Analyses of anthropometric data in the Longitudinal Study of Indigenous Children and methodological implications

by

Katherine Ann Thurber

A thesis submitted for the degree of

Master of Philosophy

The Australian National University

December 2012

Declaration

Except where otherwise indicated, this thesis is the result of my own work carried out while I was an MPhil student at the National Centre for Epidemiology and Population Health (NCEPH) at the Australian National University in Canberra.

Katherine Ann Thurber

National Centre for Epidemiology and Population Health
Australian National University
Canberra
December 2012

Acknowledgements

I would like to express my sincere thanks to all of the people who have helped me during my candidature.

First, I would like to acknowledge and celebrate the traditional custodians of the land, and pay my respects to the elders of the Ngunnawal people past and present.

I thank the 1,759 children who have generously donated their time to participate in the Longitudinal Study of Indigenous Children (LSIC), as well as their parents, carers, and teachers. Without their participation in LSIC, my research would not have been possible. I thank the LSIC Research Administration Officers for the countless hours they have spent collecting data, and for their willingness to share with me their experiences of conducting these surveys. I would like to thank the Australian Government Department of Families, Housing, Community Services and Indigenous Affairs (FaHCSIA) for funding the LSIC study, and the LSIC Steering Committee, particularly Laura Bennetts-Kneebone, Fiona Skelton, Wendy Patterson, Carole Heyworth, and Jason Brandrup, for encouraging me to use the LSIC data and for providing me with guidance along the way.

I thank the Anne Wexler Scholarship and the Australian Government Australia Awards program for funding my research, and the Australian-American Fulbright Commission for administering my scholarship and welcoming me into the Fulbright network. I would particularly like to thank Lyndell Wilson, Tangerine Holt, and Kate Lyall from the Fulbright Commission for their kindness and support, and 2012 Australian Fulbright Dr. Hamish Graham and his wife Dr. Mariam Tokhi, for generously hosting me in Alice Springs and providing me with an inimitable opportunity to contextualise my research.

I thank the National Centre for Epidemiology and Population Health (NCEPH) at the Australian National University for granting me the opportunity to pursue my MPhil, and for providing me with a supportive, encouraging, and approachable supervisory panel. Dr Cathy Banwell has been an outstanding chair of the panel, cheerfully guiding me through my MPhil and supporting my endeavour to extend my studies at ANU. I thank Dr Phyll Dance, whose careful eye and green pen have corrected countless errors, for her positive encouragement throughout my candidature. I thank Dr Gill Hall for sharing her

epidemiological expertise and for helping me to critically examine findings. I thank Dr Martyn Kirk for his practiced advice and his selfless dedication to providing guidance. I would also like to thank Dr John Boulton, from the Universities of Sydney and Newcastle, for sharing his wisdom and experience and providing a valuable perspective. I have learned so much from each of you, and am grateful for having the opportunity to work with you. Thank you for all of the time and resources you have devoted to this project.

I thank Dr Terry Neeman and Mr Bob Forrester at the Statistical Consulting Unit at the Australian National University for their statistical advice.

I thank my family: my mother Judith, my father Clifford, and my sister Mary, who have supported me from across the globe. I also thank my Australian family, the Hoskings, for their continuous support and generosity throughout my time in Canberra.

I thank my friends and colleagues at NCEPH, in particular Ellen Hart, Ellie Paige, Ray Lovett, Phil Baker, Anna Olsen, Benjawan Tawatsupa, Yanni Sun, Sarunya Sujaritpong, Wakako Takeda, Jill Guthrie, Sarah Geddes, Stephanie Davis, and Bridget O'Connor, for making my time at NCEPH so enjoyable.

Abstract

Although publications in the field of Indigenous health have increased in number in recent decades, their impact remains inadequate (1, 2). This is partially attributable to the continued reliance on descriptive studies (1, 3, 4) and the underrepresentation of urban environments in research. The Longitudinal Study of Indigenous Children (LSIC), administered by the Department of Family and Housing Community Services and Indigenous Affairs (FaHCSIA), addresses both concerns. LSIC is a cohort study of 1,759 Indigenous Australian children from environments ranging from very remote to urban. LSIC's retention rate has remained high; however, the dataset withstands a large amount of missing and implausible data.

In the first section of this thesis, I evaluated the validity of LSIC anthropometric data. I developed a data cleaning method based on World Health Organization protocols, incorporating knowledge gained from interviews I conducted with LSIC data collectors. These conversations served to depict the process of conducting surveys and to exemplify barriers impeding data collection. They shed light upon the importance of the development of a trusting relationship between participants and the LSIC team, a difficult task within the rigid structure requisite of the conduct of a longitudinal study. Based on these interviews and quantitative analysis of the accuracy of LSIC data, I provided recommendations to facilitate the collection of anthropometric data within a variety of settings. After reviewing my data cleaning methods and the final cleaned data, FaHCSIA approved the release of the cleaned anthropometric data for public use on the 4th of December, 2012.

The second part of this thesis contains analyses of the distribution of height, weight, and birth weight in the cleaned sample. In LSIC, 10% of infants were low birth weight and 11% were high birth weight; 6% of children aged three to 106 months were underweight, 74% were in the healthy weight range, 12% were overweight, and 8% were obese according to international Body Mass Index (BMI) cut-offs (5, 6).

The third segment of this thesis explores the impact of birth weight on the growth trajectories of children through eight years of age. Low and high birth weight have both been identified as risk factors for overweight and chronic disease in adulthood, and this association may be mediated by early childhood growth. Multilevel mixed-effects modelling, adjusting for the repeated measurement of children and the study's clustered

sampling, was used to examine the association between birth weight and childhood growth. Birth weight-for-gestational age z-score was a significant predictor of BMI-for-age z-score in childhood, and remained significant (coefficient = 0.166, $p < 0.001$) after accounting for age, gender, Indigenous identity (Aboriginal, Torres Strait Islander, or both), remoteness, breastfeeding duration, and maternal cigarette use during pregnancy. These findings demonstrate a long-lasting impact of birth weight on childhood growth, and suggest that interventions to improve prenatal care may have an effect beyond solely impacting birth weight. Subsequent follow-up of the LSIC cohort will enable examination of the association between of birth weight and childhood growth and later chronic disease incidence.

Table of contents

Chapter I: Introduction

The Longitudinal Study of Indigenous Children (LSIC), also known as 'Footprints in Time', remains 'the most comprehensive source of longitudinal information on early childhood development for Aboriginal and Torres Strait Islander people. It shows how early childhood experiences can impact on later life and provides policy makers with information about "what matters" and "what works" for producing improved Indigenous outcomes' (7 p. 1). This unique dataset addresses recognised gaps in the literature, advancing knowledge about the life-course development of Indigenous children across Australia. Despite the demonstrated need for this type of data, the height and weight data from this survey remain unused: scepticism surrounding the quality of this anthropometric data has prevented its release (8). Given the potential contribution of these data to public health, the quality of these data should be assessed within the reference frame of the methodological issues affecting longitudinal studies of Indigenous populations.

The longitudinal nature of the study enables analysis of the impact of early-life factors on childhood development. The height and weight data in LSIC represent a valuable and unique resource, especially given the paucity of studies examining the longitudinal growth of Indigenous Australian children. Childhood weight status is strongly associated with chronic disease (including diabetes, cardiovascular disease, and renal disease) risk in adulthood (9), and thus represents an important opportunity for intervention, requiring further research. Birth weight represents a potential risk factor, with research suggesting an association between birth weight and chronic disease in adulthood mediated by the increased deposition of fat centrally during childhood growth (10-15). The exploration of this association within the Indigenous Australian population is particularly important given the elevated rates of low birth weight and chronic disease, as well as an observed predisposition towards central adiposity (16). These issues can be examined in LSIC, using children's recorded birth weight, and measurements of height and weight taken at each wave of the study.

A) Aim

The aim of this thesis is to evaluate the validity of the anthropometric data collected in the Longitudinal Study of Indigenous Children (LSIC), and to explore the association between birth weight and childhood Body Mass Index (BMI) in this sample. This thesis will address the following research questions:

1. Are the anthropometric data collected in LSIC valid?
2. What is the distribution of height, weight, BMI, and birth weight in LSIC?
3. Does birth weight predict BMI in LSIC?

I conducted interviews and a focus group with the LSIC data collectors to illuminate the barriers hindering the collection of complete data and to inform my assessment of the representativeness of existing data. I evaluated the plausibility of the birth weight, height, and weight data collected, and the validity of LSIC as a data source. Based on the qualitative analysis of interviews and the quantitative analyses of data quality, I produced a cleaned dataset and proposed recommendations to facilitate the collection of anthropometric data in future surveys. I used World Health Organization (WHO) standards to evaluate the distribution of height and weight in the cleaned data, and I compared the cleaned birth weight data to a nationally representative Australian reference (17). I developed a multilevel mixed-effects model to explore birth weight as a predictor of BMI in childhood.

B) Thesis structure

This thesis includes preparatory sections, followed by three chapters, each addressing one of the three research questions described above. These three chapters are self-contained, including the methods, results, and discussions relevant to each research question.

Leading into these analyses, Chapter II provides a review of the literature on Indigenous research in Australia and on the height, weight, BMI, and birth weight of Indigenous children. This is followed, in Chapter III, by background on the LSIC survey and a description of the current practices for analysing anthropometric data.

In Chapter IV, I present my evaluation of the quality of the height, weight, and birth weight data collected in LSIC. First, I include a qualitative analysis of data quality based on interviews I conducted with the LSIC Research Administration Officers. Second, I conduct a quantitative analysis of data quality, examining the prevalence of missing and implausible data, and exploring the potential for any systematic biases. Based on my findings, I propose recommendations to facilitate the collection of anthropometric data in future waves of LSIC, as well as in other settings.

In Chapter V, I present the distribution of anthropometric data in LSIC. I first describe the distribution of height, weight, BMI, and their respective z-scores, exploring the variation in these indicators by demographic factors. I next examine the distribution of birth weight and gestational age, and describe the associations observed between birth weight z-scores and demographic factors.

In Chapter VI, I detail the development of a multilevel mixed-effects model to depict the association between birth weight and BMI in the LSIC sample. The findings are evaluated in the context of confounding variables, and limitations of the model are discussed.

Chapter VII contains a conclusion, tying together the three main sections of this thesis.

Chapter II: Gaps in Indigenous health research in Australia

A) Indigenous research

A research gap has been identified, with a paucity of studies examining the longitudinal development of Indigenous children across Australia (1, 3, 4). In 1992, Lake (18) identified a shortage of research into the causes of morbidity and mortality for Indigenous Australians, of research into health in urban areas, and of research that could lead to action (18, 19). A decade later, Sanson-Fisher and colleagues (1) analysed trends in the number and types of publications related to Indigenous health in Australia, Canada, New Zealand, and the United States. Since the 1980s, the number of publications regarding Indigenous health has increased in all four countries, with the most noticeable increase in Australia (1). The authors suggest that this increase may have been in response to an increased demand for, or funding of, research concerning Indigenous health (1, 20). However, this increase in research has not been succeeded by a reduction in the health gap between Indigenous and non-Indigenous populations in Australia, as has occurred in the other three countries (1, 2).

In their analysis of research trends, Sanson-Fisher and colleagues (1) noted the dominance of descriptive research in the field of Indigenous health and documented the need for strategic, methodologically-sound studies to augment the benefits arising from research. Despite the increasing number of publications about Indigenous health over time, descriptive studies have remained predominant; between 2001 and 2003, 78% of the studies examined were descriptive (1). Although there is certainly value to these descriptive studies, reliance on descriptive research 'does not provide direct evidence on how to create change, and does not produce changes as it occurs' (1 p. 505). Most research in the area of Indigenous health takes the form of cross-sectional studies, providing a view of health at just one snapshot of time (3). Priest and colleagues (4) conducted a critical analysis of studies published between 1958 and 2005 concerning the health and development of Australian Indigenous children, and found that less than ten per cent of the included studies were cohort studies. The vast majority (83%) of studies were cross-sectional. The

inadequacy of descriptive, cross-sectional research in improving health outcomes has also been voiced by the National Health and Medical Research Council (NHMRC), the Research Agenda Working Group, and the 2001 National Aboriginal and Torres Strait Islander Health Strategy Draft; these groups have indicated that a shift from descriptive to intervention or longitudinal research is a priority (3). The continued reliance on cross-sectional, rather than intervention or longitudinal, studies may be partially attributable to the difficulty in creating an Indigenous sample large enough to generate sufficient power for the analysis of these types of studies (21).

Building upon the work of Sanson-Fisher et al. (1), Derrick et al. (20) conducted a bibliometric analysis of Australian Indigenous health research published between 1972 and 2008. The authors observed a significantly higher increase in publication rate for Indigenous, compared to non-Indigenous, research in the time period, with rates of 14.1% and 8.2% per year, respectively (20). Despite this increased rate of publication, the authors found a lower 'citedness' of Indigenous, compared to non-Indigenous, research (20). Indigenous research in the fields of public health, environmental health, occupational health, healthcare science and services, nutrition and dietetics, and substance use were significantly more likely to remain uncited than non-Indigenous research in the same fields (20). Given that citation can serve as an indicator of the visibility and impact of research, these findings suggest that Indigenous research, though becoming more common, is not achieving its optimal impact (20).

Indigenous health research to date has underrepresented people living in urban and regional areas (1, 4, 18, 19, 22). For example, Priest (4) found that less than a quarter of studies were conducted in major cities or regional areas. This does not reflect the true population distribution, with 44% of Indigenous Australians inhabiting urban areas in 1971, rising to 76% in 2006 (4). The dearth of research in urban areas might be due to the 'widespread misperception that Indigenous health is just a rural and remote issue' (4 p. 56). Another contributing factor might be the 'reduced visibility' of Indigenous children living in urban, rather than remote, areas (4 p. 62). Urban Indigenous communities can be more difficult to access, as they are not geographically discrete and may lack clear structures and protocols for consultation (23). Although research addressing remote and very remote areas is unquestionably critical, examination of health in both urban and rural environments is

necessary in order to provide an accurate representation of Indigenous health and to appraise the environmental determinants of health (4).

The disparate impact of Indigenous and non-Indigenous health research, as measured by citedness or by the magnitude of the health gap, might be attributable to the appropriateness and applicability of the research conducted within Australia. The disjunction between the true Indigenous population distribution across levels of remoteness and the distribution represented in the literature may contribute to the limited impact of these studies. Further, the persisting focus on descriptive studies of Indigenous health in Australia, rather than intervention or longitudinal studies, serves as an indication that the field has not moved 'beyond describing Indigenous health issues to providing data on how to facilitate positive change' (1 p. 502). The ability of longitudinal studies to illuminate changes over time, to assess temporal relationships, to determine predictors of outcomes, and to measure the impacts of interventions is invaluable (3, 24-27). Additionally, use of a longitudinal study allows analyses of change within an individual over time, as opposed to limiting analyses to differences between individuals as in a cross-sectional survey (27). Grove et al. add that the longitudinal study design:

> enables sensitivity to history (personal or community), allows the investigation of relatively complex health-related issues and may contribute valuable information towards a 'big-picture' understanding of health and health outcomes in communities. By maintaining ongoing contact with a population, longitudinal studies are able to investigate the dynamic nature of health and to consider changes in health in relation to changes at an individual, community, and broad policy level (3 p. 638).

A shift in the conduct of research towards geographically representative and analytic studies might lead to an enhanced research impact.

LSIC, as a longitudinal study involving Indigenous Australian children from areas spanning the remoteness spectrum, addresses the above concerns. FaHCSIA acknowledge that some of the data in LSIC are accessible from other sources; 'However, as a longitudinal survey, *Footprints in Time* [LSIC] provides a unique opportunity to follow the development of a group of children and examine the factors contributing to their individual and collective outcomes' (7 p. 7). This study affords the examination of trends over time and enables a long-term, multifactorial assessment of health. In addition, the LSIC study not only adheres to, but is centred around, guidelines and principles for ethical research

conduct within Indigenous populations. The benefits of using these longitudinal data need to be weighed against the associated methodological implications.

(1) Potential limitations of the longitudinal study

There are several potential limitations specific to the use of a longitudinal survey. First, attrition or loss of participants between surveys can culminate in a large amount of missing data over the course of the study, potentially inducing bias (27). Second, over time, willingness of a community to participate may wane. Third, given the longitudinal nature of the study, there is often a significant time lag between the initiation of the study and the reporting of results (27). A fourth limitation of longitudinal studies is the confounding of aging and period effects. When one cohort of participants of the same age group is surveyed throughout a single time period, it is not possible to disentangle the effects of the aging of the participants from the effects specific to the time period (27).

Studies have shown that attrition rates can be minimised, even in transient populations, through the use of certain tools: collecting contact information, organising a tracking system, providing staff with sufficient training and support, following up participants by phone and mail, providing participants with incentives, and forming a positive relationship between staff and participants (28). The development of trust between study participants and researchers also leads to decreased missing or implausible data, and maintained community involvement. The LSIC team instituted these techniques to maximise the success of the survey.

In order to minimise the time lag between data collection and research output and to disentangle aging and period effects, a cross-sequential longitudinal design can be used. In a cross-sequential longitudinal study, several cohorts of children of different ages are followed simultaneously over time. Ideally, the age gap between cohorts is such that after a few waves, the younger cohort reaches the age of the older cohort at the first wave of the study; this allows the groups to be pieced together to represent the whole phase of development with fewer years of follow up required (27). As a result, research output can be produced in a shorter time frame, funding requirements are lessened, and attrition and loss of community engagement are minimised (27). Additionally, because multiple cohorts

are represented in the single study, the results become more widely generalizable outside of the study population (27).

This cross-sequential design was utilised in the LSIC study. Two cohorts, the Baby Cohort (Cohort B) and Child Cohort (Cohort K), were surveyed concomitantly. Infants in the younger cohort were six to 18 months of age at the first wave of the study (in 2008), and children in the older cohort were three and a half to four and a half years of age. Data were collected annually thereafter, with children in Cohort B reaching age three and a half to four and a half years of age, and children in Cohort K reaching six and a half to seven and a half years of age by the time of the fourth wave. This accelerated cross-sequential longitudinal design enables the gathering of information about Indigenous children aged six months to seven and a half years within a span of four years (7). It also allows the comparison of Indigenous children aged three and a half to four and a half years at two points in time, Cohort K in 2008 and Cohort B in 2011 (3, 29). This survey design allows for the efficient collection of data across the childhood years, minimising the number of surveys participants are asked to complete. Placing this lighter burden on participants helps minimise attrition and helps sustain engagement with the study. Additionally, the use of this method lessens the amount of time participating families must wait before learning of, and benefiting from, the outcomes of the research. Although the use of these methodologies addresses most of the limitations afflicting longitudinal surveys, there are further complications within the sphere of Indigenous health.

(2) The need for Indigenous methodologies

The difficulties facing researchers conducting longitudinal studies are compounded within the field of Indigenous health. This research cannot be disentangled from the colonial history of Australia; research 'serves as a metaphor for the colonial knowledge, for power, and for truth … In the colonial context, research becomes an objective way of representing the dark-skinned Other to the white world' (30 p. 2). The inextricable link between research and colonialism endures today (31), transcending disciplines including public health, anthropology, linguistics, sociology, history (19). Ian Anderson explains:

> The health survey, the census taker, the keeper of public hospital morbidity records, all evoke memories of the anthropologist, the missionary and those police who were actively involved in the institutionalisation of Aboriginal children and the coercive regulation of

> reserve and mission life. In such a history the anthropologist of the 1930's blends easily with the health researcher of the 1990's, although the circumstances and intent may differ greatly (32 p. 154).

Given the context of research of Indigenous Australians, reflecting the history of colonisation and exploitation, a sense of distrust towards research is widespread and, understandably, difficult to expunge (2, 3, 19, 30, 33, 34). In order to mitigate this distrust and history of exploitation, the use of appropriate Indigenous research methodologies is crucial.

In the late 1980's, the first guidelines were put in place to protect the interests of the Indigenous people and communities involved in research (31). The Research Priorities in Aboriginal Health conference, held in Alice Springs in 1986, followed by an ethics workshop in Camden (NSW) in 1987, led to the formation of research ethics guidelines (19, 31). An adapted version of these guidelines was formally released by the National Health and Medical Research Council (NHMRC) in 1991. These guidelines were 'the most obvious first step towards "transforming" research itself,' serving 'to provide the vehicle for an encounter and compromise between "Western" research practice and Indigenous political aspirations in terms that could be *understood* by non-indigenous researchers' and evoking 'a broader questioning of the ways non-indigenous intellectuals and researchers conceptualised, wrote and spoke about the Indigenous world' (19 p. 15). Although the true change brought about through the implementation of these rules may be somewhat limited (31), the widespread recognition of these principles can help move towards alleviating the mistrust so tightly bound with research.

The involvement of Indigenous people throughout the research process is a crucial component of these guidelines. However, in the analysis by Priest and colleagues (4), this involvement was not acknowledged in 71.4% of the included studies. The lack of proper consultation may partially explain the disjunction between the issues that are of interest to Indigenous communities and the issues that are the focus of research. For example, the use of a life-course approach to examine the broad social factors which can affect the health and well-being of children over time has been commended for Indigenous research (4). However, Priest et al. (4) found that, of the 217 studies included in their analysis, only 28% focused on health determinants, 3% on mental health and wellbeing, and 5% on socioeconomic, neighbourhood, community, and cultural factors. Research components

such as community consultations should not be seen as administrative criteria to meet, but should be perceived as essential to research:

> community engagement and many of the other ethical issues raised here remain as "irritating" hurdles "getting in the way" of research rather than as central aspects of the research process. The need for a supportive research infrastructure is paramount to this process: more flexible timeframes and a greater awareness of the realities of genuine consultation are essential (3 p. 640).

As Anderson explains, the involvement of Indigenous community members 'may actually enhance the project's scientific value, as people in the community will be able to identify strategies to achieve the project objectives' (32 p. 156). The involvement of the community is crucial for any research conducted, but is especially important for a longitudinal study, wherein a relationship of trust needs to be maintained throughout the course of the study.

(3) Incompatibilities between longitudinal studies and Indigenous research

Grove et al. (3) provide an extensive list of reasons why a longitudinal study would be unduly difficult to carry out within an Indigenous population: the philosophical challenge associated with divergent world views (such as different ways of viewing health for Indigenous compared to non-Indigenous Australians (35)), procedural difficulties in gathering support from communities, the inadequacy of databases to identify the pool of participants for recruitment, difficulties mediating cultural connotations surrounding the follow-up and tracking of people, and the challenge in maintaining a trusting relationship within participating communities (3). Western scientific paradigms often do not align with Indigenous views of the world, and thus input from community members is needed at each phase of the research project; collaboration, partnership, and equity through the research process function to keep research in line with Indigenous knowledge and values (2, 4). Grove and colleagues explain:

> In the push to ensure that "researchable" questions are asked, it is important to work within Indigenous paradigms to establish both what is knowable and what are valid ways of coming to that knowledge. Moreover, it is vital to appreciate that Western epistemologies carry their own social, cultural and political baggage usually unacknowledged within the borders of "value-free" scientific neutrality (3 p. 638).

The idea that an individual is the basic social unit, which underlies the use of longitudinal studies, is in itself a Western assumption, which 'may need re-thinking' (3 p. 638). This raises concern about the ability of the inherently individual-focused longitudinal study to capture the inextricable link between an individual's physiological health and the well-being (social, cultural, and emotional) of the community (3).

Despite these limitations, however, it is feasible to conduct a longitudinal study within the Indigenous Australian population, if an appropriate approach is utilised (3). There is limited evidence demonstrating the ability of factors such as trust, ownership, and consultation to improve research participation, and some authors (35-37) suggest that the benefits of these methodologies are overstated:

> There remains, however, a lack of evidence to support the proposition that the formalisation of research guidelines and protocols has resulted in an increase in the flow of benefits from research to Indigenous peoples or substantial changes to the way Indigenous peoples are positioned within the research process (36 p. 180).

However, Grove and colleagues (3) describe three successful longitudinal studies carried out within Indigenous communities in Australia. The Bibbulung Gnarneep Project is widely viewed as a successful model of research involving community participation and collaboration. The Victorian Aboriginal Health Service (VAHS) conducted a longitudinal study of urban Indigenous children living in Melbourne using a complex sampling frame, and maintained a high retention rate (3). A third successful longitudinal study was conducted by Hunter and Smith in 2000, examining Indigenous job-seekers (3). These projects demonstrate that, 'if Indigenous ownership of a project is achieved and participants and organisations that may help recruit participants see the benefit of research and feel they are engaged in a meaningful, respectful partnership with researchers and funding bodies, overcoming logistical problems is possible' (3 p. 640).

A fourth longitudinal study, the Australian Aboriginal Birth Cohort Study, has also proven successful. The study is still ongoing, following infants born between 1987 and 1990 at the Royal Darwin Hospital in the Northern Territory, with plans for a follow-up when participants reach 25 years of age (38-40). Sayers and colleagues attribute the success of the Aboriginal Birth Cohort study to the 'development of successful multiphase protocols aimed at overcoming the challenges of tracing a cohort over a widespread remote area and also to the perseverance of the study personnel' (40 p. 1). The authors postulate

that the high retention rate in the study reflects the 'good will of the wider Aboriginal community towards this study' (40 p. 1) and the formation, and maintenance, of a positive relationship between researchers and the community. As a result of the development of this positive rapport, the authors conclude, 'The continued follow-up of this life course study now seems feasible' (40 p. 1). The LSIC Steering Committee has gone to great lengths to cultivate community support for the study and maximise its potential impact, such as undergoing a dedicated, continuing process of community consultations, conversations, and feedback. The formation (and maintenance) of a positive, trusting rapport between the LSIC team and the participating families has remained a primary objective of the study. The commitment to this goal does impact on the conduct of interviews, and on the ability to collect accurate and complete data; some allowances for missing and implausible data can be made if the context of the research is considered. The ability of LSIC to maintain data integrity while ensuring the communities' 'good will' towards the study needs to be assessed.

If proven reliable, the longitudinal anthropometric data collected in LSIC have the potential to fill a wide research gap. Few studies have measured the growth of Indigenous children, and fewer yet have collected these data within a geographically-diverse sample. A standardised evaluation of the LSIC data has the capacity to provide valuable insight into the growth and development of Aboriginal and Torres Strait Islander children.

B) Height and weight status of Indigenous Australians

Given the complexities in conducting Indigenous research, little systematic information exists regarding the weight and height status of Indigenous Australian people. In the era of colonisation, anthropometry functioned to demonstrate differences between the settlers and the Indigenous population. Concern that White settlers were 'undergoing modification physically, morally, and mentally' and becoming 'cornstalks' (tall, with thin limbs and narrow chests) incited these comparisons between the Australian settlers, the Indigenous population, and European groups (41 p. 89). As a result, literature from the late 1800's and early 1900's focuses on a comparison of Indigenous and non-Indigenous body size and shape; the 'otherness,' rather than the health, of the Indigenous population was of interest. The colonisers' examinations demonstrated that the Australian 'full-bloods' were

'dolichocephalic, platyrhinic (or broad-nosed), with long legs and forearms,' short trunks, and narrow shoulders (41 p. 207). Approaching the late 1900's, anthropometric measurements were recorded as an indicator of the failed health state of the Indigenous population, the emphasis placed on the high prevalence of stunting and wasting. Only recently have studies begun to measure body size as an indicator of health, noting the double burden faced by many Indigenous communities with both underweight and overweight prevailing. However, the majority of these studies are constrained to specific geographic regions; to date, few studies have evaluated the height and weight of Indigenous children across Australia. Additionally, few studies have assessed the growth of Indigenous children over time. The methods used for evaluating height and weight status vary across studies; little consistency is observed in the choice (if any) of reference for age- and gender- standardisation, or in the choice of cut-off points for low or high height and weight.

(1) Historical studies

Although Indigenous populations have inhabited Australia for around 50,000 years (42, 43), little is recorded about the height and weight status of these populations until the mid-1900's (32); as Brough explains, 'The epidemiology of Aboriginal and Torres Strait Islander people was largely a non-existent enterprise until the 1970s' (22 p. 5). Brown and Townsend write:

> Until church missions and government settlements were established in remote regions of the continent, information on physical characteristics of Aboriginal populations in Australia was limited to field observations of anthropologists. Growth studies were hindered by the geographical isolation of traditional Aboriginal communities and their nomadic way of life (44 p. 495).

Studies conducted in the mid- to late-1900's document the status of Indigenous children living in cattle stations and mission settlements, and later in town camps and remote communities. These studies are predominantly descriptive, and function to describe differences observed between Indigenous Australians, the Australian settlers and Europeans.

Overall, these studies report a high prevalence of underweight, but describe patterns of growth in Indigenous Australian children that are not dissimilar to the growth patterns observed in European populations (see Table 1). Two distinguishing characteristics of the

Indigenous population in these studies, however, are the high prevalence of low birth weight and the reporting of faltered growth between the ages of six and 12 months. A mother's low socioeconomic status, poor nutrition, smoking and alcohol use during pregnancy have been proposed as factors explaining the significantly higher prevalence of low birth weight in Indigenous, compared to non-Indigenous, Australian infants (45). The observed growth faltering has been attributed to factors including a limited availability of food, inadequate supply of breast milk and complementary foods for children, poor hygiene, lactose maldigestion, frequency of infection, and poor socioeconomic status (46). The focus of these studies, however, is predominantly to describe the patterns of growth observed, rather than to determine factors associated with the growth of children.

The method of evaluating height and weight status varies between studies. Some studies compare raw (unadjusted) heights and weights between Indigenous and non-Indigenous groups, some calculate body proportion ratios such as the weight-stature ratio and the ratio-humeral index, while others adjust height and weight for the child's age and gender (calculating percentiles, z-scores, or per cent of the median reference value). The references chosen for this standardisation include the WHO/National Center for Health Statistics (NCHS) tables, the Tanner percentile tables, and the Boston tables. The use of these varied approaches provides distinct information, but constrains the interpretation and generalisation of findings.

Table 1: Height and weight status of Indigenous Australian children, up to 1990

Year	Authors	Location	Study Summary	Method	Outcome
1950 – 1961	Abbie (1961) (47)	Central Australia	Examined the growth pattern of 80 Aboriginal males aged six months to 16 years (ages estimated, accurate within 12 months) living in Central Australia. The growth patterns of the sample were compared to those observed within a European sample.	Children received full physical and metrical examination. Standard instruments were used to measure stature, sitting height, and limb length; ratios such as weight-stature ratio and the radio-humeral index were calculated.	The growth of Aboriginal children was comparable to that of European children up to five to six years of age. After this point, Aboriginal children underwent a rapid increase in length of the limbs, reaching the body proportions observed in European children around 12 years of age. The two groups had parallel growth trajectories through the end of the growing period (16 and 20 years of age, for males and females respectively).
1950 – 1980	Dugdale et al. (1994) (48)	Queensland	Examined the weight gain of children in five Indigenous communities, including a cohort of children from Cherbourg born between the 1950's and 1980's, a cohort of children at Yarrabah born between the 1960's and 1980's, and cohorts of children in Woorabinda, Palm Island and Doomagdee born between the 1970's and the 1980's.	Collected information on weight at birth, one year, and five years of age from various health clinics. Some children are missing data from at least one age. Weights were compared to an international reference to calculate each child's percentage of weight-for-age. ANOVA was used to calculate the difference in mean weight-for-age percentile between the three cohorts at each of age. Examined the change over time in the per cent of infants born with low birth weight.	There was a stable trend in the mean birth weight of infants over time, consistently falling below the international reference level. There were modest, though non-significant, increases in the mean per cent weight-for-age over time in most communities, approaching international reference levels. However, at Doomadgee, the reverse trend existed, with decreases in the mean per cent weight-for-age between the 1950's and 1980's.
1952 – 1983	Dugdale et al. (1990) (49)	Cherbourg Aboriginal Settlement	Studied three cohorts of children at the Cherbourg Aboriginal Settlement, born in 1952-1953, 1972-1973, and 1982-1983.	Infants were weighed at the infant health clinic; height was also measured starting in 1972 (height measurements are not available before this time). Measurements	In the 1952-1953 cohort, weight gain was comparable to reference for the first three months but faltered from three to 12 months of age, resulting in slightly low

Year	Authors	Location	Study Summary	Method	Outcome
				from the approximate ages of 3, 6, 9, 12, 24, 36, 48, and 60 months were used; most measurements were recorded within a month of the desired age. WHO/NCHS tables were used to calculate the per cent weight-for-age, per cent length-for-age, and per cent weight-for-length for each child. Per cent height-for-age measurements were not calculated for the earliest cohort. T-tests were used to compare the mean per cent values across the three cohorts.	per cent weight-for-age through five years of age. In the later cohorts, per cent weight-for-age between ages one and five years approached the international reference. Per cent height-for-age increased between the 1972-1973 cohort and the 1982-1983 cohort, approaching international reference levels. Overall, the increase in per cent weight-for-age was substantial between the 1952-1953 and 1982-1983 cohorts, despite the persistent lag in growth from three to 12 months.
1964 – 1971	Brown & Townsend (1982) (44)	Yuendumu, Central Australia	A longitudinal study of 39 boys and 23 girls of full Aboriginal descent between six and ten years of age was conducted to compare growth trends to those of a British sample.	Children's date of birth was recorded, and an anthropometer was used to measure standing height at a minimum of six consecutive visits. The Preece-Baines function was used to fit growth curves to the longitudinal data.	Compared to British boys, the Aboriginal adolescent boys were shorter at 'take off,' at peak height velocity, and in adulthood, by 2.4 to 2.9 centimetres. Aboriginal girls, however, were not significantly shorter than British girls from 'take off' to adulthood.
1960's	Propert et al. (1968) (46)	Mornington Island, Santa Teresa Mission, and communities in Western Australia	A mixed longitudinal study (some children were not measured over the whole time period) was conducted to examine growth curves and birth weights of three cohorts of Aboriginal children up to one year of age.	The rate of growth was calculated using birth weight and weights recorded at 30-day intervals. Only children with known birth date and birth weight were included. 'Mixed-blood' and 'full-blood' children were analysed separately.	Indigenous birth weights were lower than a comparative sample of English infants. The Indigenous children showed a slow rate of growth between 210 and 330 days. The 'mixed-blood' children displayed a pattern of growth in between that of the 'full-blood' children and the non-Indigenous children.

Year	Authors	Location	Study Summary	Method	Outcome
1965 - 1967	Maxwell & Elliott (1969) (50)	Central Australia (and Adelaide)	A cohort study of children up to 15 years of age in a settlement of 1,100 children (a third of whom were children); 30 Aboriginal children in Adelaide were also surveyed.	All children in the settlement were examined using standard equipment to measure height and weight. Children were excluded if date of birth was in doubt (1.5%). Medical records were obtained to provide information about previous medical history. Tanner percentile tables were used to evaluate height and weight status.	A substantial proportion of the sample fell below the third percentile for height and weight at each age. Growth faltering was observed, and was hypothesized to be associated with a decrease in maternal milk supply over the period of breast feeding.
1969 - 1993	Rousham & Gracey (1997) (51)	Western Australia: the Kimberley	A retrospective examination of the trends in the growth of Aboriginal children in the Kimberley from 1969 to 1993, using 54,000 records from 7,626 children.	191 (0.003%) weight measurements and 407 (0.007%) height measurements beyond four standard deviations away from the reference median were excluded. The percentage of infants born with low birth weight was calculated over time. Weight-for-age and height-for-age z-scores were calculated using the NCHS reference.	Across the four cohorts, mean weight-for-age z-scores were negative at birth, increased through the first month of life, and then decreased between six and 12 months of age (with the decrease sometimes beginning as early as two months). Significant trends in weight and height status across the four cohorts were not observed. There was no observed decrease in the prevalence of growth faltering from 1970 to 1993. However, a significant increase in the mean birth weight of infants was observed between 1974-1978 and 1989-1983, from 3,111 grams to 3,176 grams. From 1979-1983, 14% of infants were of low birth weight, compared to 10% in 1989-1993.
1970's	Dugdale & Lovell (1981) (52)	Cherbourg Aboriginal Settlement, South East	The growth of 145 Aboriginal children was evaluated in the first two years of life; weight, height, and adiposity status at age five to	Weight and height measurements were converted to z-scores based on the Boston tables for weight and height and the Tanner and	At age five to 10 years, height-for-age, weight-for-age, and weight-for-height z-scores of the Aboriginal children were lower

Year	Authors	Location	Study Summary	Method	Outcome
		Queensland	10 years was compared to that of white schoolchildren.	Whitehouse standards for skinfolds. In the first two years of life, growth of Aboriginal children was lower than that of white Australians, and lower than Boston standards.	than those of white students. Although the Aboriginal students had lower skinfold thickness than the white students, they had higher levels of subcutaneous fat on their bodies. They attributed the different body fat composition to differences in birth weight, nutrition, and growth in early childhood.
1984-1985	Gracey & Sullivan (1988) (53)	Six remote communities, north-west Australia	The growth of 49 Aboriginal infants in the first six months of life was assessed in a prospective study.	Birth weight was recorded; height, weight, head circumference and mid-upper arm circumference were measured every month through one year of age. Growth velocity was calculated for each child.	The mean birth weight of the sample fell below the international reference value. Growth velocity exceeded the international reference value through three months of age, but slowed after this point, with weights more than 1.04-1.35 kilograms below and heights more than 2.8-4 centimetres below sex-specific reference values at 12 months of age.

(2) Current studies

The prevalence of underweight (as defined as a weight-for-age z-score less than -2) exceeds 10% in some remote Australian Indigenous communities, the highest prevalence recorded for any Indigenous population within a developed country (54, 55). Underweight, however, is not the only weight issue facing Indigenous Australians today; the prevalence of overweight and obesity is also high, and rising. Recent studies (from the late 1900's and early 2000's) have depicted the weight status of Aboriginal and Torres Strait Islander children as heterogeneous, with high rates of underweight observed alongside high rates of overweight. Many of these studies (see Table 2) describe a marked heterogeneity in the distribution of height and weight within a single community, as well as between similar (in terms of geographical location, socioeconomic status, and cultural conditions) communities (56-58). In a study by Cunningham and Mackerras (1994), using data from the latest nationally representative National Aboriginal and Torres Strait Islander Survey (NATSIS), the variance in Body Mass Index (BMI, a proxy for body fat percentage) was large as a result of this heterogeneity, and remained large when stratifying by place of residence. Thus, homogeneity of a population should not be assumed even within the same community or region (57). As a result of this heterogeneity, the extrapolation of a single study's findings to other communities should be avoided (59); the use of a geographically-diverse sample is thus important. Additionally, caution should be exercised when judging the weight status of a community solely based on the prevalence of high (or low) BMI, as this might obscure a high prevalence of low (or high) BMI (57). Evaluating the entire distribution of BMI observed within a community is necessary.

In several studies, Indigenous adults were found to have a larger waist size than non-Indigenous Australians of a similar weight, reflecting preferential fat deposition in the central trunk region (central adiposity). It is theorized that a tendency towards central adiposity might have arisen in Indigenous Australians as an adaptation to optimize metabolic efficiency in the face of frequent food shortage or as an adaptation to a hot climate (16). Central adiposity has been associated with an increased risk of diabetes, cardiovascular disease, and renal disease in adults (16), as well as increased insulin and lipid levels, unfavorable lipid profiles, increased blood pressure, and increased left ventricular mass in children (60). Given the tendency towards central fat deposition, Indigenous people may be at a higher risk of these conditions than non-

Indigenous people of the same BMI. Rutishauser and McKay (61) found that Aboriginal women have a BMI between one and two kg/m^2 lower than Caucasian women with the same level of adiposity (as measured by skinfold thickness), indicating that Aboriginal women have higher levels of body fat than non-Indigenous women with the same BMI. Using waist circumference, waist-to-hip ratio, or skinfold thickness as measure of weight status, rather than BMI, would reveal a higher level of adiposity, and therefore a higher prevalence of overweight and obesity, in Indigenous adults.

In a sample of 486 Aboriginal children aged nine to 14 years, waist circumference, compared to BMI, was a better predictor of metabolic syndrome (MetS, the presence of several risk factors associated with chronic disease, including central obesity, impaired glucose tolerance, hypertension, high plasma triglyceride, and low HDL cholesterol) (60, 62). Over half of the children with MetS were not classified as overweight or obese by standard BMI cut-offs; thus, weight status according to BMI is not always a good predictor of chronic disease risk (60). Overall, children in the sample demonstrated a central distribution of adiposity, with a relatively low BMI, high body fat percentage, and high waist circumference; this body composition profile in adults is described as 'metabolically obese, normal-weight' (60). These findings suggest that studies using BMI as an indicator of weight status in both Indigenous children and adults may underestimate the prevalence of adiposity, and fail to identify individuals at risk of chronic disease (16). Waist circumference may be a better indicator than BMI of weight status and predictor of disease risk, as this measure has shown to be the least affected by ethnicity and sex (60).

The literature, to date, suffers from four main limitations. First, although many regional cross-sectional studies exist, 'Little systematic, longitudinal information is known about growth of Aboriginal children after five years through adolescence' (42 p. 1636S). Changes over time, or changes within individuals, cannot be examined in these cross-sectional studies. Second, the bulk of studies focus on the height and weight status of Indigenous children living in regional and remote areas. These studies cannot be assumed to be representative of children living in more urban areas, given that studies have shown that the mean weight, height, and BMI, as well as the prevalence of chronic disease risk factors, are all significantly higher in urban, compared to remote, Indigenous children (57, 59, 60). Third, there is no consistency in the methodology of assessing weight and height status, with different references, indicators, and cut-off points used across studies. The inconsistent method of standardisation, together with the

inconsistent use of indicators and cut-off points, makes it difficult to compare the findings of these studies. A standardised procedure for evaluating height and weight is necessary for the accurate comparison of height and weight status within and between groups. Fourth, most studies focus exclusively on Aboriginal children, and do not include Torres Strait Islander children, yet extrapolate their findings to all Indigenous children. A study by Leonard and colleagues (63) showed that rates of obesity, abdominal obesity, diabetes, and chronic disease and risk factors for Torres Strait Islander adults were even higher than those observed for Aboriginal adults; thus, studies of Aboriginal children might not be generalizable to Torres Strait Islander children. Differences between remote, regional, and urban areas must be considered, as well as differences between Aboriginal and Torres Strait Islander children. These limitations can all be addressed in the analysis of LSIC data.

Table 2: Height and weight status of Indigenous Australians, 1990 – present

Year	Authors	Location	Study Summary	Method	Outcome
1992 - 1998	Wang et al. (2000) (56)	Tiwi Islands	1,631 measurements of height and weight were collected from 1,138 Aboriginal Australians between the age of 5 and 77 years. Measurements were taken at multiple time points for some children, but only one measurement was used per child within each age group.	The LMS method was used to fit the data to centile curves. The BMI centiles were compared to three references: a French cohort of people up to 87 years of age, and two American cohorts, one Black and one White, of people six to 74 years of age. The French reference was chosen because of the similar age range included in the study and the similar method of calculating the centiles.	Children in the Aboriginal sample had lower BMIs, on average, than the reference populations. There was considerable heterogeneity in the BMI distribution, demonstrating a high prevalence of both under- and over-weight within each community.
1993 – 1997	Leonard et al. (2002) (63)	Torres Strait and Northern Peninsular Area Health Service District	Measured height, weight, and cardiovascular risk factors in 592 Torres Strait Islanders aged 15-87 years old, representing about half of the population of the participating communities. Standard adult BMI cut-offs of 25 and 30 kg/m^2 were used to define overweight and obesity, respectively. Results were compared to the AusDiab study of the general Australian population 1999-2000. The overall survey sample was used as the reference population for age-adjusting the sample.	Height, weight, glucose tolerance, lipid levels, lipoprotein levels, blood pressure, and urinary albumin:creatine ratios were measured. 70% of the sample had abdominal obesity (using waist-to-hip ratio cut-offs of 0.9 for men and 0.8 for women), 30% were overweight and 51% were obese according to BMI cut-offs. Less than 10% of the sample had no risk factors for cardiovascular disease and nearly 50% of the sample had at least three risk factors. Additionally, 33.1% of the age-adjusted population had diabetes and 39.2% had hypertension, compared to 5.6% and 22.7%, respectively, in the AusDiab study.	These alarmingly high percentages suggest that the burden of obesity and associated non-communicable disease for Torres Strait Islanders is significantly higher than for non-Indigenous Australians, and might exceed the burden for Aboriginal Australians. These results, after adjusting for age, were compared to the AusDiab: rates of obesity were three-fold higher and rates of diabetes were six-fold higher in this Torres Strait Islander sample. The elevated rates of diabetes in the Torres Strait Islander sample might be representative of a dose-response relationship between BMI and diabetes, wherein the risk of diabetes further increases with increasing BMI even within the overweight/obese range.

Year	Authors	Location	Study Summary	Method	Outcome
1994	Cunningham & Mackerras (1998) (57)	Across Australia	The 1994 National Aboriginal and Torres Strait Islander Survey (NATSIS) was the first nationally representative sample of Indigenous Australians; the total sample size was 15,700, including 3,221 children.	Height and weight measurements were obtained for 62% of the 3,221 children in the sample, aged five to17 years. However, other demographic information was collected for the other 38% of children, which was used to impute these measurements and confirm that there was minimal non-measurement bias. The quality of the data is unclear, and interviewers suggested that some of the measurements had been guessed, rather than measured, but the authors discount the possibility of any systematic bias.	Considerable heterogeneity was observed in the weight status of Indigenous children; almost 50% of children were at either the extreme high end or extreme low end of the BMI distribution. On average, Indigenous children aged five to nine years have higher weight-for-height and lower height-for-age z-scores than the reference median (suggesting short stature). Weight-for-age z-scores were highest in capital cities, and height-for-age z-scores were highest in rural areas. The standard deviation of the mean z-score for each indicator was greater than one, even when looking at groups separated by place of residence; this suggests that some heterogeneity of the population exists within a place of residence.
1995 - 1997	Piers, Rowley, Soares & O'Dea (2003) (64)	Remote communities in central and north-eastern Australia (Aboriginal participants); Toorak campus of Deakin University (European participants)	Comparison of the association of BMI with adiposity levels in Aboriginal and European Australians.	Height, weight, waist circumference, waist-to-hip ratio, skinfold thickness, and resistance were measured in 250 Aboriginal (130 female, 120 male) and 147 European (100 female, 47 male) Australians. Individuals with BMI exceeding 30 or with a diagnosis of diabetes were excluded to ensure a healthy sample.	Aboriginal women, compared to European Australian women, were shorter and lighter but had larger waist circumferences, waist-to-hip ratios, and skinfold thickness. The body composition of Aboriginal and European women with the same BMI is markedly different. These results cannot be generalised to other Aboriginal populations, but BMI may not be an appropriate measure for

Year	Authors	Location	Study Summary	Method	Outcome
					assessing adiposity levels in Aboriginal Australians.
1996 - 2000	Hennenberg et al. (2001) (65)	South Australia; Gerard and Raukkan	Cross sectional study, measuring height and weight of 110 children (up to age 18) and 77 adults in the Gerard and Raukkan communities. All examiners were trained and used standard equipment and measurement systems.	Some people were measured more than once, but only one data point was used for each person. Height and weight data were transformed to z-scores for analyses, using the U.S. reference. F-tests and t-tests were used to compare groups. Negative height-for-age z-scores and positive BMI-for-age z-scores were observed across all groups except females in the Gerard community (suggesting short stature).	The height findings were comparable to those reported by Abbie (66) for Central Australian Aborigines; however, the mean weights reported are heavier than those reported in most previous studies. It was concluded that weight and abdominal adiposity have significantly increased over time, reflecting a surplus of available calories. However, this abundance of food has not translated to an increase in height, possibly as a result of diet composition, psychosocial stress, or disease.
1998 - 2001	Sellers et al. (2008) (60)	Darwin Health Region, NT	Examined BMI, waist circumference, skin fold thickness, body fat percentage, insulin resistance, and prevalence of metabolic syndrome (MetS) in 486 Aboriginal children aged nine to 14 years. Children were defined as having MetS if they exceeded the cut-off point for at least three measures including: TG, HDL, waist circumference, fasting glucose, systolic or diastolic blood pressure, and glucose.	Although BMI z-scores, waist size, per cent body fat, and skin fold thickness were higher in children with MetS, more than half of children with MetS were not overweight according to standard BMI cut-offs. This suggests that waist circumference may be a useful indicator of weight status and risk factor for MetS. The prevalence of MetS did not vary by the degree of urbanisation; a surprising finding given that mean BMI is reportedly higher in urban settings. There was no association between MetS and birth weight.	Overall, the profile of children with MetS was a high body fat percentage, large waist circumference, and high skinfold measurement for a given BMI, indicating central adiposity. More than half of children in the sample with MetS had a BMI in the normal range. This is comparable to the 'metabolically-obese, normal-weight' profile of adults. The relatively low BMI of Aboriginal Australians, despite high levels of adiposity, may be an artefact of the low ratio of sitting height-to-stature observed within the population.

Year	Authors	Location	Study Summary	Method	Outcome
1998-2001	Mackerras et al. (2003) (59)	Top End of the Northern Territory	Examined the differences in growth between urban and remote children using a cross-sectional study of 482 children involved in the Aboriginal Birth Cohort Study, when they had reached the ages of eight to 14 years. The 1985 Australian Health and Fitness Survey was used to translate BMI values to percentiles. Weight-for-age and height-for-age z-scores were calculated using the CDC Growth References.	Mean weight, height, and BMI were all significantly lower in the remote versus urban sample. Additionally, systolic blood pressure, total cholesterol level, high-density lipoprotein cholesterol level, and insulin levels were higher in the urban sample. The WHO cut-off point of -2 was used to indicate low weight-for-age or height-for-age z-scores. The 15th, 85th and 95th percentiles for BMI were used as the cut-off points for underweight, overweight, and obese, respectively.	In the remote group, 37% of children fell beneath the 5th percentile for BMI-for-age, and only 9% fell above the 85th percentile, suggesting that the whole distribution of BMI was shifted towards the underweight end. In contrast, within the urban sample, 16% of children in the fell beneath the 5th percentile, and 26% above the 85th percentile. These findings demonstrate considerable heterogeneity within the sample; with a high prevalence of both underweight and overweight and associated chronic disease risk factors within the population.
2000 - 2003	Kondalsamy-Chennakesavan et al. (2008) (16)	Wadeye, Nauiyu and Borroloola	Measured height, weight, waist and hip circumferences, BMI, waist-hip ratio and waist-height ratio of 814 Aboriginal adults aged 25 to 74 years.	These findings were compared to the results of the 'AusDiab' study, a representative sample of 10,434 Australians conducted from 1999-2000, to examine differences between remote Aboriginal communities and the general Australian population.	These Aboriginal adults tended to have larger waist sizes than non-Indigenous Australians of similar weight, reflecting a preferential fat deposition in the central trunk region.
Early 2000's	Heath & Panaretto (2005) (67)	Townsville, Northern Territory	Examined the height, weight, waist and hip circumference, nutritional status, and general health of Aboriginal, Torres Strait Islander, and non-Indigenous children. The NCHS charts were used to calculate BMI percentiles, and the 5th, 90th, and 95th percentiles were used as	Children were selected from three different schools in the urban city of Townsville, Northern Territory, each with a high percentage of Indigenous students. Of the Indigenous children, 8.4% were obese, 10.3% were overweight, and 13.2% were underweight (according to BMI	The weight profile of Indigenous children was not significantly different from that of non-Indigenous children. Both subsamples demonstrated marked heterogeneity, with a higher percentage than expected in both the lower and upper tail of the BMI distribution.

Year	Authors	Location	Study Summary	Method	Outcome
			the cut-off points for underweight, overweight, and obese, respectively.	percentile cut-offs) compared to 7.1%, 11.8% and 11.8% in the non-Indigenous children.	
2010	Schultz (2012) (58)	Central Australia	Measured height and weight for a sample of 996 children aged five to 15 years attending a health check, and calculated BMI z-scores. The WHO Anthro Plus program was used to calculate BMI-for-age z-scores.	Overall, 21.4% of children were overweight (z > +1), and 5.4% were obese (z > +2) according to WHO BMI z-score cut-offs.	The prevalence of overweight in this sample does not predict the significantly higher burden of chronic disease within this population. Waist circumference might be a better indicator of risk status for this population.

(3) Overweight and obesity

Across Australia, 23% of children between the ages of five and 14 years were overweight in 2007-2008, according to international BMI cut-off points (BMI-for-age z-score greater than +1 for children over five years of age) (45). There has been a steady increase in overweight and obesity in recent decades, with the prevalence estimated at 11% in 1985 and 20% in 1995 (68). Weight status varies by socioeconomic status, region, and level of remoteness. According to data from the Australian Bureau of Statistics (ABS) National Health Survey, the prevalence of childhood overweight and obesity is 70% higher in low, compared to high, socioeconomic status areas, at 31% and 18%, respectively (45). Additionally, the prevalence of childhood overweight and obesity is reportedly 30% higher in remote, compared to non-remote, areas, at 27% and 21%, respectively (45). The prevalence of overweight and obesity also varies by ethnicity, with a higher burden faced by Indigenous versus non-Indigenous Australians (69). Among Indigenous Australians, obesity was the second leading contributor to disease burden in 2003, following tobacco use; obesity was held accountable for 16% of the health gap between Indigenous and non-Indigenous Australians (69). Further research is needed to examine if these same trends in overweight and obesity (increasing over time and varying by socioeconomic status, region, and level of remoteness) exist within this subset of the Australian population.

