore to worse recorded birth outcomes.

Ruling out the most obvious explanations, we can only speculate on the reasons for our findings. One the one hand, 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. On the other hand, the policy may have created a low mental anchor. The restriction to spend at least half of welfare income on household essentials may have caused households to reduce consumption of essential goods.

Another explanation for our negative treatment effects may be the poor administration of the rollout of the income management policy, causing 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 caused by income management and the NTER more generally. The introduction of income management is widely reported to have led to a sense of 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 Part II of the *Racial Discrimination Act 1975* a contributing factor to this perception (Australian Institute of Health and Welfare 2010). General negative attitudes towards the NTER may have contributed to reduced willingness to comply with the restrictions.

Furthermore, income management may have led to an increase in food prices and thus to lower levels of consumption. 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 (Northern Territory Department of Health 2017), an evaluation report of New Income Management in the NT suggested that prices may have increased (Bray et al. 2014).

Finally, income management may have changed the way households allocated resources within the household. 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 expectant mothers because they were no longer in full control of all of their available financial resources. In our own previous work, found that although money worries did not increase significantly in Aboriginal households that moved from unrestricted to restricted welfare payments, the probabilities of experiencing excessive demands of money (humbugging) and arguments (that affected children) doubled (Cobb-Clark et al. 2018).

Stress experienced by pregnant women can affect the development of the fetus, 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, Devereux, and Salvanes 2016; Persson and Rossin-Slater 2018), 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 2013). 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 through which economic disadvantage affects birth outcomes (Aizer and Currie 2014) – then this could help to explain our adverse treatment effects.

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. Any positive effect would need to be quite large to justify the high cost of administering the policy. The recent Cashless Debit Card trial, which is an extension of the original income management policy and which is currently trial in various locations across Australia, costs around $9,000 per recipient per year, including setup costs, but excluding payments (Department of Social Services 2017, Orima Research 2017). 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.

REFERENCES

Aizer, Anna, and Janet Currie. 2014. 'The Intergenerational Transmission of Inequality: Maternal Disadvantage and Health at Birth'. *Science* 344 (6186): 856–61. https://doi.org/10.1126/science.1251872.

Almond, Douglas, Kenneth Y. Chay, and David S. Lee. 2005. 'The Costs of Low Birth Weight'. *The Quarterly Journal of Economics* 120 (3): 1031–83.

Almond, Douglas, Janet Currie, and Katherine Meckel. 2014. 'Fetal Origins of Lifetime Health'. In *Encyclopedia of Health Economics*, edited by Anthony J Culyer, 1st ed., 309–14. Elsevier.

Almond, Douglas, Hilary W. Hoynes, and Diane Whitmore Schanzenbach. 2011. 'Inside the War on Poverty: The Impact of Food Stamps on Birth Outcomes'. *Review of Economics and Statistics* 93 (2): 387–403. https://doi.org/10.1162/REST_a_00089.

Amarante, Verónica, Marco Manacorda, Edward Miguel, and Andrea 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. https://doi.org/10.1257/pol.20140344.

Australian Human Rights Commission. 2011. 'The Suspension and Reinstatement of the RDA and Special Measures in the NTER'. 2 November 2011. https://www.humanrights.gov.au/our-work/suspension-and-reinstatement-rda-and-special-measures-nter-0.

Australian Institute of Aboriginal and Torres Strait Islander Studies. 2012. 'Guidelines for Ethical Research in Australian Indigenous Studies'. https://aiatsis.gov.au/sites/default/files/docs/research-and-guides/ethics/gerais.pdf.

Australian Institute of Health and Welfare, ed. 2010. *Evaluation of Income Management in the Northern Territory*. Occasional Paper / Department of Families, Housing, Community Services and Indigenous Affairs 34. Canberra.

Australian National Audit Office. 2018. 'The Implementation and Performance of the Cashless Debit Card Trial'. Performance Audit 1 2018-19. Canberra, Australia: The Auditor General. https://www.anao.gov.au/work/performance-audit/implementation-and-performance-cashless-debit-card-trial.

Barber, Sarah L., and Paul J. Gertler. 2008. 'The Impact of Mexico's Conditional Cash Transfer Programme, Oportunidades, on Birthweight'. *Tropical Medicine & International Health* 13 (11): 1405–14. https://doi.org/10.1111/j.1365-3156.2008.02157.x.

Beatty, Timothy K. M., and Charlotte 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. https://doi.org/10.1093/ajae/aau097.

Bitler, Marianne P., and Janet 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, 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. https://doi.org/10.1257/app.20140262.

Bray, J. Rob, Matthew Gray, Kelly Hand, and Ilan Katz. 2014. 'Evaluating New Income Management in the Northern Territory: Final Evaluation Report'. Sydney, Australia: Social Policy Research Centre. https://www.dss.gov.au/our-responsibilities/families-and-children/programmes-services/income-management/income-management-evaluations/evaluation-of-new-income-management-in-the-northern-territory.

Breunig, Robert, and Indraneel Dasgupta. 2005. 'Do Intra-Household Effects Generate the Food Stamp Cash-Out Puzzle?' *American Journal of Agricultural Economics* 87 (February): 552–68. https://doi.org/10.1111/j.1467-8276.2005.00747.x.

Brimblecombe, Julie K., Joseph McDonnell, Adam Barnes, Joanne Garnggulkpuy Dhurrkay, David P. Thomas, and Ross S. Bailie. 2010. 'Impact of Income Management on Store Sales in the Northern Territory'. *The Medical Journal of Australia* 192 (10): 549–54. https://doi.org/10.5694/j.1326-5377.2010.tb03632.x.

Brough, Mal. 2007. 'Social Security And Other Legislation Amendment (Welfare Payment Reform) Bill 2007 – Second Reading Speech'. presented at the Australian House of Representatives, Canberra, Australia, August 7. https://formerministers.dss.gov.au/2921/social-security-and-other-legislation-amendment-welfare-payment-reform-bill-2007-second-reading-speech/.

Buckmaster, Luke, and Carol Ey. 2012. 'Is Income Management Working?' Background note. Canberra, Australia: Australian Parliamentary Library. https://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/pubs/BN/2011-2012/IncomeManagement.

Cobb-Clark, Deborah, Nathan Kettlewell, Stefanie Schurer, and Sven Silburn. 2018. 'The Effect of Quarantining Welfare on School Attendance in Indigenous Communities'. https://www.lifecoursecentre.org.au/research/journal-articles/working-paper-series/the-effect-of-quarantining-welfare-on-school-attendance-in-indigenous-communities/.

Cuffey, Joel, Timothy KM Beatty, and Lisa Harnack. 2016. 'The Potential Impact of Supplemental Nutrition Assistance Program (SNAP) Restrictions on Expenditures: A Systematic Review'. *Public Health Nutrition* 19 (17): 3216–31. https://doi.org/10.1017/S1368980015003511.

Cunha, Jesse M. 2014. 'Testing Paternalism: Cash versus In-Kind Transfers'. *American Economic Journal: Applied Economics* 6 (2): 195–230. https://doi.org/10.1257/app.6.2.195.

Currie, Janet, and Douglas Almond. 2011. 'Chapter 15 - Human Capital Development before Age Five'. In *Handbook of Labor Economics*, edited by David Card and Orley Ashenfelter, 4:1315–1486. Elsevier. https://doi.org/10.1016/S0169-7218(11)02413-0.

Currie, Janet, and Firouz Gahvari. 2008. 'Transfers in Cash and In-Kind: Theory Meets the Data'. *Journal of Economic Literature* 46 (2): 333–83.

Currie, Janet, and Enrico Moretti. 2007. 'Biology as Destiny? Short- and Long- Run Determinants of Intergenerational Transmission of Birth Weight'. *Journal of Labor Economics* 25 (2): 231–64. https://doi.org/10.1086/511377.

Currie, Janet, and Maya Rossin-Slater. 2013. 'Weathering the Storm: Hurricanes and Birth Outcomes'. *Journal of Health Economics* 32 (3): 487–503. https://doi.org/10.1016/j.jhealeco.2013.01.004.

Das, Jishnu, Quy-Toan Do, and Berk Özler. 2005. 'Reassessing Conditional Cash Transfer Programs'. *The World Bank Research Observer* 20 (1): 57–80. https://doi.org/10.1093/wbro/lki005.

Davey, Calum, James Hargreaves, Jennifer A Thompson, Andrew J Copas, Emma Beard, James J Lewis, and Katherine 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). https://doi.org/DOI 10.1186/s13063-015-0838-3.

Department of Families, Housing, Community Services and Indigenous Affairs. 2011. 'Northern Territory Emergency Response: Evaluation Report'. http://webarchive.nla.gov.au/gov/20130329182840/http://www.fahcsia.gov.au/our-responsibilities/indigenous-australians/publications-articles/northern-territory-emergency-response-evaluation-report-2011.

Department of Social Services. 2017. 'Freedom of Information Request Number 16/17-123: Costs of the Cashless Debit Card Trials in Ceduna and Kununurra.' Australian Government. https://www.dss.gov.au/sites/default/files/documents/04_2017/attachment_a_cashless_debit_card_trials_costs.pdf.