Although thinness, wasting, and stunting have been the focus of studies of Indigenous populations in previous decades, the prevalence of overweight is now of increasing importance. The current epidemic of overweight and obesity in Indigenous Australians is partially attributable to social and economic determinants; a sedentary lifestyle has replaced the active historic hunter-gatherer lifestyle, and high-calorie foods (rich in saturated fats, refined sugars, and salt, and lacking fibre) have replaced healthy traditional foods (70). The high rates of overweight are implicated in the rising burden of 'lifestyle' diseases (including obesity, diabetes, cardiovascular disease, renal disease, and MetS) in Aboriginal and Torres Strait Island adults. For example, Indigenous adults are two to four times as likely as non-Indigenous adults to have diabetes (9, 70); the difference is magnified for children, with an 18-fold higher incidence of diabetes in Indigenous versus non-Indigenous children (9). Similarly, Indigenous people have mortality rates due to cardiovascular disease three to four times higher than non-Indigenous people (9). Rates of renal disease and renal failure are also elevated in this

population (10, 71, 72); in the Northern Territory, the incidence of end-stage renal disease is 20 times higher in Indigenous compared to non-Indigenous people, after adjusting for age (10, 73). In a sample of 486 Aboriginal children nine to 14 years of age, 14% had MetS (60); this is more than double the estimated prevalence of 6% among adolescents in the U.S. between 1999 and 2000 (74). The high rate of diabetes, cardiovascular disease, renal disease, and MetS observed in Indigenous Australians is largely attributable to the high, and increasing, prevalence of overweight (9).

There are many pathways, both direct and indirect, mediating the relationship between obesity and disease (75). As people gain weight, levels of leptin, free fatty acids, and insulin rise. This leads to the constriction of arterioles and increased sympathetic activity, driving up blood pressure, and causing hypertension (75). Hypertension elevates the risk of diabetes, cardiovascular disease and stroke (75). Obesity itself, due to the associated high levels of triglycerides, total cholesterol, and LDL-cholesterol, independently increases the risk of cardiovascular disease (75). Overweight status can cause insulin resistance, which puts an individual at a further elevated risk of Type 2 diabetes (75). These pathways have serious health implications; overall, the loss of Years of Disability-Free Life is significant for people with obesity, at 5.70 years for men and 5.02 years for women (75). Research has shown that Type 2 diabetes, hyperlipidaemia, and hypertension can all be moderated by weight loss (75).

Excess central adiposity, which results in large waist size, is associated with insulin resistance, high blood pressure, dyslipidaemia, albuminuria, an increase in inflammatory markers, and an increased risk of diabetes, cardiovascular disease, and renal disease (16). Indigenous Australians face a higher risk of these diseases as a result of the elevated prevalence of overweight and obesity, and this risk is further heightened by the tendency of Indigenous Australians to deposit fat centrally, as evidenced by recent studies (65).

(a) Birth weight, gestational age, and early growth patterns

The Developmental Origins of Health and Disease (DOHaD) theory proposes that the factors influencing the risk of later chronic disease begin as early as the prenatal period with maternal nutrition and weight status, and continue through early childhood with birth weight and growth patterns (54, 62, 76-78). There is a hypothesized association between both low and high birth weight (defined in the following section) and the development of obesity, diabetes and other chronic conditions in adulthood (45,

62, 72, 79). A meta-analysis of 14 studies of birth weight and diabetes (including a total of 132,180 people from Finland, Sweden, the United States, Canada, India, Taiwan, and the United Kingdom) found that both low birth weight and high birth weight, compared to normal birth weight, were associated with an increased risk of diabetes, with odds ratios of 1.47 (95% CI: 1.26, 1.72) and 1.36 (95% CI: 1.07, 1.73), respectively (80). Erikkson et al. (81) observed that two thirds of cases of Type 2 diabetes follow the pathway of low birth weight and ensuing rapid catch-up growth, but the remaining third of cases follow the pathway of high birth weight and later restricted growth in height.

Indigenous infants, on average, weigh 150 to 200 grams less than non-Indigenous infants at birth (82), and the prevalence of low birth weight is double in the Indigenous, compared to non-Indigenous, population, at 12% and 6%, respectively (45). According to the DOHaD hypothesis, this skewed birth weight distribution may underlie the high rates of chronic disease observed in the Indigenous population. Although these links have been widely explored in the general population, studies within Indigenous populations have not shown consistent results; 'The evidence for childhood growth having an impact on adult health from Indigenous populations in developed countries is increasing but remains low' (54 p. 1287). Research has shown that risk factors associated with suboptimal foetal growth (such as smoking during pregnancy) are common within Indigenous mothers (76, 83, 84), but the impact of these factors, together with intergenerational effects including maternal (and grand-maternal) nutritional deprivation, requires further examination.

Humphrey and Holzheimer (83) examined the foetal growth of 96 healthy Aboriginal infants in four remote communities, using a non-matched control group of 96 non-Indigenous infants from an urban environment. The authors observed a statistically and clinically insignificant difference in the growth characteristics of Aboriginal and non-Aboriginal infants during pregnancy, despite having adequate power to detect small differences (83). This suggests that there is no inherent difference in the growth of Indigenous and non-Indigenous foetuses, but that differences in birth weight are largely attributable to a higher prevalence of risk factors for suboptimal foetal growth within Indigenous mothers. Further, Rousham and Gracey (85) lent support to the influence of socio-economic and environmental, rather than racial, factors on birth weight by demonstrating seasonal variation in birth weight for infants born in the Kimberley between 1981 and 1993. The authors observed an increase in the prevalence of very low birth weights in the wet, compared to dry, season; these findings

suggest that the tropical monsoon climate influenced birth weight, likely through an increase in environmental health risks (85). This observed seasonal variation is inconsistent with a racial predetermination of birth weight; if race or ethnicity were solely responsible for influencing birth weight, these short-term fluctuations would not be expected, as the effect of race should be stable across seasons.

Genetic factors can impact foetal growth, especially through their interaction with environmental factors (11, 86, 87). However, specific genes affecting birth weight have yet to be discovered. McDermott (88) cautions against the use of genetic explanations for the high incidence of low and high birth weight and chronic disease in Indigenous Australian populations:

> The genetic paradigm seeks to emphasise the 'independence' of the disorder and of the group, making it a 'special problem' with no immediate ramifications for the rest of society or for specific interventions to improve the situation. It may also be simply wrong, in that what we are seeing in Aborigines has little to do with genetics but represents an extreme example of a physiological effect arising from past and continuing malnutrition (including adult obesity and lack of micronutrients), poverty and social marginalization (88 p. 1192).

Emphasis on genetics should be limited, especially given the historical connection to racism and 'otherness'; rather, factors including maternal cigarette and alcohol use, nutrition, or access to health services should be considered (52, 89).

(b) Definition of low and high birth weight and size for gestational age

Low birth weight is defined as 2,500 grams or less, and very low birth weight is defined as 1,500 grams or less (45). Low birth weight occurs when a child is born pre-term (before 37 weeks gestation), or when a child is born a small size for the gestational age (failing to meet the full potential for growth), or when both co-occur (90, 91). The etiology of low birth weight for pre-term versus full-term infants is markedly different (92). Pre-term birth primarily results from maternal hypertension or other obstetric conditions, and generally results in a low birth weight, but one that is appropriate-for-gestational age (AGA) (91). Small-for-gestational age (SGA), in contrast, often indicates a decreased fetal growth rate, described as intrauterine growth restriction (IUGR) (11). A mother's low socioeconomic status, young age, cigarette and alcohol use during pregnancy, social stress, infections, poor antenatal care, and poor nutrition,

among other factors, can all contribute to an increased risk of IUGR and SGA (12, 45, 72, 89-91, 93). These associations have also been observed within Indigenous populations (10).

The definition of high birth weight, known as macrosomia, varies; cut-offs of 4,000 or 4,500 grams are typically used (45, 94, 95). High birth weight can occur as a result of an infant being post-term (more than 42 weeks gestation), or as a result of an infant being large-for-gestational age (LGA) (96). High birth weight is strongly associated with maternal diabetes (both gestational and pre-existing), maternal overweight, excessive maternal weight gain during pregnancy, high parity, and maternal age over 35 years (94-96). Among live singleton births to Australian mothers between 1991 and 1994, one per cent of Indigenous and two per cent of non-Indigenous infants were high birth weight (89).

Examining birth weight in the context of gestational age helps to separate the consequences of pre-term birth and IUGR (89). Cut-off points for SGA and LGA are defined using percentile charts of birth weight for gestational age and gender; the 10^{th} percentile is generally used to define the cut-off point for SGA, and the 90^{th} percentile for LGA (89). Using these cut-offs, a population would be expected to have a 10% prevalence of both SGA and LGA.

(c) Pathways from low and high birth weight to chronic disease

The mechanisms underlying the association of low birth weight with later chronic disease have yet to be determined (97). The DOHaD hypothesis posits that due to developmental plasticity, environmental factors associated with low birth weight can lead to specific physiological and morphological states in later life. Similarly, the foetal origins hypothesis states that low birth weight infants adapt to foetal undernutrition by making permanent metabolic changes which are an adaptation to anticipated future nutritional paucity (88). This becomes maladaptive when the infant is exposed to adequate or excessive post-natal nutrition, leading to increased central adiposity in childhood, and putting the child at an increased risk of developing chronic diseases, such as diabetes, cardiovascular disease, and renal disease, later in life (88). This represents a developmental mismatch, wherein the infant's postnatal environment does not match up to what would have been predicted based on the foetal environment (98). This mismatch can lead to 'postnatal changes in body composition during catch-up

growth and the development of insulin resistance to reset child's growth to follow his/her initial genetic growth trajectory' (98 p. 4031).

The growth acceleration hypothesis proposes a relationship between rapid early postnatal growth and the risk of non-communicable disease, independent of birth weight (62, 77, 99). For example, there is a tendency of mothers to overfeed low birth weight babies due to their small size, leading to accelerated growth and increased risk of chronic disease (80, 100); in this case, it is not the low birth weight itself that is causing these deleterious outcomes, but the feeding patterns associated with the low birth weight. By 'malprogramming' circuits regulating metabolism, appetite, and weight, neonatal overfeeding (independent of low birth weight) may be a risk factor for disease by causing rapid catch-up growth (80).

Despite some differences, these theories are generally in agreement on the central idea that a mismatch between the foetal and postnatal environment mediates changes in body composition leading to later chronic disease risk. Research (see Table 3) suggests that rapid catch-up growth is associated with the deposition of fat centrally (11, 13, 15, 101) and increased insulin resistance (11, 102), mediating the association between low birth weight (or SGA) and disease in adulthood (11, 15). Further research is required to determine what growth trajectory is ideal for these infants; whether rapid catch-up growth is inherently dangerous or whether it only becomes damaging if children surpass their optimal size. The exact mechanism underlying the associations between birth weight, size for gestational age, early childhood growth, and chronic disease risk in adulthood is unknown. Beltrand et al. write:

> it remains to be determined whether early catch-up growth is a phenomenon primarily dependent upon postnatal nutritional environment or whether it is mostly conditioned by the pattern of fetal growth. In other words, does this early acceleration of postnatal growth result from a conflict with postnatal nutrition or is it a compensatory phenomenon intended to replace infants on their own physiological growth curves? (99 p. e5343)

Regardless of the exact mechanism, there is consensus surrounding the long-lasting impact of the prenatal and postnatal environment.

(d) Associations between birth weight and chronic disease risk in Indigenous Australians

Several studies have examined the association between birth weight and later chronic disease status in Indigenous Australians (see Table 4). Using data from the Aboriginal Birth Cohort Study (a survey of 686 Aboriginal infants born between 1987 and 1990 at the Royal Darwin Hospital in the Northern Territory), Sayers and colleagues (97) found a negative association between birth weight and blood pressure, but no other biomarkers of chronic disease, for children at a mean age of 11.4 years. Increased blood pressure in childhood for low birth weight infants has significant ramifications, as it increases the risk of cardiovascular disease (through its association with decreased artery elasticity, and increased ventricular size and mass, cardiac output, and peripheral resistance) (62). Similarly, in a cross-sectional study of a remote coastal community from 1992-1998, the relationship between birth weight and blood pressure was explored within 1,473 Aboriginal people (82% of the community) (12). For adults in the sample, the odds ratio for having high blood pressure given low (less than 2,500 grams), versus non-low (more than 2,500 grams), birth weight was 2.3 (95% CI: 1.3 – 3.5), after adjusting for current BMI, sex, and age (12). The highest blood pressures were observed in adults with low birth weight and also high current weight (12). A study of 317 Indigenous adults in the Northern Territory found that those with low birth weight (less than 2,500 grams) faced a significantly higher risk of renal disease than those without low birth weight (odds ratio = 2.82, 95% CI: 1.26 to 6.31), after controlling for BMI, blood pressure, age, and sex (10). This risk was further (odds ratio = 4.6, 95% CI: 1.3 to 11.6) elevated in people with a high BMI (greater than 25 kg/m^2 compared to less than 25 kg/m^2) at the time of the study (10).

These findings support the hypothesis that low birth weight infants who attain normal or above-average weight and height in later childhood through rapid catch-up growth (rather than maintaining small size) face the highest risk of diabetes, cardiovascular disease, and other chronic diseases (11, 14, 54, 79). However, these studies do not account for potential confounders such as maternal smoking, which might influence birth weight, childhood growth, and health indicators in adulthood. Additionally, all studies assumed a linear relationship between birth weight and health outcomes; the possibility of a U-shaped relationship (due to associations between high birth weight and chronic disease risk) was not addressed. A third limitation is that these

studies did not adjust for gestational age, confounding the effects of pre-term birth and IUGR. Further research, taking these factors into consideration, could contribute to the unfolding of the association between birth weight and chronic disease risk.

Several authors (see Table 4) propose a link between low birth weight and central fat deposition during catch-up growth, leading to disease in adulthood (10-15). Given the predisposition for Indigenous people to deposit fat centrally (16, 57, 61, 64), regardless of low birth weight and catch-up growth, they may experience a different health impact of low or high birth weight than populations without this predisposition. If central fat deposition mediates the association between low birth weight and later risk of chronic disease, how does the predisposition of Indigenous Australians towards central fat deposition affect the health consequences arising from low birth weight? Despite the high prevalence of low birth weight in Indigenous populations, and the numerous measures and initiatives implemented to reduce the prevalence of low birth weight, there have been few studies with adequate follow up to examine trends in height and weight in early childhood (103). Casey and colleagues explain that this is largely attributable to 'methodologic problems' in research to date; 'Even reports with longitudinal samples often fail to maintain enough of the cohort for long enough intervals to describe adequately the long-term status and patterns of growth in LBW pre-term infants' (103 p. 599).

The Aboriginal Birth Cohort study, however, is one example of a successful cohort study; Sayers and colleagues (38, 39, 97, 104) have used the study to examine trends in height and weight for Indigenous children of varying birth weight. A subset of 279 term infants from the original cohort was followed up at age eight to 14 years, and height, weight, insulin concentration, and glucose concentration were measured (38). At this follow-up, SGA (defined as falling in the lowest decile of birth weight for gestational age and gender) children had significantly lower z-scores for height, weight and BMI than AGA (defined as falling between the tenth and 90th percentiles of birth weight for gestational age and gender) children. No significant associations were observed with any measures of glucose or insulin concentration (38). Sayers et al. (39) conducted another study of 341 term infants from the same birth cohort, examining anthropometric indicators at a mean age of 18.3 years. Term infants born SGA, on average, had lower weight, height, BMI, mid-arm circumference, and fat percentage than term infants born AGA (39). These findings suggest that SGA infants remained smaller than AGA infants at age 18 years. There was not a significant difference in

waist-to-height or waist-to-hip ratios between the two groups, suggesting a similar distribution of fat (39).

These studies suggest that term Aboriginal infants born SGA remain smaller than term AGA infants through 18 years of age, and do not demonstrate a difference in fat distribution or other biomarkers of chronic disease. These findings do not support the DOHaD hypothesis, which would predict higher levels of central adiposity and insulin resistance in the SGA group (39). Sayers and colleagues (39) acknowledge the purported link between low birth weight, later overweight, and risk of chronic disease, and note, 'The continued study of this Aboriginal birth cohort will give us an opportunity to determine if and when in later life the effects of birth weight are modified by environmental nutritional factors' (39 pp 417). Further research is needed to examine the extent to which the postnatal environment and childhood growth mediates the future risk of central adiposity and chronic disease risk of Indigenous infants born low birth weight or SGA. Given the hypothesised association between birth weight, childhood growth, and the development of chronic disease, a life-course approach is necessitated; these associations can be explored using LSIC.

Table 3: Studies examining the association between birth weight, size for gestational age, and chronic disease risk

Year	Authors	Location	Study Summary
1934-1944	Eriksson et al. (2003) (81)	Finland	A longitudinal study of 8,760 people born between 1934 and 1944 in Helsinki. Height and weight were measured, on average, eight times before age one year, and ten times between the ages of one and 12 years. A total of 290 participants were identified as receiving diabetes medication between 1964 and 1997; this was used as an indicator of Type 2 diabetes. There was a significant interaction between the effect of birth weight and BMI at age 12 years on diabetes incidence in adulthood. For children of low and normal birth weight (less than 3,500 grams), diabetes in adulthood was unrelated to the rate of infant growth, but was associated with increased BMI and weight-for-age z-score beginning at age seven years. For children of high birth weight (more than 3,500 grams), however, the risk of diabetes in adulthood was increased for children who experienced a slow gain in length in the first three months of life. This growth faltering might be associated with impaired insulin metabolism. Maternal overweight and maternal hyperglycemia increase the risk of high birth weight; the relationship between high birth weight and later diabetes therefore might be attributable to environmental influences or to 'malprogramming' resulting from increased levels of insulin in the developing foetus of hyperglycaemic mothers.
1959-1965	Stettler et al. (2002) (105)	U.S.	A prospective cohort of 27,899 children, followed from birth to seven years of age: examining the relationship between rapid weight gain in the first four months of life and weight status at age seven years. The risk of overweight at age seven years was significantly higher for those undergoing rapid growth, after controlling for birth weight and weight attained at age one year. Each 100-gram increase in weight gain per month was associated with a 17% increased risk of overweight at age seven years. This study suggests that rapid weight gain in childhood predicts weight status at seven years of age, independent of birth weight and weight status at one year of age.
1980's	Casey (1991) (103)	U.S.	Growth rates and BMI were examined in a longitudinal study of 985 low birth weight, pre-term infants from diverse demographic and socioeconomic backgrounds. Even when adjusting for gestational age, these low birth weight infants had significantly slower growth rates and lower BMI than the NCHS standard, with no catch-up growth observed in the first three years. The authors conclude that pre-term, low birth weight infants follow a distinct growth trajectory in the first three years of life compared to term infants.
1988-1994	Hediger et al. (1998) (13)	U.S.	Body composition was examined in a sample of 4,431 infants aged two to 47 months, comparing children born SGA, AGA, and LGA (all gestational ages included). Z-scores for height and weight were calculated, and mid-upper arm circumference, triceps and subscapular skinfolds, mid-upper arm muscle areas, mid-upper arm fat areas, and the arm fat index were measured. SGA infants demonstrated a deficit in muscularity, and a relatively smaller deficit in fatness, whereas LGA infants demonstrated a surfeit in muscularity, and a relatively smaller surfeit in fatness. Per cent body fat was consistently higher in SGA vs. LGA infants. These findings support the role of fat composition as a mediator between low birth weight and chronic disease in adulthood.
Births 1989-1990	Garnett et al. (2001) (15)	Australia	In an Australian study of 255 children aged seven to nine years, there was a negative association between birth weight and the proportion of total body fat located in the abdomen (as measured directly by dual energy X-ray absorptiometry). However, birth weight was not associated with total body fat; thus, despite similar levels of body fat, low birth weight children stored a higher percentage of their body fat centrally. Additionally, weight-for-age z-

Year	Authors	Location	Study Summary
			score at age seven to eight years was positively associated with the proportion of abdominal fat. Using multiple regression models, the authors estimate that 20% of the variation in abdominal fat percentage could be explained by birth weight and weight-for-age z-scores. These findings suggest that it is this pattern of central fat deposition and the development of insulin resistance that mediates the association between low birth weight and disease in adulthood.
Births 1991-1992	Ong & Dunger (2004) (11)	Great Britain	The health impact of the interaction of low birth weight with rapid childhood growth was further supported by the Avon Longitudinal Study of Pregnancy and Childhood (ALSPAC), a study of over 10,000 children. At age five years, the children with the highest weight, BMI, and waist circumference were those who had experienced early rapid early weight gain, and this trend persisted through age eight years. In this study, rapid weight gain was defined as an increase in weight-for-age z-score greater than 0.67 within the first three years of life. This rapid weight gain was also associated with increased insulin resistance at age eight years. During this period of rapid weight gain, children of low birth weight experienced changes in body composition, with increasing deposition of fat centrally, as their BMI approached, or even exceeded, the reference median.
Births 1991-1992	Ong et al. (2000) (106)	Great Britain	A random sample of 848 infants from ALSPAC: examining the impact of postnatal catch-up growth on size and obesity at five years of age. Almost a third of infants showed catch-up growth, as defined by an increase in weight-for-age z-scores greater than 0.67 in the first two years of life, and nearly a quarter showed catch-down growth. Variation in weight-for-age z-scores was larger within the first two years of life than between the ages of two and five years. At five years of age, children who had undergone early catch-up growth had higher weight, height, and BMI z-scores, as well as a higher percentage body fat and waist circumference than those without catch-up growth. This supports the hypothesis that early catch-up growth is associated with central adiposity and overweight in later life.
Births 1991-1992	Ong et al. (2004) (14)	Great Britain	Growth and insulin levels are linked, as insulin-like growth factor-I (IGF-I) regulates childhood growth in addition to insulin sensitivity and blood glucose levels. In a study of 851 children from the ALSPAC birth cohort, Ong and colleagues found an association between early weight gain and insulin resistance. Children's height, weight, waist circumference, blood glucose, insulin, and insulin precursors were measured after an overnight fast, and glucose and insulin measures were repeated 30 minutes after consumption of a glucose drink. 'Catch-up' and 'catch-down' weight gain were defined as an increase or decrease, respectively, in weight-for-age z-score greater than 0.67 between birth and three years of age. This definition was chosen because the standard centile lines on growth charts (2^{nd}, 9^{th}, 25^{th}, 50^{th}, etc.) are 0.67 standard deviations apart, and therefore a shift in z-score greater than 0.67 would represent the crossing of centile lines. In the sample, early catch-up weight gain was associated with a higher BMI, waist circumference, and insulin resistance at age eight years. Conversely, the authors found that reduced height gain was a risk marker for Type 2 diabetes. This indicates that growth in weight and height may have differential implications for disease risk; examining both, and their relation to central adiposity, is therefore necessary.

Table 4: Studies examining the association between birth weight, size for gestational age, and chronic disease risk in Indigenous Australians

Year	Authors	Location	Study Summary	Outcome
1998-2001	Sayers et al. (2009) (97)	Northern Territory	The Aboriginal Birth Cohort Study: a prospective cohort study of 686 Aboriginal children, with follow up at age 8.9-14 years. Height, weight, sitting blood pressure, and pubertal stage were recorded; a blood sample was taken to measure plasma glucose, cholesterol, and triglycerides concentrations. The CDC 2000 reference was used to calculate z-scores for height and weight. Regression was used to examine the relationships between birth weight and the various biomarkers, also considering the interaction between birth weight and current weight.	A low prevalence of overweight was observed in the sample (9.6% and 11.5% for boys and girls, respectively), but there was a high prevalence of underweight (19%). The mean weight-for-age z-score was -0.8 and the mean height-for-age z-score was -0.5. A negative association between birth weight and blood pressure was observed, but no significant associations were found between birth weight and the other biomarkers of chronic disease. However, a positive association was observed between current child weight and blood pressure, total cholesterol, apolipoprotein B, kidney volume, lung function, fasting triglycerides, insulin, and glucose. The associations between birth weight and these biomarkers may change with increasing age.
1998-2007	Sayers, Mackerras, Singh & Reid (2004) (104)	Darwin	A study of 279 Aboriginal children aged eight to 14 years reported significantly lower z-scores for BMI, height-for-age, and weight-for-age for people in the low birth weight group compared to those in with birth weights in the normal range.	The study demonstrated no association between birth weight and later insulin concentration and glucose concentration, calling into question the association between low birth weight and diabetes in adulthood.
2006-2008	Sayers, Mott & Singh (2011) (39)	The Top End of the Northern Territory	A total of 341 Aboriginal term infants from the Aboriginal Birth Cohort Study were followed up at 18 years of age. An average decreased height of three centimetres and decreased weight of nine kilograms was noted in children with growth restriction at birth. The prevalence of BMI than 18.5 was higher, and the prevalence of BMI greater than 25 was lower, in the low birth weight group compared to the normal birth weight group.	Inconsistent with findings in non-Indigenous populations, the distribution of fat was similar in low and normal birth weight groups, with non-significant differences in waist-to-height and waist-to-hip ratios.

Chapter III: Background

A) The Longitudinal Study of Indigenous Children: methodology

LSIC, supported by the Australian Department of Families, Housing, Community Services and Indigenous Affairs (FaHCSIA), has conducted four annual surveys of Aboriginal and Torres Strait Islander children across Australia. The survey includes a total of 1,759 children, separated into two cohorts, the Baby Cohort (Cohort B) and the Child Cohort (Cohort K). The sample comprises approximately five to ten per cent of all Australian Indigenous children within each age group (107, 108). Parents or carers of the study children were also interviewed, as well as a small number of teachers. The study was conducted in order to learn about the factors that impact on the development of Aboriginal and Torres Strait Islander children, and to explore ways to allow them to grow up strong and resilient (109). The four main questions underlying LSIC include:

1. What do Aboriginal and Torres Strait Islander children need to have the best start in life to grow up strong?
2. What helps Aboriginal and Torres Strait Islander children to stay on track or get them to become healthier, more positive and strong?
3. How are Aboriginal and Torres Strait Islander children raised?
4. What is the importance of family, extended family and community in the early years of life and when growing up? (107 p. 6)

A primary purpose of LSIC is to obtain data that can be used to create targets for the 'Closing the Gap' initiative to increase the life expectancy and improve the wellbeing of Aboriginal and Torres Strait Islander children across Australia (107). Further, Professor Mick Dodson, Chair of the LSIC Steering Committee, hoped that the results of LSIC could be used to promote a sense of Indigenous pride and to inspire children:

> We all need to recognise that children's sense of themselves as Aboriginal people – who they are and where they come from – is of both practical and spiritual value. In bestowing identity we also bestow dignity. It is a good deal more than symbolic – it has profound practical effects. There are plenty of examples of Indigenous success; we just have to recognise it and replicate it … We have to see evidence of success as points of light all around us and join them up to create a universe of opportunity for our children (107 p. 3).

LSIC researchers collected information about study children's dietary habits, exercise patterns, medical history, current health status, height and weight, as well as information about socioeconomic status, household characteristics, cultural practices

and beliefs, and their parents' health status. Patterns of height and weight gain in children can be monitored over time, given the longitudinal nature of the study. Data collection was performed with full ethical approval from the Departmental Ethics Committee of the Australian Government Department of Health and Ageing. Ethical clearance was also obtained for each state and territory at their respective Human Research Ethics Committees.

(1) Participants

The target population of the LSIC study includes all Aboriginal and Torres Strait Islander children across Australia. Lists of Aboriginal and Torres Strait Islanders in the appropriate age ranges (born between December 2006 and November 2007 for Cohort B, and born between December 2003 and November 2004 for Cohort K) were provided by Centrelink and Medicare Australia, and a purposive sampling method was used to approach eligible families from the list. Research Administration Officers (RAOs) also recruited children through the use of study promotion and word of mouth (29). Only one child per family was included in the study.

The participants do not all fit into one of the two designated age ranges because of difficulties in finding children of the appropriate age, wishes of other families to have their child included in the study, and the extended time frame during which interviews were conducted. At the time of the first interview, infants in Cohort B ranged in age from three months to 34 months (two years and ten months), and children in Cohort K ranged in age from 33 months (two years and nine months) to 69 months (five years and nine months). The study was designed to include an equal number of children in each cohort, but as a result of these issues described above, the cohorts were slightly unbalanced, with 963 infants in Cohort B (58%) and 707 children in Cohort K(42%) in the first wave of the study.

The first participants were recruited in December of 2007 and interviewed between April of 2008 and February of 2009 (see Table 5). Families that were interviewed in Wave 1 were approached for a second interview the following year, with interviews conducted annually thereafter. For a variety of reasons, including an inability to locate the families, families' refusal to participate, and families' relocation to distant sites, 236 families from the original cohort were not interviewed in Wave 2. To offset this loss, 88 new families, who had been approached for participation in Wave 1 but were unable or refused to participate, were recruited to participate in Wave 2 (29). Thus,

there were a total of 1,523 interviews in Wave 2, with a retention rate of 85.9% from Wave 1. In Wave 3, surveys were conducted of 1,312 families who had completed Wave 2 (86.1% retention from Wave 2). Additionally, 92 families who had participated in Wave 1 but missed Wave 2 re-joined the study in Wave 3, for a total of 1,404 interviews (29). Of the new families entering the study in Wave 2, 73 completed the 'new entrant' questionnaire (including the questions that were asked in the first wave of the study) in Wave 2, six completed the 'new entrant' questionnaire in Wave 3, and the remaining nine did not complete a 'new entrant' questionnaire, but responded to the same questions as the general study population who had not missed the first wave (and thus some data are missing for these nine children) (108).

Table 5: Time period for each wave and number of participants interviewed (108)

Wave	Period for pilot fieldwork	Period for main fieldwork	Number of participants returning from previous wave	Per cent retention from previous wave	Number of additional interviews	Total number of interviews
1	2006 to 2007 and January 2008	April 2008 to February 2009	--	--	--	1,671
2	November 2008	March 2009 to December 2009	1,435	85.9%	88 (new entrants in W2)	1,523
3	October 2009	March 2010 to December 2010	1,312	86.1%	92 (interviewed in W1 but not W2)	1,404

The length of each interview ranged from five minutes to three hours, depending on the person completing the survey. In Wave 2, the average length of the interview was one hour for Parent 1 (both cohorts), 30 minutes for Parent 2 (both cohorts), ten minutes for children in the B Cohort, and 17 minutes for children in the K Cohort (108). This totals to nearly two hours for the average set of interviews for a family. When considering that a total of 4,598 surveys were conducted within the first two waves of the study, this constitutes an enormous contribution of time for both the participating families and the interviewers. The first four waves of the study were funded by the Federal Budget, and funding will continue as long as the study maintains a high retention rate (108).

(2) Sample characteristics

Of all the children sampled in Wave 1, 88% were Aboriginal, 6% were Torres Strait Islander, and 6% were both Aboriginal and Torres Strait Islander. A fairly similar pattern was observed with the parents and carers: 76% were Aboriginal, 7% were Torres Strait Islander, 4% were both Aboriginal and Torres Strait Islander, and 13% were neither Aboriginal nor Torres Strait Islander (107). Male and female children are approximately equally represented in the dataset. Indigenous children from Tasmania and the ACT are not represented in the study, due to the small Indigenous population in each region.

Participants in the LSIC study are not a representative sample of all Aboriginal and Torres Strait Islander children throughout Australia, but rather are selected from 11 specific sites (see Table 6), chosen to represent a wide range of environments (107). In choosing interview sites, selection criteria were established to ensure the inclusion of communities of varying levels of remoteness and to ensure the representation of Aboriginal and Torres Strait Islander children from across Australia. The aim was to select approximately 150 children from each site, for a total of 1,650 children within the two cohorts (107). However, the number of children at each site varies slightly due to difficulties in finding appropriately aged children in some locations, given the geographic spread. Thus, the percentage of Aboriginal and Torres Strait Islander children living in areas of varying levels of relative isolation in the LSIC data is not representative of the true population percentages.

Table 6: Total number of children per cohort per site, in Wave 1 (107, 108)

Site	Number of children	Per cent of total sample
Northern Territory Top End (including Darwin, Katherine, Hodgson Downs / Minyerri and Galiwin'ku)	239	14.1
South East Queensland (including Brisbane, Ipswich, Logan, Inala, Gold Coast and Bundaberg)	211	12.5
South Coast New South Wales (from Kiama to Eden)	175	10.4
Mt Isa and remote Western Queensland (Mornington Island, Doomadgee, Normanton, and Cloncurry)	172	10.2
Western Sydney (from Campbelltown to Riverston)	163	9.7

Site	Number of children	Per cent of total sample
Dubbo (including Gilgandra, Wellington and Narromine)	156	9.2
Greater Shepparton (including Wanagratta, Seymour, Bendigo, Cobram, Barmah, and surrounding areas)	143	8.5
Torres Strait (including Torres Strait Islands, Cairns, and Northern Peninsular Area)	132	7.8
Kimberley region (including Derby, Fitzroy Crossing, Broome and One Arm Point)	126	7.5
Adelaide (including Port Augusta)	106	6.3
Alice Springs and some surrounding communities	64	3.8
Total	*1,687**	*100.0*

* Some participants removed themselves from the study after interviews were conducted, leaving a final sample of 1,671 children at the first wave of the study.

(3) Role of Indigenous people in the research process

The involvement of Aboriginal and Torres Strait Islander community members in this study was crucial in order to gain community support, to minimize attrition from the study, and to maximize benefit to the communities. The formation of relationships of trust, respect, and reciprocity began with initial community consultations for study development, and continues throughout the study with the ongoing dissemination of research findings to the communities. Engaging Indigenous people in the research design enabled the creation of a project that matters to, and will contribute to, the involved communities (34). The priorities of the project were determined through discussion with community members between September 2003 and June 2004. These consultations were held to ensure that the study would address the interests of, and would produce benefit for, the Aboriginal and Torres Strait Islander people and communities involved. Consultations occurred at 23 sites, selected based on the size of the Indigenous population and the location. People at these sites were invited to participate by a letter sent three to four weeks before conversations were to occur.

From these consultations, researchers gathered that qualitative, as well as quantitative, data would be appropriate and useful; those consulted felt that these qualitative measures might provide a more accurate view of their children's development than strictly quantitative measures. In these consultations, several people expressed concern about health and nutrition (110), a need to gain 'knowledge about

healthy food and accessing healthy food' and 'Understanding getting good food and growth' (111 pp 22). Additionally, concern was raised about the best way to raise children: 'What are the warning signs/signals of poor outcomes and what are the preventive measures? … What are the things that are in place that make a difference in the quality of children's lives?' (111 pp 22). The Research Design Subcommittee considered these concerns in their development of the study, and also considered input from the Western Australian Aboriginal Child Health Survey and the National Aboriginal and Torres Strait Islander Health and Social Survey (108).

Between September 2004 and December 2005, preliminary studies were conducted in the Torres Strait and North Peninsular Area to pilot the qualitative interviews and a community engagement strategy. The Australian Bureau of Statistics (ABS) aided in developing and testing the questionnaires. FaHCSIA and the ABS trained six Indigenous RAOs to aid with the consent processes, data collection, dissemination of information, and other crucial activities, during pilot research. These RAOs conducted the majority of the first wave of interviews. Prior to beginning the study, community members and Elders at each site provided consent and approval for participation (29). The LSIC study was designed with consideration of cultural protocols and respect and acknowledgement at each site location.

(4) Consent and confidentiality

FaHCSIA went through a series of measures to ensure that parents were able to provide informed consent to participate in the study. Leading up to the interview, a letter and a DVD explaining the study and the consent process were sent to each set of parents. At the beginning of the interviews, the RAO explained the consent form to each set of parents, or parents could read a written plain language statement if they preferred. The parents were asked to consent to: participating in the study, having their interview recorded, having a second parent interviewed, having a teacher or child care worker contacted, having the study child photographed, and permitting researchers to access the study child's Medicare records. After signing the consent forms, parents were provided with a letter explaining the agreements, providing contact details for the ethics committee and for FaHCSIA, and stating that the participants were allowed to cease participation at any point of the study.

The LSIC committee was dedicated to ensuring confidentiality during the data collection phase. The names and addresses of all participants have been removed;

additionally, data users' access to geographical location, along with other potentially identifying information, has been limited to maximally protect the identity of participants.

(5) Ethics

In conducting research in Aboriginal and Torres Strait Islander communities, it was imperative that a relationship of trust was established between the research team and the participants, to enable honest discussions of sensitive topics including culture, family, religion, income, and health (112). The LSIC steering committee considered the six core values of reciprocity, respect, equality, responsibility, survival and protection, spirit and integrity, in the design and implementation of the study. Prior to proceeding with the surveys, LSIC researchers went through a deliberate process of consultation to make certain that the research methodology and aims fit in appropriately with the communities' social and cultural practices. The inclusion of community members in this process served to create a sense of equity between researchers and community members. It is important to recognize the differences between non-Indigenous people and Aboriginal and Torres Strait Islander people, but also the differences within these groups. Results from one person or one community cannot be extrapolated to a wider group.

A routine feedback system has been designed to ensure that participants benefit from the research. The study aims to uncover means to ensure that children have the best opportunity to grow up healthy, strong, and positive. Communities can use the results to inform their use of services and resources. The information can also be used to help governments improve the resources they offer Aboriginal and Torres Strait Islander families, and to help service providers improve the health outcomes of service users. The benefits of the LSIC study and ensuing research will flow to the children involved in the study, to their future children, as well as to other children not directly involved in the study.

Ethical approval for my research using LSIC was granted by the Australian National University Human Research Ethics Committee in October 2011, and a variation to the protocol to include interviews with LSIC RAOs was approved in January 2012. In my research, I adhere to the principles set out in 'Values and Ethics: Guidelines for Ethical Conduct in Aboriginal and Torres Strait Islander Health Research' (112). I have continued to uphold the principles of reciprocity, respect,

equality, responsibility, survival and protection, spirit and integrity, in the design and conduct of my research. To maintain equity and reciprocity, I have consulted RAOs and members of the LSIC Steering Committee throughout the research process to ensure that my research goals remain in line with the interests of the participating communities. Given that concern over the anthropometric data's accuracy was preventing its use (8), FaHCSIA supported and encouraged my endeavour to validate the data, and worked closely with me throughout the data validation process. Upon completion of data cleaning, the birth weight, height, and weight data have been released for analysis by other researchers, to maximise the benefits arising from the collection of these data. My research findings will be submitted to FaHCSIA to be disseminated within participating communities, helping sustain relationships with study participants (32).

(6) Study components and definitions

(a) Birth and early childhood

The primary carer who was interviewed was the mother in the majority of families (92% in Wave 1), but in some cases the primary carer was the father, a relative, a foster carer, or other person (7). Each female primary carer interviewed was asked if she was the study child's birthmother, and if so, if she was comfortable providing details about the pregnancy and birth of the child. She was asked what sources of information she sought during pregnancy, how many antenatal check-ups she attended, and what type of health care services she utilised. She was asked about the health of herself and the baby during the pregnancy, including her use of supplements, alcohol, cigarettes, and other drugs. The mother was also asked about the location of the child's birth, the birthing process, the child's birth weight, and gestational age. Mothers were asked to report this information from their Baby Health Book, as this was considered to be a reliable resource and more accurate than retrospective report. In cases that the Baby Book was not available, mothers were asked to report the birth weight and gestational age from memory. Mothers were also asked if the child was ever breast fed, if the child was still being breast fed at the time of the interview, the age of the child when breast feeding was ceased, why breast feeding cessation occurred, the age of the child at first consumption of other forms of milk, the type of non-breast milk first consumed, and the types of other drinks currently consumed by the child. Additionally, mothers of infants in Cohort B were asked at what age the infant first had solid foods, and if the infant had experienced any problems feeding.

(b) Height and weight

Interviewers measured each child's height and weight at each wave of the study. For the first four waves of the study, Homedics model SC-305-AOU-4209 digital scales were used to weigh children. In the first wave of the study, plastic height measuring sticks were used to measure the height of standing children, and tape measures were used to measure the length of small infants. This measuring equipment was replaced after the first wave of the study by Soehnle professional Model 5003 stadiometers. Measured heights were accurate to the nearest millimetre, and weights were accurate to the tenth of a kilogram. Interviewers were trained to take each measurement three times to verify that measurements were accurate.

(c) Level of Relative Isolation

LSIC uses the Level of Relative Isolation (LORI) scale to indicate the level of remoteness of the communities in which children live. The LORI scale is based on an extension of the Accessibility/Remoteness Index of Australia (ARIA) scale, an 18-point scale that uses distances from each locality to service centres to assess remoteness. This scale does not take into account factors such as socio-economic status, urban versus rural environment, or population size in its determination of remoteness; it uses a purely geographical approach (113). The ARIA scale categorizes areas into five groups: highly accessible, accessible, moderately accessible, remote, and very remote. A highly accessible community is defined as one that has 'relatively unrestricted' access to goods, services, and opportunities for social interaction. Accessible, moderately accessible, and remote communities are defined as those that have 'slightly restricted', 'significantly restricted', and 'very restricted', respectively, access to these goods, services, and opportunities for social interaction. A very remote community is defined as one that has 'very little' access to these goods, services, and opportunities for social interaction (114). LORI has five levels of relative isolation: None, Low, Moderate, High, and Extreme, paralleling these five ARIA categories. However, in LSIC the 'High' and 'Extreme' LORIs have been collapsed into one category (labelled 'High/Extreme'), due to small sample sizes in both categories.

The National Key Centre for Social Applications of Geographic Information System (GISCA) constructed the ARIA scale in 1997, by measuring the road distance from each populated locality to the nearest service centre within each of four categories

based on population size (114). The ratio of each distance to the mean distance for that service centre category was used to rate each locality on a score from zero to three, and this score was summed for the four service centre categories to get a total range of zero to 12 for each locality (114). By interpolating the scores from each locality to a nationwide grid, an ARIA score can be created for each areal unit of one square km (114). ARIA was refined by adding in an additional class of service centres for smaller populations (between 1,000 and 4,999 people), which changed the scale from a 12-point to a 15-point system; this ARIA+ scale was utilized in the 2001 census (114). Although the ARIA and the ARIA+ have been effective in categorizing the overall Australian population, these scales are less effective in describing Aboriginal and Torres Strait Islander communities (115). The ARIA scale masks important differences between groups, classifying communities together under the same ARIA score despite marked differences; for example:

> The WAACHS [Western Australian Aboriginal Child Health Survey] data showed that there were large variations in the circumstances of Aboriginal communities within this region ranging from small regional centres like Fitzroy Crossing with its own hospital servicing the surrounding region, through to truly isolated Aboriginal communities with strong ties to traditional cultures and lifestyles (115 p. 13-14).

To address the inadequacy of ARIA and ARIA+ to classify Aboriginal and Torres Strait Islander communities, the ARIA++ was created by GISCA, with the addition of yet another service centre category (for populations between 200 and 999 people) (115), resulting in an 18-point system (114). This scale was used in the WAACHS analysis, and was found to be effective in distinguishing between 'remote' communities with differing characteristics (115). For example, Halls Creek, which contains a small hospital, and Balgo, a much more isolated town, received the same score of 12 under the original ARIA scale (114). Under the ARIA++ scale, however, these two communities become distinguishable: Halls Creek has an ARIA++ score of 12 and Balgo has one of 18, a more accurate depiction of the disparate access to resources (114).

According to the WAACHS findings, the fairly homogenous set of communities with ARIA++ scores between 17 and 18, and classified under the 'Extreme' level of relative isolation category, would exhibit considerable variation from the set of communities with ARIA++ scores between 13 and 17, falling into the 'High' level of relative isolation category. However, in the LSIC dataset, these two categories are collapsed into a 'High/Extreme' category, obscuring the marked differences between

these groups. A similar level of homogeneity cannot be assumed for communities within this combined category, compared to communities within the single, original categories.

(d) Indigenous Areas

Randomised codes for Indigenous Areas (IARE) are provided in the unconfidentialised version of the data set, allowing people living within the same IARE to be grouped together. However, because the IAREs are coded randomly, it is not possible to link people to their actual geographic location. The IAREs were developed by the Australian Bureau of Statistics to improve the accuracy of mapping for Indigenous communities, replacing the Australian Indigenous Geography Classification (116). There are a total of 429 IAREs, and these spatial areas span the entire geography of Australia, with 109 across New South Wales, 41 across Victoria, 87 across Queensland, 34 across Southern Australia, 71 across Western Australia, 13 across Tasmania, 65 across the Northern Territory, five in the Australian Capital Territory, and five in other territories (including Cocos Islands, Christmas Island, and Jervis Bay) (116). Of these 429 IAREs, 197 are represented in LSIC.

(e) Conclusion

The collection of birth weight, along with longitudinal measurements of height and weight, from a diverse group of Aboriginal and Torres Strait Islander children living across Australia makes LSIC a unique resource. Professor Mick Dodson writes, 'When used wisely, *Footprints in Time* will make a difference for Indigenous children and their families, now and in the future' (7 p. 6).

B) Anthropometric methods: use of references for height and weight

Body weight by itself is not useful as an indicator of weight status; it is necessary to examine weight in the context of height and age. Similarly, height must be examined in the context of weight and age. The three most common anthropometric indices used in the assessment of children's height and weight status are weight-for-age, height-for-age, and weight-for-height (117, 118). Body Mass Index (BMI) can also be used as a proxy for body fat percentage for children.

The WHO created reference values for these four indices (weight-for-height, weight-for-age, height-for-age, and BMI-for-age), based on the growth of multiple samples of healthy children. These indices can be expressed in terms of percentiles or in terms of z-scores. From 1997-2003, as part of the Multicentre Growth Reference Study, height and weight measurements were recorded for 8440 healthy, breast fed infants across Brazil, Ghana, India, Norway, Oman and the United States (119). Longitudinal data were collected for infants from birth through 24 months of age, and cross-sectional data were collected for infants from 18 to 71 months of age. In order to create a reference based on optimal growth, only children living in conditions allowing them to reach their full growth potential were included in the study. For example, children were excluded if their mothers smoked or if there were any known health or environmental factors constraining their growth. Low birth weight (but otherwise healthy) children, however, were included in the study to avoid artificially distorting the lower end of the distribution. The WHO chose to exclude measurements that they believed represented 'unhealthy' heights and weights prior to fitting the model, in order to create a 'healthy' standard (119).

WHO Child Growth Standards (WHO CGS) were established for each of the four anthropometric indicators after an extensive review of 30 different methods of growth curve construction, selection of the most appropriate software package, and application of the selected approach to determine the best model to fit the data for each indicator (119). As a result of this process, the WHO child growth curves were constructed using the Box-Cox-power-exponential (BCPE) method with curve smoothing by cubic splines. Separate standards were created for height-for-age, weight-for-age, weight-for-height, and BMI-for-age z-scores and percentiles. The age scale, along the x-axis, was stretched using a power-transformation for all indicators involving age in order to improve the fit of the curve. The distribution of height-for-age was approximately normal at each age; however, the distribution of the weight-based indicators was skewed, and parameters were added to improve the adequacy of the fit. Goodness of fit tests, worm plots, and residual plots were used iteratively to identify biases or errors in the models (119).

(1) Height-for-age

Height-for-age, or length-for-age, provides a measure of shortness or tallness for an individual compared to the reference population (117). Length is used when referring

to individuals measured in a recumbent position; this typically occurs for individuals up to two years of age who cannot stand well on their own. Length measurements, on average, are 0.73 centimetres greater than stature (height measured in a standing position) measurements for the same individual, and thus these measurements must be adjusted (117).

Low height-for-age is described as shortness. Low height-for-age can be indicative of long-term nutritional inadequacy or health problems. However, there are other potential reasons for shortness (for example, genetics) such that stunting should not be assumed in all cases of low height-for-age without further investigation (117). Stunting is defined as 'a process of failure to reach linear growth potential as a result of suboptimal health and/or nutritional conditions,' thus implying that a pathological mechanism underlies the shortness (117 p. 164). A high height-for-age, described as tallness, is generally not associated with clinical pathology, except in the case of endocrine disorders (117).

(2) Weight-for-age

Low weight-for-age can be described as lightness; use of the term underweight implies a pathological cause for low weight-for-age (117). Weight-for-age is not a sensitive or specific measure for identifying thinness, wasting, overweight, or obesity because the index cannot account for differences in height. A tall but underweight child might have a 'normal' weight-for-age, as might a short but overweight child (118). Weight-for-height and BMI-for-age are preferred indicators of weight status, due to the inability of weight-for-age to account for the vast variation in weight attributable to variation in height (117, 118). As an alternative to these two methods, weight-for-age can be examined together with height-for-age.

(3) Weight-for-height

Weight-for-height provides a measure of relative body proportion. This measure is age-independent, a valuable feature in situations in which age is not collected or is not reliable (118). Low weight-for-height is described as thinness; use of the term 'wasting' implies that a pathological process, such as starvation, nutrient deficiency, or severe disease, is responsible for the low weight (117). High weight-for-height can be described as overweight. Overweight can be caused by high levels of adiposity

(obesity), but it can also be caused by a high lean body mass. Therefore, high weight-for-height cannot be assumed to be fatness on the individual level; on the population level, however, a high mean weight-for-height can be assumed to be an indicator of fatness, as this is significantly more common than having a high weight-for-height because of lean mass. Technically, use of the word 'obese' should be reserved for cases in which adiposity was directly measured, such as measurements of skinfold thickness (117). Although weight-for-height is a better indicator of overweight than weight-for-age, it should not be used on its own; 'Weight-for-height does not serve as a substitute for height-for-age or weight-for-age, since each index reflects a different combination of biological processes: although they may share common determinants, they cannot be used interchangeably' (117 p. 165). A limitation of weight-for-height as an indicator is that it does not account for age, which influences the relationship between weight and height (120). The lack of incorporation of age can serve as an advantage, however, in settings where children's age is inaccurate or unavailable (120). Each of the four anthropometric indicators provides distinct information about height and weight status (117).

(4) BMI-for-age

BMI, as a proxy for adiposity, is used to classify individuals into categories of thin, healthy weight, overweight, and obese. BMI is calculated by dividing an individual's weight (in kilograms) by the square of the individual's length or height (in metres). BMI for adults is age-independent; in children, however, the meaning of BMI, and therefore the cut-offs for identifying low and high BMI, vary by age. Thus, a BMI-for-age assessment is necessary to give the measure meaning (117). A low BMI-for-age indicates thinness; high BMI-for-age represents overweight (or obesity, though this term is technically reserved for cases in which body fat is directly measured) (117).

(5) Validity of the Body Mass Index

Obesity is most accurately assessed by measuring percentage body fat, but this is not feasible in most large-scale studies, so BMI serves as a proxy measure (56). Several studies have evaluated the accuracy of BMI by comparing its ability to identify adiposity to that of other indicators including skinfold thickness, waist circumference, dual-energy X-ray absorptiometry scans (which calculate adiposity by determining the

amount of absorption of X-ray beams), and bioelectrical impedance analysis (which estimate fat-free body mass by determining the amount of opposition to electric current flow through the body, and subtracting this mass from total body mass to determine body fat mass). These studies, overall have found a high correlation between levels of adiposity as measured by BMI and by these other indicators (121).

Pietrobelli et al. (121) modelled the relationship between BMI, total body fat (TBF, in kilograms) and per cent of body weight as fat (PBF) in a representative sample of 198 Italian boys and girls aged five to 19 years There was a strong association between BMI and TBF and PBF for both genders, with correlations ranging from 0.63 to 0.89. In seven similar studies using anthropometric measurements, such as skinfold thickness, to evaluate the accuracy of BMI as an indicator of adiposity, the correlation between BMI and these measures ranged from 0.68 to 0.85. Further, correlations between 0.65 and 0.82 were observed when comparing BMI to bioelectrical impedance analysis in four studies (121). Providing additional support for the use of BMI, studies have shown that the tracking of BMI through childhood and young adulthood is stronger than that of skinfold thickness (122). Many studies have also found significant associations between BMI and health outcomes, including blood pressure and serum insulin levels in children (121). BMI should not be used to predict an individual's TBF or PBF (121), but these findings suggest that BMI provides an accurate estimate of adiposity on the group level, and is appropriate as a screening tool for overweight, especially in situations when more technical measures are not feasible (121).

A common criticism of the use of the BMI to assess adiposity is that the measure cannot differentiate between the contribution of fat mass and fat-free mass to an individual's weight (123). A very lean, muscular individual and a very adipose individual could have the same BMI; interpreting BMI values on the individual level, therefore, requires caution. Additionally, the BMI does not take into account the age of puberty; although this would add considerable complexity to the measure, it would increase the meaningfulness of the BMI measurement as BMI is expected to be elevated in individuals who have gone through puberty compared to others at the same age who have not (5). The use of BMI in pre-pubescent children (as in the LSIC sample) is not affected by this concern.

The use of BMI has also been criticised due to its low sensitivity. Reilly and colleagues (123) examined the accuracy of BMI in identifying body fatness in 4,175 children aged 88-92 months, and found that it had a high specificity (99% for both

genders) as a measure of obesity, but a low sensitivity (46% for males and 72% for females) According to their analysis, the use of BMI as a screening tool will rarely incorrectly categorise non-obese children as obese, but may fail to categorize obese children as obese (123). The authors diminish the significance of the lack of sensitivity, writing:

> Historically, low sensitivity of BMI as a screening tool for clinical practice has been regarded as acceptable so long as its specificity was high… low sensitivity is a potential limitation of the BMI, but the present study suggests that, while BMI is not ideal, this limitation can be minimized in children with choice of an appropriate cut-off… there is currently no evidence that sensitivity and specificity of obesity screening differs significantly during childhood before puberty (123 p. 1626).

Similarly, Lazarus and colleagues (124) concluded that BMI was acceptable as an indicator of weight status for screening at the population level; the importance of high specificity exceeds that of high sensitivity in this case, especially given the social ramifications of labelling a child as overweight.

A third criticism of the BMI concerns its applicability to non-white populations (125). As with the use of any reference, there are likely to be differences in the accuracy of the BMI in identifying adiposity across groups of different ethnicities. In this case, the tendency of Indigenous Australians to deposit fat centrally affects the accuracy of the BMI, often leading to underestimation of adiposity level in this population (122, 126, 127). The variation in adiposity that is attributable to ethnicity, however, is small in comparison to the variation attributable to socioeconomic factors, nutrition, and health (117). In using BMI as an indicator of adiposity in this population, the potential underestimation of adiposity should be acknowledged, but the measure should not be considered invalid (57). Additionally, in the case of BMI, the escalating epidemic of obesity prevents the creation of local, current references; De Onis and Lobstein explain, ‘As soon as a new reference is produced, it is out of date. Furthermore, it is not possible to make accurate comparisons between local countries when each one has used its own local reference curve’ (128 p. 459).

The relatively low sensitivity, inability to differentiate between muscle and fat mass, and uncertain applicability to non-white populations are all valid concerns regarding BMI, but do not invalidate its use. Rather, they reiterate the importance of using BMI to make population-level, rather than individual-level assessments. Cole concludes that the BMI is a ‘simple yet “good enough” tool to compare prevalences

across populations that are inevitably heterogeneous' (5 p. 6). Acceptance of the measure is now widespread; De Onis and Lobstein write:

> There is now broad international consensus about the utility of the WHO Child Growth Standards for assessing the growth of pre-school children. Because the standards depict physiological human growth under optimal environmental conditions, they provide an improved tool for assessing growth. The WHO standards have been well received worldwide and, at the time of this writing, they have been adopted by over 110 countries and many researchers (128 p. 458).

Thus, use of BMI as a measure of weight status is generally accepted in public health research internationally.

The application of BMI within Indigenous research, however, raises further concern, given observed differences in body composition in Indigenous and non-Indigenous Australians. Wang and colleagues write, 'Due to differences in body shape and physiological and environmental factors between Aborigines and other populations, the health implication of a given BMI may be different to the reference population' (56 p. 573). Sellers and colleagues (60) propose that a BMI cut-off for overweight of 22 kg/m^2 (as opposed to the standard 25 kg/m^2) might be appropriate for Indigenous adults, as the incidence of diabetes increases significantly above this point. Modified cut-off points may also be appropriate for children; a study examining Metabolic Syndrome (MetS) in Indigenous children found that use of the standard BMI cut-offs significantly underestimated children's risk of MetS (60). Adjusted cut-offs might better reflect the deleterious health outcomes associated with elevated adiposity levels within the Indigenous population.

Cunningham and Mackerras (57) also question the appropriateness of the standard BMI cut-off points for Aboriginal and Torres Strait Islander individuals, but conclude that their use is acceptable:

> While there are certain problems with the reference curves, and they do not necessarily represent the ideal, they are based on a large sample of well-nourished children and seem to be a good general description of the growth of well-nourished pre-adolescent children for most cases, except perhaps Asians. They have not been compared to the growth of a large group of well-nourished Aboriginal or Torres Strait Islander children, but various authors have reported that the growth profile of Aboriginal children in some locations was the same as that of non-Aboriginal children. Hence there is no reason to think that the growth of well-nourished Aboriginal pre-adolescent children would differ substantially from that of other races (57 p. 6).

Despite its limitations, BMI is appropriate for the assessment of weight status if the potential for the underestimation of adiposity is acknowledged. The health outcomes associated with adiposity may arise at a lower BMI in Indigenous individuals than in non-Indigenous individuals, given the tendency towards central fat deposition. Thus, a high BMI, as defined by international standards, may have an even stronger association with disease in Indigenous populations.

(6) Use of an international reference

A reference is used as a basis for comparing measurements across different groups or populations. Debate has arisen over whether different references should be created according to factors such as country or ethnicity. Although there is evidence of variation in height and weight between different ethnic groups, the magnitude of these differences is small in comparison to the differences attributable to health, nutrition, and socioeconomic status. For example, variation in socioeconomic status alone can explain 12% of the difference in height, and 30% of the difference in weight of preschool children across different countries (117). Multiple studies have proven that children, when healthy and provided with adequate nutrition, follow similar growth curves regardless of ethnicity or country of residence (117). The value of a country- or ethnicity-specific reference would depend on the circumstance, but, 'when reference data are to be used to make decisions about populations, it is better to use statistical methods to control for differences (such as those associated with different altitudes) within or across populations than to use two different standards' (117 p. 29-30). Research has demonstrated that Indigenous and non-Indigenous Australians exhibit similar growth trends, suggesting that, 'Aboriginal children are capable of achieving expected patterns of growth-for-age' (126 p. 264). These findings support the use of international standards for Indigenous Australians.

Given the demonstrated potential for any child, regardless of ethnicity, to achieve maximum growth given proper nourishment, a universal reference, based on healthy children, is appropriate. The use of a single reference has the advantage of allowing comparison across groups and countries, while still being applicable for local populations (117). In addition, a universal reference is better able to identify cases of low (or high) weight and height than a local or ethnicity-specific reference. That is, if an entire population suffers from poor nutrition or poor health and has deficits (or surfeits)

in weight, a local reference would be based on a thin (or overweight) population, and therefore would be less able to pick up individuals with low (or high) weight (117). The Northern Territory Government Department of Health and Community Services have recommended the use of the WHO international references for all health services, for the following reasons:

> a. The data is derived from an internationally representative sample of infants and young children who received optimal nutrition including exclusive breastfeeding until at least 4 months of age.
> b. The charts can be used as a growth standard from 0-5 years.
> c. The relatively rapid rise in growth in the first few months is what we see in exclusively breast fed infants and will not lead to unnecessary intervention.
> d. Overweight children will be identified earlier and more accurately allowing interventions to be made.
> e. The charts can be used as a growth reference for children 5-19 years who still require growth monitoring.
> f. There is a strong argument for consistency in use of charts between health services (129 p. 20).

Thus, these international references are appropriate for use within Indigenous populations in Australia.

(7) Standardisation of height and weight

Z-scores are used to standardise anthropometric data, so that individuals can be compared to a reference population. A z-score indicates the magnitude by which an individual's measurement varies from the median measurement in the reference population. A z-score is calculated by finding the difference between the observed measurement for the individual and the median reference value for individuals of the same age and gender, and dividing this difference by the standard deviation of the reference population. The reference curves are normalised so the distribution of each index follows a normal distribution; 68% of the population lies within one standard deviation of the mean (with a z-score between -1 and +1), 95% of the population lies within two standard deviations of the mean (with a z-score between -2 and +2), and 99.7% of the population lies within three standard deviations of the mean (with a z-score between -3 and +3). It is important to be cautious in using z-scores to make inferences about an individual's health status; although z-scores can provide

information about height and weight status, they cannot provide any information about the cause of any deviation from the reference value (117).

The distribution of z-scores for observed data is likely to be normal, and this enables the use of analytical procedures which assume normality of data, such as t-tests and regressions (118). Percentiles, in contrast, are not normally distributed, but rather have a uniform distribution; thus, analyses based on assumptions of normality cannot be utilized (118). Another limitation of percentiles is that their range is limited: it is not possible to track children who fall outside the measurable limits, such as a child whose weight-for-age is below the first percentile (130). By using z-scores, in contrast, a precise assessment of the magnitude of their departure from the median is possible, and change over time can be monitored even within the extreme ends of the distribution. A further advantage of using z-scores over percentiles is that the mean and standard deviation of z-scores can be calculated for a group to create a population-level indicator. A mean z-score significantly different than zero usually indicates that the whole distribution has shifted relative to the reference distribution: 'all individuals, not only those below a given cut-off point, are affected' (117 p. 23). Therefore, examining differences in mean z-scores between populations provides a description of the whole population, rather than describing only the subset of the population below a defined cut-off, as occurs when comparing the prevalence of specific height and weight outcomes. Additionally, the statistical power is greater in comparing mean z-scores versus comparing the prevalence of a certain category of z-scores (117).

On the population level, interpretation of z-scores is dependent on the specific index being used (117). Each anthropometric indicator can provide unique information, and thus caution must be used in selecting the most appropriate indicator (117). 'Normal' weight-for-age, height-for-age, and weight-for-height measures include those within two standard deviations of the reference median (with $|z| \leq 2$), thus encompassing the middle 95% of values of the reference distribution (117, 118). This definition is based on the statistical distribution, by defining the outlying 5% as 'abnormal'. This cut-off point is not based on a functional or health-related outcome:

> It should be noted that the convention of using the central 90% or 95% of a given distribution to define cut-offs or reference ranges does not truly define the 'normal' range from the point of view of health or nutrition; rather, it is used as a guide to facilitate clinical screening or population-based surveillance. Obviously, a child who was originally at +2 z-scores and falls to, say -1.5 z-scores because of malnutrition will show wasting despite remaining above the -2 cut-off (117 p. 182).

It would be beneficial to determine cut-off points based on physical outcomes, such that an individual below the cut-off would be at a demonstrated higher risk of pathology. However, the definition of such outcome-based cut-offs has not yet been feasible (118). No study has obtained long enough follow-up data collection on a sample large enough, with the correct distribution of demographic variables. Thus, the use of the statistical-based z-score cut-off points is necessitated. In a longitudinal survey, change in z-score within individuals can also be examined.

The dramatic fluctuation of weight and height throughout childhood makes a BMI-based definition of overweight difficult for children, and necessitates the use of age-specific cut-offs (6). Health outcome-based cut-off points would be ideal, but the point at which the health impact of overweight becomes significant is difficult to determine in children, due to the time lag of the effect of overweight on health status (6, 131). Thus, statistical-based cut-off points are again necessary. Additionally, the health impact of 'excess' weight varies between preschool children, older children, and adolescents who have reached their full growth potential, thus necessitating different cut-offs for children under five years of age and those aged five to 19 years (128).

A 1997 workshop organized by the International Task Force on Obesity proposed using BMI percentile cut-offs for children that are comparable to those used to define the adult cut-offs of 25 kg/m^2 for overweight and 30 kg/m^2 for obese (122). In creating the reference, a representative sample of over 60,000 individuals aged six to 18 years from six countries (Brazil, Great Britain, Hong Kong, the Netherlands, Singapore, and the United States) was used (6). The WHO excluded all children with unhealthy weight-for-length/height (128). At age 19 years, the standard BMI cut-offs of 25 kg/m^2 for overweight and 30 kg/m^2 for obese approximate one and two standard deviations above the reference median, respectively (128). These cut-off points, therefore, were extended down to children five years of age (128). Thus, for children aged five to 19 years, according to WHO standards, a BMI-for-age z-score between +1 and +2 is an indicator of overweight (comparable to an adult with a BMI of 25 kg/m^2 to 30 kg/m^2), and a BMI-for-age z-score greater than +2 is an indicator of obesity (comparable to an adult with a BMI exceeding 30 kg/m^2) (132).

However, the WHO employs different standards for children younger than five years of age (see Table 7). The WHO decided to set conservative cut-offs for this age group because of the lack of information concerning the health impact of BMI during early childhood (128). WHO classifications dictate that a child whose BMI is between

one and two standards deviation above the reference median is classified as 'at risk of overweight,' a child with a BMI between two and three standard deviations above the reference median is classified as 'overweight,' and a child whose BMI is more than three standard deviations above the reference median is classified as 'obese' (133). For both age groups, a BMI-for-age z-score of -2 is the cut-off point for low BMI-for-age, equivalent to grade 2 thinness for adults (a BMI less than 17) (5).

Table 7: WHO and IOTF cut-off points for BMI-for-age z-scores

	Underweight (grade 2 thinness)	**Healthy weight**	**Overweight**	**Obese**
Children less than five years old	$z < -2$	$-2 \leq z \leq +2$	$+2 < z \leq +3$	$z > +3$
Children five to 19 years old	$z < -2$	$-2 \leq z \leq +1$	$+1 < z \leq +2$	$z > +2$

Despite limitations such as the unknown health impact of BMI for children, these BMI cut-off points are thought to represent 'a reasonable step to establish clinically relevant definitions of overweight among children and adolescents' (122 p. 2). Until further studies demonstrate the BMI level at which deleterious health outcomes occur for each age, the use of two sets of cut-off points based on z-scores remains the best option. De Onis and Lobstein explain:

> Clearly, there is need for research into the health outcomes associated with these different cut-off points. Given that childhood and adolescence are periods characterized by rapid growth and physiological change, it is entirely possible that a given centile represents varying levels of risk depending on age and stage of development (133 p. 406).

The cut-offs will be modified if future research does reveal a varying risk, but currently represent the best method for an internationally-comparable classification of weight status (133). De Onis and colleagues (130) administered a questionnaire to health authorities in 219 countries and territories to evaluate the implementation of the WHO Child Growth Standards, and found that 125 countries had already adopted the standards by April of 2011, just five years after their development. The most widely accepted indicator was weight-for-age, but the use of additional indicators has increased since 2000. Support for the use of BMI-for-age was increasing, with 36 countries reporting recent adoption of the standards.

C) Anthropometric methods: use of references for birth weight

Historically, raw values for birth weight have been analysed, using the standard cut-off points of 2,500 grams for low birth weight and 4,000 grams for high birth weight. More recently, studies have begun evaluating birth weight in the context of gender and gestational age. These studies generally employ a categorical approach, classifying infants as SGA, AGA, or LGA, rather than examining size for gestational age on a continuous scale. Various birth weight references have been constructed, enabling the comparison of an infant's birth weight to the median for other children, adjusted for an array of factors including gender, gestational age, country, ethnicity, maternal size, and parity. In the construction of birth weight references (similar to the construction of the WHO Multicentre Growth Standards), only non-pathologic births are included; 'otherwise, assessment of fetal growth is done against an inaccurate optimal weight' (90 p. 299). The inclusion of only optimal weights in the reference is necessary because, 'To be able to study and measure the effect of pathological variation, the standard should not be set by a population average, but by what should be expected under ideal circumstances… to 'escape the uncertainty of the dividing line between normal and abnormal', as the boundaries of 'optimal' are narrower than the boundaries of 'normal' (134 p. 2). This justifies the exclusion of mothers with hypertension, multiple pregnancies, diabetes, or other factors associated with low or high birth weight.

The WHO CGS standards are based on the Multicentre Growth Reference Study which aimed to only include infants facing no barriers (maternal health, environmental concerns, or economic constraints) to achieving their full growth potential. Thus these data 'are currently considered the best description of healthy growth for children worldwide from birth to 5 years of age regardless of ethnicity and socioeconomic status' (135 p. 127). Although low birth weights for term infants were included in the Multicentre Growth Reference Study (2.3% of the sample), birth weights were excluded if they were lower than 1,500 grams. The WHO CGS standards can be used to generate gender-specific z-scores for birth weights based on this reference, with no adjustment for gestational age. Although the adjustment of birth weight for gestational age has been criticised due to concerns about confounding (136), it is generally considered necessary in order to disentangle maturity and growth status, and to provide an indicator of size for gestational age and potential IUGR (137). Unfortunately, an international reference adjusting for gestational age does not yet exist.

There is a near-linear association between gestational age and birth weight, with an average 25-gram increase in birth weight for each additional day of gestation (90). Some birth weight references are based on the distribution of weights of live births at each gestational age, whereas other references are based on ultrasound-measured weight of unborn foetuses (in order to better reflect optimal foetal growth). Because pre-term births are more likely than term births to be influenced by pathology, they are more likely to have experienced IUGR, and thus do not represent the size of unborn foetuses at this age (138). Hadlock et al. (139) developed a formula to represent optimal foetal weight (in grams):

$$Foetal\ weight = exp(0.578 + 0.332 \times gestational\ age - 0.00354 \times gestational\ age^2)$$

This equation is based on ultrasound measurements at ten to 41 weeks gestation from of 392 pregnant mothers in the United States. References of birth weight for gestational age are often constructed using this type of formula combined with linear regression. Using these references, whether based on live birth weights or ultrasound foetal measurements, the standard definition of small-for-gestational age (SGA) is a birth weight falling in the lowest decile for infants of the same gender and gestational age.

(1) Use of international, country-specific, ethnicity-specific, and fully-customised birth weight references

Many researchers have proposed the use of birth weight references specific to ethnicity (135, 140). However, given the purpose of references to reflect the optimal foetal weight and the true prevalence of small for gestational age in a population, Mikolajczyk and colleagues (138) suggest that a reference should not be adapted to a particular ethnicity. In creating a reference specific to one ethnic group, there is an implicit assumption that the prevalence of small for gestational age in the population is 10%, and thus, 'that fetal growth restriction is similarly common across populations. This assumption might not be true in populations heavily affected by undernutrition' (138 p. 1860). For example, the median birth weight for gestational age should be based on a sample of mothers with adequate nutrition, rather than using a sample of undernourished mothers, even if that is common within the sub-population of interest. By using an optimal standard, rather than one specific to an ethnicity, the reference can reflect the 'local risk profile' (138 p. 1860) for SGA instead of the default prevalence of 10%. Within Australia, the prevalence of low birth weight is twice as high in the Indigenous, compared to non-Indigenous, population (45). The lower birth weight

observed in this population should not be assumed to be a genetic factor associated with Indigenous ethnicity, but rather should be considered attributable to environmental and maternal factors (enacted across her life-course) influencing both pre-natal life and developmental trajectories (141). If a reference specific to the Indigenous Australian population was used, only Indigenous infants falling in the lowest decile of birth weight for gestational age according to this Indigenous-specific reference would be identified as SGA. However, given that the prevalence of SGA exceeds 10% within this population (89), some infants categorised as AGA (falling between the 10^{th} and 90^{th} percentiles) according to an Indigenous-specific reference would actually be SGA. Therefore, categorising children based on an Indigenous reference would underestimate the true prevalence of SGA and would ignore the influence of factors beyond ethnicity itself on birth weight.

Along with variation attributable to ethnicity, birth weight increases with increasing maternal weight and height, and increases with parity (there is an average 87-110 gram increase in birth weight for a mothers' second birth, with a less notable increase for successive births) (90). Given the breadth of factors associated with birth weight, some researchers (90, 134, 142, 143) propose the use of not only ethnicity-specific, but fully-customised references in which the variation in birth weight attributable to maternal factors is accounted for through the adjustment for factors including maternal ethnicity, age, weight, and height in these references (see Appendix G). The use of these customized references improves the ability of size for gestational age category to predict perinatal morbidity (90) and metabolic disturbances in adulthood (143). Despite the enhancement of predictive validity, the practicability of the use of the customised reference is limited, and thus, the benefit of using a fully customised reference, compared to a country-specific reference, is minimal (138). In the case of LSIC, several variables of interest (including maternal weight and height) are recorded, but these data are incomplete, so the sample size would be greatly reduced if a customised reference were used. The marginal benefit of using a customised, rather than country-specific, reference is therefore further diminished for this study.

(2) Australian birth weight references

Roberts and Lancaster (89) created the first set of Australian birth weight percentile charts based on a national sample of both Indigenous and non-Indigenous infants. The charts are gender-specific, and based on a sample of 769,077 singleton

infants born between 1991 and 1994 to Australian-born mothers. The Australian Institute of Health and Welfare National Perinatal Statistics Unit (AIHW NPSU) provided the birth information; this database contains information for all births with a birth weight greater than 400 grams and a gestational age greater than 20 weeks (89).

Gestational age of births in the sample is based on the number of weeks since the last menstrual period (LMP) or clinical estimations; Roberts and Lancaster (89) propose that inaccuracy in the report of gestational age is the predominant cause of outlying birth weights in the study. The risk of inaccurately reported gestational age is increased for Indigenous compared to non-Indigenous infants because some States and Territories rely on prenatal and/or postnatal assessments to estimate gestational age, rather than calculating the weeks since the mother's LMP. However, using simulation models for infants in Queensland, the disparity in median birth weight for gestational age between Indigenous and non-Indigenous infants was not attributable to the miscalculation of gestational age (89). Thus, although it is recognised that several factors impede the collection of accurate gestational age data for Indigenous infants, gestational age should not be assumed to be unreliable in this sample (89).

Less than four per cent of mothers in the sample were identified as Indigenous (89). After excluding outliers and individuals with missing data, there were 734,145 non-Indigenous infants and 27,757 Indigenous infants included in the study. Of these Indigenous infants, 11.6% were pre-term, compared to only 5.4% of the non-Indigenous sample (89). Overall, the mean birth weight (and standard deviation) for non-Indigenous males and females was 3,476 (±550) grams and 3,345 (±516) grams, respectively, and the mean birth weight (and standard deviation) for Indigenous males and females were 3,246 (±632) grams and 3,128 (±595) grams, respectively (89). As SGA and LGA are defined by percentiles, 10% per cent of the sample would be expected to fall in each category; within the Indigenous sample, however, the prevalence of SGA birth was nearly twice as high, at 17% (89). The per cent of Indigenous infants classified as SGA increased with increasing gestational age. Thus, in addition to being more likely than non-Indigenous infants to be born pre-term, Indigenous infants in the sample were more likely to be born at a small size even if they were born at term.

Additionally, Roberts and Lancaster (89) observed significantly more heterogeneity in birth weight and gestational age for Indigenous, compared to non-Indigenous, infants, within each State and Territory. This variability in the birth weight of Indigenous infants across States and Territories may be partially attributable to the

smaller sample size for Indigenous infants, or to differences in the method of recording birth weight for Indigenous vs. non-Indigenous births (89). Alternatively, the wide variation observed within the Indigenous population might be explained by the inclusion of infants of both Aboriginal and Torres Strait Islander background, and the inclusion of infants living in households across the spectrum of urban to remote (89). Despite these observed differences in the distribution of birth weight and gestational age for Indigenous and non-Indigenous infants, the use of separate standards is not necessitated. Environmental factors such as maternal smoking, nutrition, and access to health services likely have a larger role in determining birth weight than Indigenous status itself (89). The authors conclude that 'genetic influences are relatively unimportant in explaining birth weight variation' because they are obscured by these environmental factors:

> Until the effect of such population and environmental influences is ascertained, we believe it would be unwise to publish percentiles that imply a separate population norm for Indigenous infants – to do so might lead clinicians or researchers to accept that Indigenous infants are genetically smaller, whereas the clinical characteristics suggest otherwise (89 pp 117).

Thus, an Australian representative birth weight reference is appropriate for use with an Indigenous sample, including LSIC.

Since the publication of the reference by Roberts and Lancaster (89), maternal age has increased, smoking during pregnancy has decreased, maternal overweight and obesity have increased, and the ethnic makeup of Australian mothers has changed (17). To reflect the changing characteristics of Australian mothers, Dobbins et al. (17) created an updated national reference of Australian birth weight for gestational age, using the same data source, the AIHW NPSU. The reference includes 2.53 million singleton live term infants born in Australia between 1998 and 2007. The date of the mothers' LMP was used to calculate gestational age, or if this was unavailable, gestational age was estimated from pre- or post-natal assessments (17). Birth weights were considered outliers if they fell either twice the interquartile range below the first quartile for all birth weights of the same gestational age and gender, or if they fell twice the interquartile range above the third quartile (the same exclusion criteria used by Roberts and Lancaster (89)). From the original sample of 2,539,237 infants, 8,986 were excluded because they were determined to be outliers, and an additional 1,610 were excluded because they were missing at least one of the key variables. A minimum number of birth weights were required to plot each percentile: a minimum of 100 birth

weights was necessary for each gender and ethnicity to define the 5th and 95th percentiles (17). In the creation of this reference, mothers born outside of Australia were included. Unfortunately the Indigenous status of infants was not disclosed due to ethical considerations, so comparisons of Indigenous and non-Indigenous births in the sample is not possible.

In the 1998-2007 sample, 5.9% of infants were born before 37 weeks gestation, 4.8% of infants were low birth weight, and 1.9% were high birth weight (defined as a birth weight exceeding 4,500 grams) (17). There were no significant changes in the mean birth weight for males or females from year to year (17). Compared to the findings of Roberts and Lancaster from 1991 and 1994, the current sample has a median birth weight (not adjusted for gestational age) between five and 45 grams higher for term females and between zero and 25 grams higher for term males (17). Although a small absolute increase in weight, this shift increases the birth weights associated with the 10th and 90th percentiles for each gestational age and gender, thereby increasing the categorisation of infants as SGA and decreasing the categorisation of infants as LGA compared to using the 1991-1994 reference. Despite the decreased rate of maternal smoking during pregnancy (a risk factor for pre-term birth), the prevalence of pre-term birth has increased from 6.8% of births in 1994 to 7.4% of births in 2007 (17); this might be partially attributable to the increased rate of cesarean section (144).

Limitations of the reference include the lack of identification of Indigenous status, and the potential for inaccurate recording of gestational age. Further, the reference reflects the actual, rather than optimal, birth weight distribution: the reference uses live birth weights instead of ultrasound-determined weights of unborn foetuses, and the reference includes all available birth weights, rather than excluding mothers with any known risk factors for high or low birth weight. Additionally, this reference adjusts for gestational age and gender, but no other factors. Dobbins et al. (17) acknowledge the arguments for the use of a customised reference, but conclude that these arguments are:

> based on inconsistent evidence … Whether differing points of view and fine areas of disagreement on customised and conventional birthweight percentiles for gestational age have important practical research or clinical implications is questionable (17 p. 294).

Despite the acknowledged limitations, this reference is the most up-to-date reference based on a nationally representative Australian sample, and is a valuable tool for assessing Australian birth weights. In the absence of an international reference of

birth weight for gestational age, this reference (17) is the most appropriate for the calculation of z-scores for birth weight for gestational age for the LSIC sample.

Chapter IV: Are the anthropometric data collected in LSIC valid?

The validity of the height, weight, and birth weight data in LSIC needs to be explored. Many researchers have expressed interest in using the anthropometric data, and participating families have expressed curiosity about the purpose of their collection; one LSIC interviewer explained, 'I'm getting to the point at now where I need to start showing results to our families.' However, the release of these data has been prevented by concerns over data quality (8). The LSIC team desired for the data to be transformed to a useable form, so that participants could benefit from their participation in the measurement process. I was provided with the data in raw form, and undertook an appraisal of the data's validity. I evaluated the data quality based on analysis of height-for-age, weight-for-age, BMI-for-age, and birth weight z-scores. Height, weight, and age were converted to z-scores using WHO international references, and z-scores for birth weight and gestational age were calculated using a nationally representative Australian reference. Prior to completing these analyses, I held interviews with LSIC Research Administration Officers (RAOs) who collected the height, weight, and birth weight data in LSIC. The perspectives shared with me by these key informants informed my analyses of data quality.

A) Interviewers' evaluation of LSIC anthropometric data

(1) Interview methods

To complement the quantitative analyses of data quality, I conducted interviews with Research Administration Officers (RAOs) who interviewed families in LSIC. These interviews served to examine the cultural, social, and environmental factors influencing the LSIC interviews and data collection, and to determine factors that might improve the collection of data in future waves of the study. These interviews provide the RAOs with the opportunity to tell stories of their experience conducting these interviews, contributing to a collective story (34). I regarded the RAOs as 'key informants'; experts with a detailed and extensive knowledge of the practice of conducting the LSIC interviews. Time and geographical constraints led to the

recruitment of eight RAOs (representing 73% of the RAOs currently employed) to participate.

One face-to-face interview was conducted on the 20th of May, 2012. A focus group was conducted on the 19th of June, 2012 at an RAO training session at the FaHCSIA Office in Tuggeranong, with eight of eleven current RAOs consenting to participate (including the RAO who participated in the face-to-face interview). In June and July of 2012, follow-up phone interviews were conducted with three RAOs who expressed an interest in participating in a more in-depth conversation. The face-to-face interview and focus group both ran for about one hour, and each phone conversation was approximately thirty minutes.

All of the RAOs participating in the interviews and focus groups were Aboriginal or Torres Strait Islander, although some non-Indigenous people did conduct surveys in the first wave of the study. Two males and six females participated in the conversations (in the following section, all interviewers will be referred to as females so the few males are not identifiable). RAOs ranged in age from 25 to 57 years. They had varied levels of experience; some were in their first year of work for LSIC, whereas others had been a part of the survey since its conception. The participating RAOs have conducted interviews across the country, in areas ranging from Western Sydney to Galiwinku.

Prior to commencing the interviews and focus groups, RAOs were provided with information and consent forms (see Appendix A and Appendix B). Participants read through these forms and were given an opportunity to discuss any questions or concerns. If they were comfortable proceeding with the interview, they were asked to sign the consent forms, and informed that they were able to withdraw from the study at any time. Conversations were recorded using an Olympus DS-660 Digital Voice Recorder, and transcribed by myself.

The focus group and interviews were composed of semi-structured open-ended questions (see Appendix C), aimed to reveal more about the experience of conducting LSIC interviews. The use of open-ended questions enables participants to convey their attitudes and experiences through the use of stories, rather than restricting their answers (145). The central topics discussed included: reasons for involvement in the LSIC study, families' perceptions of the LSIC study, benefits of conducting the interviews, challenges facing interviewers, measurement of children's height and weight, and collection of food recall data. The use of follow-up questions or probes depended upon the participants' interest and the natural flow of

the conversation. In the focus group, specific questions regarding the waves and sites at which RAOs conducted interviews were not asked. The questions posed were more general than those asked in the one-on-one interviews, and were designed to enable multiple interviewers to share their perspective on each issue. If interviewers were unresponsive to a question, a new question was asked. My analysis of these key informant interviews was based on the approach described by Marshall (146). Analysis was completed manually and based on responses to particular topics but also included an examination of emergent themes such as 'trust,' which was clearly an important issue, arising through the course of each interview.

(2) Interview results

Conversations with the LSIC RAOs centred on their experience conducting interviews, the unique challenges they face in their role, and the measurement of children's height and weight. RAOs shared stories of their positive and negative interview experiences, of their favourite memories and of their frustrations. The positive interaction with family members was consistently mentioned as a highlight of the job, whereas travel time, rejection, and emotional load were noted as downsides of the job, contributing to burnout. Underlying these conversations was a theme of trust; trust was described as essential for the maintenance of a positive rapport with participants, and thus for the continuation of the study. This personal insight into the interview process from these key informants is formative in understanding the prevalence of missing data in the study and in creating appropriate methods for cleaning the data.

(a) Reasons for involvement with LSIC

Overall, interviewers had a positive outlook on their jobs. One RAO stated that conducting interviews for LSIC was 'the best job in the world,' and that she hoped to continue with LSIC for at least an additional four years. Motivations for working on LSIC included: forming links to the community, making a difference for others, giving families a voice, and forming relationships. One interviewer explained, 'My Mum passed away at like fifty-nine [years old] and I can never really remember her being well … You know, I just keep working like this … maybe part of what I'm doing will help others.' Another RAO said:

> The job allows you to stay involved with people you might not otherwise, and meet people you wouldn't see otherwise … All of us get to work in our local communities;

> for me, it's about reconnecting with people I wouldn't have, not reason to, but a sort of a push to. I think that's been a really rewarding thing.

The RAO further explained that, in previous jobs:

> I loved the customer service, the one-on-one interaction you get with people, but I'd always wanted to work with my community, in my Indigenous community. This opportunity came up, so I thought I had the skill-set, and that I could adapt or train over time to become half decent at it, and it turns out I was right.

Many RAOs expressed a similar appreciation of the opportunity to work within their own community. The benefits associated with RAOs working in their own communities are numerous, not only for the RAOs themselves, but for the success of the study. Local RAOs are aware of the cultural, social, and economic issues facing their communities; this helps reduce misunderstandings and misinterpretations (28).

Providing feedback to communities is a high priority for the interviewers; one said:

> Some people wanna know, you know, where it's [the survey] going … Last year [a primary carer] she said she was sick, or she was sick, and this year I finally got her on the phone and she goes, "Uh, how is this gonna benefit my grandchild?" And I said, "Well, directly to her, it's not." I said, "It's for all Aboriginal kids." And because it's a longitudinal study it's going to take time before, you know, hopefully sooner rather than later, that there will be some sort of positive change.

The RAO expressed excitement that researchers are now starting to use the data. Another RAO said, 'I'm getting to the point at now, where I need to start showing results to our families. They are out there, it's just a matter of having the time and finances to promote them I guess.'

Many RAOs stated that the children, and their excitement to see the RAOs, would often make their day; these positive interactions are part of what makes the job so rewarding. One RAO explained that when children know the interviewers are coming, 'they'll wait up until you get to that door and the Mum will say, "Oh yes, she's been waiting for you all day."' An RAO shared a favourite memory of her experience with LSIC: 'Like one of them which always sticks in my head is, "Ah Mum, the Footprints lady is here, she's got her bag, can she sleep in my room?" And some of the other ones – we had heaps of them – really good stories from the staff, connecting.' These stories provide evidence that children gain much enjoyment from their participation in the surveys. Although interviewers did note that some families admit they maintain involvement in the study solely for the incentives, many families expressed a sense of

wanting to be part of the change, believing that this study was important. One RAO explained, 'It's enjoyable [for participants]. They don't really rush through it, they take their time. But they do open up, and it's over before you know it.'

Of course, there are some families who hold a less positive view of the study; 'There are some people who, you know, it's just not for them. And I don't even know if some people really know why they are doing the study.' However, overall, the survey has a very high retention rate, with over 80 per cent of children from each wave returning for the following wave of the study. One RAO said, 'Nobody thought that Aboriginal people would be interested in doing something like this. And they thought they'd just pull out. But because the retention rate is so high, and now … We've got a really good team now.' Although the team has experienced some 'ups and downs' from year to year, all the RAOs are 'really dedicated … which I think is what has made the study successful …' One interviewer frankly stated the value of the high retention rate: 'And I know how important it is not to lose kids from the study. Because you know, if we lost everyone, there would be no study.'

(b) Challenges faced by RAOs

The job is not without its challenges, however, as evidenced by the high burnout rate of LSIC interviewers. These challenges include: finding the time to schedule interviews, convincing parents to give up a few hours of their time for the survey, manoeuvring equipment malfunctions, mediating adverse reactions to the interviews, driving countless hours to get from interview to interview, and being positioned as an 'outsider within'. One RAO summarised her experience, 'I'd say there's more positives than there is negatives in the job. But it's not a job for everyone.'

Many interviewers stated that early on, the job was 'pretty full-on,' but with time, they felt they could settle in to the process. One interviewer described a great sense of nervousness for her first interview, explaining:

> It was like baptism of fire, because it was one of the traditional families … so it was one of those houses where there's about twenty or more people there and kids as well, and then if you're, like if you're doing the activity with one kid, of course the others want to join, and they want to do it … it's pretty nerve wracking.

Another explained that before her first interview:

> I was nervous as hell because I didn't know the participant. Even though it was in my area, it was somebody from a community that I didn't particularly associate with. So

> that was difficult. Um, that was not difficult, probably, that was nerve-wracking I guess, asking them invasive questions about their life. But the more I did it, the more confident I got with it.

Overall, however, interviewers expressed satisfaction with the training they received from FaHCSIA, and felt that they were as prepared as they could have been for their first interviews. The first few months of interviews are considered 'training on the job'; one RAO explained that after a week of training in Canberra, she received further training from her supervisor at his site for a fortnight:

> And then after that, was you know, it was kind of a case of having to figure it out myself: you have to do what is comfortable for you, what fits with your community, because as similar as most communities are, some are more sensitive than others. And you sort of only pick that up if you've got the experience in that community.

Each time new questions are added to the survey, however, interviewers experience a renewed sense of nervousness, as the reaction to each new question is unpredictable. FaHCSIA holds meetings with the RAOs every year to develop new questions, resolve issues arising from problematic questions, and address any other feedback.

Certain challenges are unique to specific sites. For example, one RAO described the strong impact of the cultural politics within one site: 'It's the cultural politics behind that that I find that to be the major issue in the community amongst the Indigenous populous.' Another RAO commented that there were significant, though surmountable, differences between the different sites at which she interviewed:

> You know, you still have your similarities, but there are a lot of differences as well. How you talk in one area you don't speak like that in another, and how you dress in one area you don't dress like that in another, and so, yeah, some areas are very different.

Navigating cultural differences and sensitivities adds another level of complexity to the job.

As well as noting differences across sites, one interviewer commented on the huge disparity in the lifestyle of families within one site: 'you get from, you know, real very high middle-class people, to very low socio economic people, and it's just amazing how great the two spectrums are.' As a result of the drastic differences between, and sometimes within, communities, some questions are not relevant to some families. One RAO explained, 'We had one question that was, "Do you have a working fridge, toilet, washing machine?" That sort of thing. I'd say, "This is taking into account the setting" … One mother goes, "Do they think we live in the dark ages?" … Some people found it insulting.' In order to mitigate any sensitivity or discomfort arising from questions

posed, one RAO explained, 'It comes back to us; we have to say why we needed those questions. It's about having good explanations as to why we are asking the questions.'

Although she described significant differences across the sites at which he conducted interviews, one RAO described the 'ability to laugh' as a trait consistently observed across sites; 'And the ability to make laughing a barrier-breaker is universal, I now believe. I mean it's not just something we do down here, it's something we do sort of all over, I've come to realise.' Several interviewers referred to the burden of dealing with 'the personal issues you see families and individuals facing.' Especially as relationships are established with participating families, it becomes more difficult to not become emotionally invested; one RAO explained that

> When you're talking to these people about that, you can't help but be human and feel for their hardship. That's a difficult part for me … It used to not affect me; I used to be able to go home, shut off, and then you know, it would be fine. But now it really has beaten me up a little bit or gotten me down. So, yeah, debriefing with people and talking with people used to work fairly well, but it's gotten to the stage now that that's not really enough, not doing what it needs to do. So I'm looking at other ways, whether that be going for a hit of golf, or speaking to somebody professional. I'm up for anything. [My coping mechanism] is ever evolving.

A senior member of the LSIC team explained the key to success at the job; 'It takes patience, persistence. I can't stress enough the value of the staff.'

Regardless of whether the RAOs are working within their own communities or in other communities, they are positioned in a complex sphere; RAOs are subject to the 'outsider within' positioning in their work for LSIC. Managing to balance research demands with their relationships and the lived reality in the community is a difficult task. As Smith, a Maori writer, explains:

> There are a number of ethical, cultural, political and personal issues that can present special difficulties for indigenous researchers who, in their own communities, work partially as insiders, and are often employed for this purpose, and partially as outsiders, because of their Western education or because they may work across clan, tribe, linguistic, age and gender boundaries. Simultaneously, they work within their research projects or institutions as insiders within a particular paradigm or research model, and as outsiders because they are often marginalized and perceived to be representative of either a minority or a rival interest group (quoted in 31 p. 5).

One interviewer mentioned that she was both 'blessed and cursed' to be working within her own community, in which her family background was known, as she found herself positioned in the middle of cultural politics. She was able to mitigate these

circumstances, but not without concerted effort. Together, these cultural issues and the complex positioning put a significant burden on RAOs conducting these interviews.

(i.) Scheduling interviews:

Each interview is unique and unpredictable, and the experience of conducting interviews varies drastically from family to family. One RAO explained:

> Look, there is no usual way about it. You could imagine what a household would be like at four thirty, five o'clock, when the kids and Mum are getting home from school and work, everybody's restless, everybody's hungry, everybody's tired, everybody's excited, all at once. And then you're fortunate enough to get some that are stay at home Mums, and the kids aren't at school, and they're normally real relaxed, cup of coffee, kitchen table, quiet, kind of set-up. Then you have the same type of set-up with a big family, in which case some kids are running around, some kids are watching TV.

The presence of other children in the household was mentioned by many interviewers as an added complication, making it harder to manage the interviews. Even the presence of the study child alone, however, can cause a significant distraction. While interviewing the primary carer, one RAO explained:

> The study child could be doing anything … playing with your camera, trying to touch the keyboard, or screaming to be fed. I've sat in houses where a five-year-old kid will come up and pick up his mother's top and start feeding on their breast milk while I'm mid-interview. It's hard to – where do you go with that, you know what I mean?

Many RAOs mentioned the increasing difficulty in scheduling an interview with children as they reach school age. One RAO explained:

> It's a hard job actually … this lady I got yesterday, she cancelled three times. And, um, because we're low priority in their life. If their kid is going to go to a birthday party or stay home and do *Footprints in Time* survey, they're going to the birthday party … Because now with kids in school, in Kindy, it's after school. And, you know, after school, kids are tired, or they've got sport, and Mum wants to get dinner ready. So … usually your first appointment is three thirty. And then Saturdays and Sundays. And public holidays …

The difficulty in finding a time to interview families was echoed by many RAOs:

> Families are getting busy, children are becoming more active … Parents are going back to work; kids are doing after school sports … You'll make an interview, you'll make an appointment, the day of. When you go back there, they've changed their mind, or something's come up. It's the nature of the game.'

One RAO described this constant cancellation of interviews as 'being in a rut.' This often happens around July or August; 'You know, we've just got to, sort of, just say, "You can do this, it's not you!" But you sort of get into a rut when you've got the last twenty or thirty to do, and you're just constantly every day, "Knock knock. Knock knock." And it's disheartening.' Another added:

> It's the same as if you've sort of booked your week out, and you're like "Yep, awesome." And say your Monday is fine, you get like two or three on the Monday, and you go, "Oh I've got four in tomorrow." You're like alright, so you go to your first one, and they're not home or whatever. So you go to the next one, and it's the same thing, same thing. It's sort of like, should I even waste my time going to the last one? You know what I mean? You get upset about it because you sort of get committed and excited about doing it, and then you can get a whole day of being shut down, and then the next day it could happen all over again.

When asked how they work through this frustration, one RAO jokingly responded, 'It's our resilience.' Another answered, 'Just got to keep it up and keep rolling till you get it.' Some interviewers turn their success rate into a game to maintain motivation: 'I've got some pretty competitive staff. When we travel together, they'll compete as to how many interviews they have got booked compared to how many interviews they actually get completed. It's a bit mental, but anyway, it works.'

(ii.) Equipment:

In the first waves of the study, the transportation of equipment (such as the stadiometers for measuring height and scales for measuring weight) was difficult for many RAOs. This was especially problematic for RAOs who had to travel to sites on small planes or on boats, given the restrictions on the size and weight of baggage. One interviewer explained, 'We've got a wheelie bag. I think it weighs about fifteen kilos. We used to have backpacks which used to absolutely kill … Yeah the wheelie bags are good.' In addition to issues with transporting equipment, several interviewers mentioned that they had faced some problems utilising the scales; 'And scales … I had to replace one. It just wouldn't work anymore.' Another said, 'I've had problems with my scales the last couple of times, but I've had mine for four years now so I think it's time to chuck them out.'

The RAOs felt that the computer programs for inputting the interviews were well-designed, enabling efficient surveying of participants. In general, they believed

that the LSIC Steering Committee and staff had gone to great lengths to make the interview process as streamlined as possible;

> We have everything we need, you know, they now have supplied us with car chargers for the laptops, because that was an issue with the laptop going flat. And if you've got say three surveys, by that third one, your laptop is going flat. And you know a lot of people don't have … an extension cord …

Of course, interviewers mentioned that there were occasionally glitches in the software, but in general, they were very happy with the materials with which they were provided.

(iii.) Travelling:

The 'travellers' of the group (those conducting interviews at sites across a large geographic area) are usually away from home three weeks out of every four. One 'traveller' explained:

> It's hard to describe. You feel as though you've never got roots I suppose … I go all over Australia … I get my week a month at home … And I sort of go there, I think [another RAO] has actually lived in my house longer than I have since January, because we usually swap over. And you sort of think, "Oh, where's home next week?" It's funny, everyone goes [after a survey]: "Oh you're going home now?" And you go "Yeah, yeah, I'm going home" … And it's to a motel.

Even for interviewers who work within one site, driving can be a significant burden. One said, 'I find one of the [most challenging] things is the driving. You feel like you're being packed all the time, and sometimes your body just wants to be stretched out.' Another added, 'The driving can be the most tiring. Constantly maintaining focus when you're on the road, driving around all day …' A third commented, 'Especially in [the city] – peak hour traffic – a lot of the times when we do home visits, or drive-bys, we try to do it after and before peak hour, depending on where you got to go… although compared to everyone else, I shouldn't be whinging about doing that [number of hours of driving], knowing what the other researchers drive.' When asked about the longest time they had spent in the car in one day, one RAO responded, 'Maybe you don't want to know.' Another answered, 'From the time I left home and the time that I got back, it was fourteen hours.' This included the time to conduct the interview, but 'mentally, you're going for fourteen hours. So although you're not driving, you're still mentally, you know, engaged … that's a fourteen-hour day. And then you go home and do your computer work, or do your emails.'

(iv.) Trust:

Trust was a dominant theme underlying the conversations with RAOs; one commented that the formation of strong connections and a relationship of trust between the LSIC staff members and the families was 'essential for the study to survive.' Another RAO mentioned, 'You know, the parents give up their time to see us, and yeah, they put their trust in us too.' As Chino explains, 'Relationship building is an essential process in tribal communities, one that is deeply embedded in history and context' (147 p. 598). Given the history of distrust between Indigenous Australians and the research process (including the researchers themselves), the development of a relationship between LSIC participants and the RAOs fundamentally necessitated the development of trust (37, 148). For some survey questions, this sense of trust is especially important. One interviewer explained that the question about smoking was not too uncomfortable for most carers, but that the question about drinking, in contrast, caused a lot of discomfort.

There was general consensus that the trust between families and the RAOs has increased over time, as families have become more familiar with the RAOs; one said, 'Now, like I think I've been going back to these same families for three years, and they know me … The kids remember you, you know.' This development of trust, of course, is strongest for families who have been visited by the same RAO for each wave of the study. One interviewer was considering relocating to a different site,

> But then I feel like I don't want to [move to another site] because I'd be like abandoning … Yeah, and one lady asked … and I said to her, you know, there might be another lady starting, and she goes, "Oh, we just got used to you."

Another RAO explained:

> It's Wave 5 now, and I've built up a fairly good relationship and rapport with most of my clients, and they are open to tell me anything because they know they can trust me. And the answers I get now, I know that they are honest. Not just because I know they are comfortable with me, but because I'm in the community, I'm involved in the community, I know what's going on, kind of thing.

Once trust has been established, many families disclose sensitive information. One RAO said,

> I'm surprised too; some of the things people tell me are things I would never tell anyone. And I think maybe it's because I'm someone they see once a year, and maybe they've got no one to talk to about this stuff. And it's never gotten out there … I know yeah, they must [have a good sense of trust in me].

This trust encourages increasingly honest responses, and enables RAOs to ask questions that otherwise would have been responded to with much apprehension. One RAO demonstrated that the relationship of trust has been formed over time; 'Early on I'd ask how often their teeth are brushed. And the standard answer was, "twice a day, every day." And now it's, "Oh, he doesn't even have a toothbrush." They are quite happy to be honest about it, and that's because of the relationship.'

Even for families who are familiar with the RAOs, however, wariness can arise about the purpose of asking all of the survey questions, and concerns about confidentiality can be evoked. An RAO described her interaction with one primary carer who has:

> … always been a bit, uh, touchy about things. She goes, "What do you ask all these questions for?" And then, you know, trying to explain to her … I think it was that she didn't want her name against that answer, and I said "Well what happens when I finish the survey is that I go back to the room, and I synchronise," and I explain to her what synchronising is. "When the data gets out," I said, "your name is already removed." That's what I've been saying, that you know your data is here and your name is there, so no one will ever be able to trace back that.

One RAO discussed the importance of the informational DVDs provided at the first visit, designed to explain the study and the process of consent; 'Everyone understands [the meaning of consent] differently. So we wanted to make sure that everyone was getting the same information.' These DVDs were especially important in areas where interpreters (members of the community) were employed, 'because then they [the families] could see that the form that was on the laptop was the same form that they were actually going to be signing. And the interpreter would interpret as the DVD was going through.' This dedicated process of informing consent aided in the formation of trust between the RAOs and the families involved in the study.

Another important aspect of the study design process was the organisation of meetings with community members in sites where the interviews were to take place. These meetings allowed community members to provide input as to how to conduct the interviews in a culturally appropriate manner, and how to design the study to provide the most benefit to the community. One RAO stated that these meetings helped the Steering Committee of LSIC realise that 'they needed to look outside of the square, to do things a different way, which they did.' These processes of community consultation, informed consent, and assurance of confidentiality are especially important within Indigenous health research, given the historical context (32). These measures, together

with interviewers' conscious efforts to create a friendly rapport, lead to a development of trust between participating families and the RAOs, facilitating the collection of accurate responses, and optimising the impact of the study.

(c) Measuring height and weight

One interviewer described the difficulty in measuring children's weight in the first wave of the study: 'I had trouble with the weights [on the carpet] at the beginning … I was kicking myself because I thought, "I can't believe it's taken me nearly a year to work out that this thing doesn't like carpet." It wouldn't catch on – it wouldn't lock on the measurement. And I think it just so happened that I was on tile [and I realised that it worked].' Another RAO explained that although, 'The kids are happy to get on it [the scale]', during the first few waves of the study:

> The equipment we used was unreliable. And then, even now, you've got to be careful – for instance, the scales will give off a different reading if you set them up on carpet than they will if they're on tile. To make sure you've got that right, and not being lazy about things like that, is annoying I guess, since you're not always on tile. But the general process of doing it, I'm fairly comfortable. The parents trust us enough to, if we have to, pick the kids up to stand on the scales, take them off the scales, put the kids down if the children don't readily want to get on it, but I haven't had to do that for two years.

To add to the difficulty of measuring children, many hung onto tables or ledges while the interviewer was attempting to record measurements, thereby distorting the measurement of their weights. In other cases, the surveys had to be conducted outside, without a flat, solid surface to weigh and measure children. An RAO explained that in one remote area:

> Generally the survey is done outside, on the ground, in the dust. And my laptop is playing up at the moment now because I think of all the dust. And I sat on a brick. That was the only thing to sit on. So then I had to look around for a hard surface. And I just said, "Oh I need, you know, cement or something." And she [the mother] just let me do it just inside the door. Because there was no way I was going to get it [the measurement] outside.

Most parents are comfortable with the idea of having their child weighed and measured but interviewers still ask permission before taking measurements; 'I always ask them – the Mums – even now. I said, you know, "Is it ok if I take their height and weight?"' One RAO noted, 'You know, we've still got kids that don't allow us to [measure them], just refuse every year. We don't force them to, we just say "Yeah that's

fine, no worries."' Some children, especially as they approach adolescence, express discomfort about being weighed:

> They're fine talking with us, sitting down, doing the computer, but as soon as you pull out [a scale] … I mean like one girl, she's really got a complex about her weight already. So her Mum says, "Oh she won't do that." I said, "Oh I'll give it a go," but no [she wouldn't].

Some interviewers also found that some children were afraid to have their height measured, despite being comfortable having their weight measured:

> I think it just depends on them personally, and whether they consider weight to be an issue. Weight I haven't had much problem with, but height – and mainly just the stadiometer. It's a little bit intimidating because you open it up and fold it out and it beeps, so for some of the little kids it's like, "What's that?" And you're trying to put it on their head and they freak out on you.

Interviewers were taught to repeat each measurement three times, to ensure that the reading is accurate. One RAO said, 'And so I've got to get them on and off [the scale], on and off again, and they think it's a game. But usually you try to get them on, write it on their contact sheet, and then get them to go off, get them to go back on.' Many interviewers reported that they purposefully made weighing and measuring children into a game, in order to ease the process. One RAO explained, 'Usually I'll get an older brother or sister to say, "Do you want to have a go?" And I'll actually let them hold the stadiometer; they'll go and measure their brothers and sisters. And usually then the study child will want to have a go.' Another RAO explained:

> Or I'll just ask the Mum to try and do it, but usually kids at this age are worse for their Mums than they are for us. So I just let them play with it [the stadiometer] sometimes, and then they come back, and I'll say "Can I see how tall you are now? Can you get on the rides?" Because they're just getting to that height where they actually can get on some of the rides, because they have to be a hundred centimetres.

Another interviewer described the children's fascination with the stadiometer, and explained, 'I just tell them there's a magic eye down there and that measures from there to there … and the bigger kids understand more.' In making a game out of weighing, one interviewer cautioned, 'You have to be careful too, because if you say, "Jump on there for me buddy", BOOM! They jump on there.' When asked how much time it usually took to do the measurements for each interview, an RAO responded, 'Depends how many kids are there – you usually end up measuring the whole family.' By the fifth wave of the study, however, RAOs comment that the process 'doesn't take

that much time at all; it's probably, if you do it right, a minute if they're already sitting there in attention, you hope; if they're running around it could be a bit of a bother, but they're rarities.' This is the last task of the interview, and children generally appreciate that the task is interactive; 'they get to do that, without sitting on a chair and stuff.'

One interviewer described her practice of leaving a card with the child's weight, height, and age behind for the families; 'And I said, "Oh you know, you can put that up on the fridge or put it with your baby stuff" if someone wants to know, or even want for their own record what their kids are.' For some children, this would be the only time they had been measured that year, and it could be an important benchmark for families. The RAO told the story of a carer whose child's weight had decreased by three kilograms since the previous survey; 'I reckon that Mum I saw yesterday, she was actually shocked at that, that she [her daughter] was that light.' Although the RAO did not comment to the mother on the meaning of this decrease in weight, she thought simply providing this information to parents was valuable;

> It's one thing you can give them, you know, without actually saying anything. You know, might just prompt them to think, maybe I should do something, maybe I haven't been with them enough … there was like a sort of like an awareness from the parents – "Oh, I might have to do something about that."

Another mentioned her practice of writing down the measurements on the contact sheet, to enable comparison to previous waves; 'So a lot of us do that now. Putting it on the contact sheet is just another way to help the data team, if they've got any glaring issues it's all on the contact sheet.'

Interviewers attributed some of the inaccuracy of the height and weight measurements in the early waves of the study to issues with the measuring equipment. Since changing the equipment, interviewers believed that measurement accuracy had improved:

> We used to have a snap-together, what do you call it, stadiometer [for measuring height]; and now we've got an automatic … laser height detector. They're a lot more accurate, and a lot more reliable, so you can guarantee that the person doing it at another site is going to be using it exactly the same way you're using it … it's got a self-levelling thing on it. It will tell you whether you need to flip it out or back to get it level. The idea is, using that, the possibility of getting faults back, or mis-readings, is minimal.

As well as being perceived as being more accurate; the measuring equipment is easy for RAOs to use; 'I really couldn't think of another way to make it any more easy. It's as

simple as it could get without being unreliable.' Especially with the new equipment, RAOs find the measurement of height and weight to be fairly easy. However, there are certain cases, such as children with autism, in which the process is more complex. The various strategies that the RAOs have developed are critical to facilitate the taking of measurements in these cases. Interviewers also pointed to the comfort of families with the RAOs as a factor leading to improved accuracy; 'Well I think because [the RAOs are] now more established within the household, so the kids aren't so scared. I don't know if everyone else agrees with that; they're not so wary, you know, because they're used to us being there.'

(d) Recording birth weight

When asked about the birth weight of the child, interviewers responded that, 'If it was their first child, most of them knew, I reckon. But if it was their fifth or sixth one …' [they were less likely to know]. This might be reflected in the accuracy of the birth weight data.