Figlio, David, Jonathan Guryan, Krzysztof Karbownik, and Jeffrey Roth. 2014. 'The Effects of Poor Neonatal Health on Children's Cognitive Development'. *American Economic Review* 104 (12): 3921–55. https://doi.org/DOI: 10.1257/aer.104.12.3921.

Fletcher, Jason M. 2011. 'The Medium Term Schooling and Health Effects of Low Birth Weight: Evidence from Siblings'. *Economics of Education Review* 30 (3): 517–27. https://doi.org/10.1016/j.econedurev.2010.12.012.

Gentilini, Ugo. 2016. 'Revisiting the "Cash versus Food" Debate: New Evidence for an Old Puzzle?' *The World Bank Research Observer* 31 (1): 135–67. https://doi.org/10.1093/wbro/lkv012.

Gresham, Ellie, Julie E Byles, Alessandra Bisquera, and Alexis J Hure. 2014. 'Effects of Dietary Interventions on Neonatal and Infant Outcomes: A Systematic Review and Meta-Analysis'. *The American Journal of Clinical Nutrition* 100 (5): 1298–1321. https://doi.org/10.3945/ajcn.113.080655.

Grieger, Jessica A, and Vicki L Clifton. 2015. 'A Review of the Impact of Dietary Intakes in Human Pregnancy on Infant Birthweight'. *Nutrients* 7: 153–78. https://doi.org/10.3390/nu7010153.

Hand, Kelly, Ilan Katz, Matthew Gray, and J. Rob Bray. 2016. 'Welfare Conditionality as a Child Protection Tool'. *Family Matters*, no. 97 (March). https://aifs.gov.au/publications/family-matters/issue-97/welfare-conditionality-child-protection-tool.

Hemming, Karla, Monica Taljaard, and Andrew Forbes. 2017. 'Analysis of Cluster Randomised Stepped Wedge Trials with Repeated Cross-Sectional Samples'. *Trials* 18 (101). https://doi.org/10.1186/s13063-017-1833-7.

Hoynes, Hilary W., Marianne Page, and Ann Huff Stevens. 2011. 'Can Targeted Transfers Improve Birth Outcomes? Evidence from the Introduction of the WIC Program'. *Journal of Public Economics* 95 (7–8): 813–27. https://doi.org/10.1016/j.jpubeco.2010.12.006.

Hoynes, Hilary W., and Diane Whitmore Schanzenbach. 2009. 'Consumption Responses to In-Kind Transfers: Evidence from the Introduction of the Food Stamp Program'. *American Economic Journal: Applied Economics* 1 (4): 109–39. https://doi.org/10.1257/app.1.4.109.

King, Janet C. 2016. 'A Summary of Pathways or Mechanisms Linking Preconception Maternal Nutrition with Birth Outcomes'. *The Journal of Nutrition* 146 (7): 1437S-1444S. https://doi.org/10.3945/jn.115.223479.

Kramer, M S. 1987. 'Determinants of Low Birth Weight: Methodological Assessment and Meta-Analysis'. *Bulletin of the World Health Organization* 65 (5): 663–737.

Lamb, David, and Martin Young. 2011. ''Pushing Buttons': An Evaluation of the Effect of Aboriginal Income Management on Commercial Gambling Expenditure'. *Australian Journal of Social Issues* 46 (2): 119–40.

Lauderdale, Diane S. 2006. 'Birth Outcomes for Arabic-Named Women in California before and after September 11'. *Demography* 43 (1): 185–201.

Northern Territory Department of Health. 2017. 'Northern Territory Market Basket Survey 2016'. Darwin: Northern Territory Government. https://digitallibrary.health.nt.gov.au/prodjspui/bitstream/10137/1428/1/2016%20MBS%20Full%20Report.pdf.

Orima Research. 2017. 'Cashless Debit Card Trial Evaluation: Final Evaluation Report'. https://www.dss.gov.au/sites/default/files/documents/10_2018/cashless-debit-card-trial-final-evaluation-report_2.pdf.

Persson, Petra, and Maya Rossin-Slater. 2018. 'Family Ruptures, Stress, and the Mental Health of the Next Generation'. *American Economic Review* 108 (4–5): 1214–52.

Savitri, AI, N Yadegari, J Bakker, RJ van Ewijk, DE Grobbee, RC Painter, CS Uiterwaal, and TJ Roseboom. 2014. 'Ramadan Fasting and Newborn's Birth Weight in Pregnant Muslim Women in The Netherlands'. *British Journal of Nutrition* 112 (9): 1503–9. https://doi.org/10.1017/S0007114514002219.

Senauer, Ben, and Nathan Young. 1986. 'The Impact of Food Stamps on Food Expenditures: Rejection of the Traditional Model'. *American Journal of Agricultural Economics* 68 (1): 37–43.

Silburn, Sven, Steven Guthridge, John McKenzie, Jiunn-Yih Su, Vincent He, and Sharon Haste, eds. 2018. *Early Pathways to School Learning: Lessons from the NT Data Linkage Study*. Darwin: Menzies School of Health Research. https://www.nintione.com.au/?p=13869.

Silburn, Sven R., Stephen R. Zubrick, David M. Lawrence, Francis G. Mitrou, John A. De Maio, Eve Blair, Adele Cox, Robin B. Dalby, Judith A. Griffin, Glenn Pearson, and Colleen Hayward (2006). The intergenerational effects of forced separation on the social and emotional wellbeing of Aboriginal children and young people. Australian Institute of Family Studies. Family Matters 2006 No. 75. 10-17.

Singh, Gurmeet, and Wendy 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): 532–35.

Southworth, Herman M. 1945. 'The Economics of Public Measures to Subsidize Food Consumption'. *American Journal of Agricultural Economics* 27 (1): 38–66. https://doi.org/10.2307/1232262.

Stein, Zena, and Mervyn Susser. 1975. 'The Dutch Famine, 1944–1945, and the Reproductive Process. I. Effects on Six Indices at Birth'. *Pediatric Research* 9: 70–76.

Su, Jiunn-Yih, Sven Silburn, Stefanie Schurer, Steven Guthridge, Vincent He, and John McKenzie. 2018. 'Early Life Health and Development'. In *Early Pathways to School Learning: Lessons from the NT Data Linkage Study*, edited by Sven Silburn, Steven Guthridge, John McKenzie, Jiunn-Yih Su, Vincent He, and Sharon Haste, 29–61. Darwin: Menzies School of Health Research.

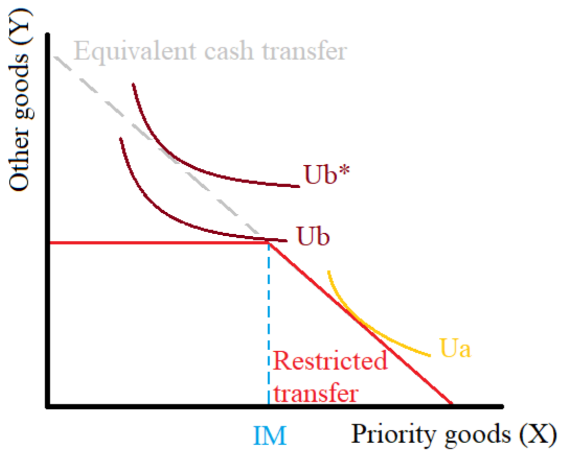
Victora, CG, L Adair, C Fall, PC Hallal, R Martorell, L Richter, HS Sachdev, and Maternal and Child Undernutrition Study Group. 2008. 'Maternal and Child Undernutrition: Consequences for Adult Health and Human Capital'. *Lancet* 371 (9606): 340–57.

Wild, Rex, and Pat Anderson. 2007. 'Ampe Akelyernemane Meke Mekarle "Little Children Are Sacred"'. Northern Territory Board of Inquiry into the Protection of Aboriginal Children from Sexual Abuse. Darwin: Northern Territory Government. http://www.inquirysaac.nt.gov.au/pdf/bipacsa_final_report.pdf.

Wilde, Parke, and Christine Ranney. 1996. 'The Distinct Impact of Food Stamps on Food Spending'. *Journal of Agricultural and Resource Economics* 21 (1): 174–85.