(e) Limitations

A limitation of this analysis is that it represents the views held by RAOs about families participating in the study, which may not completely reflect the views of the participants themselves. Participating families were not approached out of consideration of privacy; instead, the RAOs were chosen as key informants to share their knowledge from their first-hand engagement with these families.

The discussions about families' reasons for participating in the study and willingness to answer certain questions bring to light another potential for bias. The families who agreed to participate in the first wave of the study, and sustained involvement in subsequent waves, may not represent all of the families who were initially contacted. Further, there may be differences between families consenting, versus not consenting, to respond to certain questions. Although this is a limitation of the study, it should not influence the results of data analysis, as these conclusions are only applied to children participating in the survey, rather than extrapolated to all Indigenous children.

(f) Conclusion

Grove et al explain that in the past, 'the application of methods that are culturally inappropriate has produced data of questionable validity and reliability and led to research outcomes that lack meaning in a local context' (3 p. 638). The interviews with the LSIC RAOs echoed these concerns, depicting many of the difficulties that have been described in the literature surrounding the use of longitudinal studies for Indigenous research. The LSIC Steering Committee and RAOs have mediated these incompatibilities by maintaining community involvement throughout the research process and by emphasising the use of a culturally-appropriate method for the study. As described in the literature, the development of trust between study participants and the study team is critical for the success of any project. A good rapport is especially important for longitudinal studies, in order to maintain participant engagement and minimise attrition over the course of the study. This trusting relationship is even more important within Indigenous research, given the historical context leading to a widespread sense of distrust in research (2, 3, 19, 30, 33, 34). The development of trust, in order to maximise participant engagement and honesty, has been a primary objective for LISC RAOs and has contributed to a high retention rate and improved accuracy in later waves of the study. In the short-term, however, the commitment to this objective has prevented RAOs from recording measurements or responses to questions for some participants, in order to maintain a good rapport. Not pushing participants to answer sensitive questions contributes to missing data, but ensures that a trusting relationship is maintained between participants and interviewers. In some cases, participants might reject the initial request but agree to provide an answer or measurement in later waves of the interviews, after the establishment of trust and a good rapport. For example, in response to my proposal to measure children's waist circumference as a health indicator, one RAO explained that it might be possible, 'maybe now [in the fifth wave of the study], since we have a good rapport with the kids, we could do that … but if you wanted to do it in Wave 1, you can give up.' After a discussion between the RAOs and members of the Steering Committee, this proposal was rejected due to the cultural implications of taking these measurements, and the potential to breach the trust painstakingly developed between participants and the study team.

This type of appraisal – evaluating the impact of questions and measurements upon the relationship between participants and the RAOs – may not be a major concern for most studies, but it is crucial in LSIC and must be considered in the interpretation of

study results. The LSIC study team has gone to great lengths to maintain the delicate balance between creating a relationship with participants and maintaining the integrity of data, through the vigilant development of the study and interview methods. I developed data cleaning methods to compensate for a reduction in data reliably potentially arising as a consequence of the need to maintain this delicate equilibrium.

B) A quantitative evaluation of the LSIC height and weight data

Ensuring the validity and the completeness of the LSIC anthropometric data is essential prior to conducting further analyses, in order to maintain research integrity. Members of the LSIC team have acknowledged that the task of accurately weighing and measuring children (especially those that need to be measured lying down) is very difficult (8). In many homes, flat, solid surfaces were not available to use for measuring children, impeding the ability of interviewers to obtain accurate measurements. In some instances, the weighing scale was set to the wrong unit and interviewers were unable to switch it to the correct unit. In these cases, the children's weight may have been recorded in the wrong unit (pounds instead of kilograms), or interviewers may have estimated the conversion to kilograms. Additionally, the stadiometers used to measure children in Wave 1 were exceptionally heavy, and were not able to be transported on the flights and boats to some interview locations. Instead, tape measures were utilised, which might have had a lower degree of accuracy. In some cases, the interviewers recorded these conditions in the 'comments' section, providing an indication of the limited accuracy of these measurements. In other cases, however, comments were not recorded, so the circumstances affecting the measurement are unknown. As a result of these concerns, and issues brought to light through conversations with the LSIC data collectors, I undertook an intensive data cleaning process. The anthropometric data withstands a large amount of missing data and implausible measurements, as well as implausible variation within individuals. Methods were developed to distinguish these implausible values from the natural variability observed in the height and weight of children over time. All analyses were conducted using STATA version 12.

(1) Missing height and weight data

The missing data for children's height and weight is attributable to a number of factors. A large portion of the missing data results from the attrition of participants, as

around 15% of the sample was lost between successive waves of the study. Additionally, for the 88 new entrants in Wave 2, height and weight was not available for Wave 1. All parents and carers were given the option to help weigh or measure their child themselves; however, some parents remained uncomfortable and refused to have these measurements taken. These parents were asked if they would be willing to provide the weight and height measurements taken at their most recent visit to a health care centre from their Baby Book. Weights and heights from the Baby Book (if accompanied by the age at which the measurement was taken) were assumed by the LSIC team to be of high accuracy because these data were collected by health professionals in a controlled setting. These measurements from Baby Books represent up to six per cent of measurements recorded in each wave, and were included in analyses (see Table 8; see Appendix D for more detail).

Table 8: Number of study children with weight and height recorded at each wave, and source of information

	Wave 1	**Wave 2**	**Wave 3**	**Wave 4**
Height recorded	1,322	1,415	1,318	1,245
From Baby Book	38	12	18	5
(%)	(2.87)	(0.85)	(1.37)	(0.40)
Measured by RAO	1,284	1,403	1,300	1,240
(%)	(97.13)	(99.15)	(98.63)	(99.60)
Weight recorded	1,365	1,451	1,334	1,257
From Baby Book	77	59	17	8
(%)	(5.64)	(4.07)	(1.27)	(0.64)
Measured by RAO	1,288	1,392	1,317	1,249
(%)	(94.36)	(95.93)	(98.73)	(99.36)
Both height and weight recorded	1,304	1,408	1,308	1,245

Overall, parents' and carers' trust in the RAOs seemed to increase over time, as evidenced by the increased number of weight and height measurements recorded over the course of the study (see Table 9). Over time, parents were less likely to refuse measurement, and those consenting to have their child measured became more likely to allow the RAO to take the measurements than to request to take the measurement themselves (see Table 10 and Table 11). For example, the per cent of parents refusing weight measurements decreased from 22% to 1% between the first and fourth waves of the study. By Wave 4, 97% of parents and carers consented for the study child to be weighed by the RAO, and 96% consented for their child to be measured by the RAO, with only 1% and 2%, respectively, preferring to take the weight and height measurements themselves.

Table 9: Number of height, weight, and age measurements recorded in each wave of the study

	Wave 1	**Wave 2**	**Wave 3**	**Wave 4**
Height recorded	1,322	1,415	1,318	1,245
Age at height measurement recorded	1,295	1,411	1,291	1,239
Weight recorded	1,365	1,451	1,334	1,257
Age at weight measurement recorded	1,324	1,424	1,308	1,249

Table 10: Number of parents and carers consenting to have their child measured in each wave (and per cent of total sample responding to the question at each wave)

	Wave 1 (%)	**Wave 2** (%)	**Wave 3** (%)	**Wave 4** (%)
No response	19 (1.28)	9 (0.61)	--	--
Yes, RAO do it	1,063 (71.78)	1,243 (83.65)	1,194 (86.33)	1,221 (96.14)
Yes, parent do it	221 (14.92)	163 (10.97)	114 (8.24)	19 (1.50)
No	178 (12.02)	71 (4.78)	75 (5.42)	30 (2.36)
Total	*1,481 (100.00)*	*1,486 (100.00)*	*1,383 (100.00)*	*1,270 (100.00)*

Table 11: Number of parents and carers consenting to have their child weighed in each wave (and per cent of total sample responding to the question at each wave)

	Wave 1 (%)	**Wave 2** (%)	**Wave 3** (%)	**Wave 4** (%)
No response	19 (1.28)	9 (0.61)	--	--
Yes, RAO do it	964 (65.14)	1,200 (80.75)	1,178 (85.15)	1,233 (97.09)
Yes, parent do it	326 (22.03)	192 (12.92)	141 (10.20)	16 (1.26)
No	171 (11.55)	85 (5.72)	64 (4.63)	21 (1.65)
Total	*1,480 (100.00)*	*1,486 (100.00)*	*1,383 (100.00)*	*1,270 (100.00)*

(2) Accuracy of height and weight data

Within the height and weight measurements that have been recorded, there are further limitations. There is clustering, or 'heaping' of height and weight measurements around whole numbers, suggesting a lower degree of measurement accuracy than was intended (to the nearest millimetre for height, and to the nearest hundredth of a kilogram

for weight). For example, in Wave 1, 123 children (9% of recorded weights) are recorded as having a weight of 10 kg, and 86 children (6% of recorded weights) are recorded as having a weight of 15 kg. This disproportionate clustering again occurs in Wave 2, with 130 children (9% of recorded weights) listed as having a weight of 15 kilograms and 90 children (6% of recorded weights) listed as having a weight of 20 kilograms. In Wave 3, however, this clustering is less pronounced, suggesting increased reliability of measurements. Further, some recorded heights and weights appear unrealistic in the context of the other measurements for the child, and need to be excluded. In addition, some inconsistencies in age are recorded, with a younger age recorded for latter waves of the study. These inaccuracies need to be resolved prior to analysing anthropometric data.

(3) Cleaning of age data

For each wave of the study, the age of the child in months was recorded. In addition, the age at the time of height and/or weight measurement (rather than at the time of interview) was recorded if the measurements were recorded from the Baby Book, to enable the accurate calculation of height-, weight- and BMI-for-age z-scores. To calculate BMI-for-age z-scores, the age at the measurement of height must be the same as the age at the measurement of weight. Baby Book measurements were excluded from analyses if the most recently recorded measurement occurred more than 18 months prior to the interview. There was only one instance in which the recorded age at measurement decreased between successive waves of the study, and the implausible age was re-coded to missing in this case (see Table 12). This younger age of height measurement reflected the age of measurement recorded in the Baby Book, and was excluded because the measurement occurred prior to the measurement at the preceding wave of the study.

Table 12: Decrease in age between waves

Wave	**Age at height measurement** (months)	**Source of height measurement**
1	20	RAO measured
2	*6**	Baby book
3	40	RAO measured
4	52	RAO measured

* Age re-coded to missing because the age of the most recent Baby Book measurement recorded in Wave 2 was younger than age of measurement in Wave 1.

(4) Cleaning of height and weight data

The WHO has developed programs for converting raw weight, height, gender, and age data into z-scores, using to the WHO CGS as the reference (149-151). The WHO Anthro program (149) is used for LSIC children through 60 months of age, and the WHO AnthroPlus program (151) is used for children between 61 months and 19 years of age. This software is available for download from the WHO website, or macros can be downloaded for use in statistical programs including STATA (149, 151). These macros were used to calculate height-for-age, weight-for-age, and BMI-for-age z-scores for children in the LSIC sample. The WHO Anthro program also calculates weight-for-height z-scores for children through 60 months of age, but this indicator is not available in the WHO AnthroPlus program for children over five years of age. Height, rather than length, is analysed, so any recumbent length measurements for young children are converted to standing height measurements by subtracting 0.7 centimetres (150). If it is not explicitly stated in the data, the WHO Anthro program interprets height measurements for children less than 24 months of age as length measurements, and height measurements for children greater than 24 months of age as stature measurements; these measurements are adjusted accordingly. In both programs, the age in days (rather than months or years) is used in order to maximise precision; LMS values (representing the median, coefficient of variation, and skewness) are used to interpolate exact z-scores (150, 152).

According to the WHO, data points are considered extreme, or biologically implausible, if they fall outside of a specific range for each indicator (see Table 13) (153). Data that are implausible according to these standards should be excluded from analysis (150). In the LSIC sample, the prevalence of implausible data in each wave ranged from 0.25% to 12.58% for each indicator, with BMI-for-age exhibiting the highest prevalence of implausible values at each wave (see Table 14). This is logical given that BMI is a composite measure of height and weight; an implausible measurement for either height or weight would therefore lead to an implausible BMI value. Within each indicator, the prevalence of implausible data decreased from Wave 1 to Wave 4, suggesting that the data quality has improved over time. Given that z-scores are adjusted for age, they cannot be calculated for children without a recorded age at the time of measurement.

Table 13: Cut-off points for implausible data, by anthropometric indicator (153)

Indicator	**Lower bound for 'plausible' data**	**Upper bound for 'plausible' data**
Height-for-age	z = -6	z = +6
Weight-for-age	z = -5	z = +5
Weight-for-height	z = -6	z = +5
BMI-for-age	z = -5	z = +5

Table 14: Number of children with implausible measurements (and per cent of the sample with z-score recorded), by wave and anthropometric indicator

	Wave 1 (%)	**Wave 2** (%)	**Wave 3** (%)	**Wave 4** (%)	*Total* (%)
Height-for-age flagged (z < -6 or z > +6)	118 (9.12)	32 (2.27)	9 (0.70)	2 (0.16)	*161 (3.08)*
Weight-for-age flagged (z < -6 or z > +5)	46 (3.48)	29 (2.04)	11 (0.84)	10 (0.80)	*96 (1.81)*
BMI-for-age flagged (z < -5 or z > +5)	155 (12.56)	94 (6.86)	37 (2.91)	32 (2.60)	*318 (6.23)*
Any measurement flagged	*199*	*99*	*40*	*34*	*372*

Data points were also excluded if a child showed a decrease in height between waves, as this is physiologically impossible (except in the case of severe pathology). A decrease in weight between waves is physiologically possible, and plausible if a child has been sick or has experienced trauma. However, a significant decrease in weight between waves raises suspicion, given that children are expected to be gaining weight through childhood. Discerning incidents of true decreases in weight (potentially reflecting illness or trauma) from errors in measurement or data entry poses a challenge (154). Conservative criteria for exclusion were chosen in order to maintain the true biological variability represented within the sample. Decreases in weight associated with a change in weight-for-age z-score exceeding 3 were considered eligible for exclusion. This cut-off point was chosen because this magnitude of change in z-score represents a significant shift in weight status, such as from overweight to underweight or from healthy weight to severely underweight, within a year.

After removing the data deemed implausible by WHO standards, children with any decrease in height between waves or with a decrease in weight accompanied by a decrease in z-score exceeding 3 were flagged for further cleaning. Height and weight (and therefore BMI) measurements without an age of measurement recorded were re-coded to missing because the cleaning process, based on analysis of age-adjusted z-

scores, could not be applied to these data. The measurements to be excluded were determined by the following criteria:

1. If there are only two data points recorded for an individual, it is not possible to determine which is in error based on comparison to a third data point; thus, exclude both.
2. If there are three or more data points recorded for an individual, use height-for-age or weight-for-age z-scores (in the case of decreases in height and decreases in weight, respectively) to determine which data point should be excluded: remove the data point such that the sum of the differences in z-scores between successive waves is minimised.
3. If there are two non-consecutive sets of decreases in height or weight within the four data points (for example, a decrease in height between Wave 1 and Wave 2, an increase in height between Wave 2 and Wave 3, and a decrease in height between Wave 3 and Wave 4), exclude all four data points because it is not possible to determine which data points are in error.

Analyses were conducted on the data remaining after following these criteria. The final sample is described in Table 15 (see Appendix E for more detail).

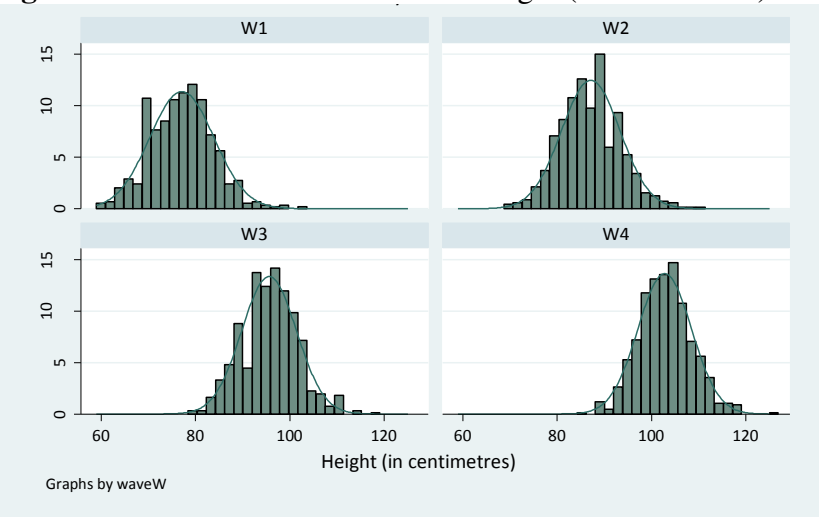
Table 15: The number of children with plausible measurements recorded points for each indicator (and per cent of the sample with a z-score recorded), by wave

Wave	**Plausible height-for-age z-score recorded** (%)	**Plausible weight-for-age z-score recorded** (%)	**Plausible BMI-for-age z-score recorded** (%)
1	1,055 (80)	1,147 (84)	996 (76)
2	1,249 (88)	1,319 (91)	1,207 (86)
3	1,167 (89)	1,255 (94)	1,149 (88)
4	1,174 (94)	1,212 (96)	1,170 (94)
Total number of measurements across four waves	*4,645* *(88)*	*4,933* *(91)*	*4,522* *(86)*

(5) Results – normality of the distribution of height

The distribution of cleaned height (in centimetres) was less skewed than that of weight. The mean height approximated the median for each cohort and wave, with a standard deviation between 5.68 and 6.81 centimetres (see Figure 1 and Figure 2; see Appendix M for more detail). The distribution of height-for-age z-scores, separated by cohort and wave, was approximately normal (see Figure 3 and Figure 4). However, there was an unexpectedly high proportion of the distribution of height-for-age z-scores in Wave 1 falling at either extreme end of the distribution, most notably at the low end, for both cohorts. The occurrence of this effect for both cohorts is suggestive of a potential systematic problem in height measurement. An elevated prevalence of low height-for-age is observed again in Waves 2, 3, and 4 but to a lesser extent. This trend is confirmed through examination of the standard deviations at each wave: the standard deviation of height-for-age z-scores for the younger cohort decreased from 1.88 in the first wave of the study to 1.11 in the fourth wave of the study; similarly, the standard deviation for the older cohort decreased from 1.35 to 1.09 (see Appendix M for more detail). The quantile-normal plots for height-for-age z-score demonstrate that the distributions became closer to normal in successive waves, conceivably suggesting increased measurement accuracy in the later waves of the study (see Appendix K).

Figure 1: The distribution of cleaned height (in centimetres) for Cohort B, by wave



* The overlaid curve represents a normal distribution with the same mean and standard deviation as the observed data.

Figure 2: The distribution of cleaned height (in centimetres) for Cohort K, by wave

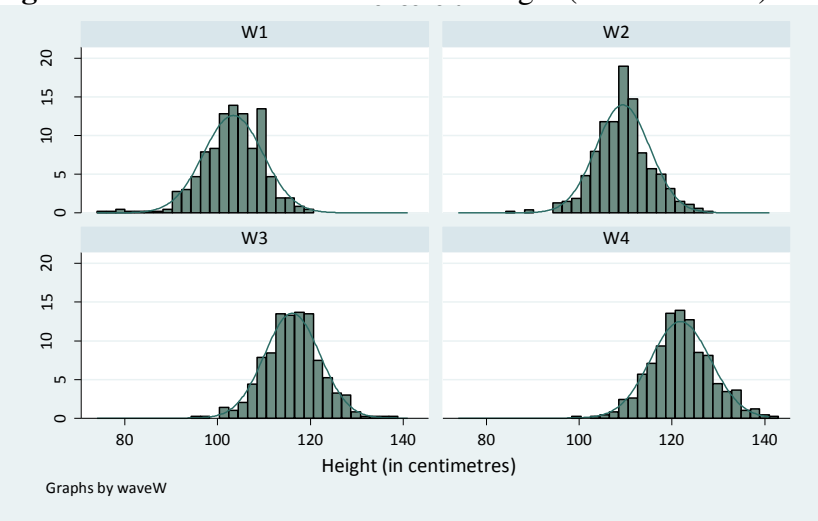
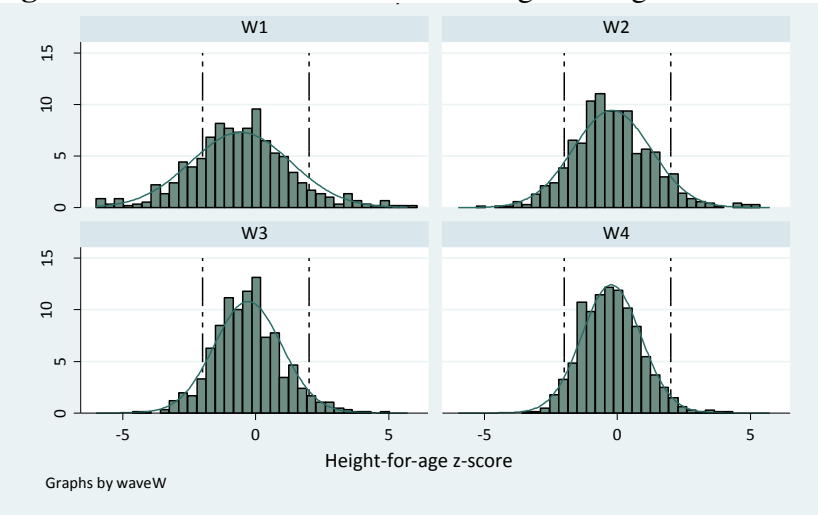
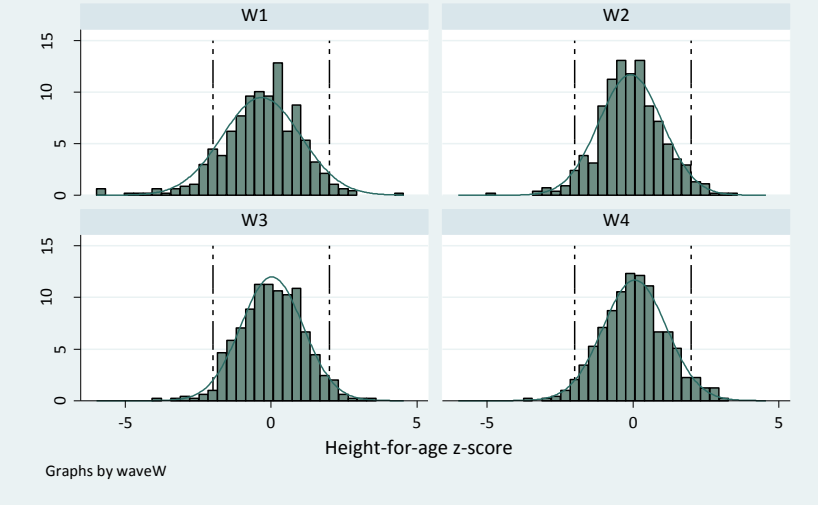


Figure 3: The distribution of cleaned height-for-age z-scores for Cohort B, by wave



* The dotted lines represent the cut-off points for low and high height-for-age.

Figure 4: The distribution of cleaned height-for-age z-scores for Cohort K, by wave



* The dotted lines represent the cut-off points for low and high height-for-age

(6) Results – normality of the distribution of weight

The variation in weight within each cohort at each wave was great, and increased with age (see Figure 5 and Figure 6). The standard deviation of weight increased from 1.96 kg to 2.39 kg, 2.50 kg, and 3.12 kg across waves for the younger cohort and from 3.38 kg to 3.82 kg, 4.31, and 5.93 kg across waves for the older cohort. The distribution showed a positive skew, as is often observed for weight (119), with the mean exceeding the median for both cohorts at each wave, with the exception of the younger cohort at Wave 3 (see Appendix M for more detail).

Figure 5: The distribution of cleaned weight (in kilograms) for Cohort B, by wave

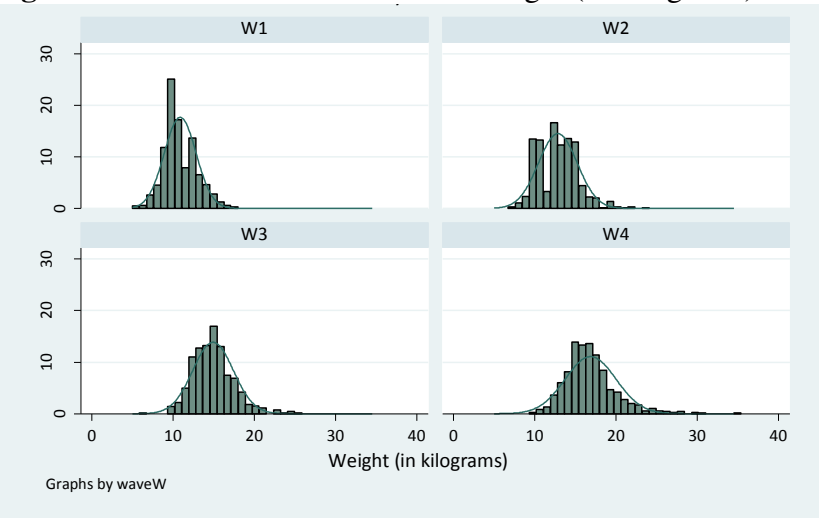
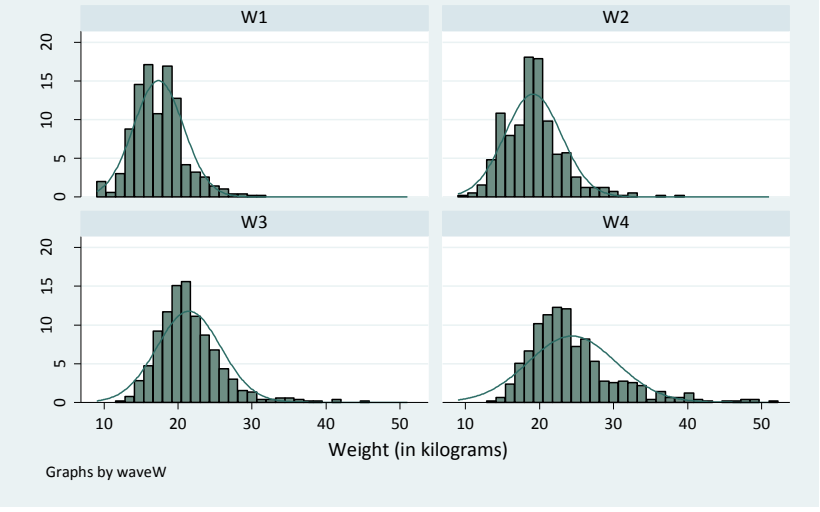


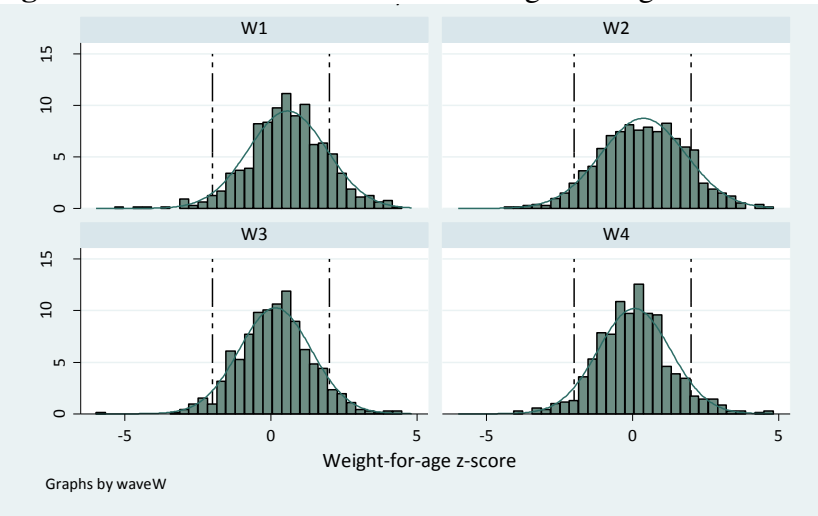
Figure 6: The distribution of cleaned weight (in kilograms) for Cohort K, by wave



The transformation of weight to weight-for-age z-scores (adjusted for age and gender) functioned to normalise the weight distribution (see Figure 7 and Figure 8).

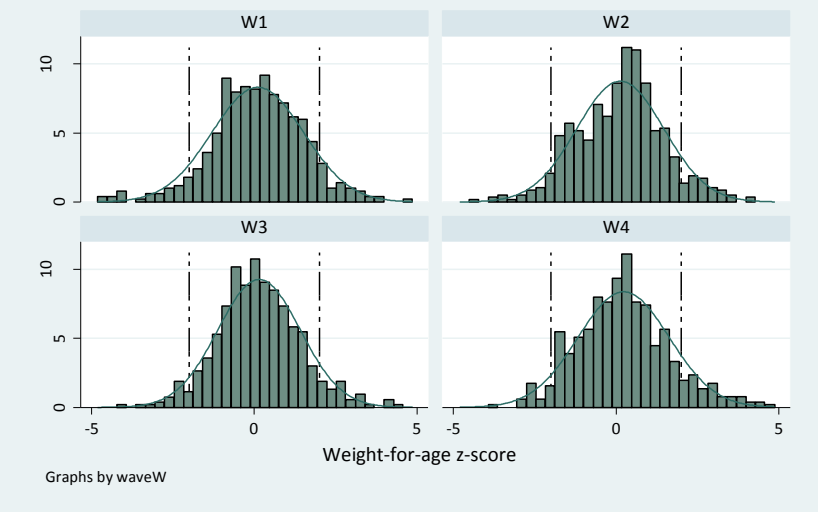
Although there were more observations in the tails at each wave (especially the first wave) than would be predicted in a normal distribution, the weight-for-age z-scores for each cohort at each wave approximated a normal distribution (see Appendix K). In the first wave of the study, there seems to be an elevated prevalence of extremely low weight-for-age for both cohorts, potentially a sign of systematic bias.

Figure 7: The distribution of cleaned weight-for-age z-scores for Cohort B, by wave



* The dotted lines represent the cut-off points for low and high weight-for-age.

Figure 8: The distribution of cleaned weight-for-age z-scores for Cohort K, by wave



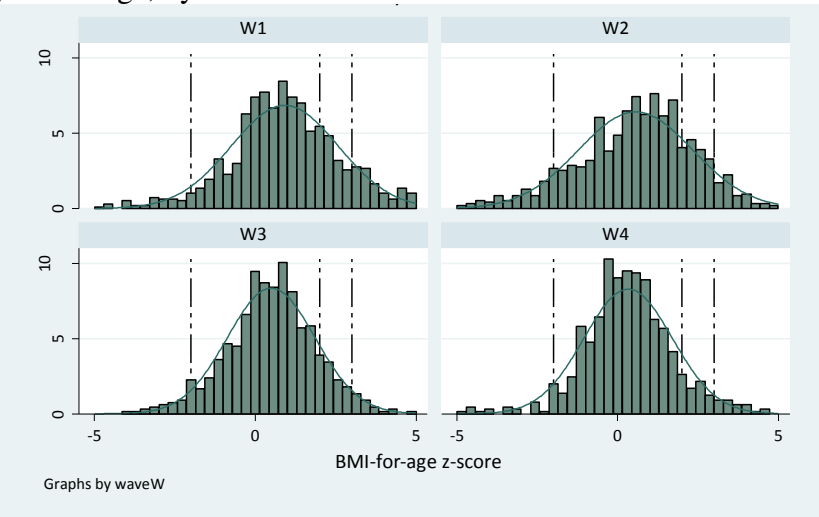
* The dotted lines represent the cut-off points for low and high weight-for-age.

(7) Results – normality of the distribution of BMI

The meaning of BMI values is strongly age-dependent for children, and therefore examining raw BMI scores is uninformative (117). BMI z-scores are analysed separately for children under five and children over five years of age, given the distinct

cut-off points used for each age group. The cut-off points are more conservative in the younger age group, with a BMI-for-age z-score between +1 and +2 indicating risk of overweight, a z-score between +2 and +3 indicating overweight, and a z-score greater than +3 indicating obesity (133). For the older cohort, a child with a z-score between +1 and +2 is considered overweight, and a child with a z-score exceeding +2 is considered obese (133). The distribution of BMI-for-age z-scores for children under the age of five years was approximately normal, but with a larger proportion in each tail than would be expected (see Figure 9). The prevalence of high BMI-for-age is especially high in the first wave of the study; this might be a result of the low height-for-age z-scores observed in this group, given that BMI is inversely proportional to height.

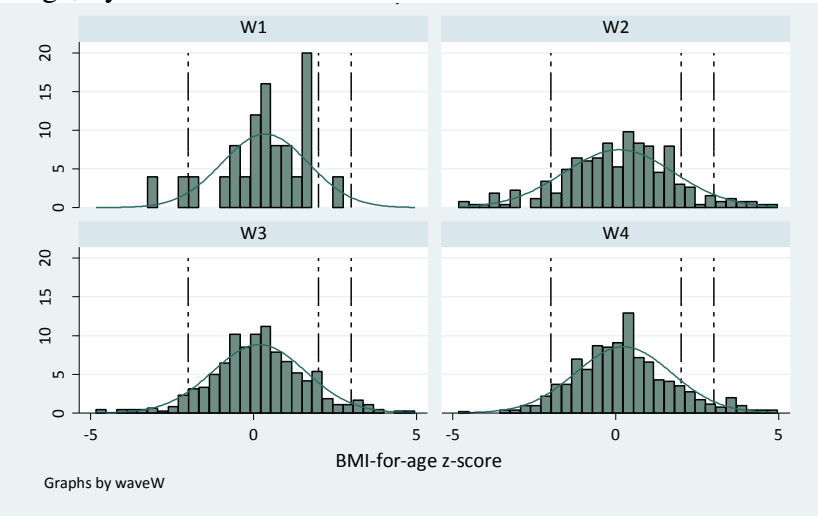
Figure 9: The distribution of cleaned BMI-for-age z-scores for children less than five years of age, by wave



* The dotted lines represent the cut-off points for underweight, overweight, and obese for children less than five years of age.

The number of children over the age of five years was very small in the first wave of the study (n = 25) because of the study's designated age cut-offs, but increased with each wave of the study (n = 265 in Wave 2, n = 483 in Wave 3, and n = 518 in Wave 4). Thus, the distribution of BMI becomes more normal in the later waves of the study with the increasing sample size (see Figure 10). However, even in later waves of the study, there remains a high proportion of the older cohort falling in the outer tails of the distribution; examination of the quantile-normal plots demonstrates that this effect persists in both cohorts through the fourth wave of the study (see Appendix K). Overall, however, the distributions appear relatively normal.

Figure 10: The distribution of cleaned BMI-for-age z-scores for children over five years of age, by wave



* The dotted lines represent the cut-off points for underweight, overweight, and obese for children over five years of age.

(8) Discussion – biases in missing and implausible height and weight data

There were no significant differences in the BMI-for-age, height-for-age, or weight-for-age z-scores at Wave 1 between children who were missing (including data missing due to both non-response and to implausibility) and children who were not missing the respective z-scores at the second wave of the study (see Table 16Table 16). There were no significant differences for any of the three indicators (BMI-for-age, height-for-age, or weight-for-age) at Wave 3 for children missing versus children not missing data at Wave 4. There was, however, a significant difference between the BMI-for-age z-scores (but not weight-for-age or height-for-age z-scores) at Wave 2 for children who were missing BMI-for-age z-scores at Wave 3 versus children who were not, with a significantly lower mean Wave 2 BMI-for-age z-score for those who were missing BMI-for-age z-score at Wave 3. Overall, this suggests that the missing and excluded data points in LSIC were, not associated with size at the previous wave, and thus bias should be minimal.

Table 16: Differences in z-scores for children with missing/implausible data at the following wave of the study (two-sample t-test with equal variances)

	Mean height-for-age z-score (at the respective wave)	**Mean weight-for-age z-score** (at the respective wave)	**Mean BMI-for-age z-score** (at the respective wave)
W1: for child not missing data at W2 (n)	-0.50 (751)	0.34 (868)	0.90 (686)
W1: for child missing data at W2 (n)	0.34 (304)	0.47 (279)	0.92 (310)
P-value of difference between two groups (two-sided)	0.15	0.18	0.84
W2: for child not missing data at W3 (n)	-0.16 (902)	0.27 (1025)	0.52 (867)
W2: for child missing data at W3 (n)	-0.13 (347)	0.23 (294)	0.29 (340)
P-value of difference between two groups (two-sided)	0.79	0.66	0.04*
W3: for child not missing data at W4 (n)	-0.17 (890)	0.17 (988)	0.40 (872)
W3: for child missing data at W4 (n)	-0.10 (277)	0.08 (267)	0.24 (277)
P-value of difference between two groups (two-sided)	0.35	0.32	0.09

* Indicates significant p-value: BMI-for-age z-score at Wave 2 was significantly higher for children with BMI-for-age z-score recorded at Wave 3 compared to those with missing data at Wave 3.

There were significant associations between children's number of missing z-scores for BMI-for-age, height-for-age, and weight-for-age and several demographic factors (LORI, Indigenous identity, and cohort). This suggests that the existing data may be biased, and not representative of the original sample; this affects the conclusions which can be drawn from these data. The mean number of waves with missing z-scores increased with increasing Level of Relative Isolation (see Table 17); for example, the mean number of missing BMI-for-age z-score measurements was 1.75 in areas with High/Extreme LORI compared to 1.14 in areas with No LORI. Additionally, children who were identified as Torres Strait Islander or as Aboriginal and Torres Strait Islander

were significantly more likely than children who were identified as Aboriginal to have missing data (see Table 18).

Children in the Baby Cohort, compared to the Child Cohort, had a higher mean number of missing and excluded data points for weight-for-age and BMI-for-age z-scores, but not for height-for-age z-scores (see Table 19). There was not a significant association between the mean number of missing data points and gender, report of the child's general health, report of the family's weekly income, or primary carer's highest education qualification attained (see Appendix F for more detail).

Table 17: Mean number of missing data points for each indicator by LORI at Wave 1

Level of Relative Isolation (n)	**Mean number of missing BMI-for-age z-score measurements** (SD)	**Mean number of missing HFA z-score measurements** (SD)	**Mean number of missing WFA z-score measurements** (SD)
None (432)	1.14* (1.05)	1.03* (1.00)	0.90* (0.97)
Low (830)	1.40* (1.11)	1.33* (1.10)	1.11* (1.06)
Moderate (253)	1.65* (1.13)	1.60* (1.13)	1.48* (1.14)
High / Extreme (156)	1.75* (1.09)	1.71* (1.10)	1.65* (1.12)
*Total (1,671)^*	*1.49 (1.11)*	*1.36 (1.09)*	*1.20* (1.08)*

* Indicates significant p-value: number of missing z-score measurements and LORI are not independent (p-value for Pearson's chi2(12) < 0.05).
^ LORI is missing in Wave 1 for the 88 children who entered the study in Wave 2.

Table 18: Mean number of missing data points for each indicator by Indigenous identity

Indigenous identity (n)	**Mean number of missing BMI-for-age z-score measurements** (SD)	**Mean number of missing HFA z-score measurements** (SD)	**Mean number of missing WFA z-score measurements** (SD)
Aboriginal (1,535)	1.37* (1.09)	1.31* (1.08)	1.14* (1.07)
Torres Strait Islander (118)	1.75* (1.12)	1.65* (1.09)	1.48* (1.07)
Aboriginal and Torres Strait Islander (106)	1.88* (1.14)	1.81* (1.16)	1.59* (1.16)
Total (1,759)	*1.43 (1.11)*	*1.36 (1.10)*	*1.20 (1.08)*

* Indicates significant p-value: number of missing z-score measurements and ethnicity are not independent (p-value for Pearson's chi2(8) < 0.05).

Table 19: Mean number of missing data points for each indicator by cohort

Cohort (n)	**Mean number of missing BMI-for-age z-score measurements** (SD)	**Mean number of missing HFA z-score measurements** (SD)	**Mean number of missing WFA z-score measurements** (SD)
Baby Cohort (1,010)	1.47* (1.11)	1.38 (1.09)	1.22* (1.07)
Child Cohort (749)	1.37* (1.10)	1.33 (1.10)	1.16 (1.09)
Total (1,759)	*1.43 (1.11)*	*1.36 (1.09)*	*1.20* (1.08)*

* Indicates significant p-value: number of missing BMI-for-age z-score and weight-for-age z-score measurements and cohort are not independent (p-value for Pearson's chi2(4) < 0.05).

Some of the missing and implausible data might be attributable to the inability of interviewers to transport measuring equipment to certain distant, remote sites. In the cases where the standard measuring equipment was not available, children might not have been measured, or if they were measured, the accuracy of measurements was likely decreased. Therefore, a higher prevalence of missing and implausible data would be expected for sites where this equipment was not available. The areas that were hardest to reach, such as the Torres Strait Islands, might have had the highest percentage of Torres Strait Islander children, and might have also been classified as a High/Extreme LORI, explaining some of the above findings. The association between LORI (at the first wave of the study) and Indigenous identity is significant (Pearson chi2(6) = 442.19, $p < 0.001$); in the first wave of the study, 52% of Torres Strait Islander children in the sample resided in areas with a High/Extreme LORI, compared to only 4% of children identifying as Aboriginal.

The higher prevalence of missing BMI-for-age and weight-for-age z-scores for children in the Baby Cohort might be attributable to parents' increased hesitation to have younger children weighed; this is supported by the observed negative correlations between age (in months) and the number of missing BMI-for-age and weight-for-age z-scores (r = -0.07 and r = -0.08, respectively). This could also reflect the increased difficulty, and therefore decreased accuracy, in measuring recumbent length for children less than 24 months of age (compared to measuring standing height for older children). Together, LORI, Indigenous identity, age, and cohort explain 5.36% of the variability in

the number of missing BMI-for-age z-scores. These findings need to be taken into consideration in the interpretation of results, as results may not be as applicable to the groups with a disproportionately high prevalence of missing data: children from areas with the highest LORI, Torres Strait Islander children, and children in the younger cohort.

C) A quantitative evaluation of the LSIC birth weight and gestational age data

Birth weight data were received in raw form from FaHCSIA. In the first wave of the study, the primary carers were asked if they were able to access the Baby Book in which the child's birth weight had been recorded, as the LSIC team presumed this to be more accurate than primary carers' self-report of their child's birth weight. Of primary carers reporting a birth weight, 81% (1,145) reported it from the child's Baby Book, and 19% (274) reported it from memory. The source of the birth weight information was not recorded for one child (less than 1% of the sample). In the first wave of the study, 1,671 primary carers were surveyed. Additionally, there were 88 new entrants to the study who were not surveyed in the first wave; 73 of these primary carers were asked about birth weight in Wave 2, six were asked in Wave 3, and nine were not asked in any wave.

(1) Missing birth weight and gestational age data

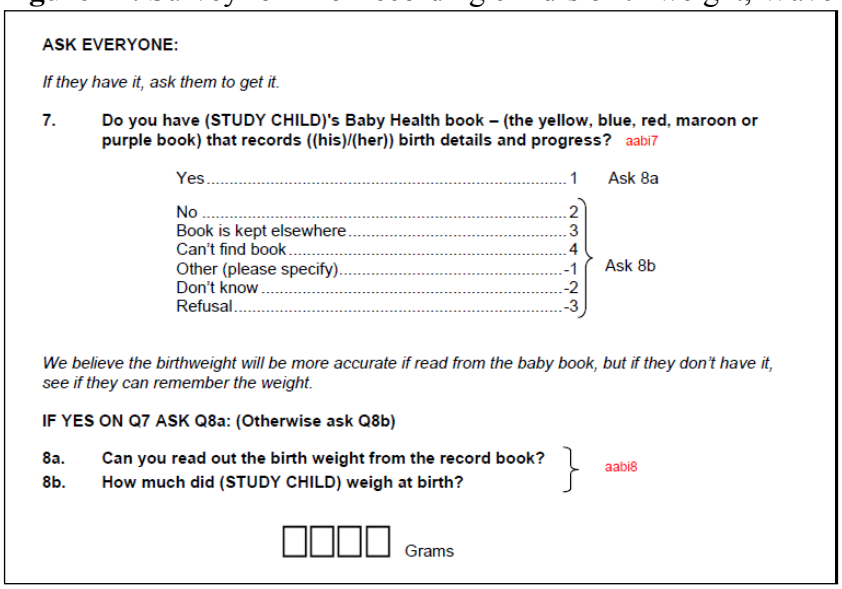
Overall, birth weights are recorded for 1,420 (81%) of the 1,759 children in the study. Some primary carers did not grant the RAOs permission to ask questions about the birth of the study child; of those willing to answer these questions, some did not have access to the Baby Book or did not know the child's birth weight (see Appendix H for more details). Some interviewers recorded additional comments about why the birth weight was not available. Gestational age was recorded for 1,619 children (92% of the sample). Eleven children had birth weights recorded but not gestational age, and 210 children had gestational age recorded but not birth weight. Overall, 1,409 children (80% of the sample) had both birth weight and gestational age recorded. For these children, z-scores of birth weight for gestational age can be calculated.

(2) Accuracy of birth weight and gestational age data

As with the height and weight data collected in LSIC, the accuracy of recorded birth weights is uncertain. Birth weights were recorded in kilograms and grams or in pounds and ounces. The recorded birth weights, using the units of measurement as specified, were converted to grams and these values were used to calculate preliminary z-scores for birth weight. The recorded birth weights, after conversion, ranged from 595 grams to 3,800,020 grams, with a mean of 20,724 grams and a standard deviation of 229,675 grams. Given the mean birth weight of 3,463 grams and 3,339 grams observed for males and females, respectively, in Australia in 2007 (17), it is clear that there are some implausible values included in the LSIC sample. The distribution of birth weights is skewed to the right by 32 birth weights exceeding 10,000 grams each. These implausible birth weights could be a result of data entry errors, incorrect specification of units of measurement, or misreporting of birth weight.

In the first wave of the study, four boxes were provided for entry of the birth weight in grams, but boxes were not available for entry of a birth weight in pounds (see Figure 11). As a result, RAOs may have entered a birth weight, in pounds, into the grams boxes. For new entrants asked about birth weight in a later wave of the study, boxes were provided to enter the birth weight in kilograms and grams, or in pounds and ounces (see Figure 12). This data entry set-up may have resulted in fewer data entry errors. Methods were developed to identify the implausible data points for exclusion.

Figure 11: Survey form for recording child's birth weight, Wave 1



ASK EVERYONE:

If they have it, ask them to get it.

7. Do you have (STUDY CHILD)'s Baby Health book – (the yellow, blue, red, maroon or purple book) that records ((his)/(her)) birth details and progress? aabi7

Yes	1	Ask 8a
No	2	Ask 8b
Book is kept elsewhere	3	
Can't find book	4	
Other (please specify)	-1	
Don't know	-2	
Refusal	-3	

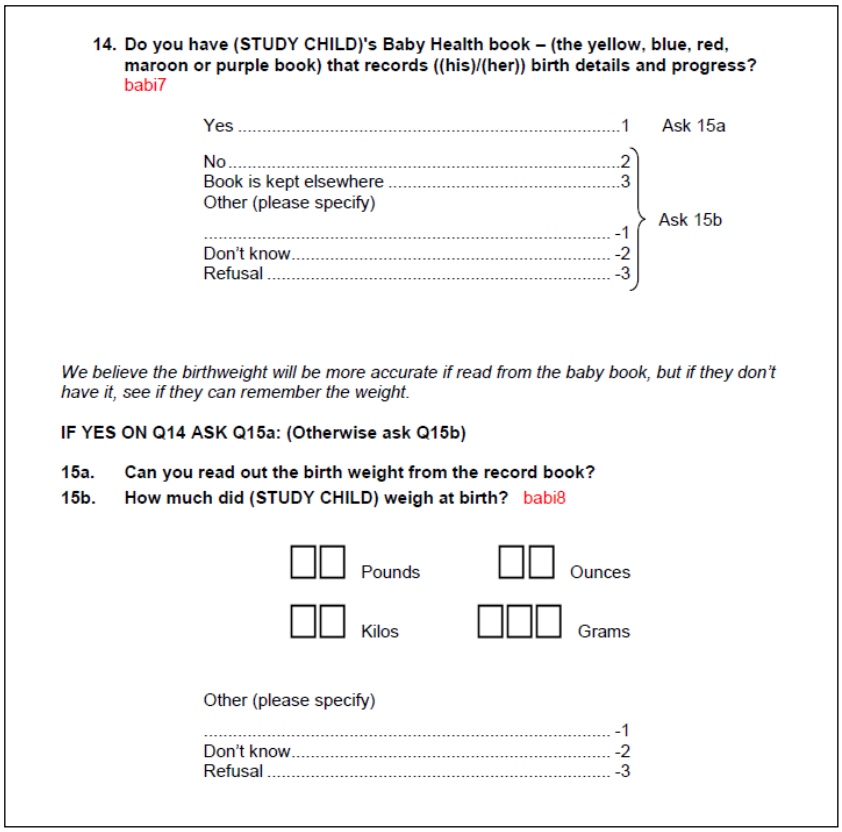
We believe the birthweight will be more accurate if read from the baby book, but if they don't have it, see if they can remember the weight.

IF YES ON Q7 ASK Q8a: (Otherwise ask Q8b)

8a. Can you read out the birth weight from the record book?
8b. How much did (STUDY CHILD) weigh at birth? aabi8

□□□□ Grams

Figure 12: Survey form for recording child's birth weight in Wave 2 or Wave 3 for new entrants



14. Do you have (STUDY CHILD)'s Baby Health book – (the yellow, blue, red, maroon or purple book) that records ((his)/(her)) birth details and progress? babi7

Yes ..... 1 — Ask 15a

No 2
Book is kept elsewhere 3
Other (please specify)
..... -1
Don't know ..... -2
Refusal ..... -3
— Ask 15b

We believe the birthweight will be more accurate if read from the baby book, but if they don't have it, see if they can remember the weight.

IF YES ON Q14 ASK Q15a: (Otherwise ask Q15b)

15a. Can you read out the birth weight from the record book?

15b. How much did (STUDY CHILD) weigh at birth? babi8

☐☐ Pounds ☐☐ Ounces

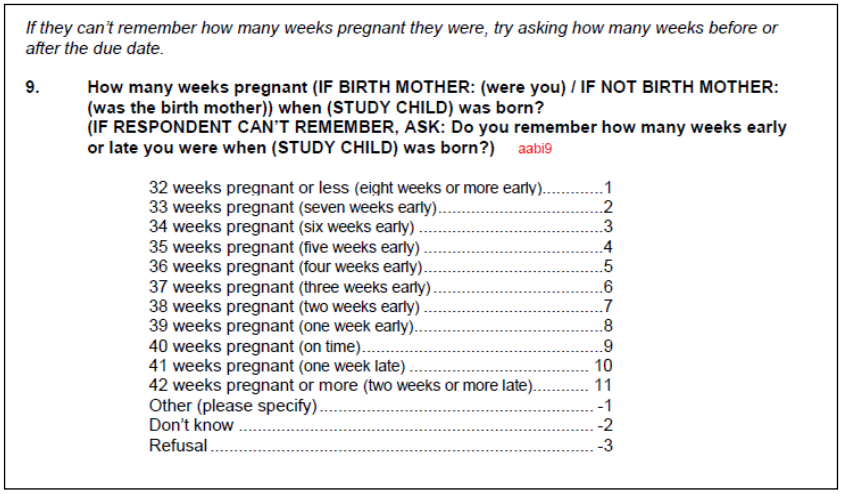
☐☐ Kilos ☐☐☐ Grams

Other (please specify)
..... -1
Don't know ..... -2
Refusal ..... -3

The accuracy of gestational age needs to be considered along with the accuracy of the birth weight. There are two limitations with the collection of gestational age in this study. First, gestational age was obtained by self-report. Primary carers were asked, 'How many weeks pregnant (IF BIRTH MOTHER: (were you) / IF NOT BIRTH MOTHER: (was the birth mother)) when (STUDY CHILD) was born?' (see Figure 13). If the primary carer could not remember, the RAO followed up with, 'Do you remember how many weeks early or late you were when (STUDY CHILD) was born?' The reliability of primary carers' recall of gestational age is unknown, and the accuracy of the response might vary by whether the primary carer was the birth mother or not. Of the 1,619 primary carers reporting gestational age, 1,529 (94%) were the birth mother, and 90 (6%) were not (see Appendix I for more details). Although a higher percentage of non-birth mothers reported that the child was born at term (47% for non-birth mothers versus 33% for birth mothers), the distribution of gestational age was not significantly different for non-birth mothers and birth mothers ($t(1617) = 0.7526$, $p = 0.4518$ for two-sample t-test with equal variance). This suggests that the gestational ages reported by birth and non-birth mothers represent the same distribution of

gestational ages, and therefore the same degree of accuracy will be assumed for the reports from non-birth mothers.

Figure 13: Survey form for recording child's gestational age, Wave 1 (a similar format was used in Wave 2)



If they can't remember how many weeks pregnant they were, try asking how many weeks before or after the due date.

9. How many weeks pregnant (IF BIRTH MOTHER: (were you) / IF NOT BIRTH MOTHER: (was the birth mother)) when (STUDY CHILD) was born?
(IF RESPONDENT CAN'T REMEMBER, ASK: Do you remember how many weeks early or late you were when (STUDY CHILD) was born?) aabi9

32 weeks pregnant or less (eight weeks or more early)	1
33 weeks pregnant (seven weeks early)	2
34 weeks pregnant (six weeks early)	3
35 weeks pregnant (five weeks early)	4
36 weeks pregnant (four weeks early)	5
37 weeks pregnant (three weeks early)	6
38 weeks pregnant (two weeks early)	7
39 weeks pregnant (one week early)	8
40 weeks pregnant (on time)	9
41 weeks pregnant (one week late)	10
42 weeks pregnant or more (two weeks or more late)	11
Other (please specify)	-1
Don't know	-2
Refusal	-3

The mechanism by which mothers determined their gestational age is unknown, increasing the uncertainty of reliability. The most commonly used indicator of gestational age is LMP, although this is less accurate than sonographic assessment (ultrasound) (90, 155). The use of LMP as a method of determining gestational age suffers from systematic error, with the common overestimation of gestational age leading to a positive skew of the distribution. Some primary carers in LSIC may have estimated their child's gestational age based on LMP, whereas others might have undergone sonographic testing, so the accuracy of gestational age may be inconsistent across individuals. The method by which gestational age was assessed is likely to vary by location, with use of sonographic testing less likely in areas with higher levels of relative isolation.

Although the accuracy of gestational age cannot be assessed, the association of explanatory variables with the absence or presence of recorded gestational age can be evaluated. There was a significant association between the presence of missing data for gestational age and the level of relative isolation (Pearson's $chi2(10) = 48.3806$, $p < 0.001$) (see Appendix I for more details). The risk ratio for missing data in urban (defined as No or Low LORI) compared to remote (defined as Moderate or High/Extreme LORI) was 0.70 (95% CI: 0.60, 0.83); the significantly lower prevalence of missing gestational age data in the most urban areas hints that the accuracy of reported gestational ages might be higher in these areas as well. Additionally, there was

a significantly higher percentage of missing gestational age data for non-birth mothers compared to birth mothers (risk ratio: 7.71, 95% CI: 5.84, 10.17) (see Appendix I for more details). This is unsurprising given that birth mothers are more likely than other carers to be aware of the gestational age of the child. There was not a significant difference in the prevalence of gestational ages at the extreme ends of the distribution (32 or fewer weeks or 42 or more weeks gestation) by LORI, but the difference approached significance for birth mothers compared to non-birth mothers (risk ratio: 1.04, 95% CI: 1.01, 1.06), with birth mothers more likely to report gestational ages at either extreme. Overall, there do not seem to be systematic biases in the reporting of gestational age, but these factors should be considered in analyses.

A second limitation of the gestational age data in LSIC is that the response choices for gestational age were limited; any gestational age 32 weeks or fewer was classified as '32 weeks pregnant or less,' and any gestational age 42 weeks or more was classified as '42 weeks pregnant or more.' Although these gestational ages are outside the normal range for births, there may have been children in the sample with gestational ages outside of the 32 to 42 week range, and information is lost by grouping them into this broader category. For the calculation of birth weight for gestational age category, any children in the '32 weeks or less' gestational age group were assigned a gestational age of 32 weeks, and any children in the '42 weeks or more' gestational age group were assigned a gestational age of 42 weeks. This has the potential to induce a skew in the data. If a child was born at 29 weeks, therefore falling into the '32 weeks or less' gestational age category and assigned a gestational age of 32 weeks, the child's birth weight for gestational age z-score would be smaller than if the true gestational age of 29 weeks was used. Similarly, if a child is born at 44 weeks, but is assigned a gestational age of 42 weeks, the child's birth weight for gestational age z-score would be artificially inflated. Fifty-six primary carers (3% of the sample) report that their child was born at a gestational age of 32 weeks or fewer, and 137 P1s (8% of the sample) report that their child was born at a gestational age of 42 weeks or more. Thus, for 12% of the sample, there is added uncertainty in gestational age, and therefore birth weight for gestational age z-score.

However, given the documented systematic error and positive skew of gestational age (155), birth weights at the post-term end (42 weeks gestation or greater) of the distribution are often incorrectly interpreted as low for their gestational age. The study design's categorisation of all gestational ages exceeding 42 weeks as 42 weeks

unintentionally compensates for this skew, decreasing the probability of post-term infants being incorrectly classed as SGA due to the overestimation of gestational age. Additionally, the prevalence of gestational ages less than 32 weeks or exceeding 42 weeks is very low in the general population. Of the 2.53 million births included in the 1998-2007 Australian birth weight reference, less than 1% had a gestational age less than 32 weeks, and less than 2% had a gestational age exceeding 42 weeks (17). Thus, although the gestational age classification system used is not ideal, it should not have a significant impact on the calculation of size for gestational age, potentially compensating for biases in the classification of gestational age.

(3) Cleaning of birth weight and gestational age data

To calculate birth weight for gestational age z-scores in the LSIC sample, I subtracted the birth weight of each child from the median gender- and gestational age-specific birth weight from the 1998-2007 Australian reference (17), and divided the difference by the gender- and gestational age-specific standard deviation (see Table 20Table 20). Z-scores are missing for children in the LSIC sample missing birth weight, gender, or gestational age.

Table 20: The median, standard deviation and number of birth weights included in the 1998-2007 Australian reference, by gender and gestational age (17)

	Males			**Females**		
Gestational age (weeks)	**Median birth weight** (grams)	**Standard Deviation** (grams)	**Number in Australian reference**	**Median birth weight** (grams)	**Standard Deviation** (grams)	**Number in Australian reference**
32	1,880	331	3,895	1,780	322	3,119
33	2,106	371	5,599	2,011	356	4,421
34	2,340	385	9,824	2,240	375	8,108
35	2,578	408	16,054	2,480	403	13,104
36	2,820	428	32,747	2,710	420	28,386
37	3,080	449	73,986	2,965	439	66,928
38	3,330	439	230,003	3,200	425	214,002
39	3,470	430	293,109	3,340	415	282,046
40	3,620	434	409,976	3,480	416	398,257
41	3,755	438	192,154	3,605	424	181,434
42	3,820	462	19,804	3,650	445	177,701

Implausible birth weight for gestational age z-scores can arise from errors in the recording of birth weight, gestational age, or both. I employed a flagging system to identify extreme or implausible birth weights. Birth weights with a z-score greater than

3 or less than -3 were flagged for review, given that they lie in the outer 0.13% of the reference data. Birth weights were also flagged for review if the units specified were pounds and ounces, and the number of ounces recorded exceeded 16. There are only 16 ounces in a pound, so the entry of a number greater than 16 in the ounces column is indicative of an error in data entry. Similarly, birth weights were flagged if a number exceeding 999 was entered into the grams column, as any number in the 1,000s of grams should have been entered into the kilograms column, and thus this might reflect an error in data entry. If the number of kilograms recorded was not a single digit, the birth weight was also flagged, given that a birth weight greater than nine kilograms would be associated with a z-score exceeding 15 for infants at term. These flags are intended to highlight potential errors in data entry, such as the incorrect specification of units or the recording of values into the wrong boxes.

Of the sample with recorded birth weights and gestational ages, 8% (115 children) were flagged for having birth weight for gestational age z-scores greater than +3 or less than -3, 2% (25 children) were flagged for having a birth weight entered with the number of ounces exceeding 16, and 1% (9 children) were flagged for having a birth weight with the number of grams exceeding 999 (see Table 21). Overall, birth weights were flagged by at least one indicator for 9% (124 children) of the sample with birth weight and gestational age recorded. Birth weights for infants at extreme ends of the gestational end spectrum (less than 32 weeks or more than 42 weeks gestation) were significantly more likely to be flagged than birth weights for infants born between 33 and 41 weeks gestation (Pearson chi2(10) = 60.9244, $p < 0.001$) (see Appendix J for more detail).

Table 21: Number of children with flagged birth weights

Flag	**Number of birth weights flagged** (% of sample with recorded birth weight and gestational age)
1. Birth weight z-score < -3 or > +3	115 (8.16)
2. Ounces > 16	25 (1.77)
3. Grams > 999	27 (1.92)
4. Flagged for comment	10 (0.71)
Flagged for one or more	124 (8.80)

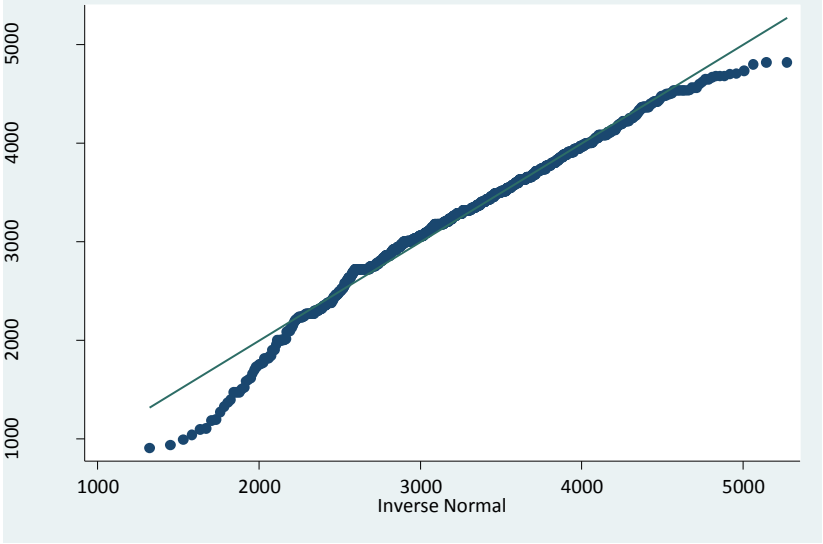
No flags	1,285 (91.20)
Total	*1,409* *(100.00)*

If the same four-digit birth weight was recorded in both the kilograms and the grams columns, this four digit number was interpreted to be the birth weight in grams, and the number in the kilograms column was ignored. When two two-digit numbers were recorded, one in the kilograms column and one in the grams column, this was assumed to be a data entry error, with the digits entered into the incorrect boxes. The two digits in the kilograms column were multiplied by 100 (rather than 1,000) and added to the two digits in the grams column to provide a birth weight in grams. If a four digit number was recorded in the kilograms column, and nothing was recorded in the grams column, the incorrect specification of units was assumed, and the weight was divided by 1,000 (to create a birth weight in the 1,000s of grams, rather than 1,000s of kilograms). If four digits were recorded in the grams column, and the first digit was repeated in the kilograms column, this was considered an error in data entry, and the digit in the kilograms column was ignored (rather than added to the four-digit birth weight in the grams column). After making these adjustments, z-scores were re-calculated. If the absolute value of these newly calculated z-scores exceeded 3, the birth weights and associated z-scores were re-coded to missing. After the completion of data cleaning, 1,304 birth weights remained.

(4) Results – normality of the distribution of birth weight

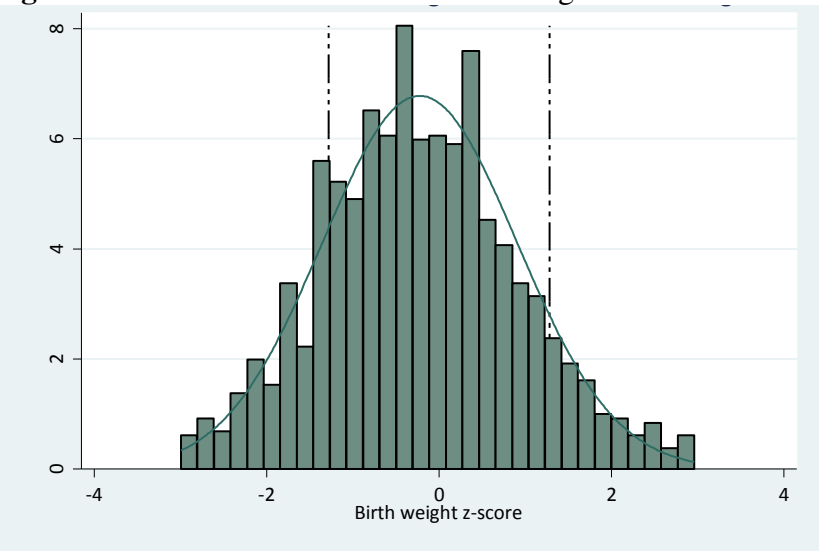
After the birth weight data were cleaned, birth weights remained for 1,315 infants. The distribution of birth weight and the associated quantile-normal plot of birth weight (see Figure 14) depict the lack of normality of the distribution, with an excess of values in the upper and, more notably, lower tails. Given the variation in birth weight attributable to gestational age, however, a normal distribution should not be expected for raw birth weight data. Adjusting birth weight for gender and gestational age can help normalise the data.

Figure 14: Quantile-normal plot of cleaned birth weight (in grams), for both cohorts



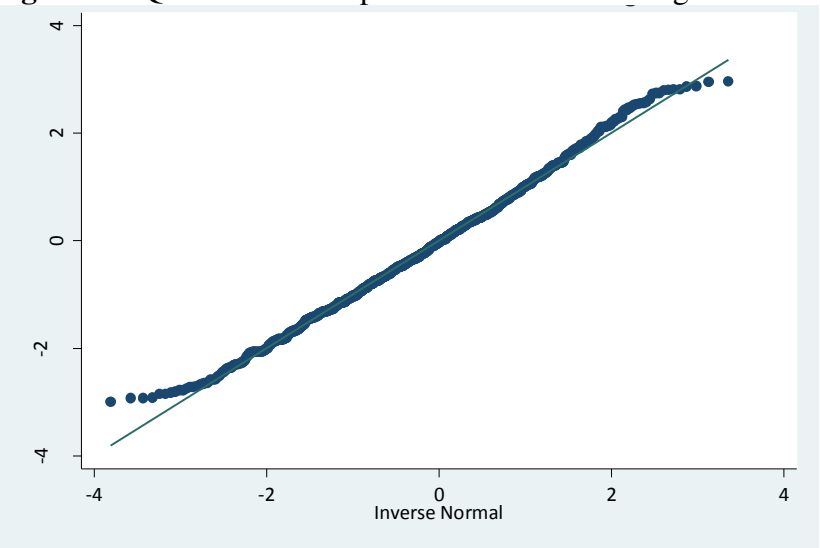
Transforming the birth weight, gender, and gestational age data to z-scores using the 1998-2007 Australian reference (17) increased the normality of the distribution (see Figure 15). Birth weight, gestational age, and gender were all required to calculate z-scores; 1,304 infants remained with z-scores recorded after cleaning. As demonstrated by the quantile-normal plot of birth weight z-score, there remains a larger than expected proportion of birth weight data in the outer tails of the distribution after adjusting for gender and gestational age (see Figure 16). Given the elevated prevalence of extreme birth weights among Indigenous, compared to non-Indigenous Australian infants, however, this lack of normality might be expected when using a reference predominantly based on non-Indigenous infants. Thus, this lack of normality does not represent a limitation of the data.

Figure 15: Distribution of cleaned birth weight z-scores in LSIC, for both cohorts



* Reference population: 1998-2007 Australian national birth weight centiles (17). The dotted lines represent the cut-off points for SGA and LGA.

Figure 16: Quantile-normal plot of cleaned birth weight z-scores, for both cohorts



(5) Discussion – biases in missing and implausible birth weight data

As with the missing data for height and weight, there were associations between some variables and the presence of missing birth weight data. Birth weight was significantly more likely to be missing or implausible when reported by non-birth mothers compared to birth mothers, with a risk ratio of 4.18 (95% CI: 3.07, 5.69). Similarly, birth weight was significantly more likely to be implausible when it was recalled from the primary carer's memory rather than reported from the Baby Book (risk ratio: 2.37, 95% CI: 2.02, 2.77). Birth weight was also significantly more likely to

be missing or implausible for children living in rural (defined as areas with Moderate or High/Extreme LORI) compared to urban (defined as areas with No or Low LORI) areas, with a risk ratio of 2.18 (95% CI: 1.85, 2.56). There was not a significant association between the prevalence of missing birth weight and gender or whether the child's gestational age was at either extreme (32 weeks or 42 weeks) versus in the middle range (33 to 41 weeks) (see Appendix L for more detail).

The decreased reliability of birth weight for non-birth mothers versus birth mothers and for birth weights recalled from carers' memory versus reported from the Baby Book is unsurprising. The observed association between LORI and missing birth weight z-scores is similar to that observed between LORI and missing BMI-for-age z-scores, suggesting that the accuracy of data collection was decreased in areas with higher levels of relative isolation.

The association between missing birth weight and the outcome (BMI-for-age z-score) was explored. The association between missing birth weight z-score and the number of waves at which BMI-for-age z-score was missing was not significant (Pearson chi2(4) = 4.03, p = 0.402), suggesting that different factors were associated with the collection of plausible birth weight data and the collection of plausible height and weight data. There was, however, a significant difference in the mean BMI-for-age z-score at Wave 1 for children with versus without a plausible birth weight z-score recorded (t(994) = -3.68, two-tailed p = 0.0002 for two-sample t-test with equal variances). This difference was also significant for BMI-for-age z-scores at Waves 2, 3, and 4, with significantly lower BMI-for-age z-scores for children missing birth weight z-scores (see Appendix L for more detail). There was also a significant difference in the weight-for-age z-scores at each wave for the two groups, and a significant difference in the height-for-age z-scores at Waves 1, 2, and 4. These differences persisted even after adjusting for LORI. This suggests that the children with birth weights recorded in LSIC are not representative of the entire sample of LSIC children. This impacts the conclusions that can be drawn from analyses using these data.

D) Conclusions

These analyses depict the limitations of the height, weight, and birth weight data collected in LSIC. I evaluated the anthropometric data quality in light of the information garnered from key informants and from the literature on Indigenous research methodologies. The development of a data cleaning method based on standardised

WHO protocols greatly benefited from the incorporation of this insight, resulting in an informed analysis of data quality.

There is a high prevalence of missing and implausible data for height, weight, and birth weight in LSIC; however, concern over this is lessened if the data are examined within a framework of understanding the necessity of forming and maintaining a relationship of trust with participants. The success of RAOs in gaining the trust of participants is demonstrated by the increased willingness of carers to have their child weighed and measured in later waves of the study. Additionally, examination of the distributions of height and weight and of the prevalence of implausible measurements across waves of the study reveals an improvement in measurement accuracy.

Some potential biases in the LSIC data do present themselves. First, the distribution of height at the first wave of the study is suggestive of a possible systematic bias, given the high prevalence of low height-for-age for both cohorts. This is confirmed by statements made by the RAOs describing the difficulty in collecting height at the first wave of the study due to the unreliable equipment. Their report of increased ease of measurement after the first wave of the study is reflected in the increased normality of the distributions of height at Waves 2, 3, and 4. Second, the prevalence of missing or implausible height-, weight-, and BMI-for-age z-scores was not independent of LORI, Indigenous identity, or cohort. RAOs' comments suggest that the elevated prevalence of missing data in the more remote areas and among Torres Strait Islander children is partially attributable to difficulties in transporting measuring equipment to these sites. The higher prevalence of missing data for the younger cohort might be attributable to carers' concerns about having their young children measured. Third, biases are observed in the prevalence of missing data for birth weight z-scores. Data were more likely to be missing or implausible if the primary carer was not the birth mother, if the information was provided by recall rather than from the Baby Book, and if the level of relative isolation was high. These biases should be considered in analysis, but do not pose a threat to the overall validity of the data. The significant difference observed in weight-, height-, and BMI-for-age z-scores for children with recorded versus missing or implausible birth weight z-scores poses a larger concern. This demonstrates that the subsample of LSIC children with birth weights recorded does not accurately represent the height and weight of the entire sample. This limitation needs to be considered in analyses of birth weight and BMI.

On the basis of my data cleaning analyses, FaHCSIA approved the anthropometric data for release for public use on the 4th of December, 2012. Keeping the described limitations in mind, the cleaned anthropometric data in LSIC should be considered a valid source of information about the sample's birth weight, height, and weight. Although the data are not representative of all Indigenous Australian children, they represent the longitudinal growth of two cohorts of geographically diverse children, and as such, are a valuable resource. With FaHCSIA's approval, analyses of these data can now be conducted.

Based on the data validity analyses I conducted, I propose recommendations to facilitate the collection of anthropometric data in future waves of LSIC, as well as in other settings. These recommendations aim to benefit study participants, data collectors, and researchers by simplifying the process of recording measurements and minimising the need for data cleaning.

(1) Recommendations specific to LSIC

1. Given the acknowledged limitations to the representativeness of the birth weight data, it would be beneficial to ask primary carers again to report the study child's birth weight. As a result of the formation of a trusting relationship between RAOs and the participating families, more accurate data may be collected at this point in time.
2. Given the higher prevalence of missing or implausible data in areas with higher levels of relative isolation and among Torres Strait Islander children, explicit efforts should be made to collect accurate data from these underrepresented groups.
3. Careful consideration should go into the design of the survey forms for recording height, weight, and birth weight. The form should provide a space to record measurements in every unit that is an option on the measuring instrument (for example kilograms and grams, and stones, pounds and ounces for weight). On the survey form for the first wave of LSIC, the only boxes available for recording weight were kilograms and grams (see Appendix D). If weight was measured in pounds, RAOs needed to convert the weight to kilograms themselves, or alternatively they could record the weight in the comment section (for example, one RAO wrote in, 'Please note weight done in Lbs not Kilos as scales playing up'). If the survey form includes options for recording measurements in every unit, there remains an opportunity to accurately record the measurement even in the event that

the scale is shifted to the wrong unit of measurement and the data collector is unable to change the scale back to the desired unit of measurement.

4. Additionally, the measurement accuracy of the equipment should be reflected in the number of significant figures available for input on the survey form. The stadiometers used in later waves of LSIC measured height to the nearest millimeter, but the form did not include a space to record this level of accuracy, only offering a space for the number of centimetres. As a result, the data recorded did not reflect the full accuracy of the measurement. Some RAOs recorded the number of millimetres in the comments section, and I manually added these to the heights recorded in centimetres. For example, one RAO wrote, 'Height was actually 96.8 centimetres but laptop will only accept whole number'. In the case of weight, the survey form allowed the input of weights (including birth weight) to a degree of accuracy higher than that which was possible on the scale (see Appendix D), resulting in data entry error.
5. To improve the process of data cleaning, data collectors should be given the option to identify any measurements that they have needed to estimate, or that they believe have compromised accuracy. With the current recording system, many RAOs recorded comments indicating that the reliability of the specific measurement was limited, such as 'This may not be accurate as my scales are playing up'. These comments allowed some of the less reliable measurements to be identified for exclusion; however, the recording of comments was not consistent across RAOs, and thus some measurements might not have been identified as unreliable. In the Longitudinal Study of Australian Children (LSAC), administered by the Australian Institute of Family Studies, data collectors also record children's height and weight measurements (156). After taking each measurement, the data collectors are asked to record if the measurement was an estimate or not. Inclusion of this question on the LSIC survey form would allow for a more systematic process of data cleaning. A space for comments about each measurement should still be included, as comments such as 'I know this seems like not a lot, but I did check and re-check the weight. She is a very slight child' can be informative for data cleaning.
6. To reduce the recording of implausible measurements, the survey form could be designed to immediately calculate the z-score of each height and weight measurement; extreme z-scores could alert the data collector of possible

measurement or data entry error, and encourage them to take the measurement again.

(2) General recommendations

1. Choice of measuring equipment is critical, and should take into consideration the setting in which measurement will occur.
2. Extensive training in the use of measuring equipment is important in order to minimise errors resulting from the improper use of the equipment.
3. Equipment should be calibrated regularly to ensure unbiased measurement.
4. Data collectors should be taught about the process of analysing anthropometric data (such as standardisation) in addition to their training of how to take these measurements. Learning about how the data will be used may improve their data collection skills. Reflecting on a presentation I gave to the LSIC team about my methods of evaluating the height and weight data, one RAO commented:

> That's the bit that I was really impressed with. As I told you before, it's not that we didn't know the data was being used, but how it was being broken down and put into studies and stuff, we hadn't really had examples of before. But, I mean, it definitely made me go, "Well, I'll be a bit more cleaner with my data now."

Chapter V: What is the distribution of height, weight, BMI, and birth weight in LSIC?

A) Age, height, and weight

(1) Results – distribution of age, height, and weight

After data cleaning, over 1,000 ages, weights and heights remained at each wave of the study (see Appendix M for more detail). The mean age of the younger cohort was 15.73 months in the first wave, 25.80 months in the second wave, 38.00 months in the third wave, and 49.47 months in the fourth wave. Children in the older cohort were approximately 36 months older, with a mean age increasing from 51.31 months to 60.71 months to 72.99 months to 84.56 months across waves.

The mean height of the younger cohort was 77.10 cm, 87.13 cm, 95.59, and 102.89 cm at successive waves of the study, and the mean height of the older cohort was 103.39 cm, 109.41 cm, 116.10 cm, and 121.84 cm. The mean height-for-age z-score for the younger cohort remained negative at each wave, moving from -0.56 in the first wave to -0.20, -0.29, and -0.23 in the second through fourth waves of the study, respectively. The older cohort demonstrated a trend of increasing height-for-age z-scores across waves, with mean z-scores of -0.33, -0.08, 0.03, and 0.06, respectively.

For the younger cohort, the mean weight increased from 10.86 kg in Wave 1 to 12.84 kg in Wave 2, 12.91 kg in Wave 3, and 16.87 kg in Wave 4. For the older cohort, the mean weight increased from 17.37 kg in Wave 1 to 19.13 kg in Wave 2, 21.47 kg in Wave 3, and 24.39 kg in Wave 4. The mean weight-for-age z-score decreased from 0.57 to 0.36 to 0.15 to 0.05 across waves for the younger cohort, and increased from 0.11 to 0.13 to 0.15 to 0.23 across waves for the older cohort.

BMI was assessed separately for children under the age of five years (up to and including 60 months) and children over the age of five years, given the disparate set of BMI-for-age z-score cut-offs implemented for these age groups (5, 6). The mean BMI-for-age z-score for the younger age group decreased from 0.92 to 0.56 to 0.50 to 0.32 across waves. The older cohort fluctuated, with the mean BMI-for-age z-score changing from 0.36 in Wave 1 to 0.10 in Wave 2, 0.18 in Wave 3, and 0.24 in Wave 4.

(2) Prevalence of low and high height-, weight-, and BMI-for-age

The prevalence of low height-for-age (defined as a z-score smaller than -2) for the younger cohort decreased in successive waves, from 20% in Wave 1 to 9% in Wave 2, 7% in Wave 3, and 5% in Wave 4. There was a similar decrease in the prevalence of low height-for-age in the older cohort, at 10% in the first wave of the study, 4% in the second wave of the study, and 2% in the third and fourth waves of the study. The drastic reduction in the prevalence of low height-for-age after the first wave of the study may be indicative of a systematic underreporting of height in the first wave of the study. Given that the decrease was observed within both cohorts, the decrease cannot be attributed to an age effect. The prevalence of high height-for-age was also elevated in the first wave of the study for the younger cohort, at 8%, decreasing to 7% in the second wave of the study, 5% in the third wave of the study, and 3% in the fourth wave. Among the older cohort, the prevalence of high height-for-age was 2% in the first wave, 3% in the second and third waves, and 5% in the fourth wave of the study. The inconsistency of trends within the two cohorts makes this less indicative of a systematic bias for heights at the tall end of the spectrum.

The prevalence of high weight-for-age (defined as a z-score greater than +2) for the younger cohort was highest in the first two waves of the study at 13%, compared to 6% in the latter two waves of the study. In contrast, the prevalence of high weight-for-age in the older cohort was highest in the fourth wave of the study, at 10%, compared to 8% in the first two waves of the study and 7% in the third wave of the study. The prevalence of low weight-for-age (defined as a z-score smaller than -2) for the younger cohort was 3% in the first wave of the study, 5% in the second wave of the study, and 4% in the third and fourth waves of the study. The prevalence of low weight-for-age in the older cohort was similar, at 6%, 5%, 5%, and 4% in successive waves.

The prevalence of low BMI-for-age (defined as a z-score smaller than -2) for the children under five years of age was 4% in the first wave, 8% in the second wave, and 4% in the third and fourth waves of the study. The prevalence of low BMI-for-age was higher among children over five years of age, at 8% in Wave 1, 11% in Wave 2, and 5% in waves 3 and 4. The membership of children within the two age categories changes between waves, and thus analysis of temporal trends is not possible. The prevalence of overweight for children under five years of age (defined as a BMI-for-age z-score between +2 and +3) was 13% in the first wave of the study, 14% in the second wave, 9% in the third wave, and 6% in the fourth wave. An additional 11% of children under

five years of age were obese (defined as a BMI-for-age z-score exceeding +3) at the first wave of the study, 7% at the second wave of the study, 3% at the third wave of the study, and 4% at the fourth wave of the study. For children over five years of age, 28% were overweight (defined as a BMI-for-age z-score between +1 and +2) and 4% were obese (defined as a BMI-for-age z-score exceeding +2) at the first wave of the study (with n = 25), 18% were overweight and 10% were obese at the second wave, 16% were overweight and 9% obese at the third wave, and 14% overweight and 12% obese at the fourth wave of the study. The incongruently high prevalence of high BMI-for-age for young children at the first wave of the study is likely associated with the high prevalence of low height-for-age at the first wave of the study, given the inverse relationship between BMI and height.

Overall, 32% of the younger cohort is classified as overweight or obese according to age-specific cut-offs at Wave 1, 22% at Wave 2, 12% at Wave 3, and 10% at Wave 4. Of the older cohort, 15% are classified as overweight or obese at Wave 1, 22% at Wave 2, 25% at Wave 3, and 26% at Wave 4. The prevalence of overweight is lower for the both cohorts when evaluating weight status using weight-for-age z-scores (13%, 13%, 6%, and 6% for the younger cohort and 8%, 8%, 7%, and 10% for the older cohort across the four waves). Although the distribution of children across the weight categories was not congruent, there was a significant association between the category membership assigned by the two indicators for each wave and cohort. For example, the per cent of children identified as overweight according to their weight-for-age z-score that were also identified as overweight according to their BMI-for-age z-score ranged from 80% in Waves 1 and 2 to 84% in Wave 3 to 92% in Wave 4 (see Table 22). Each indicator provides unique information about children's weight status; BMI will be used in further analyses given its inclusion of both weight and height information.

Table 22: The association between weight classification by weight-for-age and BMI-for-age z-scores in Wave 4

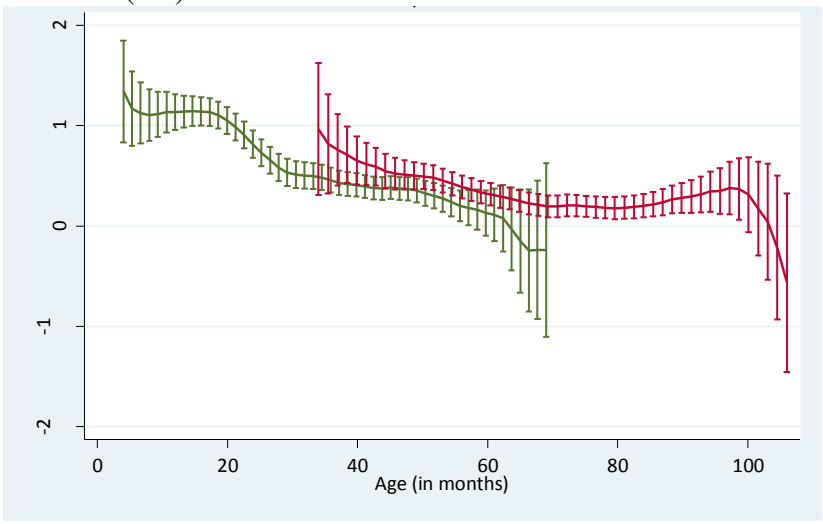
	BMI-for-age-category, Wave 4		
Weight-for-age category, Wave 4	**Low** (z < -2)	**Normal** (-2 ≤ z ≤ +2 for children < 5; -2 ≤ z ≤ +1 for children > 5)	**High (overweight or obese)** (z > +2 for children < 5; z > +1 for children > 5)
Low (z < -2) (%)	28 (66.67)	14 (33.33)	0 (0.00)
Normal (-2 ≤ z ≤ 2) (%)	24 (2.33)	901 (87.39)	106 (10.28)
High (z > 2) (%)	0 (0.00)	8 (8.25)	89 (91.75)

(3) Variation of BMI-for-age z-scores by demographic factors

(a) Age and cohort

There was a negative association between BMI-for-age z-score and age (in months) in the sample for both cohorts (see Figure 17). For the younger cohort, the mean BMI-for-age z-score decreased from 1.13 for children less than two years of age (n = 763) to 0.26 for children over four years of age (n = 416). This represents a shift from a mean BMI status of 'at risk of overweight' in the early years to a mean BMI in the healthy range by age four years. For the older cohort, the mean BMI-for-age z-score decreased from 0.52 for children less than five years of age (n = 655) to 0.21 for children over six years of age (n = 743), falling in the healthy range for both groups. The distribution of BMI-for-age z-scores at each age is higher in the older cohort compared to the younger cohort, as shown by the gap between their respective means on the graph at overlapping ages (from 34-69 months). Looking at all ages together, however, the mean BMI-for-age z-score was significantly higher for the younger cohort (0.62) compared to the older cohort (0.31) ($t(4,520) = -6.70$, two-sided $p < 0.0001$ for two-sample t-tests with equal variances), as would be expected given the observed negative association between BMI-for-age z-score and age.

Figure 17: BMI-for-age z-score versus age (in months), for Cohort B (green) versus Cohort K (red)



* The green line shows the mean and 95% confidence interval for the BMI-for-age z-scores of the younger cohort at each age; the red line shows the same for the older cohort. Data from all waves are included, so each individual has between zero and four BMI-for-age measurements included.

(b) Indigenous identity

There was not a significant association between Indigenous identity (Aboriginal, Torres Strait Islander, or Aboriginal and Torres Strait Islander) and BMI-for-age z-scores for either cohort at any wave ($p > 0.05$ for each one-way ANOVA) (see Figure 18 and Figure 19).

Figure 18: BMI-for-age z-score versus Indigenous identity for Cohort B, by wave

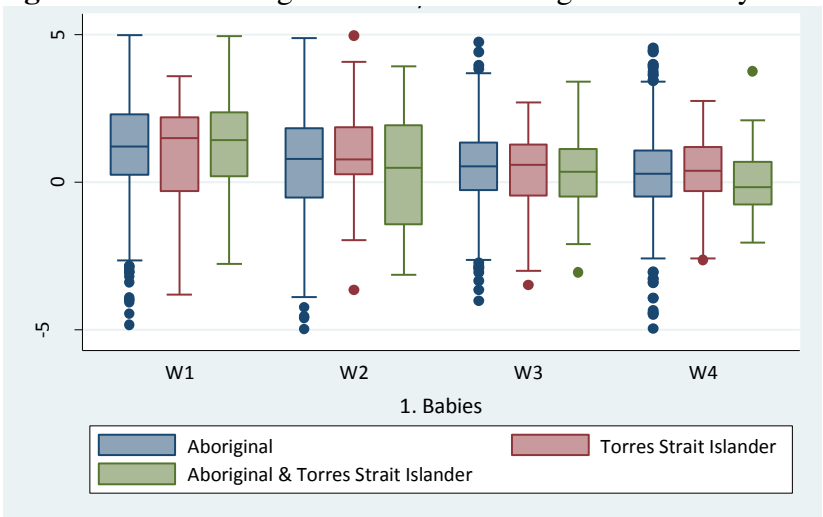
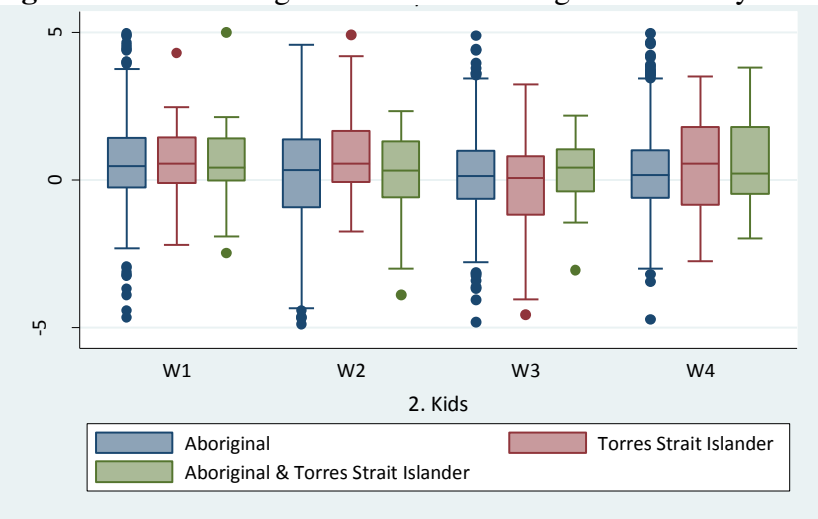


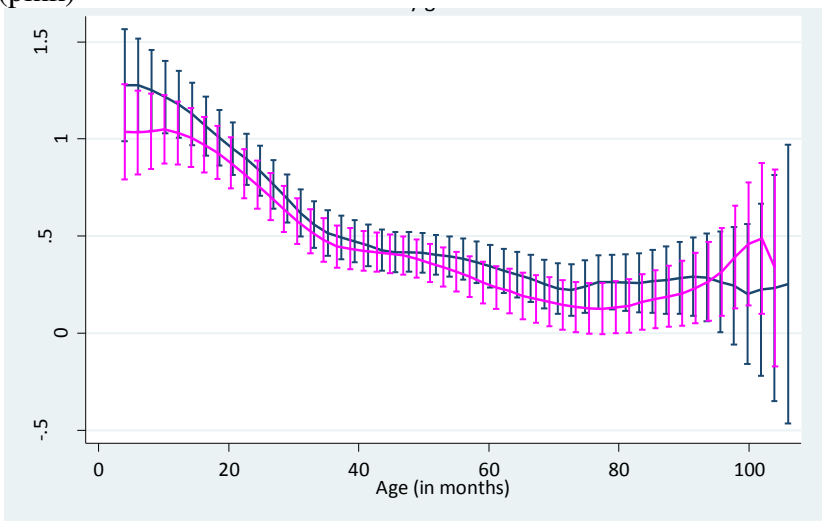
Figure 19: BMI-for-age z-score versus Indigenous identity for Cohort K, by wave



(c) Gender

The mean BMI-for-age z-score was not significantly different for males and females at any wave of the study, for either cohort ($p > 0.05$ for each two-sample t-test with equal variance) (see Figure 20).

Figure 20: BMI-for-age z-score by age (in months), for males (navy) versus females (pink)



* The navy line shows the mean and 95% confidence interval for the BMI-for-age z-scores for males at each age; the pink line shows the same for females. Data from all waves are included, so each individual has between zero and four BMI-for-age measurements included.

(d) LORI

Overall, the mean BMI-for-age z-score decreased with increasing level of relative isolation (see Figure 21 and Figure 22). The mean BMI-for-age z-score at each wave was significantly higher for urban children (living in areas with No or Low LORI) compared to remote children (living in areas of Moderate or High/Extreme LORI) ($p < 0.001$ for each two-sample t-test with equal variances).

Figure 21: BMI-for-age z-score versus Level of Relative Isolation for Cohort B, by wave

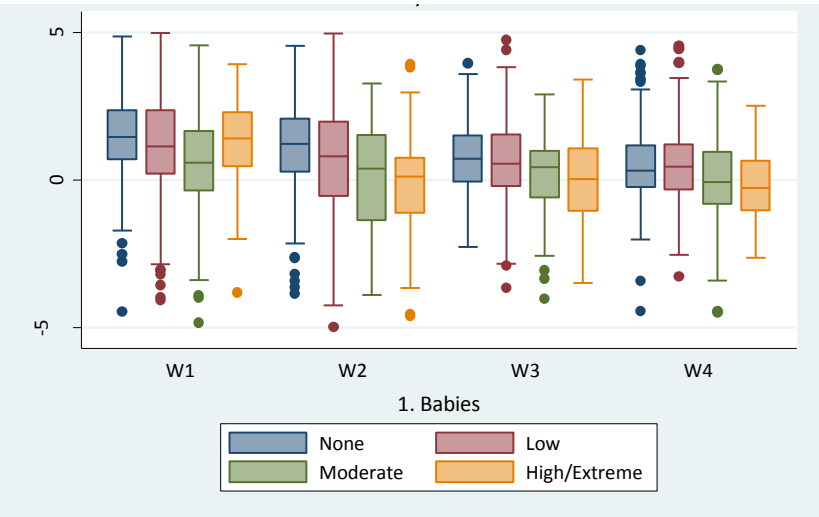
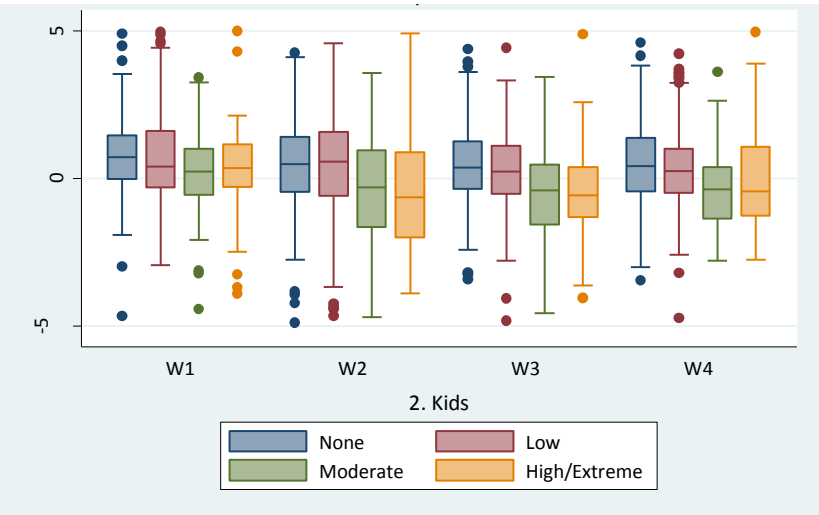


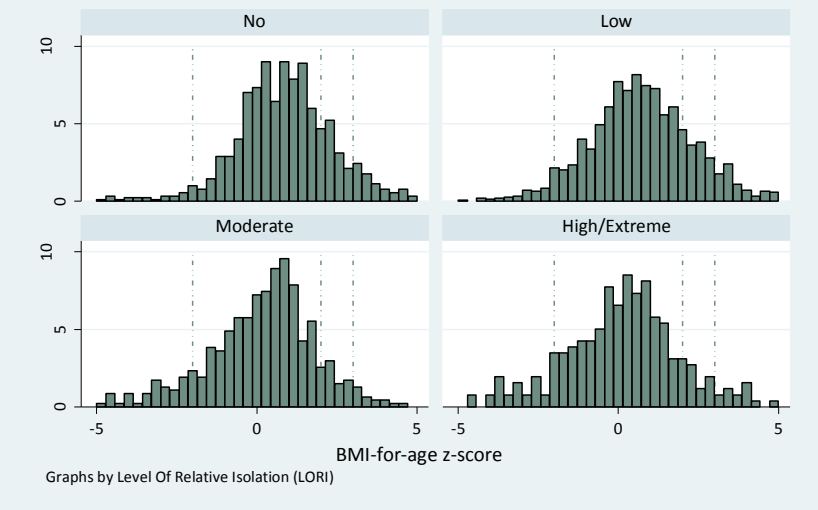
Figure 22: BMI-for-age z-score versus Level of Relative Isolation for Cohort K, by wave



Examining the entire distribution of BMI-for-age z-scores demonstrates the heterogeneity of weight status, with underweight and overweight observed within each

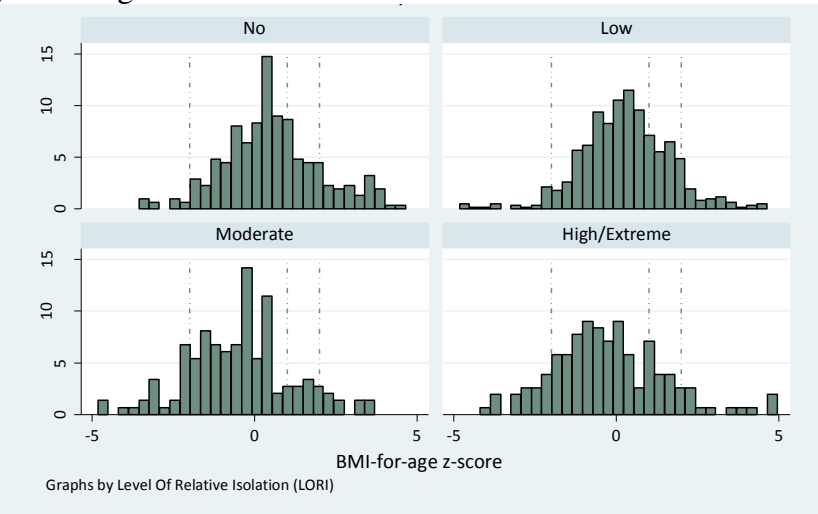
LORI (see Figure 23 and Figure 24). BMI category was significantly associated with LORI for each wave of the study (with $p < 0.05$ for each Pearson chi2 test); the prevalence of underweight was higher in the more remote areas, and the prevalence of overweight and obesity was higher in the more urban areas. However, even in the most remote areas (with High/Extreme LORI), there was a high prevalence of overweight and obesity (according to age-specific BMI-for-age z-score cut-offs). At the fourth wave of the study, the prevalence of underweight, overweight and obesity was equivalent, at 8%, in areas with High/Extreme LORI (for both cohorts together) (see Table 23). In contrast, the prevalence of underweight was 3% in areas with No LORI, and the prevalence of overweight and obesity was 11% and 10%, respectively (for both cohorts together).

Figure 23: Distribution of BMI-for-age z-scores by wave, for children less than five years of age



* The dotted lines represent the cut-off points for underweight (z = -2), overweight (z = +2), and obese (z = +3) for children under five years of age.

Figure 24: Distribution of BMI-for-age z-scores by wave, for children more than five years of age



* The dotted lines represent the cut-off points for underweight (z = -2), overweight (z = +1), and obese (z = +2) for children over five years of age.

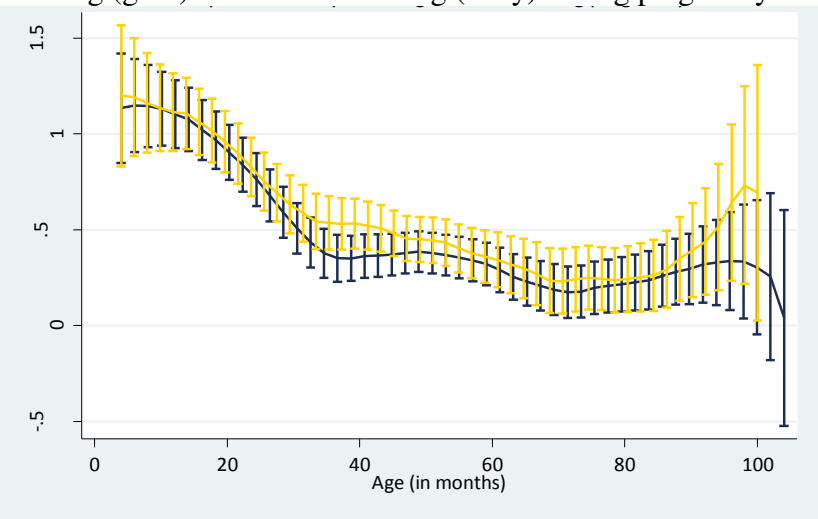
Table 23: Number (and per cent) of children in each BMI category (according to age-specific cut-off points) at Wave 4, by LORI

LORI	Underweight	Healthy weight	Overweight	Obese	*Total*
No	8	240	34	31	*313*
(%)	(2.56)	(76.68)	(10.86)	(9.90)	*(100)*
Low	15	406	57	26	*504*
(%)	(2.98)	(80.56)	(11.31)	(5.16)	*(100)*
Moderate	13	116	9	7	*145*
(%)	(8.97)	(80.00)	(6.21)	(4.83)	*(100)*
High/Extreme	7	70	7	7	*91*
(%)	(7.69)	(76.92)	(7.69)	(7.69)	*(100)*
Total	*43*	*832*	*107*	*71*	*1,053*
(%)	*(4.08)*	*(79.01)*	*(10.16)*	*(6.74)*	*(100)*

(e) Smoking during pregnancy

Half of mothers in LSIC reported smoking cigarettes while they were pregnant with the study child. Overall, the mean BMI-for-age z-score was 0.208 units higher for children whose mothers reported smoking during their pregnancy than those whose mothers did not report smoking. However, there was not a significant difference in the mean BMI-for-age z-scores of children of smoking and non-smoking mothers when looking at each wave and cohort individually. The two groups seem to demonstrate a similar growth trajectory until they diverge around the age of 90 months, with increasing BMI-for-age z-scores in the group whose mothers reported smoking during pregnancy, and decreasing BMI-for-age z-scores in the other group (see Figure 25).

Figure 25: BMI-for-age z-score by age (in months), for children whose mothers report smoking (gold) versus not smoking (navy) during pregnancy



* The navy line shows the mean and 95% confidence interval for the BMI-for-age z-scores for children whose mothers report not smoking during their pregnancy; the yellow line shows the same for children whose mothers report smoking during their pregnancy. Data from all waves are included, so each individual has between zero and four BMI-for-age measurements included.

B) Birth weight and gestational age

(1) Calculating birth weight for gestational age z-scores

The standard definition of size for gestational age uses the 10th percentile as the cut-off point for SGA and the 90th percentile as the cut-off point for LGA. This definition of SGA, however, is not outcome-based, and 'need not be used to define SGA for all applications. For epidemiologic studies of fetal growth retardation, the fifth or even the third percentile might be preferable' (137 p. 46-47). The somewhat arbitrary selection of the cut-off points dividing SGA, AGA, and LGA, is a limitation of this categorical approach; researchers and clinicians have described the inadequacy of these categories in highlighting differences between infants, especially for those falling below the 1st or 3rd percentile (157). The use of z-scores for birth weight provides more information about the child's size than the standard categorical approach; thus, this approach will be used in this study.

Calculating a z-score for birth weight for gestational age, rather than simply classifying an infant as small, appropriate, or large for gestational age, enables the assessment of an infant's size for gestational age on a continuous scale. The magnitude of the deviance from the median birth weight for infants of the same gestational age and gender is more apparent through the use of z-scores than through the classification of

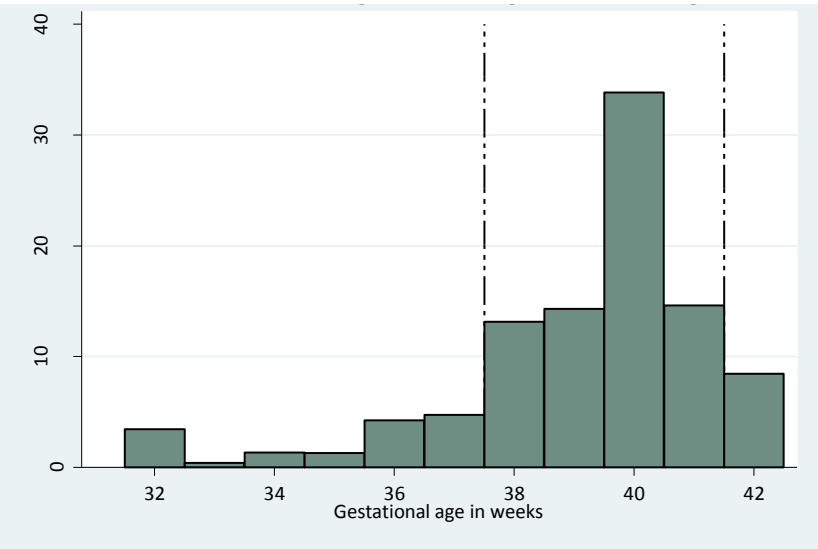
infants into size for gestational age categories based on arbitrary cut-off points. Thus, analysis of birth weight z-scores (adjusted for birth weight for gestational age) maximises the use of available information regarding birth status. Where size for gestational age category is evaluated in this study, SGA will be defined as a birth weight z-score less than -1.28, equivalent to a birth weight falling below the tenth percentile for infants of the same gestational age and gender. Similarly, LGA will be defined as a birth weight z-score exceeding +1.28, equivalent to a birth weight falling above the 90th percentile for gender and gestational age.

(2) Results – distribution of birth weight and gestational age

After the birth weight data were cleaned, birth weights remained for 1,315 infants. The mean birth weight of the LSIC sample was 3,372 grams (with a standard deviation of 636 grams) for males, and 3,217 grams (with a standard deviation of 599 grams) for females (see Appendix N for more detail). The prevalence of low birth weight (defined as a weight less than 2,500 grams) in the LSIC sample was 10%, and the prevalence of high birth weight (defined as a birth weight exceeding 4,000 grams) was 11%. The majority of infants (80.73%) in LSIC were born at term (see Figure 26).

Around 11% of births were recorded as pre-term, and just over 8% were recorded as post-term. The mean birth weight z-score of the sample was -0.2, suggesting that the LSIC sample was shifted to the left of the 1998-2007 Australian reference (17). The prevalence of SGA (defined as birth weight $z < -1.28$) in LSIC was 18%, and the prevalence of LGA (defined as birth weight $z > +1.28$) was 10% (see Appendix N for more detail).

Figure 26: Distribution of gestational age in LSIC, for both cohorts



* The dotted lines represent the cut-off points for pre-term and post-term birth.

(3) Prevalence of low and high birth weight

The mean birth weights observed in the LSIC sample (3,372 grams for males, and 3,217 grams for females) are lower than the mean birth weight of 3,476 grams (with a standard deviation of 550 grams) for Indigenous males and 3,345 (with a standard deviation of 516 grams) for Indigenous females observed in the 1991-1994 Australian reference (89). Information about Indigenous identity was not available in the 1998-2007 Australian reference, so a comparison cannot be made to the Indigenous subsample of this more recent reference. The low birth weight prevalence of 10% observed in LSIC is comparable to that of Indigenous infants in the 1991-1994 Australian reference. In that reference, the prevalence of low birth weight was more than twice as high within the Indigenous, compared to non-Indigenous, subsample, at 11% and 4%, respectively (89).

The prevalence of high birth weight (defined in this study as a birth weight exceeding 4,000 grams) was 11% in LSIC. In the 1991-1994 Australian reference (89), the prevalence of high birth weight among Indigenous infants was 1.3%, but high birth weight was defined as a weight exceeding 4,500 grams, rather than 4,000 grams. Using this definition, the prevalence of high birth weight in LSIC is 2.8%. The prevalence of high birth weight (again using the cut-off of 4,500 grams) in the 1998-2007 Australian reference was 1.9%. Information about the Indigenous identity of infants in the 1998-2007 Australian reference is unavailable, so direct comparison to the 1991-1994 figure for Indigenous infants is not possible. The prevalence of high birth weight was lower

among Indigenous (1.3%) compared to non-Indigenous (1.8%) infants in the 1991-1994 sample, so the prevalence of high birth weight among Indigenous infants in the 1998-2007 sample might be expected to be lower than the overall prevalence of 1.9%.