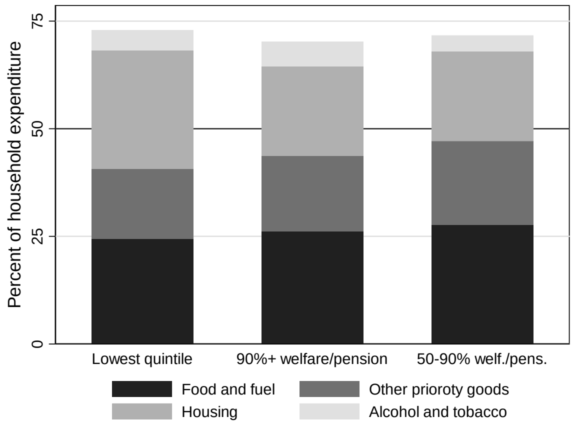
FIGURES

Figure 1: Effect of restricting welfare payments for inframarginal and extramarginal households



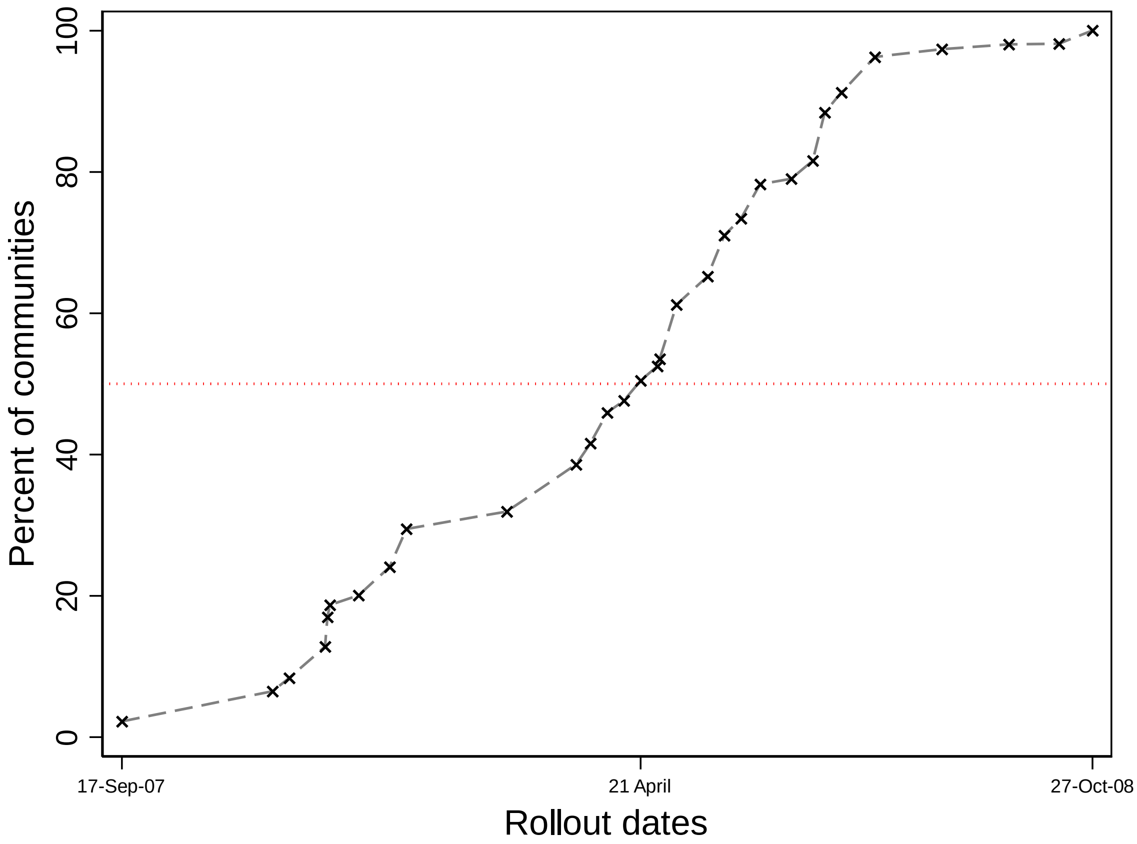
Note: Figure adapated from Southworth (1945). Ub indicates utility from extramarginal households, whose behavior is changed by the income management policy. Ua indicates utility of inframarginal households whose spending patterns are unchanged by the policy.

Figure 2. Household expenditures in the Northern Territory



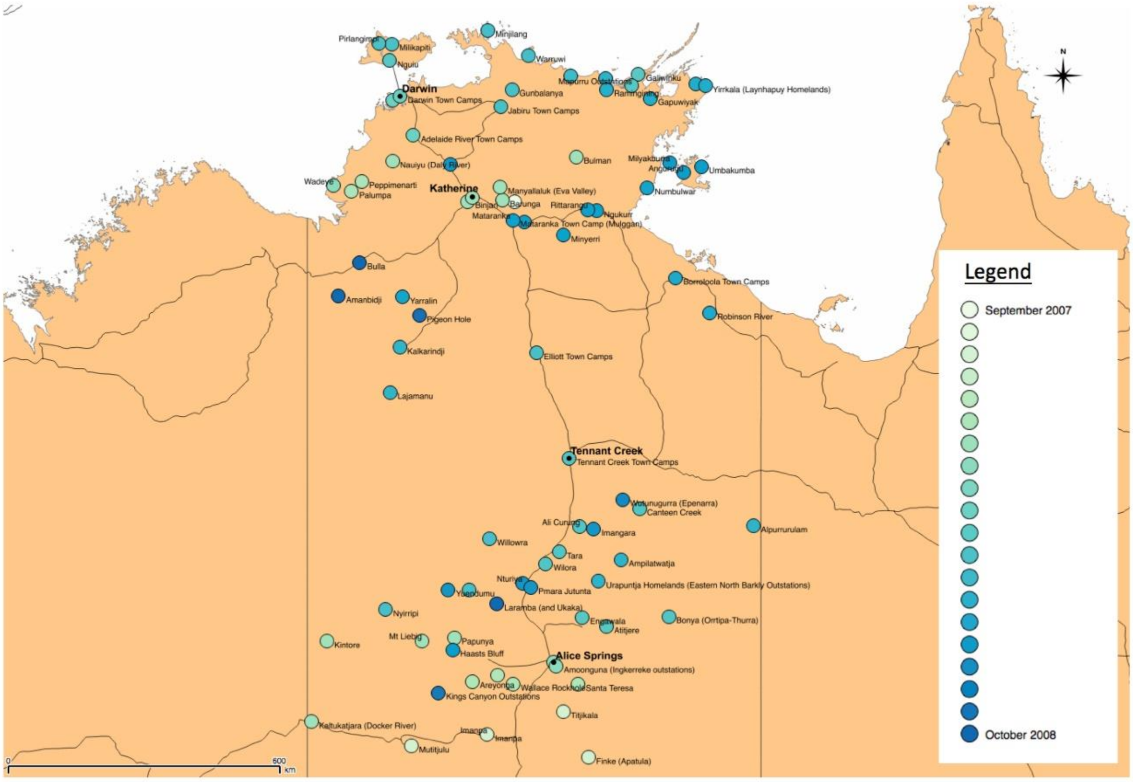
Note: The figure shows spending on food, fuel, housing and 'other priority goods', and spending on alcohol and tobacco, as a share of total household expenditures. The data do not relate specifically to residents of remote or Aboriginal communities, but give estimates of expenditures by other low-income, welfare-dependent households in the NT.
Lowest quintile refers to households in the lowest 20 per cent of the distribution of household income in the Northern Territory; 90%+ welfare pension refers to households where 90% or more of all household income comes from welfare payments or pensions. 50-90%+ welf./pens. refers to households where 50-90% of all income comes from welfare payments or pensions. 'Other priority goods' includes clothing and footwear, household furnishing, medical care expenses, and transport. The income management policy legislates that at least 50% of household income must be spent on 'priority' goods.
Source: Australian Bureau of Statistics *Household Expenditure Survey* 2003-04.

Figure 3. Rollout of the Income Management Policy



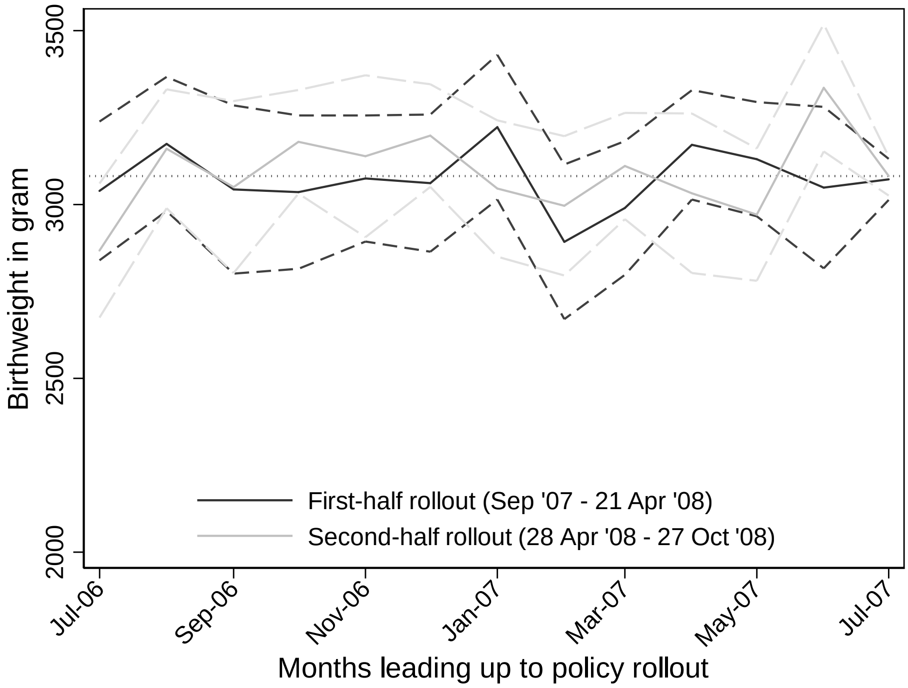
Note: The graph shows the cumulative share of NTER communities that were covered by income management on each of 31 implementations dates (indicated by crosses). Each cross includes one or more community, and data are weighted by number of births in each community. The rollout commenced on 17 Septemer 2007 and was completed by 27 October 2008. By 21 April 2008, 50% of all communities had received income management. For full details on the rollout schedule, see AIHW (2010).