More recent data for Indigenous infants is not available to determine if the elevated prevalence of high birth weight in LSIC (2.8% using 4,500 grams as the cut-off for high birth weight) reflects a national secular trend of increasing birth weight over time. However, an increase in birth weight has been observed among Cree (First Nation) infants, with a 36.5% prevalence of high birth weight observed in a sample of 2,127 infants (135). This has been attributed to increasing maternal weight and prevalence of gestational diabetes, and decreasing cigarette use during pregnancy (135). Thus, although a secular trend in increasing birth weight has not been identified in the general Australian population, an elevated prevalence of high birth weight within an Indigenous sample is feasible given the high rates of maternal overweight and diabetes.

The prevalence of pre-term birth in LSIC was comparable to that observed among Indigenous infants in the 1991-1994 Australian reference, at 11% and 12%, respectively (89). The prevalence of post-term birth was just over 8% in LSIC, much higher than the prevalence of 3% observed among Indigenous infants in the 1991-1994 Australian reference (89). This elevated prevalence of post-term birth might partially explain the elevated prevalence of high birth weight in the sample, given the positive association between birth weight and gestational age.

(4) Prevalence of small-, appropriate-, and large-for-gestational age

In this study, SGA was defined as a birth weight z-score less than -1.28, and LGA was defined as a birth weight z-score greater than +1.28. These z-score cut-off points were chosen to reflect the standard percentile definitions of SGA and LGA, representing the lowest and highest deciles of birth weight for gestational age, respectively. According to this definition, the expected prevalence of both SGA and LGA would be 10%, and the expected prevalence of AGA would be 80%. Thus, the prevalence of LGA in LSIC (10%) is approximately what would be expected; this is consistent with the finding in the 1991-1994 Australian reference of a similar prevalence of high birth weight among Indigenous and non-Indigenous infants (the prevalence of LGA for Indigenous versus non-Indigenous infants was not calculated for the sample, so the prevalence of high birth weight can be used as an indirect measure).

In contrast, the prevalence of SGA in LSIC (18%) was much higher than the expected 10%. This is consistent with the 1991-1994 Australian birth weight reference (89); the prevalence of SGA in the overall sample (of both Indigenous and non-Indigenous infants) was defined as 10%, but the prevalence among the Indigenous infants was 17%. Thus, using the 1998-2007 nationally representative Australian reference (17) to evaluate birth weights in LSIC, it is unsurprising that the prevalence of SGA exceeds 10%. The lack of normality in the distribution of birth weight z-scores does not represent a limitation of the birth weight data.

Within the LSIC sample, there was a significant association between birth weight category (low, normal, or high) and size for gestational age category (SGA, AGA, or LGA). However, given the contribution of both intrauterine growth and gestational age to birth weight, not all SGA infants are low birth weight, not all AGA infants are normal birth weight, and not all LGA infants are high birth weight (see Table 24). The majority (59%) of low birth weight infants were born pre-term, suggesting that their low birth weight was partially attributable to their premature birth (see Table 25). The remaining 41% of low birth weight infants were born at term, indicative of a failure to reach optimal foetal growth, potentially as a result of intrauterine growth restriction. The majority (86%) of normal birth weight infants were born at term, but 5% were born pre-term and 9% were born post-term. The majority (75%) of high birth weight infants were born at term, and 24% were post-term.

Table 24: Association between birth weight category and size for gestational age category

	Size for gestational age category			
Birth weight category	Small for gestational age (%)	Appropriate for gestational age (%)	Large for gestational age (%)	*Total*
Low birth weight	81 (61.83)	49 (37.40)	1 (0.76)	*131*
Normal birth weight	152 (14.77)	847 (82.31)	30 (2.92)	*1,029*
High birth weight	0 (0.00)	51 (35.42)	93 (64.58)	*144*
Total	*233*	*947*	*124*	*1,304*

Table 25: Association between term of birth and birth weight category

	Term of birth			
Birth weight category	Pre-term (%)	Term (%)	Post-term (%)	*Total*
Low birth weight	78 (59.54)	53 (40.46)	0 (0.00)	*131*
Normal birth weight	51 (4.96)	890 (86.49)	88 (8.55)	*1,029*
High birth weight	1 (0.69)	108 (75.00)	35 (24.31)	*144*
Total	*130 (9.97)*	*1,051 (80.60)*	*123 (9.43)*	*1,304*

There was also a significant association between term of pregnancy and size for gestational age category (Pearson chi2(4) = 10.3392, p = 0.035) (see Table 26). Of SGA infants, 81% were born at term and 6% were born post-term; these infants demonstrate a failure to reach their expected growth potential. Thirteen per cent of SGA infants were born pre-term; the small size of these infants is thus attributable to both intrauterine growth restriction and pre-term birth. The majority (81%) of AGA infants were born at term. Nine per cent of AGA infants were born pre-term; despite being born prematurely, they did not exhibit intrauterine growth restriction. An additional 10% of AGA infants were born post-term, but maintained a size within the normal range for their gestational age. Three-quarters of LGA infants were born at term, 12% were born pre-term, and 13% were born post-term. Thus, SGA categories, birth weight categories, and gestational age categories all provide distinct information. The use of z-scores, adjusting birth weight for gestational age, maximises the incorporation of available data.

Table 26: Association between term of birth and size for gestational age category

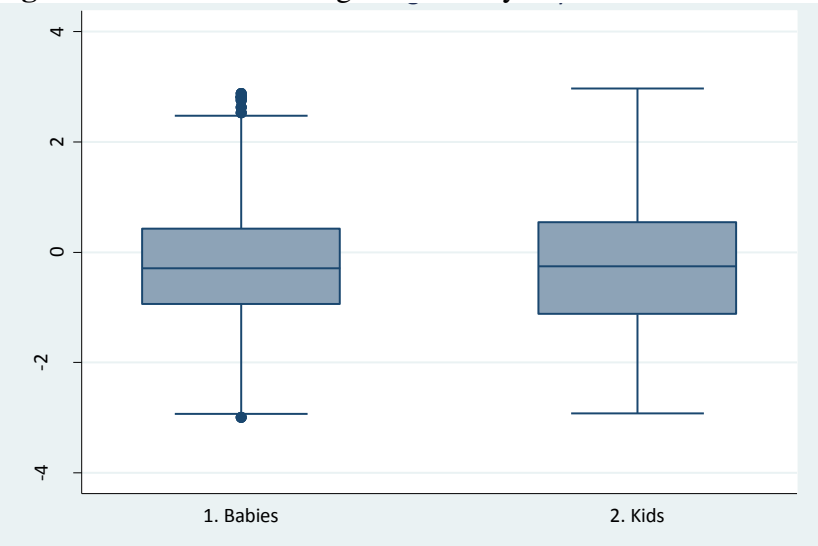
	Term of birth			
Size for gestational age category	Pre-term (%)	Term (%)	Post-term (%)	*Total*
Small for gestational age	31 (13.30)	189 (81.12)	13 (5.58)	*233*
Appropriate for gestational age	84 (8.87)	769 (81.20)	94 (9.93)	*947*
Large for gestational age	15 (12.10)	93 (75.00)	16 (12.90)	*124*
Total	*130 (9.97)*	*1,051 (80.60)*	*123 (9.43)*	*1,304*

(5) Variation of birth weight z-scores by demographic factors

(a) Cohort

The mean birth weight z-score was slightly lower for the older cohort, at -0.24 (95% CI: -0.29, -0.14) compared to -0.21 (95% CI: -0.34, -0.13) but the difference was not significant (see Figure 27).

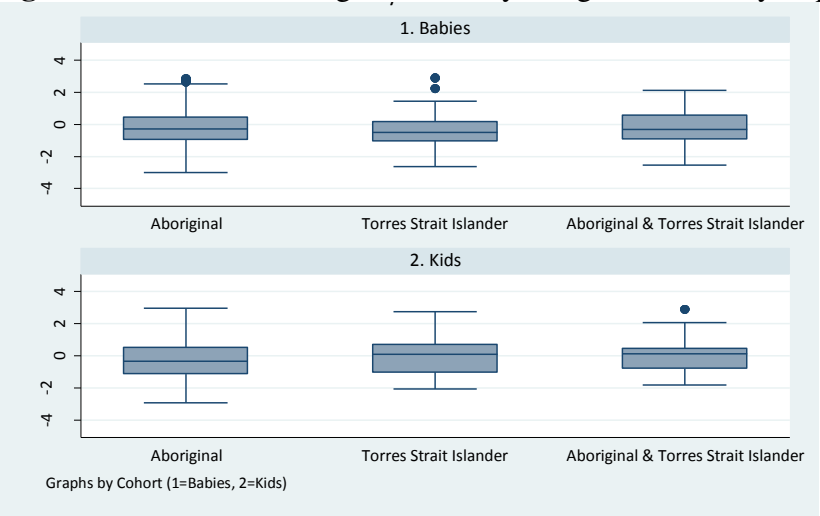
Figure 27: Mean birth weight z-score by cohort



(b) Indigenous identity

There was not a significant association between birth weight z-score and Indigenous identity for either of the two cohorts, and each cohort displayed a different pattern across the three groups (see Figure 28). For Cohort B, the mean birth weight z-score was lowest for Torres Strait Islander infants at -0.38 (95% CI: -0.66, -0.10), followed by Aboriginal infants at -0.20 (95% CI: -0.29, -0.12), and Aboriginal and Torres Strait Islander infants at -0.18 (95% CI: -0.47, 0.11). In contrast, for Cohort K, the mean birth weight z-score was lowest for Aboriginal infants at -0.28 (95% CI: -0.39, -0.17), and higher for Torres Strait Islander and Aboriginal and Torres Strait Islander infants, at -0.01 (95% CI: -0.36, 0.34) and 0.02 (95% CI: -0.45, 0.49), respectively.

Figure 28: Mean birth weight z-score by Indigenous identity, separated by cohort



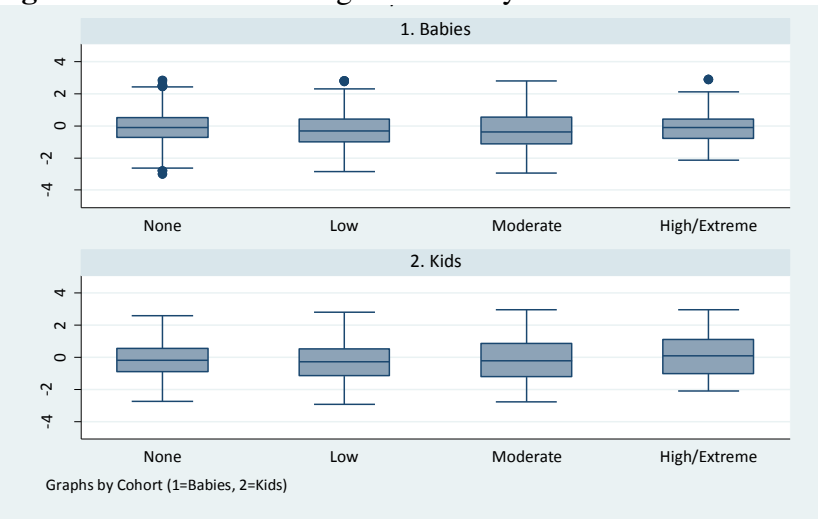
(c) Gender

The difference in birth weight z-scores between males and females was not significant for either cohort ($p > 0.05$ for each two-sample t-test with equal variances). In the younger cohort, the mean birth weight z-score was higher for females than males, at -0.20 (95% CI: -0.31, -0.09) and -0.23 (95% CI: -0.33, -0.12), respectively. In the older cohort, the opposite trend emerged, with a mean birth weight z-score of -0.34 (95% CI: -0.49, -0.19) for females and -0.15 (95% CI: -0.29, -0.01) for males.

(d) LORI

Birth weight z-score was not significantly associated with LORI for the older cohort ($F(518) = 1.54$, $p = 0.2036$ for one-way ANOVA), but the association was significant for the younger cohort ($F(749) = 2.89$, $p = 0.0346$ for one-way ANOVA). For the younger cohort, the mean birth weight z-scores was lowest in areas with Low and Moderate LORI, at -0.31 (95% CI: -0.42, -0.20) and -0.30 (95% CI: -0.55, -0.04) respectively, and higher in areas of High/Extreme and No LORI, at -0.13 (95% CI: -0.39, 0.14) and -0.06 (95% CI: -0.20, 0.08), respectively (see Figure 29). A similar trend was observed in the older cohort, but the mean birth weight z-score was highest for areas with High/Extreme LORI at 0.09 (95% CI: -0.32, -.49), rather than areas with No LORI (mean = -0.19, 95% CI: -0.38, -0.01).

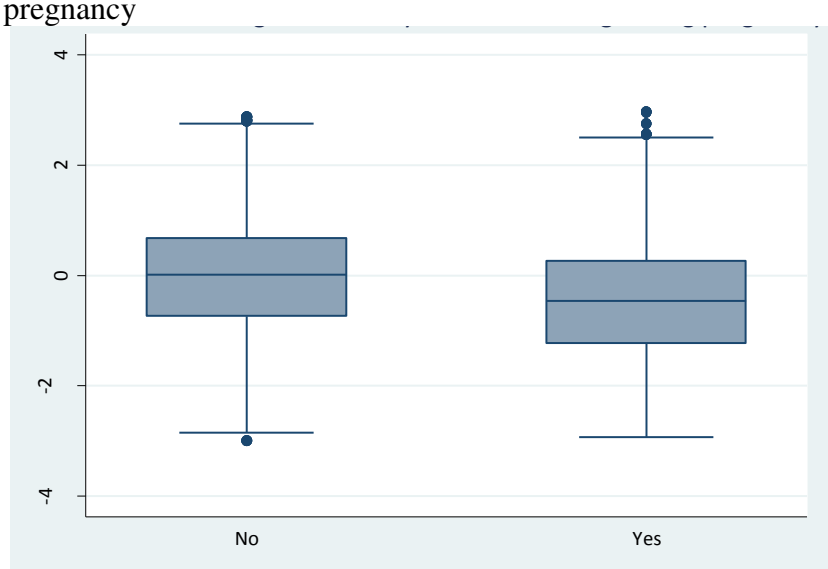
Figure 29: Mean birth weight z-score by Level of Relative Isolation



(e) Smoking during pregnancy

For both cohorts, the mean birth weight z-score was significantly lower for children whose mothers reported smoking during pregnancy ($t(740) = 4.79$, two-sided $p < 0.0001$ for Cohort B, and $t(499) = 4.86$, two-sided $p < 0.0001$ for Cohort K for two-sample t-tests with equal variances) compared to children whose mothers reported not smoking during pregnancy (see Figure 30).

Figure 30: Mean birth weight z-score by mother's report of smoking cigarettes during pregnancy



(f) BMI-for-age z-score

For the older cohort, there was a significant difference in BMI-for-age z-score of children with recorded versus missing birth weight z-scores at each wave (see Figure 31). The association was significant for the younger cohort at all waves except Wave 1 (see Figure 32; see Appendix L for more detail). The lower BMI-for-age z-score observed for children without a recorded birth weight z-score might be partially attributed to the higher prevalence of missing birth weight z-scores in areas with higher LORI, as the mean BMI-for-age z-scores are also lower in these areas.

For the younger cohort, the difference in BMI-for-age z-scores for those with and without birth weight z-scores recorded is attenuated after stratifying by urban (No or Low LORI) or remote (Moderate or High/Extreme LORI) area of residence, with a significant difference only observed at one wave (see Appendix L for more detail). However, for the older cohort, the difference in BMI-for-age z-scores between groups persists after stratifying by urban versus remote environment, with a significant difference observed at most waves. This suggests that, for the older cohort, there is a bias in the BMI-for-age z-scores of the two groups, independent of the association with LORI. For the younger cohort, much of the difference between the two groups can be attributed to LORI. For the older cohort in particular, these findings demonstrate that the distribution of BMI-for-age z-scores for children with birth weight z-scores recorded is not representative of the distribution of BMI-for-age z-scores for the whole sample.

Figure 31: Mean BMI-for-age z-score, by wave, for Cohort B children with missing versus not missing birth weight z-scores

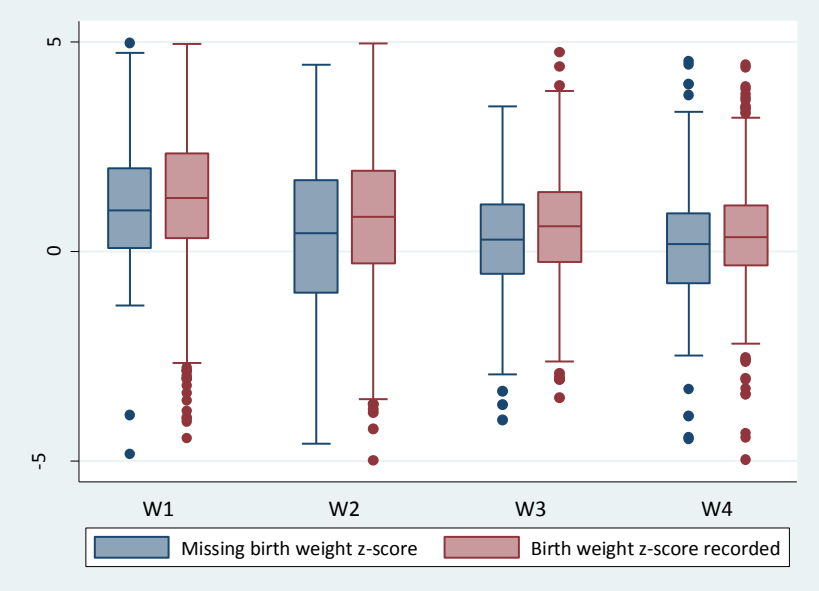
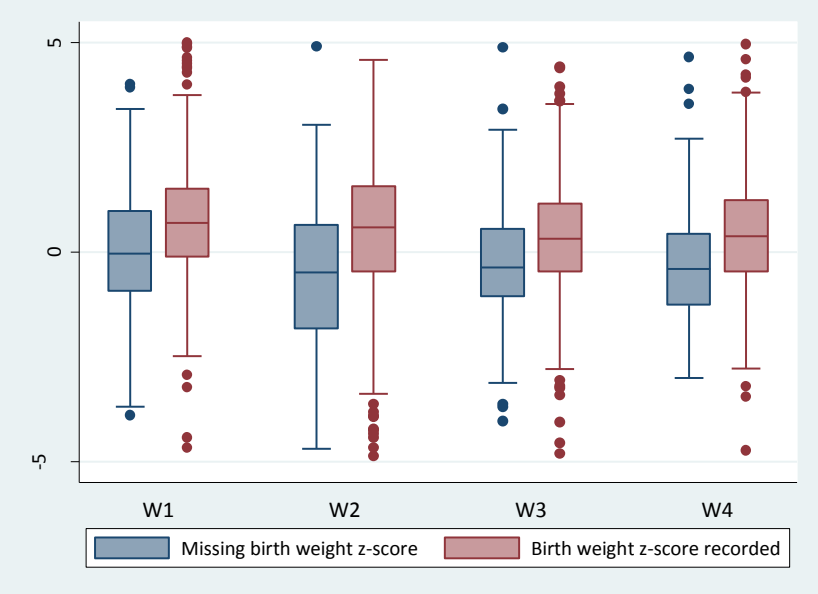


Figure 32: Mean BMI-for-age z-score, by wave, for Cohort K children with missing versus not missing birth weight z-scores



C) Conclusion

Overall, the majority of children in the LSIC cohorts fall into the healthy range for height, weight, and BMI. The distribution of each measure is heterogeneous, with children falling at both extreme ends of each distribution. The prevalence of overweight and obesity was higher at each wave when using BMI-for-age, rather than weight-for-age, as an indicator of weight status. BMI-for-age z-scores will be used in further analyses given that the indicator includes information about weight, height, and age, and given that the indicator is highly correlated with body fat percentage (121), and therefore with the risk of developing chronic disease (16).

The distribution of birth weight z-scores in LSIC was shifted to the left, with a higher than expected proportion of infants born small for their gestational age. The prevalence of infants born large for their gestational age, in contrast, was 10%, the expected prevalence in a population given the definition of LGA. This is consistent with the literature describing an elevated prevalence of SGA, but not LGA, in the Indigenous population (89).

Neither BMI-for-age z-score nor birth weight z-score demonstrated a significant association with gender or with Indigenous identity. The lack of significant association with gender is unsurprising given that z-scores are adjusted for gender. Research has suggested that body size differs across Indigenous identities (63), but the lack of significance observed in this sample might be attributable to the heterogeneity observed

within each group. LORI was not strongly associated with birth weight z-score, but it was significantly associated with BMI-for-age z-score, with higher z-scores recorded in areas with lower levels of relative isolation (more urban areas). This might suggest that the impact of LORI on prenatal growth differs from the impact of LORI on childhood growth. Similarly, maternal report of smoking during pregnancy demonstrated a disparate association with birth weight and with BMI in childhood. Children whose mothers reported smoking during pregnancy had significantly lower birth weight z-scores but higher BMI-for-age z-scores than children whose mothers reported not smoking during pregnancy. This is consistent with the literature on the effects of smoking on birth weight and childhood weight (158-162). These associations should be considered in the exploration of the association between birth weight and BMI in childhood.

The difference in BMI-for-age z-scores observed for children with and without birth weight z-scores recorded needs to be considered in analyses of the association between birth weight and BMI. These analyses include only the subsample of children with both birth weight and BMI-for-age z-scores, and thus biases in the distribution of BMI between those with and without birth weight z-scores will influence the results. For the younger cohort, the difference in BMI for the two groups loses significance once stratifying by LORI; thus, the impact of the bias is minimal in this cohort. For the older cohort, however, the difference remains significant after stratifying by LORI, with lower BMI-for-age z-scores among those who do not have birth weight z-scores recorded. Thus, the representativeness of findings about the association between birth weight and BMI within the older cohort is compromised; this represents a limitation of the LSIC data. This concern could be reconciled by repeating the question about birth weight at the next wave of the study.

Chapter VI: Does birth weight predict BMI in LSIC?

A) Multilevel mixed-effects modelling

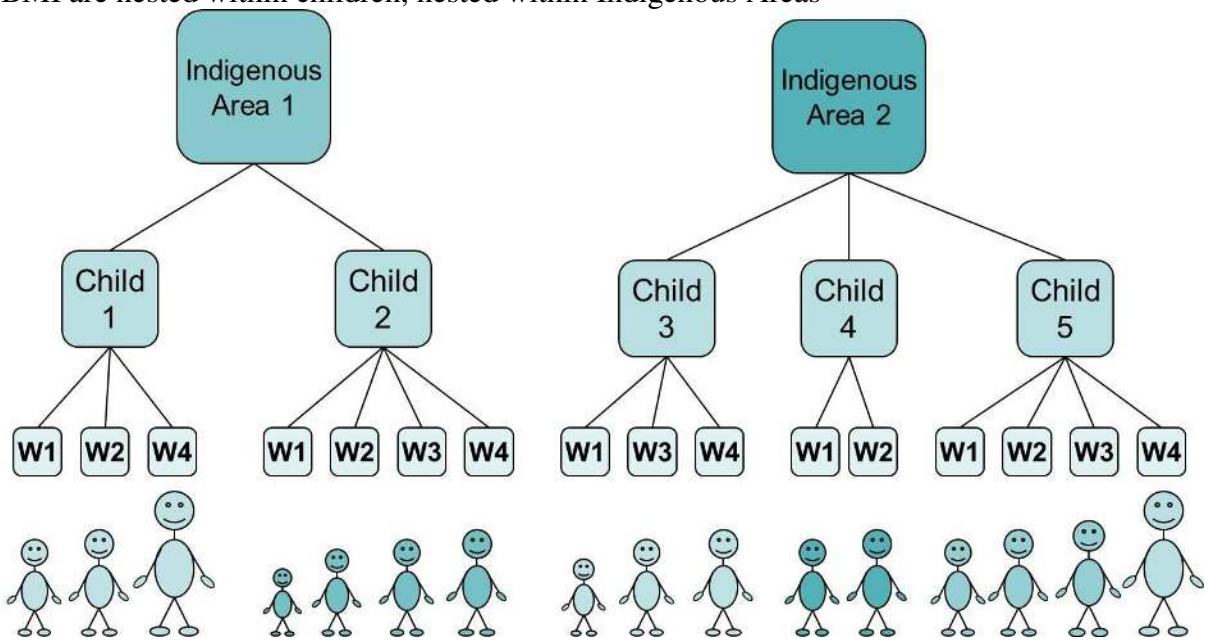
In order to explore the association between birth weight and BMI in LSIC, multilevel mixed-effects modelling was used. This method (also known as hierarchical, random-coefficient, repeated-measures, or growth-curve modelling) is well suited for the analysis of longitudinal data, as it can account for the inherent correlation of repeated measurements of the same individual (163-165). The hierarchical (multilevel) nature of the method enables the modelling of correlation at multiple levels, such as within-subject and within-cluster (163, 164). These correlations are modelled using the covariance structure, which can be defined to best fit the specific model. An adapted regression method, this approach can explicitly model, and distinguish between, fixed and random effects (165). Fixed effects are estimated directly, as in a standard regression. Random effects are described by their estimated variances and covariances (they are not estimated directly).

Another strength of the method is that it does not require balance: individuals are not required to all have the same number of measurements recorded. This is in contrast to methods such as ANOVA, in which individuals missing any data are excluded; in multilevel mixed-effects modelling, all available data points are included. Additionally, the time between measurements can vary across individuals in multilevel mixed-effects analysis. These features of the method are important for analyses of longitudinal studies, as attrition of participants across waves is virtually unavoidable, and maintaining a consistent follow-up time across individuals can be difficult. The method is especially suitable for analyses of LSIC, a longitudinal study which has a hierarchical correlation structure, a high prevalence of missing data, and wide variation in follow-up time across participants.

In the current study, BMI-for-age z-score, calculated at up to four time points, is the outcome variable, and birth weight z-score is an explanatory variable (a fixed effect). The repeated measurements of BMI-for-age z-score represent the first level of the model (see Figure 33). The child, represented by a random identification number, was included as a random effect, denoting the second level of the model, within which repeated measurements of BMI are nested. The Indigenous Area (the randomised code

for the child's location of residence) was included as a third-level random effect, within which children are nested. This set-up is designed to model the correlation structure inherent to the LSIC study design: BMI measurements on the same child are likely to have a stronger correlation than BMI measurements between different children, and similarly, BMI measurements of children living within the same Indigenous Area are likely to have a stronger correlation than BMI measurements of children living in different Indigenous Areas.

Figure 33: Structure of the multilevel mixed-effects model: repeated measurements of BMI are nested within children, nested within Indigenous Areas



The likelihood-ratio test can be used to compare the fit of two competing nested models by comparing their respective log-likelihood values. A model with an additional parameter will always have an equivalent if not better fit, and therefore will have an equivalent or lower log-likelihood; the likelihood-ratio test can be used to determine if the enhancement of fit is significant. The likelihood-ratio statistic has a chi-squared distribution, with degrees of freedom equal to the difference in the number of free parameters between the null and alternative models (equal to the number of additional variables included in the alternative model). This test was used to first evaluate the proposed hierarchical structure of the model, and then to measure the effect of the addition of variables to the model.

(1) Development of the structure of the model

BMI-for-age z-score, representing the distance (in standard deviations) from the median BMI of children of the same gender and age, was recorded for 996 children at Wave 1, 1,207 children at Wave 2, 1,149 children at Wave 3, and 1,170 children at Wave 4, for a total of 4,522 observations (see Table 27). Birth weight z-scores, representing the distance (in standard deviations) from the median birth weight of Australian infants of the same gender and gestational age, were recorded for 1,304 children. Only BMI-for-age z-scores for children with recorded birth weight z-scores can be used for analyses; this leaves a total of 3,391 measurements of BMI-for-age z-score. An initial single-level model was fit via maximum likelihood, with BMI-for-age z-score as the outcome and birth weight z-score as the exposure variable (essentially a simple linear regression). Within this single-level model, there was a significant positive association between birth weight z-score and BMI-for-age z-score; each one-unit increase in birth weight z-score was associated with a 0.163-unit increase in BMI-for-age z-score ($p < 0.001$).

Table 27: Number of recorded and missing data points, by wave

	Wave 1		**Wave 2**		**Wave 3**		**Wave 4**	
	# recorded	# missing	# recorded	# missing	# recorded	# missing	# recorded	# missing
BMI z-score	996	763	1,207	552	1,149	610	1,170	589
Birth weight z-score*	1,304	455	--	--	--	--	--	--
Indigenous Area	1,671	88	1,523	236	1,404	355	(1,404)~	355
Age at BMI measurement	1,237	522	1,374	385	1,281	478	1,238	521
Gender*	1,759	0	--	--	--	--	--	--
Ethnicity*	1,759	0	--	--	--	--	--	--
Cohort*	1,759	0	--	--	--	--	--	--
LORI	1,671	88	1,523	236	1,404	355	(1,404)~	355
Reported family weekly income	1,563	196	1,381	378	(1,381)^	378	(1,381)^	378
Primary carer currently smokes	1,669	90	1,522	237	1,402	357	(1,402)~	357
Birth mother smoked during pregnancy	1,554	205	--	--	--	--	--	--

	Wave 1		Wave 2		Wave 3		Wave 4	
Age study child stopped breastfeeding (days)*	1,613	146	--	--	--	--	--	--
Age study child first consumed solid food (days)*	928	831	--	--	--	--	--	--

* Variable was only collected at the first wave of the study.
~ Wave 4 data (except height and weight) has not yet been released, so this information is unavailable; value at Wave 3 is imputed for Wave 4.
^ Variable was collected at Wave 1 and Wave 2; value at Wave 2 is imputed for Waves 3 and 4.

The inclusion of the children's identification number as a random-effects parameter was statistically significant. The significant positive association between birth weight z-score and BMI-for-age z-score persisted ($\beta = 0.151$, $p < 0.001$) in this two-level model. The likelihood-ratio test comparing this two-level model to the initial single-level model demonstrated that the two-level model was a significantly better fit to the data (LR chi2(1) = 443.13, $p < 0.0001$). When the randomised code for Indigenous Area was included as the level variable, the improvement of the model's fit was also significant, and a similar association was observed between birth weight z-score and BMI-for-age z-score ($\beta = 0.168$, $p < 0.001$). A likelihood-ratio test could not be conducted comparing this model to the initial single-level model because the inclusion of Indigenous Area reduced the sample size (from n = 3,391 to n = 3,303). Next, a three-level model was tested, with identification number nested within Indigenous Area (both as random effects). A likelihood-ratio test confirmed that this three-level model demonstrated an enhanced fit compared to the two-level model including Indigenous Area as a level variable (LR chi2(1) = 403.78, $p < 0.0001$; n = 3,303). Again, the association between birth weight z-score and BMI-for-age z-score remained significant ($\beta = 0.141$, $p < 0.001$). The identity covariance structure was used (assuming independent residuals with constant variance) in each model to maintain parsimony.

(2) Addition of a priori explanatory variables to the model

Next, the *a priori* explanatory variables were individually added into the three-level model (repeated measurements of BMI nested within individuals, nested within Indigenous Areas). Variables for inclusion in the *a priori* model were birth weight z-

score, age, cohort, Indigenous identity (Aboriginal, Torres Strait Islander, or Aboriginal and Torres Strait Islander), gender, and Level of Relative Isolation. BMI-for-age z-scores are inherently adjusted for age and gender, to give a z-score that represents variation from average growth. Age (in months) was included as a covariate because the sample showed age-related trends in BMI-for-age z-score. Regressing age against BMI z-score demonstrated a significant linear relationship between the variables for each cohort ($\beta = -0.023$, $p < 0.001$, $r^2 = 0.0404$ for the Baby Cohort, and $\beta = -0.008$, $p = 0.002$, $r^2 = 0.0050$ for the Child Cohort) and thus age was included as a continuous variable. Cohort (Baby or Child) was included in the model, as each cohort represents a unique sample, and differences have been observed between groups. LORI was included in the model given its strong association with BMI. The child's self-reported identity (Aboriginal, Torres Strait Islander, or Aboriginal and Torres Strait Islander) is included in the model because, although differences in BMI z-scores across groups were insignificant, homogeneity across identities should not be assumed. Likewise, gender will be included in the *a priori* model. The inclusion of these variables did not reduce the sample of measurements available for inclusion in the model (n = 3,303).

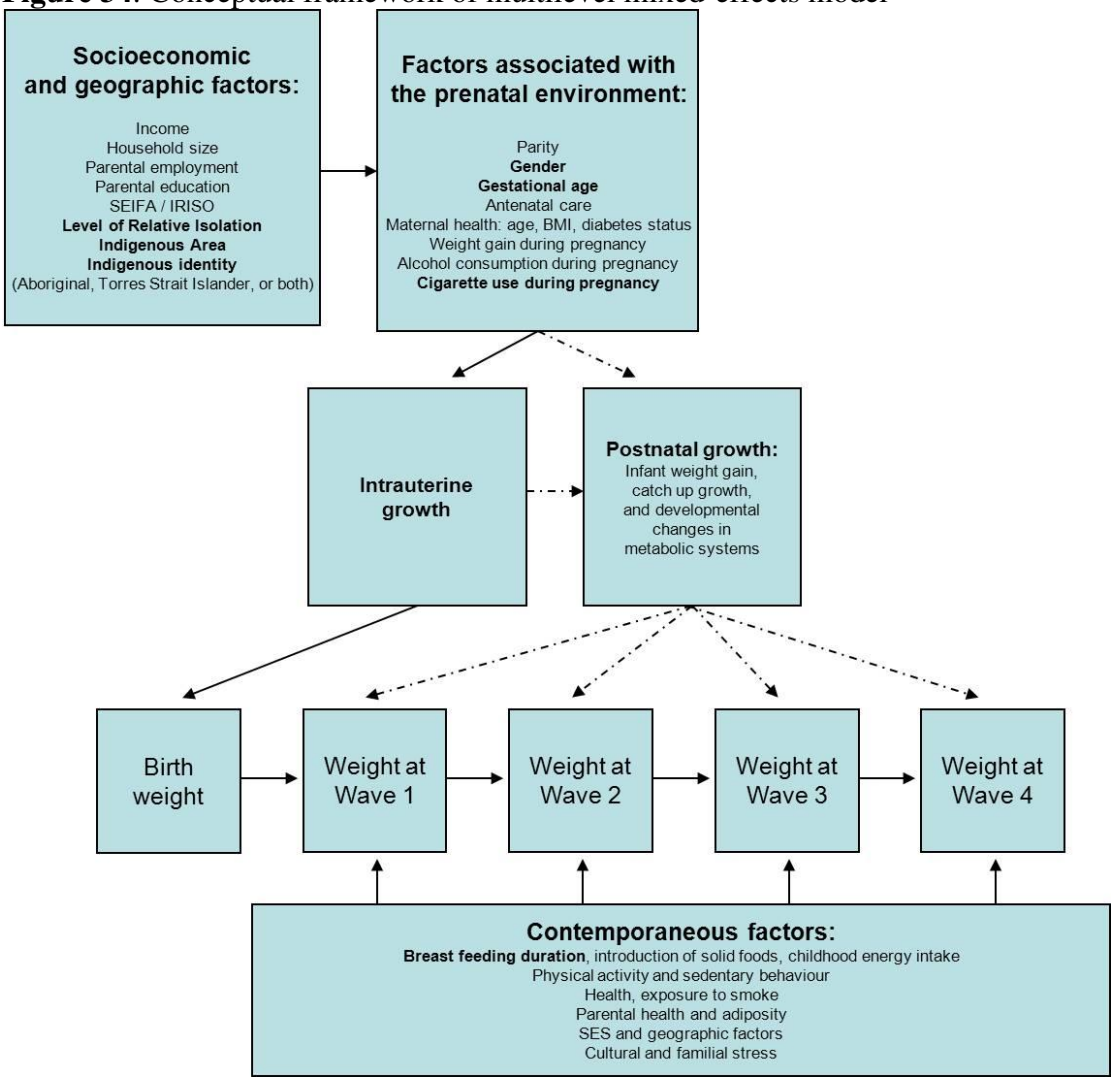
The inclusion of these variables significantly improved the fit of the model (LR chi2(8) = 134.33, $p < 0.0001$ compared to initial model including only birth weight z-score, identification number, and Indigenous Area). In the *a priori* model, there was a significant positive association between BMI-for-age z-score and birth weight z-score ($\beta = 0.133$, $p < 0.001$). There was a significant negative association between BMI-for-age z-score and age in months ($\beta = -0.017$, $p < 0.001$); however, membership in the older cohort was associated with a significantly higher BMI-for-age z-score ($\beta = 0.370$, $p < 0.001$) than membership in the younger cohort. Inhabiting an area with a Low ($\beta = -0.245$, $p = 0.007$), Moderate ($\beta = -0.647$, $p < 0.001$), or High/Extreme LORI ($\beta = -0.466$, $p = 0.005$) was associated with a significantly lower BMI z-score than living in an area with No LORI. There was not a significant association between BMI-for-age z-score and Indigenous identity or gender.

(3) Addition of confounding variables to the model

Additional variables were chosen for inclusion in the model, based on the conceptual framework of the model (see Figure 34). Variables were selected to represent a subset of the wide range of factors associated with the prenatal environment, postnatal health, and childhood growth. The variables considered for inclusion were:

reported weekly family income, primary carer's concerns about money, household size, exposure to cigarette smoke in the household, primary carer's current cigarette smoking status, birth mother's cigarette use during pregnancy, birth mother's alcohol consumption during pregnancy, birth mother's weight gain during pregnancy, birth mother's diagnosis of diabetes (including gestational diabetes), primary carer's highest educational qualification obtained, study child of a single or multiple birth, term of study child's birth (pre-term, term, or post-term), study child ever breast fed, study child's age at the cessation of breast feeding, study child's age at the first consumption of solid foods, and report of study child's general health.

Figure 34: Conceptual framework of multilevel mixed-effects model



*Bolded variables are included in the final model. Framework based on images from (100, 166-170).

Each variable was added to the *a priori* model individually, and the associated improvement to the model was assessed using the likelihood-ratio test. The model was

significantly improved through the independent addition of weekly income, primary carer's current smoking status, maternal smoking during pregnancy, study child's age at the cessation of breast feeding, and study child's age at the first consumption of solid foods (see Table 28). From this group of five variables, inclusion in the final model was determined on the basis of minimisation of multicollinearity and maximisation of sample size.

Table 28: Addition of explanatory variables (one at a time) to *a priori* model

Variable added to *a priori* model	**Log likelihood**	**n**	**β**	**P-value**	**L-R test vs. a priori model** (df)	**Significant improvement to model?**
(*a priori* model)	-5,791	3,303	--	--	--	--
Reported family weekly income (*vs. less than $150 per week*)	-5,228	2,964*	*$150-249:* -0.372	0.026	13.4 p = 0.037 (6)	Yes
			$250-399: -0.018	0.906		
			$400-599: -0.103	0.483		
			$600-799: -0.085	0.575		
			$800-999: -0.078	0.620		
			$1000 +: -0.296	0.061		
P1 worries about money (*vs. no*)	-5,761	3,286*	*Yes:* 0.019	0.740	0.11 p = 0.741 (1)	No
Household size	-5,791	3,303	-0.012	0.477	0.50 p = 0.4787 (1)	No
Anyone smokes in the house (*vs. no*)	-5,585	3,183*	*Yes:* 0.106	0.187	1.71 p = 0.191 (1)	No
P1 currently smokes (*vs. no*)	-5,784	3,301*	*Yes:* 0.162	0.011	6.47 p = 0.011 (1)	Yes
Birth mother smoked during pregnancy (*vs. no*)	-5,524	3,156*	*Yes:* 0.199	0.005	7.74 p = 0.005 (1)	Yes
Consumed alcohol during pregnancy (*vs. no*)	-5,513	3,148*	*Yes:* -0.138	0.095	2.78 p = 0.095 (1)	No
Weight gain during pregnancy (*vs. too much*)	-4,808	2,758*	*Okay:* -0.089	0.469	1.14 p = 0.566 (1)	No
			Not enough: 0.044	0.823		

Variable added to *a priori* model	**Log likelihood**	**n**	**β**	**P-value**	**L-R test vs. a priori model** (df)	**Significant improvement to model?**
Any diabetes diagnosis (*vs. no*)	-5,791	3,303	0.044	0.734	0.12 p = 0.734 (1)	No
P1 highest qualification obtained (*vs. never attended school*)	-5,372	3,078*	*< Year 8:* -0.519	0.585	18.08 p = 0.154 (13)	No
			Year 9: -0.322	0.731		
			Year 10: -0.338	0.717		
			Year 11: -0.442	0.636		
			Year 12: -0.257	0.783		
			Certificate of completion: 0.134	0.903		
			Other non-school qualification: -0.037	0.972		
			Certificate I/II: -0.041	0.965		
			Certificate III/IV: -0.536	0.568		
			Advanced diploma: -0.326	0.731		
			Bachelor degree: 0.047	0.960		
			Graduate diploma: -0.424	0.665		
			Post graduate degree: 0.097	0.923		
Single birth *(vs. multiple birth)*	-5,790	3,303	*Single birth:* 0.266	0.257	1.28 p = 0.257 (1)	No
Term (*vs. pre-term*)	-5,791	3,303	*Term:* -0.013	0.910	0.83 p = 0.659 (1)	No
			Post-term: 0.097	0.536		
Study child ever breast fed (*vs. no*)	-5,763	3,290*	*Yes:* 0.016	0.856	0.03 p = 0.857 (1)	No

Variable added to *a priori* model	**Log likelihood**	**n**	**β**	**P-value**	**L-R test vs. a priori model** (df)	**Significant improvement to model?**
Age study child stopped breast feeding (days)	-5,505	3,165*	-0.000	0.006	7.26 p = 0.007 (1)	Yes
Age study child first consumed solid food (days)	-3,300	1,844*	0.001	0.024	5.08 p = 0.024 (1)	Yes
Study child's general health (*vs. excellent*)	-5,783	3,299*	*Very good:* -0.024	0.684	5.55 p = 0.235 (4)	No
			Good: -0.084	0.220		
			Fair: -0.257	0.127		
			Poor: 0.489	0.197		

* Likelihood-ratio test was conducted against the *a priori* model only including the subset of the sample without missing data for the exposure variable of interest.

To allow for an accurate interpretation of predictors, a group of explanatory variables was chosen such that pairwise correlations were minimised. The pairwise correlation between the birth mother's report of smoking during the study child's pregnancy and the primary carer's report of smoking status at the first wave of the study was significant ($r = 0.731$, $p < 0.0001$), as might be expected. There were fewer missing data for smoking during pregnancy (versus primary carer's current smoking status) (see Table 29), so this variable was chosen for inclusion in the final model and primary carer's current smoking status was dropped.

There was a significant correlation between the age of breast feeding cessation and the age of introduction of solid foods ($r = 0.194$, $p < 0.0001$); the age of breast feeding cessation was chosen for inclusion over the age of introduction of solid foods because the former variable contained fewer missing values. There was also a significant correlation between maternal smoking during pregnancy and the age of breast feeding cessation ($r = -0.058$, $p < 0.0001$), with mothers who did not smoke during pregnancy likely to breast feed their child for a longer time period. The added influence of the age of breast feeding cessation after including maternal smoking during pregnancy in the model might therefore be limited. However, research suggests that the direction of the association with childhood weight status is opposite for the two variables (158-161), so the inclusion of both age of breast feeding cessation and maternal smoking during pregnancy is important.

There was a significant correlation between the primary carer's report of the family's weekly income and maternal smoking during pregnancy ($r = -0.206$, $p < 0.0001$). Given the high prevalence (19%) of missing data for weekly income, together with its strong correlation with smoking during pregnancy (as well as LORI and Indigenous identity) reported weekly income was dropped from the model. The final model, with a total of 3,019 observations, included the following variables: birth weight z-score, age, cohort, Indigenous identity, gender, LORI, maternal smoking during pregnancy, and age of breastfeeding cessation (see Table 29). The included variables act at the level of the individual (birth weight z-score, age, cohort, Indigenous identity, age of breastfeeding cessation), the family (smoking during pregnancy), and the neighbourhood (LORI), and represent a range of exposures associated with childhood growth. The concurrent addition of age of breastfeeding cessation and maternal smoking during pregnancy to the *a priori* model significantly improved the fit (LR chi2(3) = 14.18, $p = 0.0008$).

Table 29: Sample size for model including *a priori* variables and confounding variables

Variables included in model, in addition to *a priori* variables	**Number of BMI-for-age z-score measurements included in model**
--	3,303
Age study child stopped breast feeding (days)	3,165
Birth mother smoked during pregnancy	3,156
Age study child stopped breast feeding (days) and birth mother smoked during pregnancy	3,019

(4) Results

The independent addition of maternal smoking during pregnancy to the *a priori* model did not change the significance or the direction of association of any variables, but slightly increased the coefficient for birth weight z-score (from $\beta = 0.133$ to $\beta = 0.159$) and slightly altered the coefficient for each LORI (from $\beta = -0.245$ to $\beta = -0.238$ for Low LORI, from $\beta = -0.647$ to $\beta = -0.606$ for Moderate LORI, and from $\beta = -0.466$ to $\beta = -0.504$ for High LORI, compared to no LORI). There was a significant association between maternal smoking during pregnancy and children's BMI status,

with a higher BMI observed for children whose mothers reported smoking during pregnancy (β = 0.199, p = 0.005).

There was no alteration to the significance or direction of any variables with the addition of age at breast feeding cessation to the *a priori* model. The coefficient for birth weight z-score increased, from β = 0.133 (p < 0.001) to β = 0.141 (p < 0.001). The magnitude of the coefficient for each LORI decreased (from β = -0.245 to β = -0.194 for Low LORI, from β = -0.647 to β = -0.595 for Moderate LORI, and from β = -0.466 to β = -0.356 for High LORI, compared to no LORI). Age of breast feeding cessation was significantly associated with BMI-for-age z-score (β = -0.0003, p = 0.006).

When both confounding variables were added into the final model, there was a significant association between BMI-for-age z-score and birth weight z-score, age at BMI measurement, cohort, LORI, study child's age of breast feeding cessation, and maternal smoking during pregnancy. The concurrent addition of these two variables did not alter the significance or direction of any associations from the *a priori* model, but increased the coefficient for birth weight z-score (from β = 0.133 to β = 0.166) and for each level of LORI, compared to No LORI (from β = -0.245 to β = -0.187 for Low LORI, from β = -0.647 to β = -0.593 for Moderate LORI, and from β = -0.466 to β = -0.398 for High LORI) (see Table 30).

Table 30: Coefficient, p-value, and 95% confidence interval for variables included in each model

	Random-effects model			***A priori* model**			**Final model**		
Fixed-effects variables	**β**	**p**	**95% CI**	**β**	**p**	**95% CI**	**β**	**p**	**95% CI**
Birth weight z-score	0.141	< 0.001	0.08, 0.20	0.133	< 0.001	0.07, 0.19	0.166	< 0.001	0.10, 0.23
Age (months)	--	--	--	-0.017	< 0.001	-0.02, -0.01	-0.017	< 0.001	-0.02, -0.01
Gender (female vs. male)	--	--	--	-0.056	0.416	-0.19, 0.08	-0.060	0.398	-0.20, 0.08
Indigenous Identity (vs. Aboriginal) *Torres Strait Islander*	--	--	--	0.064	0.683	-0.24, 0.37	0.072	0.660	-0.25, 0.39
Aboriginal & Torres Strait	--	--	--	0.041	0.795	-0.27, 0.35	0.174	0.280	-0.14, 0.49

	Random-effects model			***A priori* model**			**Final model**		
Islander									
Cohort (Child vs. Baby Cohort)	--	--	--	0.370	< 0.001	0.19, 0.55	0.391	< 0.001	0.21, 0.57
LORI (vs. No) *Low*	--	--	--	-0.245	0.007	-0.42, -0.07	-0.187	0.035	-0.36, -0.13
Moderate	--	--	--	-0.657	< 0.001	-0.92, -0.37	-0.593	< 0.001	-0.88, -0.31
High/ Extreme	--	--	--	-0.466	0.005	-0.79, -0.14	-0.398	0.022	-0.74, -0.06
Mother smoked during pregnancy (yes vs. no)	--	--	--	--	--	--	0.208	0.004	0.07, 0.35
Age at breast feeding cessation (days)	--	--	--	--	--	--	-0.000	0.024	-0.00, -0.00

In the final model, a one-unit increase in birth weight z-score was associated with a 0.166-unit increase in BMI-for-age z-score ($p < 0.001$). Maternal smoking during pregnancy was associated with a 0.208-unit increase ($p = 0.004$), and an additional month of breast feeding was associated with a 0.009-unit decrease ($p = 0.024$), in BMI-for-age z-score. A one-month increase in age was associated with a 0.017-point decrease in BMI-for-age z-score ($p < 0.001$), but membership in the older cohort was associated with a 0.391-unit increase ($p < 0.001$) compared to membership in the younger cohort. Compared to living in an area with no level of relative isolation, living in an area with a Low LORI was associated with a 0.187-unit decrease in BMI-for-age z-score ($p = 0.035$), living in an area with a Moderate LORI was associated with a 0.593-unit decrease in BMI-for-age z-score ($p < 0.001$), and living in an area with a High/Extreme LORI was associated with a 0.398-point decrease in BMI-for-age z-score ($p = 0.022$).

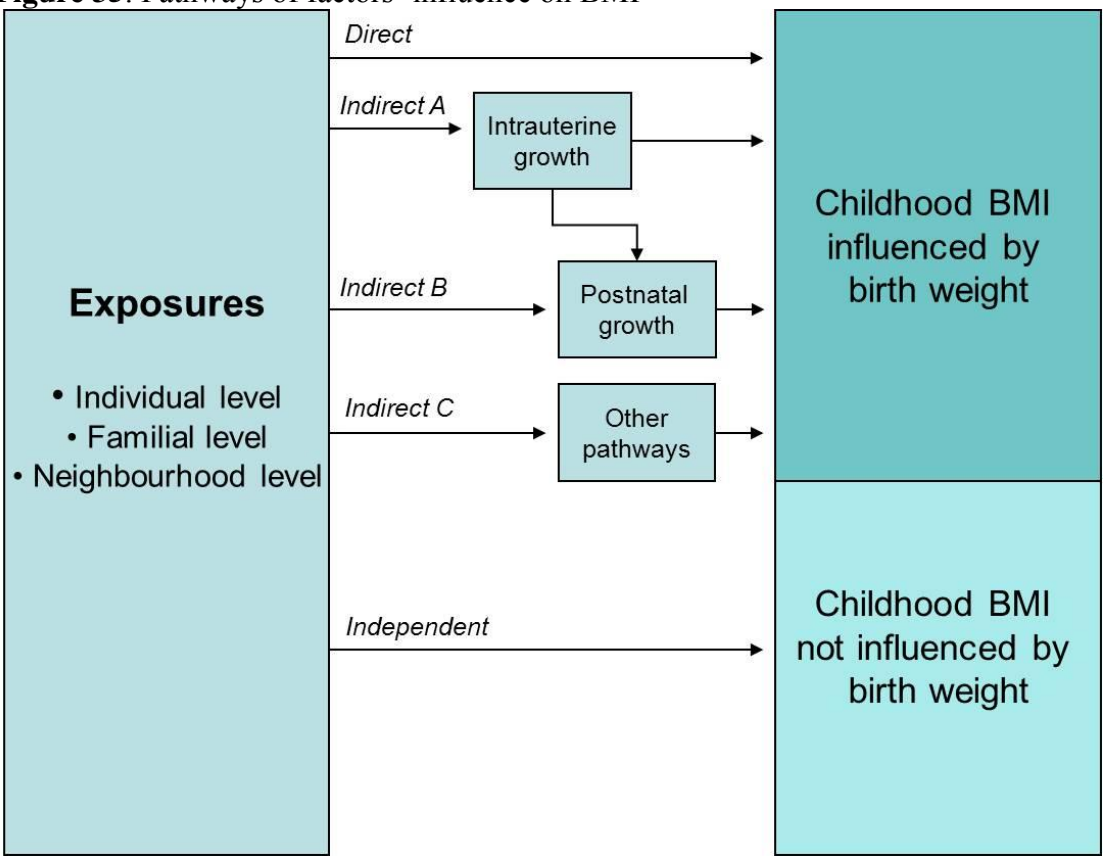
(5) Discussion

The multilevel mixed-effects modelling approach enables exploration of the association between birth weight and BMI in a longitudinal study with missing data and

clustered sampling. The inclusion of identification number and Indigenous Area as random effects accounts for the correlation between repeated measurements of BMI-for-age z-score on the same individual, and for the correlation between measurements of children living in the same Indigenous Area. The addition of age, cohort, Indigenous identity, gender, LORI, maternal smoking during pregnancy, and age of breastfeeding cessation to the model as explanatory variables enriches the interpretation of results. These variables represent a subset of the factors influencing birth weight and BMI at the individual, familial, and neighbourhood levels.

These factors act at various points on the pathway from birth weight to childhood BMI (see Figure 35). Factors influencing birth weight can impact BMI both directly and indirectly; additionally, factors can influence BMI independently of birth weight. For example, age at the measurement of BMI and age at the cessation of breastfeeding would be expected to influence only BMI, and not birth weight, as these factors take place after birth. Indigenous identity and gender might have an impact on birth weight and/or on BMI in childhood; however, neither association proved significant in this model. Each cohort represents a unique sample, and thus differences in both birth weight and BMI might be expected; in addition, the age difference of the cohorts may influence BMI. Maternal smoking during pregnancy has been shown to impact birth weight (45), which can influence later BMI. Research has also demonstrated an independent effect of smoking during pregnancy on BMI in childhood (160); this finding was replicated in the present study. LORI can exert an influence both on birth weight (influencing the child's prenatal environment through its association with maternal health and well-being,) and on childhood BMI (influencing the postnatal environment through its association with the health and well-being of the child).

Figure 35: Pathways of factors' influence on BMI



*Bolded variables are included in the final model. Diagram based on images from (169, 171, 172).

(a) Gender

In this study, gender was not significantly associated with birth weight or BMI-for-age z-score. This is largely attributable to the adjustment for gender in the calculation of z-scores. In the final model, being female was associated with a 0.06-unit decrease in BMI-for-age z-score compared to being male, but this difference was not significant.

(b) Age and cohort

A negative association was observed between the age (in months) at the time of height and weight measurement and the BMI-for-age z-score for both cohorts, and this persisted in the final model. Paradoxically, membership in the older cohort was associated with a 0.39-point increase in BMI-for-age z-score in the final model. The age gap between the two cohorts is approximately three years, or 36 months, so when measurements of the older cohort at one time point are compared to measurements of the younger cohort three years later, the mean BMI-for-age z-score of the older cohort would be expected to exceed that of the younger cohort by 0.22 (accounting for the

coefficient associated with cohort as well as the coefficient associated with age). For example, this gap is observed when comparing Cohort K at Wave 1 to Cohort B at Wave 2, with mean BMI-for-age z-scores of 0.57 and 0.30, respectively. This might reflect a secular trend (with decreasing BMI of children in recent years) or, more likely, a cohort effect.

(c) Indigenous identity

Identifying as Torres Strait Islander or as Aboriginal and Torres Strait Islander, compared to identifying as Aboriginal, was associated with an increased in BMI-for-age z-score ($\beta = 0.174$, $p = 0.280$ for Aboriginal and Torres Strait Islander children, and $\beta = 0.072$, $p = 0.660$ for Torres Strait Islander children); however, these differences were not significant.

(d) LORI

LORI appears to exert a disparate influence on birth weight z-scores and BMI-for-age z-scores. There was not a consistent trend in the association between birth weight z-score and LORI. Overall, the mean birth weight z-score was lowest for children living in areas with a Low LORI (-0.311), followed by Moderate LORI (-0.260) and No LORI (-0.110), and was highest in areas with High/Extreme LORI (-0.039). The only significant difference between groups was between areas with Low and No LORI, with lower birth weight z-scores observed in areas with Low LORI. There was a more consistent trend in the association between LORI and BMI-for-age z-score, with a significantly lower mean BMI-for-age z-score for children living in Low, Moderate, and High/Extreme LORI compared to children living in an area with No LORI. The largest decrease in BMI-for-age z-score was associated with living in an area with a Moderate, compared to No, LORI ($\beta = -0.593$, $p < 0.001$), followed High/Extreme LORI ($\beta = -0.398$, $p = 0.022$) and Low LORI ($\beta = -0.187$, $p = 0.035$), though only the difference between Low and No LORI was statistically significant. Thus, the impact of LORI on birth weight is not the same as the impact of LORI on BMI; given these associations, it is not surprising that birth weight and LORI both remain significant predictors of BMI in the final model. Further research is needed to untangle this interaction.

(e) Age of breast feeding cessation

The model depicts a negative association between breast feeding duration and BMI; each additional month of breast feeding was associated with a 0.01-unit decrease in BMI-for-age z-score in childhood. This finding is consistent with the literature which describes a slight protective effect of breast feeding on overweight in childhood (173). The addition of breast feeding duration to the model alters the coefficients associated with LORI; this is intuitive given the significantly longer mean duration of breast feeding reported in remote, compared to urban, areas. The inclusion of both variables helps disentangle the impact of LORI on BMI-for-age z-score from the impact attributable to its association with breastfeeding duration.

(f) Birth mother's cigarette use during pregnancy

The addition of maternal cigarette use to the model strengthens the association between birth weight z-score and BMI-for-age z-score. This is consistent with the literature, which describes a negative association between smoking during pregnancy and birth weight (162), but a positive association between smoking during pregnancy and childhood weight (158-161). The mechanism underlying the latter association is unknown, and may be largely attributable to the long-lasting effects of intrauterine growth restriction (158). However, one study found that the association between maternal smoking and overweight in childhood was independent of intrauterine growth restriction, suggesting that there may also be a direct impact of prenatal smoke exposure to on the development of overweight and obesity (160). This interaction can be further explored as additional waves of data are collected, increasing the sample size and the range of ages available.

(6) Limitations

As described previously, the presence of missing or implausible BMI-for-age z-scores is associated with several demographic factors (LORI, Indigenous identity, and cohort); similarly associations exist with the presence of missing or implausible birth weight z-scores and other factors (LORI, whether the primary carer was the birth mother, and whether the information was provided from the baby book or by recall). These associations demonstrate that the birth weight and BMI measurements in the study do not fully represent the entire sample. Further, there is a significant difference in

the BMI-for-age z-scores of children who have birth weight z-scores recorded compared to those who do not. BMI-for-age z-scores were significantly lower for children without birth weight z-scores recorded, and therefore for the children who are not included in analyses of birth weight as a predictor BMI. As a result, examinations of the associations between birth weight and BMI in LSIC are based on a sample of LSIC children that do not fully represent the BMI distribution in the sample.

Additionally, bias is induced when additional variables are added to the model and the sample size is further reduced. The final model included only children with recorded birth weight, gestational age, height, weight, age at height measurement, age at weight measurement, LORI, maternal cigarette use during pregnancy, and age of breastfeeding cessation. The initial model including only BMI-for-age and birth weight z-scores included a total of 3,391 observations; the sample was reduced to 3,303 after the addition of individual identification number and Indigenous Area as level variables and age, gender, cohort, and Indigenous identity as explanatory variables. When maternal smoking during pregnancy and age of breast feeding cessation were added to the final model, the sample size was reduced to 3,019. Thus, 284 observations of BMI-for-age z-scores from the *a priori* model were not included in the final model. Two-sample t-tests with equal variance were used to explore differences between the sample (n = 3,303) from the *a priori* model included in the final model (n = 3,019) versus those not included in the final model (n = 284). There was not a significant difference in the mean BMI-for-age z-score for the observations included versus not included in the final model ($t(3,301) = -1.174$, two-sided $p = 0.2403$) but the mean birth weight z-score was significantly lower for the 284 observations not included in the final model ($t(3,301) = -2.8217$, two-sided $p = 0.0040$). This must be taken into consideration in the interpretation of the results of the model.

(7) Conclusion

A significant association between birth weight z-score and BMI-for-age z-score persists after adjusting for the correlation between repeated measurements and between individuals living in the same Indigenous Area, and for variables including age, cohort, Indigenous identity, gender, LORI, maternal smoking during pregnancy, and age of breastfeeding cessation. Accounting for these factors, a one-unit increase in birth weight z-score is associated with a 0.166-unit increase in BMI-for-age z-score ($p < 0.001$). A child born LGA with a z-score of +1.28 (90^{th} percentile) would be predicted to have a

BMI-for-age z-score 0.213 units higher than a child born AGA with a z-score of 0 (50th percentile), keeping other factors constant. Given that a one-unit increase in BMI-for-age z-score represents a shift from healthy weight to overweight (from z = 0 to z = +1), or overweight to obese (from z = +1 to z = +2) for a child older than five years of age, a 0.213-unit increase in BMI-for-age z-score attributable to birth weight z-score alone is significant.

Chapter VII: Conclusion

The data contained in LSIC are a valuable resource, representing the first large-scale study of the longitudinal development of Australian Indigenous children (174). The dataset is not intended to be representative of all Indigenous children across Australia, but it shares the story of nearly 2,000 children from diverse backgrounds. These data, through their incorporation into evidence-based policy, can be used to support the healthy growth of Aboriginal and Torres Strait Islander children across Australia. The birth weight, height, and weight data collected in LSIC are an unmatched resource, given the study's large sample size and geographical diversity; they have the potential to fill a wide gap in the literature. Birth weight and childhood size are both potentially entwined in the escalating rates of diabetes, cardiovascular disease, and renal disease observed within Indigenous Australian adults. Despite the wealth of research examining the links between birth weight, early childhood growth, and chronic disease risk in non-Indigenous populations, few studies have investigated these associations within an Indigenous sample. The dearth of research in the area has been attributed to 'methodologic problems'; this type of life-course analysis requires a longitudinal study, and few studies of Indigenous populations are able to maintain an adequate sample size across waves of the study (103).

The LSIC team has attempted to circumvent the limitations threatening longitudinal studies through the implementation of an accelerated cross-sequential longitudinal design and through the development of a trusting relationship with study participants. As a result of these efforts, participants have remained engaged in the study and the retention rate has exceeded 85% between successive waves. One unintended consequence of the emphasis placed on maintaining a positive rapport with families participating in the study has been compromised data integrity. A high prevalence of missing data, together with a high prevalence of implausible data (to some extent attributable to inadequate measuring equipment in the first wave of the study) has prevented FaHCSIA from releasing the anthropometric data for analysis. By integrating insight I gained through interviewing the LSIC RAOs into established WHO protocols, I developed a method to clean the data and assess their validity. The proportion of the data remaining after data cleaning is smaller than would be ideal, but few systematic biases are apparent, and the quality of data can be seen to improve across waves. When considering the validity of the data within the context of Indigenous research, the

elevated prevalence of missing and implausible data can be perceived as a limitation to the data, rather than preclusive to its use.

On the basis of this evaluation, FaHCSIA approved the cleaned data for public release on the 4th of December, 2012. Given this sanction, I analysed the distribution of birth weight, height, and weight in LSIC, and explored the association between birth weight and BMI throughout childhood. Using multilevel mixed effects modelling to account for the study's design, I found that birth weight, adjusted for gestational age, significantly predicts BMI in the LSIC sample. This association persisted after adjusting for the child's age, gender, Indigenous identity, cohort, and breastfeeding duration, the mother's cigarette use during pregnancy, and the area's level of relative isolation.

These findings suggest that children who are born small for their gestational age (with a negative z-score for birth weight) will remain smaller than their appropriate-for-gestational age counterparts through eight years of age. The level of adiposity or the pattern of fat distribution could not be examined in this study, but BMI can serve as an indicator of percentage body fat. The LSIC findings are consistent with those of Sayers and colleagues, who found that children born small for gestational age in the Aboriginal Birth Cohort study remained smaller than children born appropriate for gestational age through 18 years of age (38). There was no difference in the fat distribution between the two groups (39), contrary to what the DOHaD hypothesis would have predicted. The LSIC results lend support to the Aboriginal Birth Cohort Study findings, and reinforce the need for further examination of these associations. A 25-year follow-up of the Aboriginal Birth Cohort study is planned (38-40); this would provide the opportunity to investigate if the same trend is observed through age 25 years. Similarly, continued follow-up of the LSIC sample will enable examination of the association between birth weight and BMI across an expanding age range.

Measurement accuracy in future waves of the study could be improved if anthropometric data are collected in adherence to the recommendations set forth in this thesis. In designing a survey form, consideration must be given to the setting in which measurements will take place, and the equipment which will be used to take these measurements. Matching the survey form to the specific survey setting will facilitate the recording of measurements and decrease the occurrence of data entry errors. Teaching data collectors about the use of measuring equipment as well as about the processes of data analysis can contribute to improved data accuracy. Analyses of the quality of LSIC data, aided by consideration of the RAOs' perspectives, suggest that these simple

changes could have a substantial impact on data quality. Heightened measurement accuracy and increased representativeness would lead to an increased sample size (due to fewer data points being excluded) and would enhance the validity of research findings, lending the study more power to influence policy.

Glossary of acronyms

ABS	Australian Bureau of Statistics
AGA	Appropriate-for-gestational age
AIHW NPSU	The Australian Institute of Health and Welfare National Perinatal Statistics Unit
ALSPAC	Avon Longitudinal Study of Pregnancy and Childhood
ARIA	Accessibility/Remoteness Index of Australia
BMI	Body Mass Index
Cohort B	Baby Cohort
Cohort K	Child Cohort
DOHaD	Developmental Origins of Health and Disease
FaHCSIA	Australian Government Department of Families, Housing, Community Services and Indigenous Affairs
GISCA	The National Key Centre for Social Applications of Geographic Information System
IARE	Indigenous Areas
IGF-I	Insulin-like growth factor-I
IUGR	Intrauterine growth restriction
LBW	Low birth weight
LGA	Large-for-gestational age
LMP	Last menstrual period
LORI	Level of Relative Isolation
LSIC	Longitudinal Study of Indigenous Children, also known as Footprints in Time
MetS	Metabolic syndrome
NATSIS	National Aboriginal and Torres Strait Islander Survey
NCHS	National Center for Health Statistics
NHMRC	National Health and Medical Research Council
P1	Primary carer
PBF	Per cent of body weight as fat
RAOs	Research Administration Officers
SC	Study Child
SGA	Small-for-gestational age
TBF	Total body fat, in kilograms
VAHS	Victorian Aboriginal Health Service
W1	Wave 1
W2	Wave 2
W3	Wave 3
W4	Wave 4
WAACHS	Western Australian Aboriginal Child Health Survey
WHO	World Health Organization
WHO CGS	World Health Organization Child Growth Standards

Indices

A) Index of figures

B) Index of tables

C) Index of appendices

Appendices

Appendix A: Consent form for interviews with the LSIC RAOs

Consent Form –
Examination of the interview process in the Longitudinal Study of Indigenous Children

THE AUSTRALIAN NATIONAL UNIVERSITY

Ethics Approval No:

*Participant ID No*_____

In relation to this study I, _____ *(name of participant)* have read the Patient Information Sheet and have been informed of the following points:

1. Approval for this study has been granted by the ANU Human Research Ethics Committee.
2. The aim of this study is to examine the experience of individuals who conducted interviews for the Longitudinal Study of Indigenous Children (LSIC).
3. The study procedure will involve the researchers asking me questions about the time and place of the LSIC interviews I conducted, my experience conducting the interviews, and any difficulties I may have faced in the interview process.
4. Possible adverse effects or risks related to this study include the loss of time taken to participate in the conversation, and the possibility (though unlikely) of personal or psychological discomfort from being asked to reflect on difficult or complex situations.
5. Should I develop a problem which I suspect may have resulted from my involvement in this project, I am aware that I may contact the lead investigator Katie Thurber (by email: katherine.thurber@anu.edu.au or by phone: 6125 5615) or Dr Cathy Banwell (by email: cathy.banwell@anu.edu.au or by phone: 6125 5615).

   Should I have any problems or queries about the way in which the study was conducted, and I do not feel comfortable contacting the research staff, I am aware that I may contact the ANU Human Research Ethics Committee (by email: human.ethics.officer@anu.edu.au or phone: 6125 3427) or Dr Simon Bain, Director of the Office of Research Integrity (by email: simon.bain@anu.edu.au or phone: 6125 0422).
6. I can refuse to take part in this project or withdraw from it at any time.
7. Participation in this project will not result in any costs to me.
8. I understand that while the results of this research will be made accessible, my involvement and my identity will not be revealed.

After considering all of these points, I accept the invitation to participate in this study.

Name (please print): _____ **Date:** _____

Signature (Participant): _____

Signature (Investigator): _____ **Date:** _____

Appendix B: Information form for interviews with the LSIC RAOs

Examination of the interview process in the Longitudinal Study of Indigenous Children

Ethics approval No:

What is this study about?
You are invited to participate in a pilot study looking at the experience of individuals conducting interviews for the Longitudinal Study of Indigenous Children (LSIC). You were selected as a possible participant because you were identified as a member of the LSIC interview team.

If you decide to participate, we will ask you to:
Participate in a conversation about your experience conducting interviews for the Longitudinal Study of Indigenous Children (LSIC). Questions will relate to the time and place of the interviews you conducted, your experience conducting the interviews, and any difficulties you may have faced in the interview process. Conversations will be tape recorded in order to ensure accurate interpretation of the conversation.

You will be reimbursed with a small gift to show my appreciation for your time.

Information about your involvement:
- The research is confidential.
- Any information you give us will not be associated with your name or phone number.
- Your information will be de-identified and will not be made available outside the study team, except as required by law.
- You have a right not to participate in, or to withdraw from the study at any time, both during the study and after the study.

How will your information be used?
If you give us your permission by signing the consent form, your involvement in this research will assist us to better understand the experience of conducting interviews for the Longitudinal Study of Indigenous Children (LSIC). This insight will enlighten our use of the LSIC dataset to examine the nutrition and development of Indigenous children. However, we cannot and do not guarantee or promise that you will receive any direct benefits from this study.

The information collected may be used at scientific meetings and conferences, and to publish the findings in academic books and journals. In any publication, information will be provided in such a way that you cannot be identified.

If you have any complaints or questions:
Complaints may be directed to the ANU Human Research Ethics Committee (by email: human.ethics.officer@anu.edu.au or phone: 6125 3427) or to Dr Simon Bain, the Director of the Office of Research Integrity (by email: simon.bain@anu.edu.au or phone: 6125 0422).

If you have any questions, please feel free to ask us now. If you have any additional questions later, please contact Katie Thurber (email: katherine.thurber@anu.edu.au or phone: 6125 5615) or Dr Cathy Banwell (email: cathy.banwell@anu.edu.au or phone: 6125 5615).

Appendix C: *Interview questions for conversations with the LSIC RAOs*

For the one-on-one interviews, questions included a selection from the following:

1. For which wave(s) of the study did you conduct interviews?
2. At what site(s) have you conducted interviews?
    a. Do you usually know the families that you interview?
    b. Do you feel you are making connections with these families?
    c. Have you noticed differences between communities or between sites?
    d. Do these differences influence the way in which you conduct interviews?
3. How did you come to be an interviewer for LSIC?
    a. What appealed to you about the study?
4. How was your training for LSIC?
5. How did you feel going into your first interview?
6. How do you think the families you interview perceive the LSIC study?
7. What has been your favourite memory so far?
8. How would you describe the process of measuring the height and weight of children?
    a. How has the equipment changed since the first wave of the study?
    b. Are parents usually comfortable with you taking the height and weight measurements?
    c. Are parents usually interested or concerned about their child's weight or height?
    d. How long does it usually take to record these measurements for each interview?
    e. If children are hesitant to be measured, what type of strategies do you use?
    f. Is there anything that would make it easier for you to weigh or measure the children?
9. How would you describe the process of collecting the food recall data?
    a. Do you feel that mothers have a good sense of what their children had eaten the previous day?
    b. Do you think that mothers report their child's intake honestly?
    c. Do you feel that this information is a good representation of children's food intake?
    d. How much time does it take to go through the food recall questions?
10. Are there any questions that participants seem to be sensitive in responding to?
    a. How do you try to mediate that?
11. Is there anything that you'd like to share or that you think is important?

For the focus group, questions asked included the following:

1. How would you describe the process of measuring the height and weight of children?
    a. For one interview, how long does it usually take you to measure the child?
    b. Do children usually understand that they are meant to stand still?
    c. Have you faced any issues taking these measurements?
    d. Has the equipment been adequate for measuring and weighing children?
    e. What changes have you made to your technique in measuring height and weight?

    f. Do you have a sense of why some children refuse to have these measurements taken? Has this changed from wave to wave?
2. When recording birth weight and gestational age, did most parents or carers know this information?
3. How would you describe the process of collecting the food recall data?
    a. Do you think that mothers were happy to share this information with you?
    b. Did you feel that mothers had a good sense of what their children had eaten the previous day?
    c. Do you think that mothers reported their child's intake accurately?
    d. How has your relationship with the families affected their reporting of food intake?
4. In general, how do you think that families perceive the interviews?
5. As interviewers, why do you choose to work for LSIC?
    a. What is the most rewarding part of the experience?
6. How is the process of scheduling interviews?
    a. How do you manage the frustration of being turned away?
    b. Do you face any other challenges?
    c. How much time do you spend traveling to interview sites?
7. Is there anything you would like to share?

Appendix D: *Raw weight and height data*

Wave 1:

Weight: In the first wave of the study, 964 primary carers were happy for their child to be weighed by the RAOs, and an additional 326 carers were happy to take the measurements themselves. Carers of 171 children refused to have their child weighed, and data are missing for 19 individuals. There were 85 children with weights recorded, in kilograms, from the Baby Book. For eight of these 85, a weight from the Baby Book was recorded in addition to a weight that was measured by the RAO. The weight measured by the RAO will be utilised for analyses, but the Baby Book weight can be used as an indicator of the consistency of the RAOs' weight measurements and those from the Baby Book (see Table 31). This leaves 77 weights from Baby Book eligible for inclusion.

Table 31: Weight measured by RAO versus from Baby Book, Wave 1

SC weight as measured at interview (grams)	**Age at time of RAO measurement** (months)	**SC weight recorded from Baby Book** (grams)	**Age at time of Baby Book measurement** (months)
6,500	9	7,000	--
10,006	16	10,000	--
12,000	14	11,500	13.31
14,004	18	11,400	--
14,300	19	14,000	18.08
13,200	11	15,000	10.77
15,000	52	15,000	--
18,000	51	18,000	--

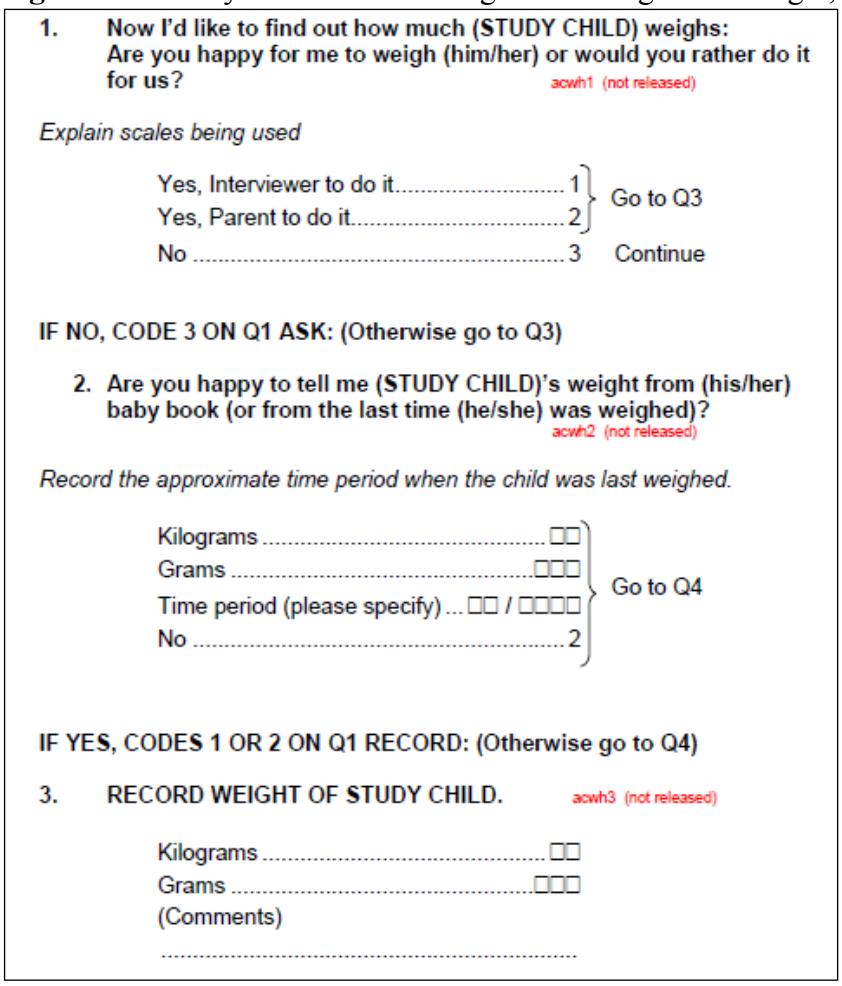
For 35 of the remaining 77 Baby Book measurements, there was no age recorded for the time of the measurement (see Table 32); in these cases, the weight can be included in analyses but z-scores, adjusted for age, cannot be calculated. For an additional two cases, there was a comment recorded about the time period of the weight measurement, but it was ambiguous and did not enable the determination of the child's age at the time of measurement. Thus, there were 40 cases in which a weight was recorded from the Baby Book, together with a non-ambiguous age at the time of measurement. For these 40 cases, weight-for-age and BMI-for-age z-scores can be calculated.

Table 32: Weight and time period of measurement recorded in Baby Book?

	Frequency	**Per cent**
No age at time of measurement recorded.	35	41.18
Age at time of measurement is unclear.	2	2.35
RAO also measured child's weight – Baby Book weight not included in analyses.	8	9.41
Can calculate age at time of measurement.	40	47.06
Total	*85*	*100.00*

The survey program for recording weight included two groups of boxes, one labelled 'Kilograms' and one labelled 'Grams' (see Figure 36 for image of survey instrument). Interviewers could type the measured number of kilograms into a set of two boxes, and type the measured number of grams into a set of three boxes. The set-up of the boxes induced some confusion; for example, if an RAO wished to record a weight of 12.1 kilograms, the correct input would be to type 1 and 2 into the two 'Kilograms' boxes and 1, 0, and 0 into the 'Grams' boxes. However, in many cases, interviewers typed the single digit, in this case 1, representing the first decimal place displayed on the scale, into one of the 'Grams' boxes. This entry, intended to represent 12.1 kilograms, would be incorrectly recorded as a weight of 12.001 kilograms. Given that the scales provided to the RAOs for weighing children were not accurate to the thousandth of a kilogram (to the nearest gram), a measured weight of 12.001 kilograms is implausible. The WHO recommendations also state that weight measurements should be recorded to two digits (150). Thus, in cases where a weight was recorded in kilograms and grams, and a one- or two-digit number (not ending in zero) was recorded in the grams column, the number in the grams column was re-coded to represent hundreds or tens of grams, respectively. This occurred in 325 instances in the measured weight data in Wave 1.

Figure 36: Survey form for measuring child's height and weight, Wave 1



1. **Now I'd like to find out how much (STUDY CHILD) weighs: Are you happy for me to weigh (him/her) or would you rather do it for us?** acwh1 (not released)

Explain scales being used

Yes, Interviewer to do it 1
Yes, Parent to do it ..... 2 } Go to Q3
No ..... 3 Continue

IF NO, CODE 3 ON Q1 ASK: (Otherwise go to Q3)

2. **Are you happy to tell me (STUDY CHILD)'s weight from (his/her) baby book (or from the last time (he/she) was weighed)?** acwh2 (not released)

Record the approximate time period when the child was last weighed.

Kilograms ..... □□
Grams ..... □□□
Time period (please specify) ... □□ / □□□□ } Go to Q4
No 2

IF YES, CODES 1 OR 2 ON Q1 RECORD: (Otherwise go to Q4)

3. **RECORD WEIGHT OF STUDY CHILD.** acwh3 (not released)

Kilograms ..... □□
Grams ..... □□□
(Comments)
.....

RAOs used the same template to record the weights recorded from the Baby Book, and the same type of data entry errors occurred. There were a total of nine changes to the weights recorded from the Baby Book (see Table 33).

Table 33: Changes to weights from Baby Book after cleaning

Originally recorded weight (grams)	**Re-coded weight** (grams)
5,091	5,910
7,005	7,500
8,016	8,160
11,016	11,160
11,027	11,270
12,001	12,100
14,006	14,600
15,009	15,900
16,012	16,120

Height: In the first wave of the study, the majority (96%) of parents consenting for their children to be weighed also consented for their children to be measured, and the

majority (71%) of parents who were unwilling for their child to be weighed were also unwilling for their child to be measured (see Table 34). Of those who agreed to have measurements taken, more parents were comfortable having their child measured by the RAO (rather than measuring the child themselves) than having their child weighed by the RAO (rather than weighing the child themselves), at 83% and 75%, respectively. In the first wave of the study, 1,063 primary carers were happy for their child to be measured by the RAOs, and an additional 221 carers were happy to take the measurements themselves. Of the 171 refusing measurement, 40 primary carers were willing to share the child's most recent height measurement from the Baby Book. Data are missing from 19 children for both height and weight.

Table 34: Willingness to have weight and height measured in Wave 1

	Happy to be weighed in Wave 1?				
Happy to be measured in Wave 1?	No answer	Yes, RAO to do it	Yes, parent to do it	No	*Total*
No answer	19	0	0	0	*19*
Yes, RAO to do it	0	904	120	39	*1,063*
Yes, parent to do it	0	27	183	11	*221*
No	0	33	23	121	*177*
Total	*19*	*964*	*326*	*171*	*1,480*

One of the heights taken from a Baby Book was recorded in feet and inches, and the remaining 39 were recorded in centimeters. Of these 40 measurements recorded, half had no age listed for the time of measurement, four children had an ambiguous time period listed, and 16 children had a non-ambiguous time period listed which allowed the calculation of the child's age at the time of the measurement (see Table 35). For these 16 children, height-for-age and BMI-for-age z-scores can be calculated. There are 23 cases in which the age of measurement of weight is not the same as the measurement of height; height-for-age and weight-for-age z-scores can each be calculated, but BMI-for-age z-scores cannot be calculated in this case. In total, there are 1,257 children with weight, height, and age at the time of measurement recorded in Wave 1 (see Table 36).

Table 35: Height and time period of measurement recorded in Baby Book?

	Frequency	**Per cent**
No age at time of measurement recorded.	20	50.00
Age at time of measurement is unclear.	4	10.00
Can calculate age at time of measurement.	16	40.00
Total	*40*	*100.00*

Table 36: Weight, height, and age recorded in Wave 1

	Both weight and age recorded	**Weight, but no age recorded**	**Age, but no weight recorded**	**Neither weight nor age recorded**	*Total*
Both height and age recorded	1,257	22	2	13	*1,294*
Height, but no age recorded	16	9	0	3	*28*
Age, but no height recorded	1	0	0	0	*1*
Neither height nor age recorded	48	11	0	98	*157*
Total	*1,322*	*42*	*2*	*114*	*1,480*

Wave 2:

Weight: In the second wave, 1,200 primary carers were happy for their child to be weighed by the RAOs, and an additional 192 carers agreed to take the measurements themselves. These figures seem to depict an increase in parents' trust in RAOs between the first and second waves of the study; 129 parents who refused to have their child weighed in the first wave consented to have their child weighed in the second wave of the study. Additionally, the per cent of consenting parents requesting to weigh the child themselves (rather than allowing the RAOs to weigh the child) decreased from 25 in Wave 1 to 14 in Wave 2. Carers of 85 children refused to have their child weighed, but 69% of these carers agreed to provide the child's most recent weight from the Baby Book. Data are missing for nine individuals. In the comment fields, some RAOs wrote down the reason why the child was not weighed (see Table 37). Additionally, ten comments included the children's measured weight (with units specified), and these weights will be included in analyses. Four of these were listed in kilograms, and the remaining six were listed in stones, pounds, and ounces. These weights were all listed as measured by the RAO (not taken from the Baby Books); RAOs may have entered the weights in the wrong section or in the case of the measurements recorded in stones, the

appropriate boxes for entry did not exist. In one instance, a weight recorded as measured by the RAO was re-coded to missing on the basis of the comment which stated, 'Respondent advised Weight,' as the accuracy of this weight is uncertain.

Table 37: Free text comments about weight measurement in Wave 2

Comments about weight recording, Wave 2:
Child became a bit distressed when tried to get go him to go near scales
Child is sleeping
Are speaking Portuguese
Child at Childcare at time of this interview
Child asleep
Child was tired and upset and was judged best not to inflame situation
Left scales at another families place
Respondent advised Weight*
Child not at home when interview was conducted
Child sick
Child very upset and would not settle to do measurement
Weighed 1st 14
SC asleep
At 18 month needles, wouldn't get on scales
Child was asleep we used last weight details from Baby Book

* Child's weight was re-coded to missing given lack of verification of measurement accuracy.
~ Names, dates, and locations removed for the protection of privacy; spelling edited from original comment.

There were 59 children with weights recorded from the Baby Book. For 18 (31.03%) of these children, there was no age recorded for the time of the measurement (see Table 38). There were two cases (3.45%) in which an ambiguous comment was recorded for the time period at which the measurement was taken. For the remaining 38 children (65.52%), the age at the time of measurement could be determined, enabling the calculation of z-scores for weight-for-age and BMI-for-age.

Table 38: Weight and time period of measurement recorded in Baby Book?

	Frequency	**Per cent**
No age at time of measurement recorded	18	31.03
Age at time of measurement is unclear	2	3.45
Can calculate age at time of measurement	38	65.52
Total	*58*	*100.00*

There were six cases in which a one- or two-digit number was entered into the grams column by the RAO, indicative of a data entry error; these weights were re-coded according to the criteria described earlier (see Table 39). Overall, there were 1,425 children with both weight and age of weight measurement recorded in the second wave of the study.

Table 39: Changes to weights recorded by RAOs after cleaning

Originally recorded weight (grams)	**Re-coded weight** (grams)
1,001	1,100
9,037	9,370
11,076	11,760
15,055	15,550
55,006	55,600
79,001	79,100

Height: Nearly 99% of parents consenting for their children to be weighed in Wave 2 also consented for them to be measured: 1,243 primary carers were happy for their child to be measured by the RAOs, and an additional 163 carers preferred to take the measurements themselves. Three of these measurements were re-coded to missing on the basis of comments left by the interviewers suggesting their limited accuracy (see Table 40). Other comments describe why height measurements were not taken for some children. In total, 71 parents refused to have their child's height measured; 51 of whom had also refused to have their child's weight recorded, but 20 of whom had consented to their child being weighed in the same wave of the study (see Table 41). Of the 71 refusing measurement, 12 primary carers were willing to share the child's most recent height measurement from the Baby Book. The age at which the measurement was taken was recorded for ten of the 12 entries from the Baby Book. Data are missing from nine children for both height and weight. In total, 1,408 children have both a height and an age at the time of measurement recorded (see Table 42).

Table 40: Comments about the height measurements in Wave 2

Comment about height measurement, Wave 2:
Child not available during interview period
Child was having lunch and did not want to be interrupted
Sleeping
Child gone out with father
SC asleep
Respondent advised Height*
Child sleeping
Tried, but ran away
SC kept moving so height reading is a close estimate*
Child not at home
SC was wriggling around and was not standing straight at the time of reading*
SC 102; went on first plane flight
Child asleep and using last height measure 6 months ago from Baby Book

Child at Childcare Centre at time of this interview.
This interview is being conducted at mother's workplace and child is at Childcare
[Name] is asleep
Details taken from last check-up [date]
Child asleep

* Child's height was re-coded to missing given lack of verification of measurement accuracy.
~ Names, dates, and locations removed for the protection of privacy; spelling edited from original comment.

Table 41: Willingness to have weight and height measured in Wave 2

	Happy to be weighed in Wave 2?				
Happy to be measured in Wave 2?	No answer	Yes, RAO to do it	Yes, parent to do it	No	*Total*
No answer	9	0	0	0	*9*
Yes, RAO to do it	0	1,177	38	28	*1,243*
Yes, parent to do it	0	14	143	6	*163*
No	0	9	11	51	*71*
Total	*9*	*1,200*	*192*	*85*	*1,486*

Table 42: Weight, height, and age recorded in Wave 2

	Both weight and age recorded	**Weight, but no age recorded**	**Age, but no weight recorded**	**Neither weight nor age recorded**	*Total*
Both height and age recorded	1,392	9	0	7	*1,408*
Height, but no age recorded	0	7	0	0	*7*
Age, but no height recorded	3	0	0	0	*3*
Neither height nor age recorded	30	10	0	28	*68*
Total	*1,425*	*26*	*0*	*35*	*1,486*

Wave 3:

Weight: In Wave 3, 1,178 primary carers consented for their children to be weighed by the RAOs, and 141 primary carers chose to weigh their child themselves; 64 primary carers refused to have their child weighed (see Table 43). Four weight measurements were re-coded to missing on the basis of comments recorded by the interviewer (see Table 44). The comment section also included 47 weights (including units of measurement); these were included as the child's measured weight. There were 743

cases where the recorded weight was modified to adjust for data entry errors in the recording of the number of grams (as described previously).

Table 43: Consent to be weighed in Wave 3

	Frequency	**Per cent**
Yes, RAO to do it	1,178	85.18
Yes, parent to do it	141	10.20
No	64	4.63
Total	*1,383*	*100.00*

Table 44: Comments about weighing child in Wave 3

Comments about permission to weigh, Wave 3:
Child at Day care
Child not at home
Child was asleep and I had to return to [place]
Child would not hop on scales for RAO or mother
Come back to complete interview, [Name] was at day care. Mum has recorded [Name]'s weight a few weeks ago, he weighed 16kg*
I didn't have the scales
[Name] at Preschool. Mum said she was around 18kgs two weeks ago*
Question not answered on the hard copy form - data entered by [Name] [date]
Refused; remove from sample
SC is absent
SC didn't want to be weighed
With assistance from P1 as SC was reluctant to stand on scales
Child asleep during interview
Child had fallen asleep during interview
Child not present
Child refused
Child would not stand or consent to this being done
Gone to [place] for Christmas
[Name] fell asleep during interview
[Name] is asleep
[Name] refuses / too shy
Not doing interview this year
P1 refused last year
SC didn't want to weighed
Comments about being weighed by RAO [excluding numeric weights]:
Child ran off with her friends
Child wasn't there, so P1 just guessed*
Had blister on foot not able to stand properly
I know this seems like not a lot, but I did check and re check the weight. She is a very slight child
Sore foot, did not balance well on scales
The last time I got her weighed she was 5.7 kilos, but that was three weeks ago. They said she was very underweight for her age but we're not sure why, because she eats like a horse
Used my clients scales, mine need new batteries*

Weighed three weeks ago - child refused

* Child's weight was re-coded to missing given lack of verification of measurement accuracy.
~ Names, dates, and locations removed for the protection of privacy; spelling edited from original comment.

Of the 64 children who refused to be weighed in Wave 3, only 22 (34.38%) provided a weight (in kilograms and grams) from the Baby Book. The age at the time of measurement could only be determined for 11 children. In two cases, no age at the time of measurement was recorded. The time between the survey and measured weight exceeded two years in three cases, and thus the Baby Book measurements were considered out-dated; for the remaining six cases, the age information was non-specific (see Table 45). Of the 22 weights recorded from the Baby Book, there were four instances in which the number of grams needed to be multiplied by 100 to account for errors in data entry, as described previously (see Table 46). Additionally, five weights from the Baby Book were recoded to missing as the comments relating to these measurements implied a low degree of reliability (see Table 47). A total of 1,334 weight measurements remain (see Table 48).

Table 45: Weight and time period of measurement recorded in Baby Book?

	Frequency	**Per cent**
No age at time of measurement recorded	2	9.09
Age at time of measurement is unclear	9	40.91
Can calculate age at time of measurement	11	50.00
Total	*22*	*100.00*

Table 46: Changes to weights from Baby Book after cleaning

Originally recorded weight (grams)	**Re-coded weight** (grams)
11,001	11,100
22,001	22,100
15,005	15,500
7,009	7,900

Table 47: Free text comments about weight reported from Baby Book in Wave 3

Comments about weight from Baby Book, Wave 3:
Child asleep
Child did not want to do
Child not home at time of interview
Child not home at time of interview, so parent had a guess as to weight*
Child sleeping
Child would not get on scales
[Name] doesn't know when child was last weighed
Participant moved to another area not covered by a RAO; try again next year in Wave 4
Participants moved to another area; try again next year in Wave 4
Participants moved to another area; try again next year in Wave 4
SC would stand on the scales
Can't find Baby Book, roughly 15kgs*
Child at day care - unable to height and weight
Child not at home at time of interview
Child not at home at time of interview
Child sleeping at time of interview
Child would not let parent weigh him either we had to go to the clinic and get his birth weight*
Child would not stand on scales and parent does not remember last time weighed
Interview was completed by another interviewer
[Name] fell asleep
Not doing interview this year

* Child's weight was re-coded to missing given lack of verification of measurement accuracy.
~ Names, dates, and locations removed for the protection of privacy; spelling edited from original comment.

Table 48: Source of weight information in Wave 3, after cleaning

Source of weight information	**Frequency**	**Per cent**
RAO weighed	1,317	98.73
Baby book	17	1.27
Total	*1,334*	*100.00*

Height: In Wave 3, 1,194 families consented to having the child measured by the RAO, and 114 primary carers asked to take the measurements themselves. Measurement was refused for 75 children. Of those refusing to be measured, 19 provided heights (in centimetres) from the Baby Book, only five of which had a usable age of measurement recorded. Based on the comments recorded by the RAOs, eight measured heights, as well as one height recorded from the Baby Book, were re-coded to missing (see Table 49). There are 1,272 children for which weight, height, and the age of measurement was recorded (see Table 50).

Table 49: Comments about child height measurements in Wave 3

Comments about permission to measure, Wave 3:
Child got too upset and didn't want to
Child not at home
Child not present
Child refused and was upset. Measurement not taken
Child was asleep and I had to return to [place]
Child would not come near parent or RAO to get height measurement
[Name] is at day care. Mum has recorded [Name]'s height a few weeks ago* *[Baby Book height]*
[Name] at preschool, Mum unsure of her height
No response on hard copy of the questionnaire
SC would not stand still, so the parent had to do a 'guess-timate'*
Child refused
Child refused
Child refused
Child refused
Child would not consent to this being done
Could not do height
[Name] fell asleep
[Name] is asleep
[Name] wants dad to do it
Not doing interview this year
Not interviewed this wave
SC didn't want to be measured
SC no longer in the program
Comments about being measured by RAO [excluding numeric heights], Wave 3:
Child did not want to be measured
Child upset and did not want to cooperate
Child was too shy to come near RAO, some mother / father / older sister all had a hand in doing the height and weight measures
Child wasn't present, but mother said she's about this big*
[Name] had shoes on so would be around 101cm*
Only a guess, child ran off!*
[Name] was not standing still*
Child ran off upset. this was an estimate*
Child refused - measured against myself*
Was moving a little*
Comments about height from Baby Book, Wave 3:
Child did not want to do
Child had fallen asleep
Child not available at time of measurements
Child refused - [name] did not know height of child
Child sleeping
Child would not allow RAO or mother to do this
Child would not consent to height and weight measures
[Name] at grandmother's
Measured about a month ago, can't remember how tall she was.
Participant moved to another area not covered by a RAO; try again next year in Wave 4
Participants moved to another area; try again next year in Wave 4

Participants moved to another area; try again next year in wave 4
Cant find Baby Book
Child at day care - unable to height and weight
Child not at home at time of interview
Child not home at time of interview
Child not home at time of interview
Child not home at time of interview
Child not present
Child refused/too shy
Child sleeping at time of interview
Had to get height from clinic records
[Name] fell asleep
Not doing interview this year
SC no longer in the program

* Child's height was re-coded to missing given lack of verification of measurement accuracy.
~ Names, dates, and locations removed for the protection of privacy; spelling edited from original comment.

Table 50: Weight, height, and age recorded in Wave 3

	Both weight and age recorded	**Weight, but no age recorded**	**Age, but no weight recorded**	**Neither weight nor age recorded**	*Total*
Both height and age recorded	1,272	2	2	6	*1,282*
Height, but no age recorded	5	29	0	2	*36*
Age, but no height recorded	8	0	1	0	*9*
Neither height nor age recorded	18	0	2	36	*56*
Total	*1,303*	*31*	*5*	*44*	*1,383*

Wave 4:

Weight: In Wave 4, 1,233 children consented to be weighed by the RAOs, 16 children consented to be weighed by their parents, and 21 children refused to be weighed (see Table 51). Eight of these 21 children provided a weight from the Baby Book, as well as the age at the time of measurement. Two weight measurements were re-coded to missing on the basis of comments recorded by the interviewer (see Table 52). In one case, the comment advised that the weight was actually recorded in pounds, not kilograms as stated, so this weight was adjusted. There were 473 cases where the recorded measured weight was modified to adjust for data entry errors in the recording of the number of grams (as described previously), and an additional three cases for

weights recorded from the Baby Book. After data cleaning, 1,255 weights remained for analysis (see Table 53).

Table 51: Consent to be weighed in Wave 4

	Frequency	**Per cent**
Yes, RAO to do it	1,233	97.09
Yes, parent to do it	16	1.26
No	21	1.65
Total	*1,270*	*100.00*

Table 52: Comments about weighing child in Wave 4

Comments about permission to weigh, Wave 4:
As child is shy, her mother weighed her on the scales for RAO
Child too distressed to complete task; started crying uncontrollably and clinging to mother. Asked parents to assist but child still too distressed
[Name] was not happy on the day of my visit and refused to do any of the activities with me. P1 was also unable to get [Name] to do any of the activities
SC at school did not get to weigh him
SC wouldn't get weighed
Child did not want to participate, was upset and didn't want to do it
SC refused
SC wouldn't get weighed
Too shy, wouldn't even let Mum.
Comments about weight from Baby Book, Wave 4:
Child too distressed to complete task; started crying uncontrollably and clinging to mother. Asked parents to assist but child still too distressed
Approximately, parent's best guess*
Comments about being weighed by RAO [excluding numeric weights]:
Please note weight done in Lbs not Kilos as scales playing up ^
This may not be accurate as my scales are playing up*
Battery flat on the scales*

*Child's weight was re-coded to missing given lack of verification of measurement accuracy.
^ Child's weight was converted from pounds to grams.
~ Names, dates, and locations removed for the protection of privacy; spelling edited from original comment.

Table 53: Weights and ages recorded in Wave 4

Recorded in Wave 4	**Frequency**	**Per cent**
Both weight and age	1,255	98.82
Weight, but no age	2	0.16
Neither weight nor age	13	1.02

Height: In Wave 4, 1,221 families consented to having the child measured by the RAO, and 19 primary carers asked to take the measurements themselves. Measurement was refused for 30 children. Of those refusing to be measured, six provided heights (in

centimetres) from the Baby Book, each with the age of measurement recorded. Based on the comments recorded by the RAOs, two heights measured by RAOs were re-coded to missing (see Table 54). Additionally, in seven cases, the number of millimetres was recorded in the comment section, so heights were adjusted to address these comments. In total, 1,243 heights remain for analyses (see Table 55).

Table 54: Comments about child height measurements in Wave 4

Comments about permission to measure, Wave 4:
Child too distressed to complete task; started crying uncontrollably and clinging to mother. Asked parents to assist but child still too distressed
[Name] didn't cooperate and wouldn't let mum put stadiometer on his head
P1 took child's measurement as child is very shy
SC was at school; didn't measure him
Stadiometer is flat
Child was too upset to complete
Measure and weigh together
SC refused
SC wouldn't get measured
SC wouldn't get measured
Stadiometer not giving accurate reading*
Stadiometer not giving accurate reading*
Too shy
Too shy and disruptive
Comments about height from Baby Book, Wave 4:
Child too distressed to complete task; started crying uncontrollably and clinging to mother. Asked parents to assist but child still too distressed
Did not get information from P1
Approximately 1 metre
Can't find blue book
Comments about being measured by RAO, Wave 4:
.5 ^
5 ^
98.7 centimetres ^
99.4 centimetres was his correct height ^
I measured him 97.5 centimetres for his height ^
SC was 93.5 centimetres ^
Height was actually 96.8 centimetres but laptop will only accept whole number ^
Approximate*
Last year recorded 115 centimetres, but could have been 105 centimetres
Stadiometer flat*

*Child's height was re-coded to missing given lack of verification of measurement accuracy.
^ Height changed to reflect fractions of centimetres.
~ Names, dates, and locations removed for the protection of privacy; spelling edited from original comment.

Table 55: Weight, height, and age recorded in Wave 4

	Both weight and age recorded	**Weight, but no age recorded**	**Age, but no weight recorded**	**Neither weight nor age recorded**	*Total*
Both height and age recorded	1,244	0	0	0	*1,244*
Height, but no age recorded	0	1	0	0	*1*
Age, but no height recorded	1	0	0	0	*1*
Neither height nor age recorded	10	1	0	13	*24*
Total	*1,255*	*2*	*0*	*13*	*1,270*

***Appendix E**: Anthropometric data remaining after cleaning*

Table 56: Summary of weight and height after cleaning

	Wave 1	**Wave 2**	**Wave 3**	**Wave 4**
Age at interview recorded	1,671	1,523	1,404	1,283
Weight				
Weight recorded	1,365	1,451	1,334	1,257
Age at weight measurement recorded	1,324	1,425	1,308	1,249
Weight-for-age z-score (WFAz) recorded	1,310	1,417	1,293	1,245
WFAz flagged (z <-6 or z >5)	46	29	11	10
WFAz re-coded to missing because BMIz flagged (z < -5 or z > 5)	116	68	26	22
WFAz remaining after cleaning (% of recorded WFAz measurements)	1,147 (87.6)	1,319 (93.1)	1,255 (97.1)	1,212 (97.3)
Height				
Height recorded	1,322	1,415	1,318	1,245
Age at height measurement recorded	1,295	1,410	1,291	1,239
Height-for-age z-score (HFAz) recorded	1,254	1,344	1,206	1,206
HFAz flagged (z <-6 or z >6)	118	32	9	2
HFAz re-coded to missing because BMIz flagged (z < -5 or z > 5)	81	64	30	30
HFAz remaining after cleaning (% of recorded HFAz measurements)	1,055 (84.1)	1,249 (92.9)	1,167 (90.4)	1,174 (97.3)
Body Mass Index (BMI)				
BMI recorded	1,304	1,408	1,308	1,241
Age at height measurement equal to age at weight measurement (can calculate BMI-for-age z-score)	1,237	1,374	1,281	1,238
BMI-for-age z-score (BMIz) recorded	1,234	1,371	1,270	1,233
BMIz flagged (z < -5 or z > 5)	155	94	37	32
BMIz remaining after cleaning (% of recorded BMIz measurements)	996 (80.7)	1,207 (88.0)	1,149 (90.5)	1,170 (94.9)

Table 57: Number of anthropometric measurements recorded at Wave 1, Wave 2, Wave 3, and Wave 4, before and after cleaning

	Wave 1	**Wave 2**	**Wave 3**	**Wave 4**
Before cleaning				
Height	1,322	1,415	1,318	1,245
Weight	1,365	1,451	1,334	1,257
BMI	1,304	1,408	1,308	1,241
HFA z-score	1,254	1,344	1,206	1,206
WFA z-score	1,310	1,417	1,293	1,245
BMI-for-age z-score	1,234	1,371	1,270	1,233
After cleaning				
Height	1,055	1,249	1,167	1,176
Weight	1,147	1,319	1,255	1,214
BMI	996	1,207	1,149	1,170
HFA z-score	1,055	1,249	1,167	1,174

	Wave 1	Wave 2	Wave 3	Wave 4
WFA z-score	1,147	1,319	1,255	1,212
BMI-for-age z-score	996	1,207	1,149	1,170

Table 58: Number of children with data recorded in four, three, two, one, and zero waves of the study, before and after cleaning

	Data recorded in four waves	Data recorded in three waves	Data recorded in two waves	Data recorded in one wave	Data recorded in zero waves
Before cleaning					
Height	723	553	294	161	28
Weight	775	544	265	145	30
BMI	707	555	303	162	32
HFAz	567	528	346	166	52
WFAz	700	566	311	145	37
BMIz	637	571	339	169	43
After cleaning					
Height	414	657	412	196	80
Weight	542	623	359	180	55
BMI	380	632	450	206	91
HFAz	413	657	413	196	80
WFAz	542	622	359	181	55
BMIz	380	632	450	206	91

Table 59: Decreases in weight (associated with a decrease in z-score exceeding 3) between waves; the bolded weights and weight-for-age z-scores indicate those that were re-coded to missing

WFAz W1	Weight W1 (kg)	WFAz W2	Weight W2 (kg)	WFAz W3	Weight W3 (kg)	WFAz W4	Weight W4 (kg)
-0.85	14	**1.79**	**23**	-1.28	16.8	-0.88	19.6
1.93	**12.8**	-1.14	10			-0.62	15.4
		2.66	**19**	**-0.72**	**14.1**		
0.75	**18**	-2.43	13	-1.71	16.6		
1.19	12	**3.16**	**18.4**	-0.66	13.4		
3.48	**14**	**-0.11**	**12**	**-3.58**	**9**	**0.39**	**17**
4.57	**17**	**0.37**	**13.5**	**1.05**	**16.5**	**-2.14**	**12.6**
		1.4	**22**	-3.45	13	-1.67	18.4
0.74	12	**-4.13**	**8.5**	0.53	16.3	-0.25	16.3
4.14	**12**	-0.2	9.37	-1.15	11.2		
-2.36	12	0.28	19	**-2.98**	**13.6**	-0.05	22.4
2.16	**20**	-2.21	12	-0.36	18	-0.01	21.2
-1.28	10	**2.07**	**18**	-1.38	13.6	-0.69	16

WFAz W1	**Weight W1** (kg)	**WFAz W2**	**Weight W2** (kg)	**WFAz W3**	**Weight W3** (kg)	**WFAz W4**	**Weight W4** (kg)
-0.16	9	**1.19**	**13.2**	**3.83**	**21.1**	0.46	16.8
3.49	**14.7**	-1.13	10	-1.75	12	-1.65	13
1.31	**20**	-1.94	15	-1.3	17.8	-1.33	19.2
4.61	**14.3**	-1.29	8.4	-1.31	11.6	-1.16	13.6
0.07	**12**	-3.13	9			-3.16	12.4
4.4	**16**	1.27	14.3	1.3	16.8	1.32	19.6
0.7	10.7	0.92	14.3	**-2.22**	**11.4**	1.19	20
0.91	11.4	**4.21**	**19**	0.85	15.7	0.31	17
4.75	**18**	0.21	12			-0.07	16.8
2.01	**12**	-1.19	10.1	-0.44	13.6	0.36	16.4
1.11	12	1.41	15	**-2.3**	**10.8**	2.94	26.7
0.5	12	0.94	15	**-2.25**	**11**	1.48	21.4
-0.51	8.9	-1.28	10	**2.35**	**18.1**	-3.32	10.4
2.36	**15**	**-3.6**	**8**				
		0.92	20	**-3.57**	**12**	2.26	29.4
0.15	10			1.01	16.04	**-2**	**12.8**
		-0.29	12.5	**-4.88**	**8**	-0.38	16.2

*There were 183 decreases in height between waves; data not shown.