Figure 4. Geographic pattern of the income management rollouts



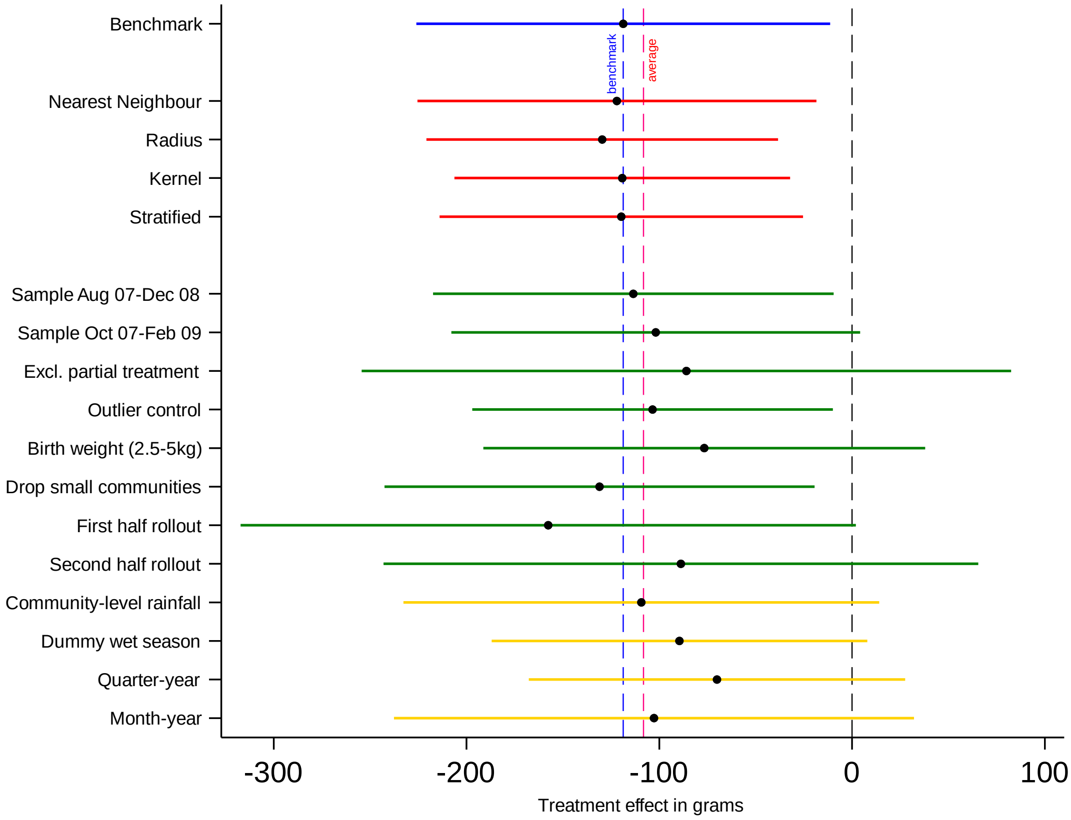
Note: Color coding on communities covered by income management reflects the date the polciy started in the relevant community, as indicated in the Legend. Major cities or towns in the Northern Territory are displayed in bold font. People living in the municipal parts of those towns were not subject to IM, but people living in associated town camps were. Lines connecting communities represent highways and arterial roads.

Figure 5. Birthweight trends by timing of the rollout



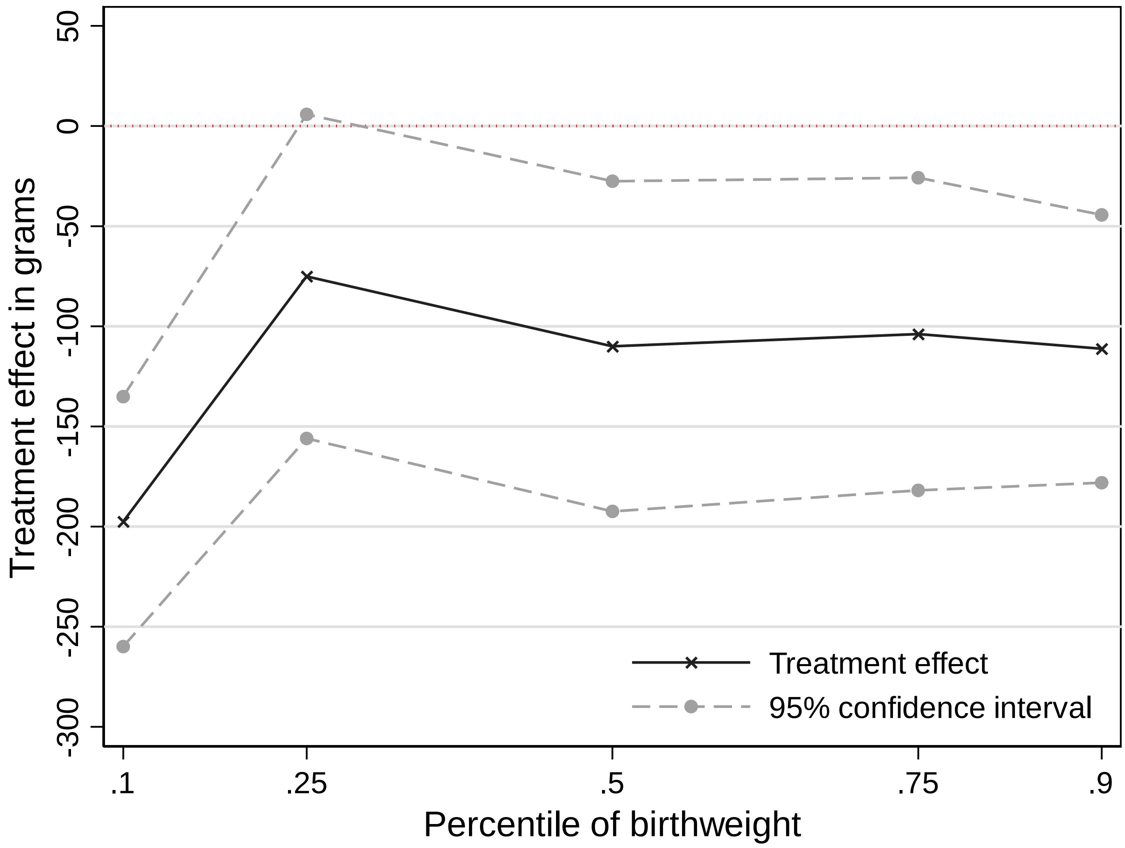
Note: The graphs displays the unadjusted mean birthweight in NTER communities in each month leading up to the Income management policy, separately for communities that received IM in the first half of the rollout (weighted by the number of births) and the half of all communities. Dashed lines indicated 95% confidence intervals.

Figure 6. Robustness checks on benchmark average treatment effect



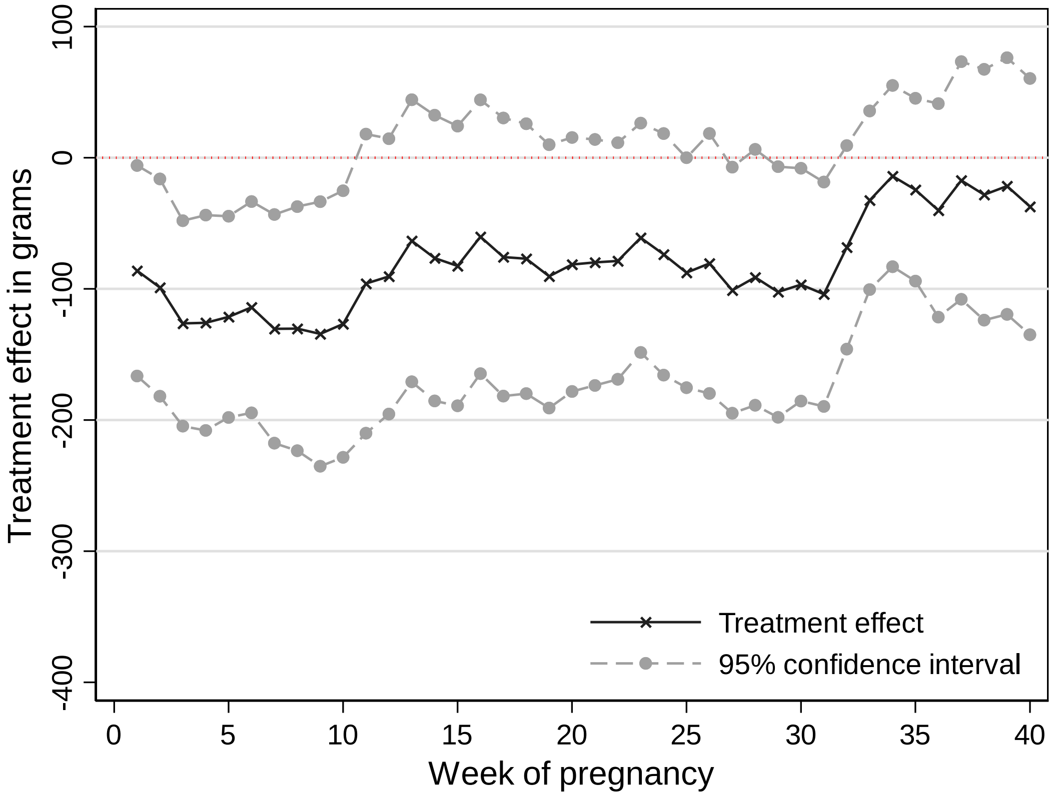
Note: The figure depicts the treatment effect of income management on average birthweight obtained from our benchmark model (Eq. (1)), which controls for premature birth, community fixed effects, year of birth and rainfall. Each row is the estimated treatment effect (dot point) obtained from a separate regression model (results tables available in Appendix C). Vertical lines represent 95 confidence intervals.

Figure 7. Treatment effect across the birthweight distribution



Note: Estimated coefficients obtained from a quantile regression model, where conditional treatment effect is estimated at the 10th, 25th, 50th, 75th, and 90th percentile of the birthweight distribution.

Figure 8. Treatment effect by timing of the introduction of the income management policy



Note: This graph depicts the treatment effect of the income management policy on birthweight, by the timing of the policy introduction relative to pregnancy. Each cross represents the treatment effect from a different regression, in which infants are considered 'treated' if the policy was introduced in or before that week or pregnancy. In our main results, infants are considered 'treated' if the policy was introduced in or before the 28^{th} week of pregnancy (beginning of third trimester). The sample in each regression is selected relative to the treatment definition (for instance, in the regression where treatment occurs in week 1 of pregnancy, the sample includes births occurring 9 months after the rollout was complete).

8 TABLES

Table 1 – Pre-rollout outcomes and community characteristics
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)	3072	3082	9.89	3354
	(30)	(29)	(41.67)	(11)
Low birthweight (%)	14.44	13.95	-0.49	6.83
	(1.62)	(1.61)	(2.28)	(0.48)
Obstetric complications				
Premature (%)	15.5	15.02	-0.48	7.91
	(1.67)	(1.66)	(2.35)	(0.51)
Due to intrauterine growth restriction (%)	5.1	3.22	-1.88	1.44
	(1.01)	(0.82)	(1.3)	(0.23)
Due to anaemia (%)	9.55	9.87	0.317	2.01
	(1.35)	(1.38)	(1.94)	(0.27)
Due to gestational diabetes (%)	7.22	9.01	1.79	6.83
	(1.19)	(1.33)	(1.78)	(0.48)
Any complication (%)	43.52	45.92	2.4	24.19
	(2.28)	(2.31)	(3.25)	(0.81)
Other characteristics				
Age of mother	23.88	23.74	-0.14	28.56
	(0.28)	(0.28)	(0.4)	(0.12)
Aboriginal mother (%)	91.72	94.21	2.49	21.28
	(1.27)	(1.08)	(1.67)	(0.78)
APGAR 1	8.04	8.03	-0.01	8.17
	(0.08)	(0.09)	(0.12)	(0.03)
APGAR 5	8.79	8.88	0.09	8.97
	(0.07)	(0.07)	(0.1)	(0.02)
Community characteristics[(a)]				
Community size	388.84	486.45	97.61	na
	(55.96)	(67.09)	(88.77)	
Female share of population (%)	50.83	50.96	0.13	48.49
	(0.6)	(0.75)	(0.98)	
Median age	22.8077	22.1429	-0.67	31
	(0.38)	(0.41)	(0.58)	
Population aged 65+ (%)	3.3647	3.3821	0.02	4.8
	(0.25)	(0.25)	(0.37)	
People per household	5.3947	6.5294	1.14	2.9
	(0.23)	(0.22)	(0.32)	

Median personal income	214.62 (10.45)	206.61 (3.1)	-8.01 (10.77)	549
Median household income	727.43 (47.93)	874.00 (42.64)	146.60 (65.17)	1192
Median rent payments	43.91 (3.17)	42.21 (6.15)	-1.69 (7.22)	140
Labour force share of population (%)	39.85 (3.24)	36.43 (2.81)	-3.42 (4.36)	47.27

Standard errors in parentheses. First half of rollout defined as communities where income management was introduced from 17 September 2007 to 21 April 2008, second half defined as communities where income management was introduced from 28 April 2008 to 27 October 2008. APGAR stands for Appearance, Pulse, Grimace, Activity and Respiration and is measured at 1 minute and 5 minutes after birth. Each of the five categories is scored 0, 1 or 2, for a maximum total score of 10.
(a) Community characteristics are from the Australian Bureau of Statistics 2006 Census community profile data; most variables available for 54 NTER communities; rest of NT Census data is the average of all of NT.

Table 2 – Impact of income management on birthweight and probability of low birthweight

Regression results

	(1)	(2)	(3)	(4)
Panel A: Birthweight (grams, OLS)				
Income management	-60.64	-119.54	-163.86	-118.67
	(35.95)	(44.77)	(57.58)	(54.76)
Rainfall in 3 months to birth (mm)		-0.18	-0.10	-0.09
		(0.05)	(0.06)	(0.05)
Year (base category = 2007)				
2008		70.94	91.16	53.14
		(52.96)	(56.28)	(52.43)
2009		160.85	169.80	64.08
		(97.14)	(97.96)	(94.39)
Premature				-932.92
				(54.42)
Constant	3,161.44	3,175.29	3,458.45	3,464.62
	(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	0.00	0.01	0.09	0.34
Panel B: Probability of low birthweight (probit, average marginal effects)				
Income management	0.028	0.060	0.081	0.048
	(0.017)	(0.021)	(0.027)	(0.022)
Rainfall in 3 months to birth (mm)		0.000	0.000	0.000
		(0.000)	(0.000)	(0.000)
Year (base category = 2007)				
2008		-0.039	-0.054	-0.021
		(0.035)	(0.045)	(0.038)
2009		-0.077	-0.099	-0.028
		(0.036)	(0.045)	(0.043)
Premature				0.591
				(0.040)
Community fixed effects	No	No	Yes	Yes
Number of communities	83	83	50	50
Observations	1,153	1,153	991	991

Community-clustered standard errors in parentheses.
Note: Low birthweight is defined as below 2500 grams.

Table 3 – Placebo test

Years lead	Date range			Birthweight	Low birthweight
Actual sample	17-Sep-07	to	31-Jan-09	-118.7	0.0480
				(54.76)	(0.0220)
1	17-Sep-05	to	31-Jan-07	-35.56	0.0249
				(43.93)	(0.0268)
2	17-Sep-04	to	31-Jan-06	-6.786	-0.0241
				(37.43)	(0.0339)
3	17-Sep-03	to	31-Jan-05	-60.19	0.0228
				(52.47)	(0.0334)
4	17-Sep-02	to	31-Jan-04	-18.43	0.0258
				(37.36)	(0.0318)
5	17-Sep-01	to	31-Jan-03	54.07	-0.00782
				(46.02)	(0.0287)
6	17-Sep-00	to	31-Jan-02	4.357	0.00935
				(43.90)	(0.0227)

Community-clustered standard errors in parentheses.
Note: Regressions include controls for year, rainfall in 3 months to birth, community fixed effects, and prematurity. Low birthweight column shows average marginal effects from a probit model.

Table 4 – Fertility and mother's medical history

	Treatment average	Control average	Difference (treatment - control): No covariates	Difference (treatment - control): With covariates[(a)]	Observations
Panel A: Fertility rate in NTER communities [(b)]					
Births per 1000 residents, per quarter	8.88	10.09	-1.19 (0.94)	-0.23 (1.13)	337
Births per 1000 women, per quarter	15.48	15.61	-0.26 (2.91)	-0.51 (1.47)	297
Panel B: Previous pregnancies					
First pregnancy (%)	25.43	26.68	-0.475 (2.72)	1.14 (4.61)	1153
Number of pregnancies carried to >20 weeks	1.70	1.60	0.0764 (0.1)	0.00 (0.14)	1153
Total number of previous pregnancies	2.97	2.86	0.0758 (0.11)	-0.05 (0.18)	1153
Panel C: Mother's history of medical complications					
Anaemia (%)	6.88	6.41	0.46 (1.54)	-0.76 (2.61)	1057
Cardiac disease (%)	8.13	6.76	1.37 (1.62)	-3.14 (2.32)	1057
Epilepsy (%)	1.25	1.39	-0.14 (0.71)	0.61 (1.03)	1057
Pre-existing hypertension (%)	1.25	1.04	0.21 (0.66)	0.89 (1.36)	1057
Pre-existing diabetes (%)	4.58	3.12	1.46 (1.18)	1.67 (1.58)	1057
Gestational diabetes (%)	3.13	3.47	-0.34 (1.11)	-0.37 (1.73)	1057
Renal disease (%)	3.96	3.12	0.84 (1.14)	1.05 (1.68)	1057
Syphilis (%)	5.21	3.81	1.40 (1.27)	3.59 (1.81)	1057
Urinary tract infection (%)	6.04	3.29	2.75 (1.28)	3.70 (2.20)	1057
Any complication (%)	57.29	54.59	2.70 (3.07)	-2.17 (4.45)	1057

Number of previous complications	0.81	0.72	0.09 (0.05)	0.03 (0.09)	1057
Medical history unknown (%)	8.92	7.83	1.09 (1.63)	-0.39 (2.05)	1153

Community-clustered standard errors in parentheses.
(a) Covariates are: community fixed effects, rainfall in 3 months to birth, and year of birth.
(b) Regression conducted on data averaged at the community-quarter level. A cell is defined as 'treated' if the majority of infants born in that community during that quarter were treated (i.e., income management was introduced before the third trimester of pregnancy).

Table 5 – Impact of Income Management on Smoking and Drinking Behaviour at First Antenatal Visit

	Treatment	Control	Difference (treatment - control)		
			No covariates	With covariates[a]	Pr missing
Panel A: Smoking					
Smoking at first antenatal visit	49.36	51.28	-1.91 (4.37)	4.70 (5.00)	4.93 (2.68)
Observations	393	353	746	707	955
Panel B: Drinking					
Drinking at first antenatal visit	12.68	10.08	2.60 (2.24)	-1.22 (3.79)	4.06 (2.39)
Observations	426	377	803	586	955

The ‘difference’ columns show probit average marginal effects, with community-clustered robust standard errors in parentheses.
In this table, treatment timing is defined relative to the date of the first antenatal visit, not relative to the date of birth (as in the main analysis). An observation is defined as 'treated' if the first antenatal visit occurred on or after the date that income management was introduced in the mother's community. The sample is limited to births for which the first antenatal visit occurred during the rollout period (17 September 2007 to 27 October 2008).

(a) Covariates are: rainfall in 3 months to birth, year, and community-level fixed effects. The sample size is lower in this column because the probit model drops communities (and the observations contained within those communities) if there is no variation in the outcome measure within that community.

Table 6 – Indicators of antenatal care and hospital care

	Level		Difference (treatment - control)		Observations	
	Treatment	Control	No covariates	With covariates[(a)]	N	Pr missing (T-C)
Panel A: Antenatal care						
Number of antenatal visits	8.66	9.05	-0.39 (0.30)	0.00 (0.40)	1139	-0.84 (0.65)
Gestational age at first visit (weeks)	16.18	16.44	-0.26 (0.51)	-1.41 (0.67)	1129	-1.04 (0.84)
Gestational age at first ultrasound (weeks)	19.56	20.63	-1.07 (0.50)	-1.52 (0.65)	1091	-0.12 (1.33)
Had a dating ultrasound (%)[(b)]	23.00	19.00	3.98 (2.52)	9.48 (3.43)	1091	-0.12 (1.33)
Panel B: Hospital care						
Born in main hospital (%)[(c)]	72.00	65.00	6.77 (2.74)	0.40 (2.95)	1153	na
Emergency delivery (%)	20.00	19.00	1.45 (2.34)	2.12 (3.09)	1153	na
Length of stay in hospital (days)	6.83	6.29	0.54 (0.63)	1.20 (0.8)	1137	0.01 (0.69)
Admitted to special care nursery (%)	25.00	21.00	4.31 (2.49)	5.24 (3.25)	1152	-0.16 (0.17)

Standard errors in parentheses.
(a) Covariates are: community fixed effects, year of birth and rainfall in 3 months to birth.
(b) Date of first ultrasound was during weeks 6-13.
(c) Indicates the infant was born in one of the NT's three largest hospitals: Royal Darwin Hospital, Darwin Private or Alice Springs Hospital.

Table 7: Impact of income management on mortality and very low birthweight

	(1)	(2)	(3)	(4)
Panel A: Linear probability models on Pr(stillbirth)				
Income management	0.011	0.004	0.015	0.001
	(0.007)	(0.005)	(0.01)	(0.007)
Born before 3rd trimester		0.36		0.359
		(0.106)		(0.109)
Rainfall in 3 months to birth			7.89e-06	-2.84e-06
			(1.38e-05)	(9.20e-06)
Year				
2008			-0.014	-0.003
			(0.014)	(0.003)
2009			0.004	0.025
			(-0.004)	(-0.025)
Fixed effects	no	no	yes	yes
Sample size	1,172	1,172	1,172	1,172
Probability of still birth in sample	0.0128			
Panel B: OLS regression on birthweight (grams)				
Income management			-166.9	-116.7
			(62.40)	(54.62)
Born before 3rd trimester				-1,736
				(94.87)
Rainfall in 3 months to birth			-0.124	-0.079
			(0.058)	(0.052)
Year				
2008			92.56	49.97
			(62.50)	(52.50)
2009			132.2	59.76
			(109.2)	(93.73)
Premature			-1,120	-933.8
			(57.25)	(54.07)
Fixed effects			yes	yes
Sample size			1,172	1,172
Panel C: Linear probability model on Pr(very low birthweight (<1500))				
Income management	0.030	0.049	0.057	0.042
	(0.01)	(0.011)	(0.013)	(0.013)
Rainfall in 3 months to birth		0.000	0.000	0.000
		(0)	(0)	(0)
Year				
2008		-0.045	-0.052	-0.039
		(0.017)	(0.018)	(0.016)
2009		-0.047	-0.058	-0.026

		(0.034)	(0.038)	(0.033)
Fixed effects	no	no	yes	yes
Sample size	1,172	1,172	1,172	1,172

Note: Community-clustered standard errors in parentheses. The sample in this table includes 19 infants born before the beginning of the third trimester of pregnancy, who are dropped from main analysis.

APPENDIX A: OTHER NTER MEASURES

Table A.1 Other NTER policies: cumulative number of communities covered by each policy, July 2007-July 2008

Policy measure [a]	Jul-Sept 07	Oct-Dec-07	Jan-Mar 08	Apr-Jul 08	Target communities
Income management	4	23	33	78	83
Child health checks	22	48	69	81	83
School nutrition	3	7	25	68	73
Accelerated literacy	0	0	0	30	73
Quality teacher package	0	0	0	34	73
Leases	27	27	65	68	68
Store license	2	8	18	54	54
Safe house	0	0	0	10	73
Night patrols	0	0	1	14	43
Extra police	6	12	16	17	73
THEMIS police station	6	12	16	17	73
Remote Area Exemptions lifted	15	65	65	65	65
Community Development Employment Projects transition	3	30	30	30	83
Community Employment Brokers	25	38	54	69	83
Banning alcohol	73	83	83	83	83
Banning pornography	73	83	83	83	83
Remote Aboriginal Family and Community Workers	0	0	0	12	83
Child special services	0	0	0	12	83
Make safe works	2	24	44	68	68
Minor repairs	0	1	7	68	68
Asbestos survey	0	5	22	73	73
All Community Clean Up works completed	0	0	0	72	73
Government Business Managers	12	67	81	81	81

(a) See NTER Review Board (2008) for details of each policy.
Source: NTER Review Board, 2008

APPENDIX B: DESCRIPTIVE STATISTICS

Table B.1 – Pre-rollout outcomes and community characteristics

	Year prior to NTER (1 July 2006 - 30 June 2007)			
	Earliest and latest NTER communities to receive income management			Rest of NTER
Outcome variables	First 10 communities	Last 10 communities	Difference	
Birthweight (grams)	3193.54	3162.84	-30.70	3056.83
	(71.3797)	(60.5082)	(95.17)	(23.0612)
Low birthweight (%)	14.63	9.46	-5.17	14.6
	(3.91)	(3.4)	(5.26)	(1.26)
Obstetric complications				
Premature (%)	14.63	6.76	-7.88	16.13
	(3.91)	(2.92)	(4.99)	(1.32)
Due to intrauterine growth restriction (%)	4.88	2.7	-2.18	4.23
	(2.38)	(1.89)	(3.1)	(0.72)
Due to anaemia (%)	2.44	4.05	1.62	11.01
	(1.7)	(2.29)	(2.84)	(1.12)
Due to gestational diabetes (%)	12.2	8.11	-4.09	7.68
	(3.62)	(3.17)	(4.88)	(0.95)
Any complication (%)	50	54.05	4.05	43.28
	(5.52)	(5.8)	(8.06)	(1.77)
Other characteristics				
Age of mother	22.61	23.65	1.039	23.96
	(0.6)	(0.64)	(0.88)	(0.22)
Aboriginal mother (%)	93.9	94.59	0.692	92.7
	(2.64)	(2.63)	(3.76)	(0.93)
APGAR 1	8.24	8.15	-0.0953	8.01
	(0.15)	(0.2)	(0.26)	(0.07)
APGAR 5	8.98	8.81	-0.165	8.82
	(0.12)	(0.19)	(0.22)	(0.05)
Community characteristics[(a)]				
Community size	318.80	277.33	-41.47	485.50
	(68.85)	(58.75)	(96.73)	(55.65)
Female share of population (%)	49.48	52.81	3.33	0.51
	(1.11)	(1.29)	(2)	(0.01)
Median age	22	22.75	0.750	22.5116
	(0.73)	(0.82)	(1.27)	(0.33)
Population aged 65+	3.3117	3.75	0.438	3.3488
	(0.57)	(0.76)	(1.05)	(0.19)
People per household	5.4216	5.95	0.528	6.057
	(0.39)	(0.22)	(0.63)	(0.21)
Median personal income	209.14	211.00	1.857	210.63
	(2.71)	(7.27)	(7.26)	(6.62)
Median household income	838.90	742.50	-96.40	803.33
	(55.81)	(88.59)	(110.2)	(40.05)
Median rent payments	41.79	32.50	-9.292	44.21

	(6.75)	(8.39)	(12.11)	(4.2)
Labour force share of population (%)	39.86	22.75	-17.11**	39.21
	(4.35)	(4.76)	(7.51)	(2.46)

Standard errors in parentheses. APGAR stands for Appearance, Pulse, Grimace, Activity and Respiration and is measured at 1 minute and 5 minutes after birth. Each of the five categories is scored 0, 1 or 2, for a maximum total score of 10.

(a) Community characteristics are from the Australian Bureau of Statistics 2006 Census community profile data; most variables available for 54 NTER communities; rest of NTER is the average across communities that were not the first or last 10 to receive income management.

Table B.2 – Pre-rollout outcomes and community characteristics

Sample limited to communities with at least one treatment and one control observation during rollout period

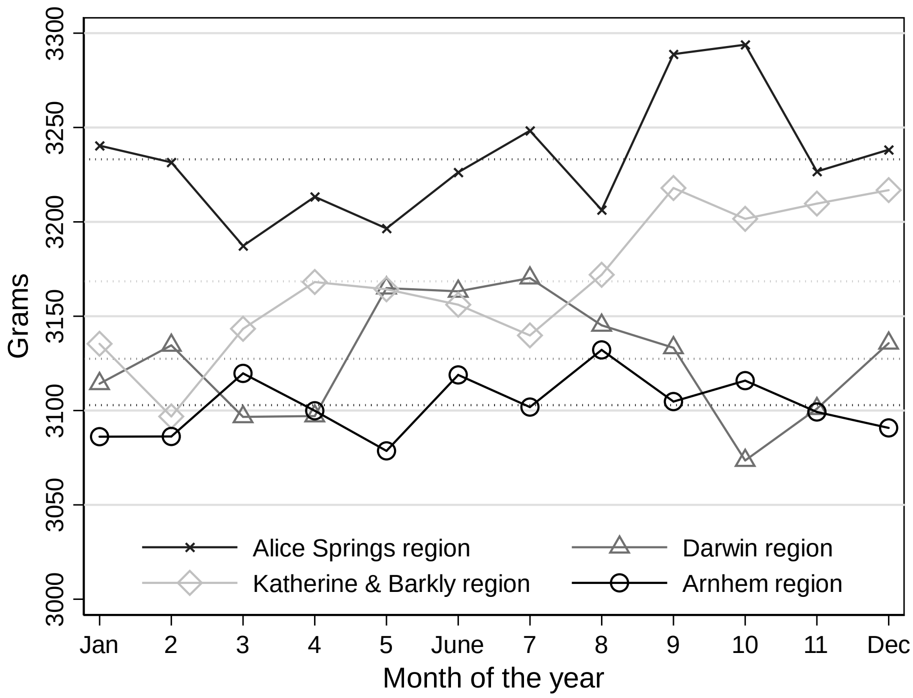
Year prior to NTER (1 July 2006 - 30 June 2007)

	NTER communities			Rest of NT
Outcome variables	Communities in first half of rollout	Communities in second half of rollout	Difference	
Birthweight (grams)	3071.85	3087.68	15.84	3354.30
	(30.9768)	(30.175)	(43.33)	(10.9835)
Low birthweight (%)	14.09	14.22	0.13	6.83
	(1.65)	(1.69)	(2.36)	(0.48)
Obstetric complications				
Premature (%)	14.77	15.85	1.09	7.91
	(1.68)	(1.76)	(2.44)	(0.51)
Due to intrauterine growth restriction (%)	5.37 (1.07)	3.03 (0.83)	-2.34 (1.36)	1.44 (0.23)
Due to anaemia (%)	9.84	9.56	-0.286	2.01
	(1.41)	(1.42)	(2)	(0.27)
Due to gestational diabetes (%)	6.94	9.32	2.39	6.83
	(1.2)	(1.4)	(1.84)	(0.48)
Any (%)	44.07	47.09	3.01	24.19
	(2.35)	(2.41)	(3.37)	(0.81)
Other characteristics				
Age of mother	23.79	23.63	-0.160	28.56
	(0.28)	(0.29)	(0.41)	(0.12)
Aboriginal mother (%)	91.72	94.64	2.92	21.28
	(1.3)	(1.09)	(1.71)	(0.78)
APGAR 1	8.00	8.03	0.0212	8.17
	(0.09)	(0.09)	(0.13)	(0.03)
APGAR 5	8.77	8.87	0.0999	8.97
	(0.07)	(0.07)	(0.1)	(0.02)
Community characteristics (a)				
Community size	396.56	566.13	169.6	na
	(58.14)	(78.42)	(98.39)	
Female share of population (%)	51.27	50.79	-0.478	48.49
	(0.58)	(0.86)	(1.06)	
Median age	22.7083	22.0769	-0.631	31.00
	(0.41)	(0.44)	(0.61)	
Population aged 65+	3.4326	3.4808	0.0482	4.80
	(0.27)	(0.26)	(0.38)	
People per household	5.5663	6.5538	0.99	2.90
	(0.22)	(0.24)	(0.33)	
Median personal income	204.33	206.96	2.628	549.00
	(3.74)	(3.33)	(5.09)	
Median household income	692.68	862.81	170.10	1192.00
	(44.52)	(44.97)	(64.7)	

Median rent payments	43.61	37.77	-5.837	140.00
	(3.43)	(5.09)	(6.36)	
Labour force share of population (%)	38.75	35.85	-2.904	47.27
	(3.42)	(3)	(4.62)	

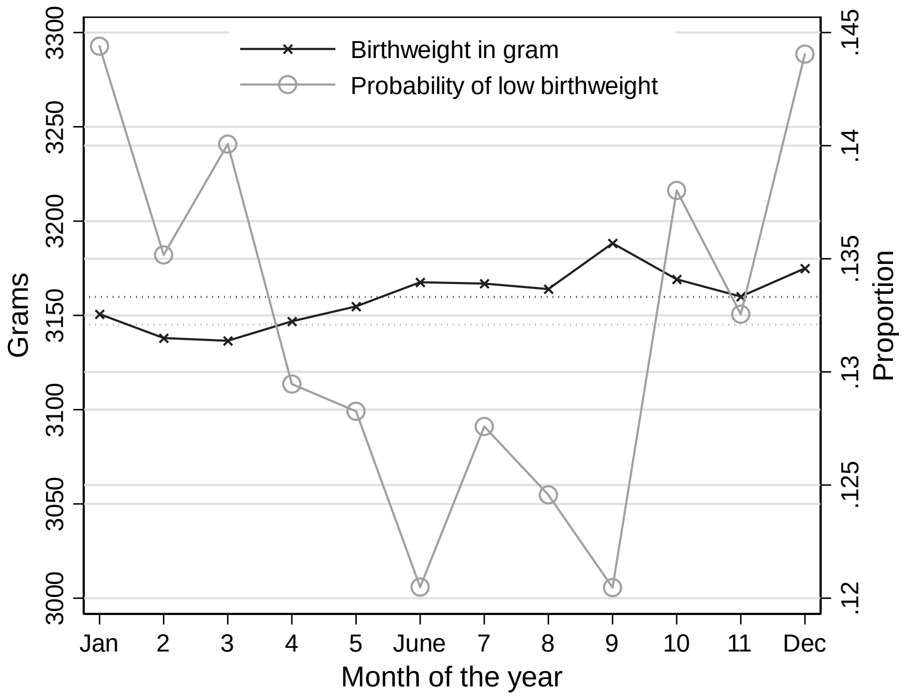
Standard errors in parentheses. First half of rollout defined as communities where income management was introduced from 17 September 2007 to 21 April 2008, second half defined as communities where income management was introduced from 28 April 2008 to 27 October 2008. APGAR stands for Appearance, Pulse, Grimace, Activity and Respiration and is measured at 1 minute and 5 minutes after birth. Each of the five categories is scored 0, 1 or 2, for a maximum total score of 10.
(a) Community characteristics are from the Australian Bureau of Statistics 2006 Census community profile data; most variables available for 54 NTER communities; rest of NT Census data is the average of all of NT.

Figure B1. Seasonal trends in birthweight by geographic region, 1996-2003



Note: The graph reports the average birthweight (left axis) by month of births, averaged between 1996 and 2013, with each obsevation assigned to one of the four regions. Each line demonstrates the seasonal variation by region. Averages are Winsorized at the 10% level to limit the influence of extreme outliers.

Figure B2. Seasonal trends in birthweight and low birth weight, 1996-2003



Note: The graph reports the average birthweight (left vertical axis) and the probability of low birthweight (<2500 g) (right vertical axis) by month, averaged between 1996 and 2013. Data series are Winsorized at the 10% level to limit the influence of extreme outliers.

APPENDIX C: ROBUSTNESS TESTS

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

	Sample size	Birthweight	Low birthweight
Matching on: Aboriginal status, whether mother is under 20, whether mother has complications in medical history, year of birth, whether born in main hospital, sex of baby, regional rainfall in 3 months to birth			
Matching method			
Nearest neighbour	1,153	-150.4	0.0826
		(55.23)	(0.0284)
Radius of 0.1	1,153	-134.5	0.0672
		(48.05)	(0.0256)
Kernel	1,153	-118.4	0.0577
		(45.63)	(0.0236)
Stratified	1,153	-117.7	0.0551
		(49.90)	(0.0230)

Bootstrapped standard errors in parentheses.
(a) Where 'treated' is defined as being in utero in a community covered by income management. We are unable to identify welfare recipient status.

Table C.2 – Treatment effect, varying sample definitions

	Sample size	Birthweight	Low birthweight
Main sample (17 Sept 2007 to 31 Jan 2009)	1,153	-118.67 (54.76)	0.048 (0.022)
Earlier sample (17 Aug 2007 to 31 Dec 2008)	1,139	-113.473 (53.009)	0.052 (0.019)
Later sample (17 Oct 2007 to 28 Feb 2009)	1,192	-101.829 (54.104)	0.042 (0.021)
Excluding partial treatment (main sample)[(a)]	1,006	-85.943 (57.621)	0.016 (0.031)
5% Winsorised (main sample)	1,153	-103.480 (47.718)	na

Community-clustered standard errors in parentheses. Regressions include controls for year of birth, community fixed effects, rainfall in 3 months to birth and premature birth.
(a) Drops infants for whom income management was introduced during the third trimester (in main regression, these observations are included in the 'control' group).

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

	(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, with community-clustered standard errors in parentheses.

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

	Birthweight	Low birthweight
Benchmark regression model	-118.667	0.047
	(54.760)	(0.021)
Dummy for wet season, year FE	-89.544	0.029
	(51.583)	(0.019)
Quarter & year interacted	-70.061	0.011
	(51.687)	(0.019)
Month & year interacted	-102.729	0.043
	(71.453)	(0.034)
Steps between rollout dates	-95.960	0.029
	(80.446)	(0.038)

Models are estimated using ordinary least squares and probit (average marginal effects). Community-clustered standard errors in parentheses.

Table C. 5 – Main regressions, with controls for rainfall at closest weather station

	(1)	(2)	(3)	(4)
Panel A: Birthweight				
Income management	-60.64	-117.9	-159.4	-109.3
	(35.95)	(50.93)	(68.32)	(62.97)
Rainfall 3 months to birth		-0.222	-0.152	-0.0868
		(0.0477)	(0.0739)	(0.0602)
Year (base category = 2007)				
2008		49.97	70.31	23.57
		(59.07)	(65.72)	(58.34)
2009		122.5	137.4	4.936
		(107.7)	(114.8)	(99.60)
Premature				-911.9
				(59.94)
Constant	3,161	3,198	3,475	3,484
	(23.28)	(45.14)	(44.71)	(42.82)
Community fixed effects	No	No	Yes	Yes
Observations	1,153	1,024	1,024	1,024
R-squared	0.002	0.014	0.093	0.325
Pane B: Low birthweight (average marginal effects)				
Income management	0.0285	0.0524	0.0741	0.0395
	(0.0167)	(0.0251)	(0.0361)	(0.0286)
Rainfall in 3 months to birth		9.31e-05	0.000110	6.54e-05
		(2.65e-05)	(3.88e-05)	(3.05e-05)
Year (base category = 2007)				
2008		-0.0254	-0.0435	0.00165
		(0.0378)	(0.0512)	(0.0421)
2009		-0.0562	-0.0832	0.0112
		(0.0446)	(0.0593)	(0.0515)
Premature				0.312
				(0.0179)
Community fixed effects	No	No	Yes	Yes
Observations	1,153	1,024	874	874

This table uses R package 'Bomrang' to locate and download data from the Australian Bureau of Meteorology weather station closest to each individual community, based on its longitude and latitude coordinates. Some communities' weather stations have unreliable or missing data for all or part of the sample period, which is why there are fewer observations in columns 2-4. For this reason, our main analysis uses region-level, instead of community-level, weather controls.

Table C.6 – Treatment effect for subsets of sample period

	Sample size	Birth weight	Low birthweight
Main regression	1,153	-118.667	0.047
		(54.760)	(0.021)
Drop communities with <10 births	956	-134.570	0.057
		(58.964)	(0.021)
First half of rollout	576	-177.973	0.137
		(123.650)	(0.076)
Second half of rollout	577	-110.457	0.067
		(81.372)	(0.039)

Models are estimated using ordinary least squares and probit (average marginal effects). Community-clustered standard errors in parentheses.
First half of rollout are those born on or before 20 May 2008, second half are those born after 20 May 2008. Note that Table 1 defines the first and second half of the rollout by number of births within the rollout period, while this table defines first and second half by infants that were in utero in their third trimester during the sample period.

Table C.7 – Treatment effect controlling for mother's medical history

	Birth weight	Low birthweight
Main regression	-118.67 (54.76)	0.048 (0.022)
Including covariates for previous pregnancies[(a)]	-117.32 (53.91)	0.046 (0.022)
Including covariates for previous pregnancies and pre-existing medical conditions[(b)]	-113.53 (54.88)	0.039 (0.022)

Models are estimated using ordinary least squares and probit (average marginal effects). Community-clustered standard errors in parentheses. All models include the standard covariates (community fixed effects, year of birth, premature birth and rainfall in 3 months to birth).

(a) Covariates are: indicator for whether first pregnancy, total number of pregnancies carried to over 20 weeks, total number of previous pregnancies.

(b) Controls are: indicators of medical history of anaemia, cardiac disease, epilepsy, hypertension, diabetes, gestational diabetes, renal disease, syphilis, urinary tract infection, and number of pre-existing medical complications.

Table C.8 – Treatment effect controlling for mother's behaviour and characteristics

	Birthweight			Low birthweight		
	(1)	(2)	(3)	(4)	(5)	(6)
Income management	-108.89	-51.00	-125.01	0.036	0.006	0.043
	(50.16)	(64.13)	(50.49)	(0.020)	(0.027)	(0.019)
Mother's age	12.89	15.74	13.39	0.001	0.000	0.000
	(2.68)	(3.32)	(2.69)	(0.001)	(0.001)	(0.001)
Aboriginal or Torres Strait Islander mother	-246.02	-179.45	-158.49	0.096	0.059	0.065
	(66.09)	(87.12)	(61.61)	(0.025)	(0.043)	(0.026)
Female baby	-122.97	-154.01	-118.46	0.048	0.049	0.049
	(33.40)	(49.90)	(32.81)	(0.018)	(0.021)	(0.018)
Controls						
Smoking at first antenatal visit, drinking at first antenatal visit, smoking at 36 weeks, drinking at 36 weeks	No	Yes	Yes	No	Yes	Yes
Missing smoking and drinking data	No	No	Yes	No	No	Yes
Community fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Premature, year and rainfall	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,153	775	1,153	1,153	775	1,153
R-squared	0.37	0.37	0.37	0.393	0.372	0.407

Models are estimated using ordinary least squares and linear probability models. Community-clustered standard errors in parentheses.

Table C.9 – Treatment effect on birthweight, controlling for quality of care

	(1)	(2)	(3)
Income management	-118.2	-118.2	-118.2
	(54.32)	(54.30)	(53.55)
Main hospital	71.14	71.53	51.75
	(50.03)	(49.68)	(51.83)
Emergency delivery		-2.086	-102.2
		(44.93)	(62.27)
Spontaneous delivery			-123.9
			(47.32)
Community fixed effects	Yes	Yes	Yes
Premature, year and rainfall controls	Yes	Yes	Yes
Observations	1,153	1,153	1,153
R-squared	0.340	0.340	0.344

Models are estimated using ordinary least squares. Community-clustered standard errors in parentheses.