Appendix F: Associations between missing and implausible z-scores and demographic variables

Table 60: Mean number of missing data points for each indicator by gender

Gender (n)	**Mean number of missing BMI z-score measurements** (SD)	**Mean number of missing HFA z-score measurements** (SD)	**Mean number of missing WFA z-score measurements** (SD)
Male (887)	1.42 (1.12)	1.34 (1.11)	1.19 (1.09)
Female (872)	1.44 (1.09)	1.37 (1.08)	1.20 (1.07)
Total (1,759)	*1.43 (1.11)*	*1.36 (1.09)*	*1.20 (1.08)*

Table 61: Mean number of missing data points for each indicator by primary carer report of child's general health at Wave 1

P1's report of child's general health at W1 (n)	**Mean number of missing BMI z-score measurements** (SD)	**Mean number of missing HFA z-score measurements** (SD)	**Mean number of missing WFA z-score measurements** (SD)
Excellent (753)	1.44 (1.12)	1.36 (1.10)	1.20 (1.10)
Very good (522)	1.37 (1.12)	1.32 (1.11)	1.14 (1.07)
Good (340)	1.34 (1.08)	1.28 (1.08)	1.13 (1.05)
Fair (41)	1.41 (1.22)	1.27 (1.18)	1.10 (1.16)
Very Poor (6)	1.5 (1.22)	1.17 (1.17)	1.17 (1.33)
Total (1,662)	*1.40 (1.11)*	*1.33 (1.10)*	*1.16 (1.09)*

Table 62: Mean number of missing data points for each indicator by primary carer report of family's weekly income at Wave 1

P1's report of weekly income at W1 (n)	**Mean number of missing BMI z-score measurements** (SD)	**Mean number of missing HFA z-score measurements** (SD)	**Mean number of missing WFA z-score measurements** (SD)
Less than $150 per week (156)	1.32 (1.07)	1.26 (1.05)	1.12 (1.07)
$150 - $249 per week (226)	1.49 (1.08)	1.43 (1.08)	1.29 (1.09)
$250 - $399 per week (282)	1.45 (1.17)	1.39 (1.15)	1.27 (1.15)
$400 - $599 per week (333)	1.42 (1.06)	1.36 (1.04)	1.15 (1.04)
$600 - $799 per week (218)	1.40 (1.14)	1.31 (1.33)	1.16 (1.10)
$800 - $999 per week (139)	1.32 (1.07)	1.20 (1.03)	0.98 (1.05)
$1,000 or more per week (209)	1.33 (1.14)	1.22 (1.13)	1.01 (1.00)
Total (1,563)	*1.40 (1.11)*	*1.33 (1.09)*	*1.16 (1.08)*

Table 63: Mean number of missing data points for each indicator by primary carer's highest educational qualification attained

P1's highest educational qualification attained (n)	**Mean number of missing BMI z-score measurements** (SD)	**Mean number of missing HFA z-score measurements** (SD)	**Mean number of missing WFA z-score measurements** (SD)
Never attended school (10)	1.10 (0.88)	1.00 (0.82)	0.70 (0.82)
Year 8 or below (76)	1.16 (0.82)	1.08 (0.84)	1.00 (0.85)
Year 9 or equivalent (157)	1.19 (1.01)	1.12 (0.99)	0.99 (0.98)
Year 10 or equivalent (408)	1.24 (0.97)	1.18 (0.93)	1.00 (0.90)

P1's highest educational qualification attained (n)	**Mean number of missing BMI z-score measurements** (SD)	**Mean number of missing HFA z-score measurements** (SD)	**Mean number of missing WFA z-score measurements** (SD)
Year 11 or equivalent (248)	1.29 (0.92)	1.22 (0.90)	1.03 (0.89)
Year 12 or equivalent (249)	1.18 (1.01)	1.10 (0.96)	0.93 (0.94)
Certificate of completion (5)	1.80 (0.84)	1.80 (0.84)	1.60 (0.89)
Other non-school qualification (6)	1.00 (1.10)	1.00 (1.10)	0.67 (0.82)
Certificate I/II (67)	1.28 (1.04)	1.18 (1.01)	0.94 (0.83)
Certificate III/IV (including trade certificate) (154)	1.21 (1.01)	1.14 (1.00)	0.95 (0.93)
Advanced diploma (47)	1.34 (1.11)	1.17 (1.11)	0.91 (1.04)
Bachelor degree (with or without honours) (62)	1.15 (0.99)	1.08 (0.95)	0.95 (0.95)
Graduate diploma/Graduate certificate (16)	1.56 (0.96)	1.50 (1.03)	1.38 (0.96)
Postgraduate degree (13)	1.31 (1.18)	1.31 (1.18)	0.62 (0.87)
Total (1,518)	*1.23 (0.98)*	*1.16 (0.95)*	*0.98 (0.92)*

***Appendix G**: Use of customised birth weight references*

Kierans et al. (140) conducted a study of 865,968 births recorded by the British Columbia Vital Statistics Agency between 1981 and 2000. Infants were classified as Chinese (40,092), South Asian (38,670), First Nation (56,097), and other (731,109; predominantly Caucasian). The prevalence of SGA was calculated using both a general British Columbia reference and by using the ethnicity-specific subsample distribution. Given that SGA is intended to serve as an indicator of an adverse intrauterine environment, and therefore should be associated with perinatal mortality, the authors examined the concordance between SGA and perinatal mortality using the two references. They found that the use of the ethnicity-specific reference resulted in concordance between rates of SGA and rates of perinatal mortality, but that discordance resulted from the use of the British Columbia reference with no adjustment for ethnicity (140). This would suggest that the reference non-specific to ethnicity was not an accurate tool for categorising infants into size for gestational age categories. The authors concluded that a physiologic, rather than pathologic, difference must underlie the observed differences in birth weight between the ethnic groups examined:

> In our view, this evidence justifies the consideration of ethnic-specific standards of birth weight for gestational age, at least for Chinese, South Asian, and North American Indian ethnicities… If differences in fetal growth (as reflected by GA-specific mean birth weights and revealed SGA rates) were truly pathologic, rather than physiologic, we would expect patterns that were more coherent with those observed for perinatal mortality when the definition of SGA was based on a single population standard, rather than ethnic-specific standards (140 p. 6).

Although Kierans et al. (140) did identify differences in concordance between SGA and perinatal mortality when using an ethnicity-specific versus a non-specific reference, they did not explore if this disparity might be attributable to other factors correlated with ethnicity, rather than ethnicity itself. For example, maternal characteristics such as height and weight, which are associated with birth weight, might vary widely between ethnic groups. Further, Gardosi et al. (134) explain:

> A second methodological flaw [of ethnicity-specific standards] is that pathological factors are not excluded, yet each ethnic specific population average will be affected by an unknown, and probably varying, extend of pathology, for example, due to differing rates of smoking in pregnancy (p. 7-8) .

These authors, among other researchers, propose the use of a birth weight for gestational age reference customised to a range of maternal factors, rather than adjusting for ethnicity alone (134).

The construction of references based solely on ethnicity would be insufficient, as it would 'neglect the effect on fetal growth of other maternal and fetal characteristics secondarily associated with ethnicity. Not only ethnicity could be considered a measure of 'biological' difference, since other social and environmental variables greatly differ between different ethnicities' (90 p. 298). These variables (gender, parity, maternal height, weight and ethnicity) explain between 20 and 60% of the variance in birth weight, and factors such as maternal age, education, socioeconomic status, and marital status, as well as paternal height, contribute to further variability (134, 138). Given the significant variation in foetal growth potential attributable to maternal factors, consideration of these variables is important in disentangling cases of IUGR from cases of SGA (90, 134, 143, 175). IUGR is representative of a pathological failure to reach full foetal growth potential, so the health trajectory of infants born SGA without IUGR is thus markedly different from the trajectory of those born SGA with IUGR (142). The clinical standard for differentiating SGA and IUGR is umbilical artery Doppler; however, this is not feasible in large-scale studies (90). Soothill and colleagues (175) explain,

> Recently, with improved ultrasound imaging and the advent of Doppler studies, it has become obvious that, within the descriptive term SGA, there are separate groups with quite distinct etiologies and prognoses, each therefore requiring different management. Various terms and abbreviations for these subgroups are now adding further confusion … a standardized classification system is imperative (p. 225).

Ideally, Doppler assessment would be used to determine whether cases of SGA are also IUGR, but as this information is rarely accessible, an alternative approach is to 'customise' the birth weight reference, adjusting the optimal birth weight for each child according to physiological and pathological factors (90). Customisation, as argued by Gardosi and colleagues (90), enables the discrimination of infants who are constitutionally small from those whose smallness results from pathology (90).

Additionally, the use of customised centiles enables the identification of infants who have experienced IUGR and face a high risk of morbidity and mortality in adulthood, despite being recognised as AGA by population references (142). These infants, often born to mothers of larger weight and height, have a birth weight considered in the normal range for their gestational age, but have not reached their full

potential of intrauterine growth, as estimated by maternal factors such as size, parity, and ethnicity (142). The use of these customised centiles also enables the identification of infants who are classified as SGA according to the standard population centiles, but are just constitutionally small, having experienced no IUGR and thus facing no increased risk of pathology in adulthood (142).

For example, Verkauskiene and colleagues (143) compared the association between metabolic outcomes in adulthood and size at birth, as defined using a population-based reference and using a customised reference. The customised reference was based on a sample of 8,199 singleton live term (gestational age between 273 and 287 days, based on ultrasound) births without congenital abnormalities. Linear regression was used to create a model for the optimal term birth weight, adjusting for maternal height, weight, parity, and ethnicity (European vs. non-European). Their resulting equation (143) was:

> Term optimal weight (grams) = 3,438.9 + 105.7 (if parity = 1) + 204.8 (if parity = 2) + 183.3 (if parity = 3 or more) + 6.3 (height from 162 cm) + 0.0203 (height from 162 cm) ^{3} + 10.5 (weight from 56 kg) – 0.164 (weight from 56 kg) ^{2} – 68.3 (if sex = female) + 68.3 (if sex = male).

To create a reference accounting for gestational age as well as these predictors, this equation was combined with a Hadlock-inspired proportionality equation, resulting in a reference customised for gestational age, parity, maternal height and weight, and gender. A total of 825 individuals were classified as AGA under both references ($AGA_{pop+cust}$), and 575 were classified as SGA by both ($SGA_{pop+cust}$); 131 were classified as SGA by the population reference but AGA by the customised reference (SGA_{pop}), and 22 were classified as SGA by the customised reference but AGA by the population reference (SGA_{cust}) (143). According to this customised reference, the SGA_{cust} group had not met their optimal growth potential in utero; this group faced an increased risk of metabolic problems in adulthood, despite being characterised as AGA by the population reference. This group had an increased insulin resistance than the $AGA_{pop+cust}$ group, and a worse lipid profile than the $AGA_{pop+cust}$ group and the SGA_{pop} group(143). The contribution of fat to total body mass was higher within the SGA_{cust} versus $AGA_{cust+pop}$ group, consistent with theories on the programming of fat deposition resulting from IUGR. Thus, the use of the customised centiles seemed to more accurately identify people at risk of metabolic disease, compared to the use of population-based centiles (143). The homogeneity of the sample limited analysis of the influence of ethnicity.

Because of the method by which centiles are defined for each reference, pre-term infants are more likely to be identified as SGA using customised references, and term infants are more likely to be identified as SGA using population references (142). Population references are based on the observed distribution of weights of infants born at each gestational age, and therefore do not reflect optimal foetal growth for infants born pre-term (89, 137, 138, 142). The incidence of pathology and IUGR is higher for pre-term versus term infants (138), and thus infants who are born pre-term are less likely than term infants to have met their full growth potential. The customised centiles, in contrast, are based on optimal foetal growth, calculating birthweight percentiles from ultrasound measurements of the weights of healthy unborn foetuses at each gestational age. These customised centiles 'are calculated from an adjusted birthweight range expected at 40 weeks and extrapolated back using a standard, longitudinal, ultrasound-derived curve of intrauterine weight gain' (142 p. 239.e4).

Thus, the customised references identify additional pre-term children as SGA, compared to a population reference (142). Further, because the customised centiles account for the variation in size attributable to maternal factors, children who are constitutionally small are no longer classified as SGA, thus decreasing the portion of children born at term who are classified as SGA (142). The infants classified as SGA by these customised centiles are shown to experience higher perinatal morbidity and mortality (90), and higher rates of metabolic disturbances in adulthood (143), than those classified as SGA by population centiles. The stronger predictive value of these customised, versus population, centiles for these conditions suggests that they are better at identifying cases of IUGR and pathological smallness for age (90, 134). Gardosi et al. (134) conclude:

> … local population standards are also not enough, as they are unable to account for variation *within* populations. Instead, the individually determined growth potential is emerging as an internationally applicable standard. Support for this concept comes from the finding that physiological factors seem to affect growth similarly in different countries and continents, and that a 'standard mother' (same height, weight, parity and ethnic origin) can expect a baby with a similar birthweight, whether she is living in the UK, Australia, New Zealand, or the US. The task now is to add to the existing international data and derive coefficients for additional ethnic groups in different geographic areas (p. 8).

Mikolajczyk et al (138) compared the predictive value of SGA on adverse perinatal outcomes in the 2004-2008 WHO Global Survey on Maternal and Perinatal

Health (WHOGS) using three different size for gestational age references. First, they examined the association in Hadlock's original foetal growth equation, which only adjusted for gestational age and gender. Second, they examined the association in a country-customised reference. Third, they examined the association in a fully-customised reference, adjusting for factors including country, gender, maternal height and weight, and parity. The WHOGS sample included 237,025 singleton pregnancies from Africa, Latin America, and Asia, with information on birthweight, sex of the infant, maternal weight and height, and birth outcomes (138).

Mikolajczyk et al (138) found that the prevalence of SGA decreased, and the incidence of adverse perinatal outcomes in the SGA group increased, when using the country-specific reference versus the Hadlock reference. This was associated with a significant increase in the Odds Ratio for adverse perinatal events for SGA vs. non-SGA infants for the country-specific reference. Additional customisation, through the inclusion of factors such as maternal height and weight, however, did not significantly influence the results. The authors conclude:

> Our analysis showed that country as a proxy for the local ethnic mix was much more important than other variables. Although further customisation beyond country or ethnic origin could be theoretically appealing, additional gains were few… Although an improved prediction of perinatal mortality by individual customisation was reported in specific strata defined by maternal characteristics, this result is apparently not generalisable to the population level (138 p. 1859).

Thus, the benefits arising from the use of a fully-customised, versus country-specific, reference are minimal in the analyses of birth weight in LSIC.

Appendix H: *Missing birth weight and gestational age data*

When RAOs requested permission to ask the primary carer (P1) questions about the birth of the study child, 113 P1s were unwilling (see Table 64). Of the 1,646 P1s who were willing to answer these questions, a birth weight was recorded for 1,417 (86%) children; for the remaining 14%, P1s refused to answer the question about birth weight, said they did not know, or the information was missing. The vast majority of these birth weights (1,378; 97%) were recorded in Wave 1, with 38 additional birth weights recorded in Wave 2 and one additional birth weight recorded in Wave 3. Of the 1,417 recorded birth weights, 471 (33%) were recorded in kilograms and grams, and 946 (67%) were recorded in pounds and ounces. In addition, 11 P1s responded 'Other' and written comments were recorded. Of these, five included numeric values for birth weight, and six explained why the birth weight information was not available (see Table 65). If the numeric values also included units of measurement (e.g. kilograms or pounds) and were unambiguous, they were entered as the child's birth weight. Any non-numeric or ambiguous comments were coded as a missing birth weight. In total, a numeric birth weight was recorded for 1,420 children (80.73% of the sample).

Table 64: Recorded and missing birth weight

	Number with birth weight recorded (%)	**Number without birth weight recorded** (%)	**Total**
Total sample	1,420 (80.73)	339 (19.27)	*1,759*
Wave of study in which asked about birth?			
Asked in Wave 1	1,381 (82.65)	290 (17.35)	*1,671*
New entrant, asked in Wave 2	38 (52.05)	35 (47.95)	*73*
New entrant, asked in Wave 3	1 (16.67)	5 (83.33)	*6*
New entrant, not asked	0 (0)	9 (100)	*9*
Primary carer (P1) is birth mother?			
P1 is birth mother	1,354 (84.15)	255 (15.85)	*1,609*
P1 is not birth mother	66 (44.00)	84 (56.00)	*150*
P1 willing to answer questions about birth?			
P1 is willing to answer questions	1,420 (86.27)	226 (13.73)	*1,646*

	Number with birth weight recorded (%)	Number without birth weight recorded (%)	Total
P1 is not willing to answer questions (or responded other or don't know)	0 (0)	113 (100.00)	*113*
Birth weight reported from Baby Book?			
Yes	1,145 (92.49)	93 (7.51)	*1,238*
No	274 (65.87)	142 (34.13)	*416*
Source of birth weight information not recorded	1 (0.95)	104 (99.05)	*105*
Gestational age recorded?			
Yes	1,409 (87.03)	210 (13.97)	*1,619*
No	11 (7.86)	129 (92.14)	*140*
Level of Relative Isolation?			
None	416 (96.30)	16 (3.70)	*432*
Low	706 (82.57)	149 (17.43)	*855*
Moderate	165 (69.92)	71 (30.08)	*236*
High / Extreme	132 (56.17)	103 (43.83)	*235*
LORI not recorded	1 (100.00)	0 (0)	*1*

Table 65: Comments recorded about birth weight

Comment	Birth weight coded as
2 pounds, 6 ounces	2 pounds, 6 ounces = 1,077.28 grams
4180 grams	4 kilograms, 180 grams = 4,180 grams
4kgs - tumour was 1.5 kilograms	Missing
6 Pounds	6 pounds, 0 ounces = 2,721.55 grams
8	Missing
Baby Book kept at health clinic	Missing
Can't find book	Missing
Can't remember	Missing
Don't remember	Missing
P1 doesn't take her to child health service	Missing
Packed away	Missing

* Spelling edited from original comment.

Appendix I: *Accuracy of birth weight and gestational age data*

The distribution of reported gestational age was not significantly different for P1s who were birth mothers compared to P1s who were not birth mothers ($t(1617) = 0.7526$, two-tailed $p = 0.4518$ for two-sample t-test with equal variance) (see Table 66). Of the 150 non-birth mothers asked about the child's birth, 66 provided a birth weight 140 provided a gestational age, and 65 provided both. Of the 65 non-birth mothers providing both birth weight and gestational age, 53 (82%) reported the birth weight from the Baby Book, and 12 (18%) reported the birth weight from memory or another source of information. The accuracy of these 12 gestational ages is of concern given that the P1 is not the birth mother and the Baby Book was not accessible; however, there was not a significant difference in the distribution of gestational ages for non-birth mothers who reported the birth weight from the Baby Book versus from recall ($t(63) = 0.82167$, two-tailed $p = 0.4172$ for two-sample t-test with equal variance). Thus, all gestational ages will be included in analyses.

Table 66: Distribution of gestational age, in weeks, for P1s who are birth mothers versus non-birth mothers

Gestational age	**P1 is birth mother** (%)	**P1 is not birth mother** (%)	*Total sample with gestational age recorded* (%)
32 weeks pregnant or less (eight weeks or more early)	56 (3.66)	0 (0)	*56 (3.46)*
33 weeks pregnant (seven weeks early)	6 (0.39)	1 (1.11)	*7 (0.43)*
34 weeks pregnant (six weeks early)	21 (1.37)	1 (1.11)	*22 (1.36)*
35 weeks pregnant (five weeks early)	20 (1.31)	1 (1.11)	*21 (1.30)*
36 weeks pregnant (four weeks early)	65 (4.25)	4 (4.44)	*69 (4.26)*
37 weeks pregnant (three weeks early)	73 (4.77)	4 (4.44)	*77 (4.76)*
38 weeks pregnant (two weeks early)	205 (13.41)	8 (8.89)	*213 (13.16)*
39 weeks pregnant (one week early)	215 (14.06)	17 (18.89)	*232 (14.33)*
40 weeks pregnant (on time)	506 (33.09)	42 (46.67)	*548 (33.85)*
41 weeks pregnant (one week late)	230 (15.04)	7 (7.78)	*237 (14.64)*

Gestational age	**P1 is birth mother** (%)	**P1 is not birth mother** (%)	*Total sample with gestational age recorded* (%)
42 weeks pregnant or more (two weeks or more late)	132 (8.63)	5 (5.56)	*137 (8.46)*
Total	*1,529 (100.00)*	*90 (100.00)*	*1,619 (100.00)*

The prevalence of missing data for gestational age was not independent of LORI or whether the P1 was the birth mother or not. The prevalence of missing data was significantly higher in areas with higher levels of relative isolation (Pearson chi2(1) = 44.5964, $p < 0.001$) (see Table 67) and the risk of missing data nearly 70% higher for non-birth compared to birth mothers (95% CI: 1.43, 1.91) (see Table 68).

Table 67: Missing gestational age and LORI (for both cohorts)

	Missing gestational age? (%)		
LORI	No	Yes	*Total*
None	422 (97.69)	10 (2.31)	*432*
Low	791 (92.51)	64 (7.49)	*855*
Moderate	210 (88.98)	26 (11.02)	*236*
High / Extreme	195 (82.98)	40 (17.02)	*235*
Total	*1,618 (92.04)*	*140 (7.96)*	*1,758*

Table 68: per cent of mothers versus birth mothers with missing gestational age for study child (for both cohorts)

	Missing gestational age?		
Birth mother	Yes	No	*Total*
No	60 (40)	90 (60)	*150*
Yes	80 (4.97)	1,529 (95)	*1,609*
Total	*140 (7.96)*	*1,619 (92.04)*	*1,759*

***Appendix J**: Cleaning birth weight and gestational age data*

There is a significant association between flagged birth weight and gestational age, with a higher percentage of birth weight z-scores exceeding 3 or falling below -3 at the low end of the gestational age range (Pearson chi2(10) = 60.9244, $p < 0.001$) (see Table 69). For example, 29% of infants with a gestational age of 32 weeks and 33% of infants with a gestational age of 33 weeks were flagged for having extreme z-scores, compared to less than 10% in all other gestational age groups.

Table 69: per cent of birth weights with z-score greater than 3 or less than -3 for each gestational age (for both cohorts)

Gestational age (in weeks)	**Not flagged for |z| > 3** (%)	**Flagged for |z| > 3** (%)
32	30 (71.43)	12 (28.57)
33	4 (66.67)	2 (33.33)
34	20 (100.00)	0 (0.00)
35	18 (94.74)	1 (5.26)
36	58 (96.67)	2 (3.33)
37	67 (94.37)	4 (5.63)
38	178 (92.71)	14 (7.29)
39	195 (93.75)	13 (6.25)
40	418 (94.57)	24 (5.43)
41	215 (96.85)	7 (3.15)
42	126 (99.21)	1 (0.79)
Total	*1,329 (94.32)*	*80 (5.68)*

* Per cent refers to the sample of 1,409 children with both birth weight and gestational age recorded.

***Appendix K**: Quantile-normal plots of height-, weight-, and BMI-for-age z-scores*

Figure 37: Quantiles of height-for-age z-scores plotted against a normal distribution, for each wave and cohort

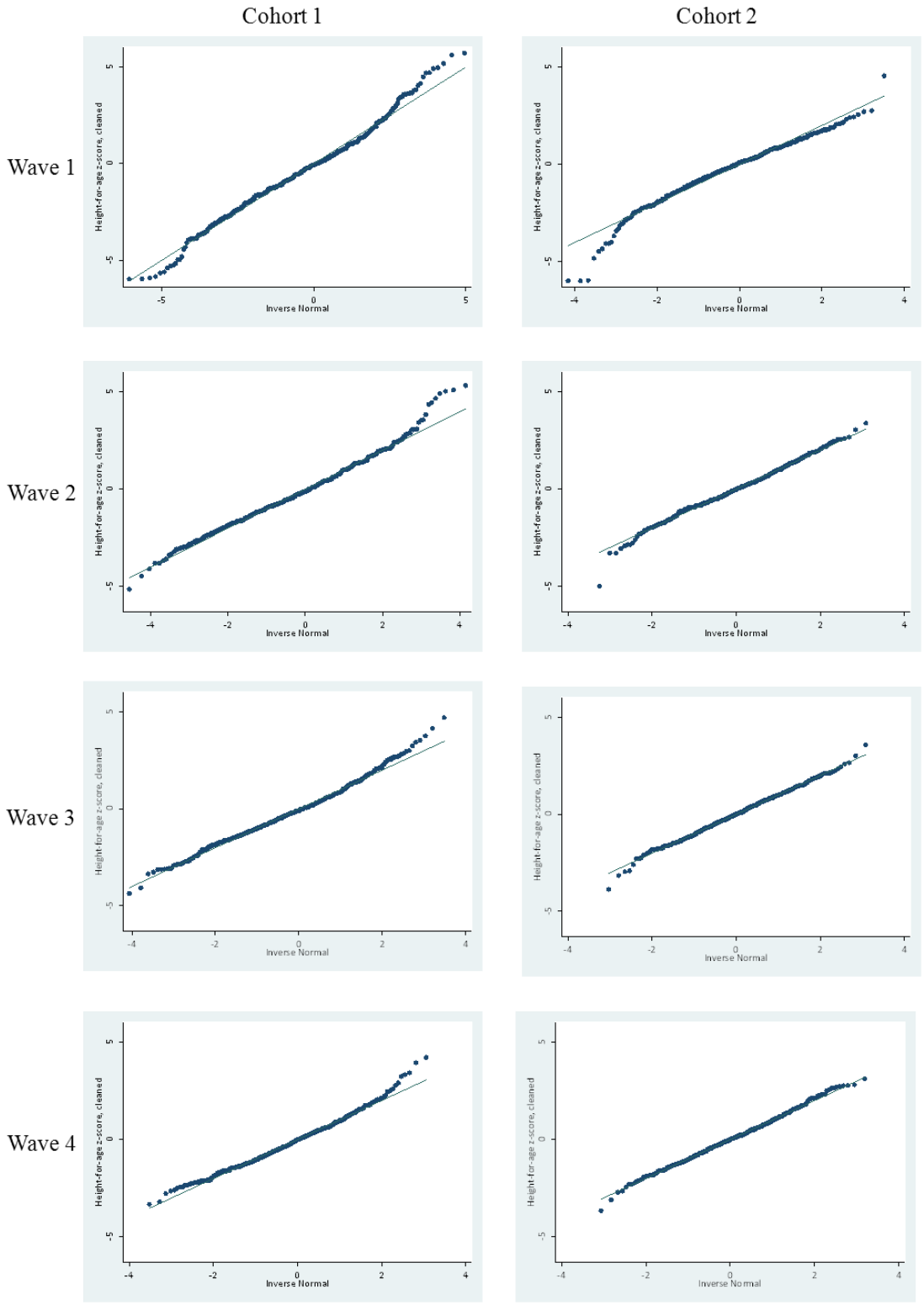


Figure 38: Quantiles of weight-for-age z-scores plotted against a normal distribution, for each wave and cohort

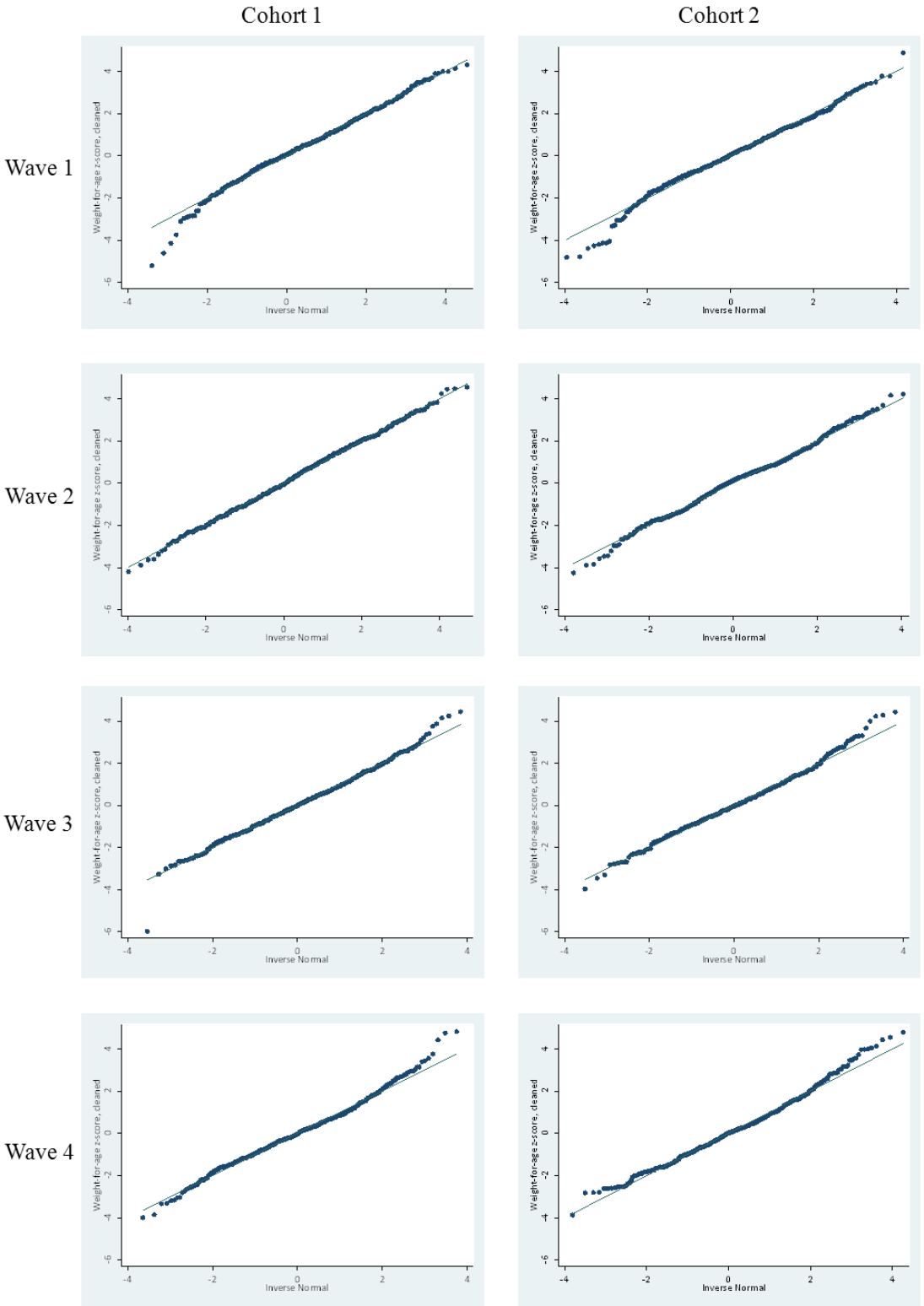
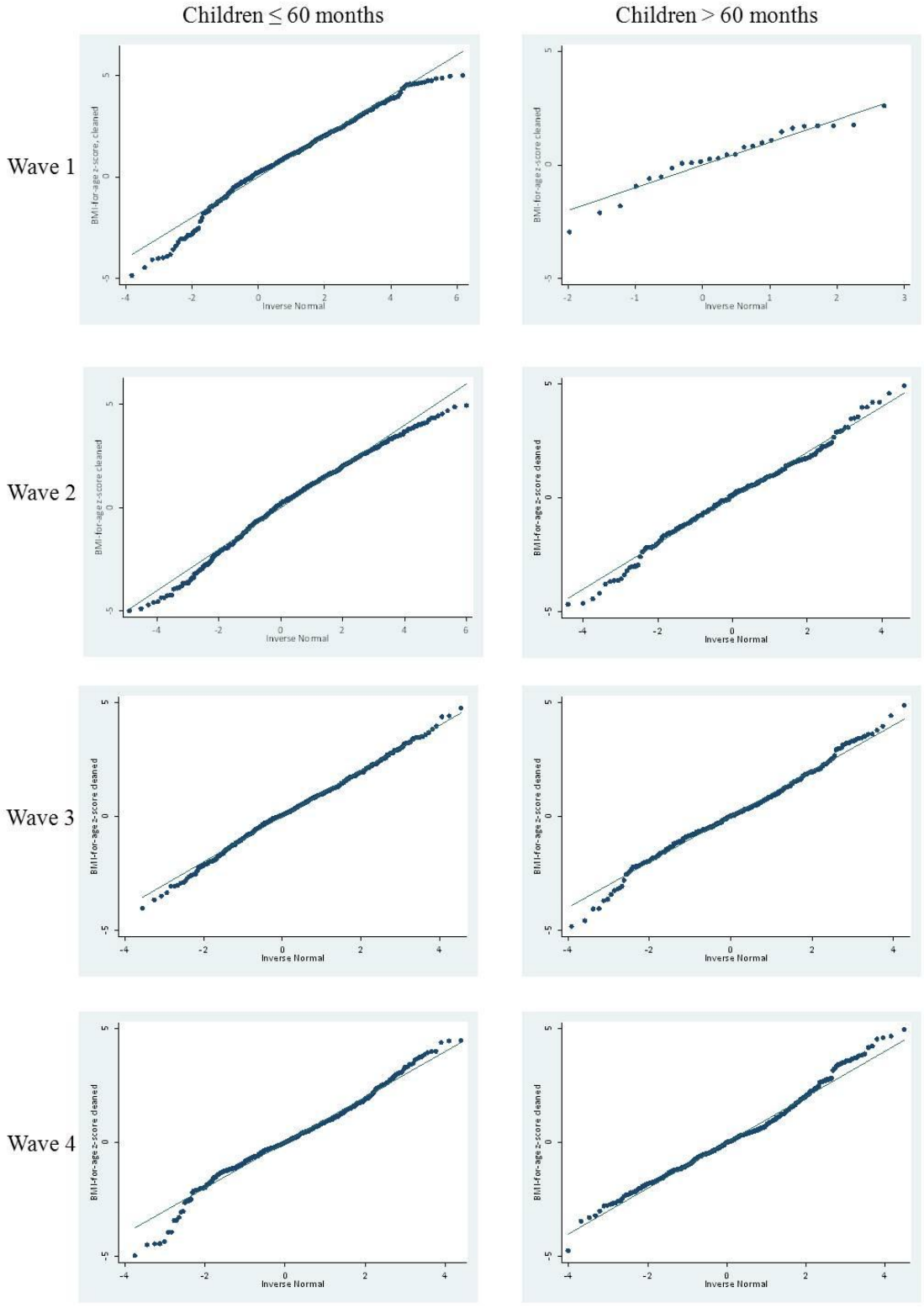


Figure 39: Quantiles of BMI-for-age z-scores plotted against a normal distribution, for each wave and age group



Appendix L: Biases in missing and implausible birth weight data

The presence of missing or implausible birth weight z-scores was significantly associated with whether the P1 was the birth mother or not, whether the birth weight was reported from the Baby Book or reported from memory, and urban versus rural location (see Table 70). There was not a significant association between the presence of missing or implausible birth weight z-scores and gender or gestational age. The mean BMI-for-age z-score at each wave was lower for children who were missing birth weight z-scores compared to those who were not missing birth weight z-scores; the difference was significant for each cohort at each wave, with the exception of the younger cohort at Wave 1 (see Table 71 and Table 72). After stratifying by LORI, the difference in the mean BMI-for-age z-score persisted for children with and without birth weight z-scores (see Table 73). For the younger cohort, however, the difference was no longer significant for most groups (see Table 74).

Table 70: Factors associated with missing and implausible birth weight

	Birth weight missing or implausible (%)	**Birth weight included in analyses** (%)
Total	*455* *(25.87)*	*1,304* *(74.13)*
Birth mother? (n = 1,759)		
No	89 (59.33)	61 (40.67)
Yes	366 (22.75)	1,243 (77.25)
Reported from Baby Health Book? (for n = 1,654 who answered questions about Baby Book)		
No	162 (38.94)	254 (61.06)
Yes	189 (15.27)	1,049 (84.73)
Gender? (n = 1,759)		
Male	216 (24.35)	671 (75.65)
Female	239 (27.41)	633 (72.59)
Gestational age at either extreme end of distribution (<32 weeks or >42 weeks)**?** (for n = 1,619 with gestational age recorded)		
No	275 (19.28)	1,151 (80.72)
Yes	40 (20.73)	153 (79.27)
Level of Relative Isolation? (for n = 1,671 with LORI recorded at Wave 1)		
Rural	167	242

	Birth weight missing or implausible (%)	Birth weight included in analyses (%)
(Moderate or High/Extreme LORI)	(40.83)	(58.17)
Urban (No or Low LORI)	235 (18.62)	1,027 (81.38)

* Denotes significant association.

Table 71: Difference in BMI-for-age z-scores for Cohort K children with missing versus recorded birth weight z-scores, by wave (two sample t-test with equal variance)

Wave	T-statistic	Number of observations	P-value (two-sided)
1	-4.49	451	< 0.0001*
2	-6.32	530	< 0.0001*
3	-4.26	493	< 0.0001*
4	-4.98	494	< 0.0001*

* Denotes significant difference between groups.

Table 72: Difference in BMI-for-age z-scores for Cohort B children with missing versus recorded birth weight z-scores, by wave (two sample t-test with equal variance)

Wave	T-statistic	Number of observations	P-value (two-sided)
1	-0.62	545	0.5330
2	-3.12	677	0.0019*
3	-2.38	656	0.0176*
4	-1.99	676	0.0468*

* Denotes significant difference between groups.

Table 73: Difference in BMI-for-age z-scores for Cohort K children with plausible versus missing birth weight z-scores, by wave and urban versus rural environment (two sample t-test with equal variance)

	Urban (No or Low LORI)			Remote (Moderate or High/Extreme LORI)		
Wave	T-statistic	Number of observations	P-value (two-sided)	T-statistic	Number of observations	P-value (two-sided)
1	-3.04	355	0.0025*	-2.96	96	0.0039*
2	-3.38	393	0.0008*	-3.59	135	0.0005*
3	-1.15	384	0.2489	-3.48	107	0.0007*
4	-1.67	345	0.0950	-4.74	95	< 0.0001*

* Denotes significant difference between groups.

Table 74: Difference in BMI-for-age z-scores for Cohort B children with plausible versus missing birth weight z-scores, by wave and urban versus rural environment (two sample t-test with equal variance)

	Urban (No or Low LORI)			Remote (Moderate or High/Extreme LORI)		
Wave	T-statistic	Number of observations	P-value (two-sided)	T-statistic	Number of observations	P-value (two-sided)

1	0.12	425	0.9037	-0.16	120	0.8714
2	-0.64	500	0.5222	-2.20	177	0.0294*
3	-0.10	497	0.9197	-1.69	159	0.0930
4	-1.19	472	0.2356	0.33	141	0.7422

* Denotes significant difference between groups.

There were several significant differences in the characteristics of children with and without plausible birth weight z-scores recorded (see Table 75). There was an association between Indigenous identity and the prevalence of plausible birth weight z-scores, with the highest percentage of missing birth weight z-scores among Aboriginal, compared to Torres Strait Islander or Aboriginal and Torres Strait Islander, children (Pearson chi2(2) = 10.14, p = 0.006). There was also a significant association with Cohort, with children in the Baby Cohort significantly more likely to have a plausible birth weight z-score recorded (Pearson chi2(1) = 5.48, p = 0.019). Children whose P1 reported a high weekly income or a high level of education attained were more likely to have a plausible birth weight z-score recorded (Pearson chi2(6) = 80.01, p < 0.001 and Pearson chi2(13) = 75.51, p < 0.001, respectively). There was also a significantly older mean age of breast feeding cessation for infants with missing versus non-missing birth weight z-scores (t(1,613) = 4.73, two-tailed p < 0.0001 for two-sample t-test with equal variances). The prevalence of a plausible birth weight z-score was independent of gender, P1's report of the child's general health at Wave 1, and maternal report of cigarette use during pregnancy. These differences were considered in analyses.

Table 75: Characteristics of children (both cohorts) without and with plausible birth weight z-score recorded

	Children without a plausible birth weight z-score recorded	**Children with plausible birth weight z-score recorded**
*BMI-for-age z-score**		
Mean BMI-for-age z-score at Wave 1 (n)	0.54 (216)	1.00 (780)
Mean BMI-for-age z-score at Wave 2 (n)	-0.09 (318)	0.65 (889)
Mean BMI-for-age z-score at Wave 3 (n)	0.04 (306)	0.48 (843)
Mean BMI-for-age z-score at Wave 4 (n)	-0.06 (291)	0.40 (879)
Gender		

	Children without a plausible birth weight z-score recorded	Children with plausible birth weight z-score recorded
Male (n = 887)	24.35%	75.65%
Female (n = 872)	27.41%	72.59%
*Indigenous identity**		
Aboriginal (n = 1,535)	27.04%	72.96%
Torres Strait Islander (n = 118)	14.41%	85.59%
Aboriginal and Torres Strait Islander (n = 106)	21.70%	78.30%
*Cohort**		
Baby Cohort (n = 1,010)	23.76%	76.24%
Child Cohort (n = 749)	28.70%	71.30%
P1's report of child's general health at Wave 1		
Excellent (n = 753)	22.05%	77.95%
Very good (n = 522)	25.10%	74.90%
Good (n = 340)	26.47%	73.53%
Fair (n = 41)	21.95%	78.05%
Very Poor (n = 6)	33.33%	66.67%
*P1's report of weekly income at Wave 1**		
Less than $150 a week (n = 156)	25.00%	75.00%
$150 - $249 a week (n = 226)	39.82%	60.18%
$250 - $399 a week (n = 282)	27.30%	72.70%
$400 - $599 a week (n = 333)	21.02%	78.98%
$600 - $799 a week (n = 218)	12.84%	87.16%
$800 - $999 a week (n = 139)	14.39%	85.61%
$1,000 or more a week (n = 209)	10.05%	89.95%
*P1's highest educational qualification attained**		
Never attended school (n = 10)	80.00%	20.00%
Year 8 or below (n = 76)	51.32%	48.68%

	Children without a plausible birth weight z-score recorded	Children with plausible birth weight z-score recorded
Year 9 or equivalent (n = 157)	32.48%	67.52%
Year 10 or equivalent (n = 408)	24.75%	75.25%
Year 11 or equivalent (n = 248)	31.85%	68.15%
Year 12 or equivalent (n = 249)	21.29%	78.71%
Certificate of completion (n = 5)	0.00%	100.00%
Other non-school qualification (n = 6)	16.67%	83.33%
Certificate I/II (n = 67)	14.93%	85.07%
Certificate III/IV (including trade certificate) (n = 154)	21.43%	78.57%
Advanced diploma (n = 47)	14.89%	85.11%
Bachelor degree (with or without honours) (n = 62)	9.68%	90.32%
Graduate diploma/Graduate certificate (n = 16)	6.25%	93.75%
Postgraduate degree (n = 13)	15.38%	84.62%
*Age study child stopped breast feeding (days)**		
Mean a*ge study child stopped breastfeeding (days*) (n)	310.57 (381)	227.61 (1,232)
Birth mother cigarette use during pregnancy		
Yes (n = 777)	165 (21.24%)	612 (78.76%)
No (n = 777)	146 (18.79%)	631 (81.21%)

* Indicates a significant association between the variable and the prevalence of missing or implausible birth weight z-scores.

***Appendix M**: Mean age, weight, height, BMI, and z-scores across waves, after data cleaning*

Age:

Table 76: Distribution of age at BMI measurement across waves: Cohort B

	Wave 1	**Wave 2**	**Wave 3**	**Wave 4**
n	700	775	741	721
Mean (months)	15.73	25.80	38.00	49.47
Standard deviation (months)	5.27	5.83	6.14	5.82
25th percentile (months)	12	21	33	45
Median (months)	16	25	38	49
75th percentile (months)	19	30	42	53
Minimum (months)	3	13	24	33
Maximum (months)	33	46	58	69

Table 77: Distribution of age at BMI measurement across waves: Cohort K

	Wave 1	**Wave 2**	**Wave 3**	**Wave 4**
n	537	599	540	517
Mean (months)	51.31	60.71	72.99	84.56
Standard deviation (months)	5.52	5.77	6.05	6.01
25th percentile (months)	48	57	69	80
Median (months)	51	61	73	84
75th percentile (months)	55	65	77	89
Minimum (months)	34	42	51	64
Maximum (months)	72	79	96	106

Table 78: Distribution of age at BMI measurement across waves: both cohorts

	Wave 1	**Wave 2**	**Wave 3**	**Wave 4**
n	1,237	1,374	1,281	1,238
Mean (months)	31.18	41.02	52.75	64.13
Standard deviation (months)	18.44	18.26	18.33	18.30
25th percentile (months)	15	24	37	48
Median (months)	22	33	44	56
75th percentile (months)	50	60	72	83
Minimum (months)	3	13	24	33
Maximum (months)	72	79	96	106

Height:

Table 79: Distribution of height across waves, after data cleaning: Cohort B

	Wave 1	**Wave 2**	**Wave 3**	**Wave 4**
n	587	706	670	680
Mean (cm)	77.10	87.13	95.59	102.80
Standard deviation (cm)	6.81	6.21	5.80	5.68
25th percentile (cm)	72	83	92	99
Median (cm)	77	87	96	103
75th percentile (cm)	82	91	99	106
Minimum (cm)	59	70	80	85
Maximum (cm)	97	110	119	125

Table 80: Distribution of height across waves, after data cleaning: Cohort K

	Wave 1	**Wave 2**	**Wave 3**	**Wave 4**
n	468	543	497	494
Mean (cm)	103.39	109.41	116.10	121.84
Standard deviation (cm)	6.43	5.81	5.98	6.49
25th percentile (cm)	100	106	112	118
Median (cm)	104	110	116	122
75th percentile (cm)	108	112	120	126
Minimum (cm)	74	85	95	106
Maximum (cm)	120	127	137	141

Table 81: Distribution of height across waves, after data cleaning: both cohorts

	Wave 1	**Wave 2**	**Wave 3**	**Wave 4**
n	1,055	1,249	1,167	1,174
Mean (cm)	88.8	96.8	104.3	110.8
Standard deviation (cm)	14.7	12.6	11.7	11.2
25th percentile (cm)	76	86	95	102
Median (cm)	85	94	102	109
75th percentile (cm)	103	109	115	120
Minimum (cm)	59	70	80	85
Maximum (cm)	120	127	137	141

Table 82: Distribution of height-for-age z-scores across waves, after data cleaning: Cohort B

	Wave 1	**Wave 2**	**Wave 3**	**Wave 4**
n	587	706	670	680
Mean	-0.56	-0.20	-0.29	-0.23
Standard deviation	1.88	1.46	1.27	1.11
25th percentile	-1.63	-1.14	-1.14	-1.03
Median	-0.57	-0.26	-0.34	-0.31
75th percentile	0.50	0.66	0.47	0.49
Minimum	-5.99	-5.14	-4.36	-3.33
Maximum	5.72	5.33	4.70	4.22
Low height-for-age	115	65	46	33

	Wave 1	Wave 2	Wave 3	Wave 4
(z < -2)	(19.59%)	(9.21%)	(6.87%)	(4.85%)
Normal height-for-age (-2 ≤ z ≤ +2)	424 (72.23%)	593 (83.99%)	592 (88.36%)	627 (92.21%)
High height-for-age (z > +2)	48 (8.18%)	48 (6.80%)	32 (4.78%)	20 (2.94%)

Table 83: Distribution of height-for-age z-scores across waves, after data cleaning: Cohort K

	Wave 1	Wave 2	Wave 3	Wave 4
n	468	543	497	494
Mean	-0.33	-0.08	0.03	0.06
Standard deviation	1.35	1.09	1.06	1.09
25^{th} percentile	-1.06	-0.80	-0.68	-0.67
Median	-0.24	-0.08	0.01	0.06
75^{th} percentile	0.56	0.58	0.76	0.76
Minimum	-5.99	-4.99	-3.86	-3.37
Maximum	4.55	3.39	3.59	3.13
Low height-for-age (z < -2)	47 (10.04%)	19 (3.50%)	11 (2.21%)	12 (2.43%)
Normal height-for-age (-2 ≤ z ≤ 2)	410 (87.61%)	508 (93.55%)	472 (94.97%)	458 (92.71%)
High height-for-age (z > +2)	11 (2.35%)	16 (2.95%)	14 (2.82%)	24 (4.86%)

Table 84: Distribution of height-for-age z-scores across waves, after data cleaning: both cohorts

	Wave 1	Wave 2	Wave 3	Wave 4
n	1,055	1,249	1,167	1,174
Mean	-0.46	-0.15	-0.15	-0.11
Standard deviation	1.67	1.31	1.20	1.11
25^{th} percentile	-1.43	-0.93	-0.97	-0.91
Median	-0.38	-0.19	-0.17	-0.13
75^{th} percentile	0.53	0.64	0.62	0.59
Minimum	-5.99	-5.14	-4.36	-3.67
Maximum	5.72	5.33	4.70	4.22
Low height-for-age (z < -2)	162 (15.36%)	84 (6.73%)	57 (4.88%)	45 (3.83%)
Normal height-for-age (-2 ≤ z ≤ +2)	834 (79.05%)	1,101 (88.15%)	1,064 (91.17%)	1,085 (92.42%)
High height-for-age (z > +2)	59 (5.99%)	64 (5.12%)	46 (3.94%)	44 (3.75%)

Weight:

Table 85: Distribution of weight across waves, after data cleaning: Cohort B

	Wave 1	**Wave 2**	**Wave 3**	**Wave 4**
n	645	738	724	699
Mean (kilograms)	10.86	12.84	12.91	16.87
Standard deviation (kilograms)	1.96	2.39	2.50	3.12
25th percentile (kilograms)	9.8	11.0	13.2	15.0
Median (kilograms)	10.6	12.8	14.8	16.4
75th percentile (kilograms)	12.0	14.4	16.2	18.4
Minimum (kilograms)	5.0	7.0	6.0	10.0
Maximum (kilograms)	17.4	24.0	25.0	34.5

Table 86: Distribution of weight across waves, after data cleaning: Cohort K

	Wave 1	**Wave 2**	**Wave 3**	**Wave 4**
n	502	581	531	513
Mean (kilograms)	17.37	19.13	21.47	24.39
Standard deviation (kilograms)	3.38	3.82	4.31	5.93
25th percentile (kilograms)	15.0	16.3	18.6	20.5
Median (kilograms)	17.0	19.0	21.0	23.0
75th percentile (kilograms)	19.5	21.0	23.3	26.6
Minimum (kilograms)	9.0	9.0	12.0	13.8
Maximum (kilograms)	31.0	39.0	45.8	51.0

Table 87: Distribution of weight across waves, after data cleaning: both cohorts

Weight (both cohorts together)				
	Wave 1	**Wave 2**	**Wave 3**	**Wave 4**
n	1,147	1,319	1,255	1,212
Mean (kilograms)	13.7	15.6	17.7	20.1
Standard deviation (kilograms)	4.2	4.4	4.7	5.9
25th percentile (kilograms)	10.1	12.0	14.4	16.0
Median (kilograms)	13.0	15.0	16.8	18.6
75th percentile (kilograms)	16.4	19.0	20.4	22.8
Minimum (kilograms)	5.0	7.0	6.0	10.0
Maximum (kilograms)	31.0	39.0	45.8	51.0

Table 88: Distribution of weight-for-age z-scores across waves, after data cleaning: Cohort B

	Wave 1	**Wave 2**	**Wave 3**	**Wave 4**
n	645	738	724	699
Mean	0.57	0.36	0.15	0.05
Standard deviation	1.34	1.45	1.23	1.24
25th percentile	-0.26	-0.65	-0.68	-0.76
Median	0.56	0.40	0.16	0.06
75th percentile	1.42	1.41	0.91	0.75

	Wave 1	Wave 2	Wave 3	Wave 4
Minimum	-5.21	-4.2	-5.98	-3.99
Maximum	4.28	4.54	4.45	4.81
Low weight-for-age ($z < -2$)	19 (2.95%)	38 (5.15%)	26 (3.59%)	28 (4.01%)
Normal weight-for-age ($-2 \leq z \leq +2$)	539 (83.57%)	603 (81.71%)	653 (90.19%)	627 (89.70%)
High weight-for-age ($z > +2$)	87 (13.49%)	97 (13.14%)	45 (6.22%)	44 (6.29%)

Table 89: Distribution of weight-for-age z-scores across waves, after data cleaning: Cohort K

	Wave 1	Wave 2	Wave 3	Wave 4
n	502	581	531	513
Mean	0.11	0.13	0.15	0.23
Standard deviation	1.41	1.34	1.26	1.40
25^{th} percentile	-0.75	-0.76	-0.67	-0.67
Median	0.15	0.24	0.10	0.18
75^{th} percentile	1.01	0.89	0.92	1.03
Minimum	-4.81	-4.26	-3.96	-3.88
Maximum	4.87	4.22	4.44	4.77
Low weight-for-age ($z < -2$)	31 (6.18%)	28 (4.82%)	25 (4.71%)	18 (3.51%)
Normal weight-for-age ($-2 \leq z \leq +2$)	431 (85.86%)	509 (87.61%)	471 (88.70%)	442 (86.16%)
High weight-for-age ($z > +2$)	40 (7.97%)	44 (7.57%)	35 (6.59%)	53 (10.33%)

Table 90: Distribution of weight-for-age z-scores across waves, after data cleaning: both cohorts

	Wave 1	Wave 2	Wave 3	Wave 4
n	1,147	1,319	1,255	1,212
Mean	0.37	0.26	0.15	0.13
Standard deviation	1.39	1.41	1.25	1.31
25^{th} percentile	-0.47	-0.70	-0.68	-0.73
Median	0.40	0.30	0.12	0.12
75^{th} percentile	1.27	1.17	0.92	0.87
Minimum	-5.21	-4.26	-5.98	-3.99
Maximum	4.87	4.54	4.45	4.81
Low weight-for-age ($z < -2$)	50 (4.36%)	66 (5.00%)	54 (4.06%)	46 (3.80%)
Normal weight-for-age ($-2 \leq z \leq +2$)	970 (84.57%)	1,112 (84.31%)	1,124 (89.56%)	1,069 (88.20%)
High weight-for-age ($z > +2$)	127 (11.07%)	141 (10.69%)	80 (6.37%)	97 (8.00%)

BMI:

Table 91: Distribution of BMI (kg/m^2) across waves, after data cleaning: Cohort B

	Wave 1	**Wave 2**	**Wave 3**	**Wave 4**
n	545	677	656	676
Mean (cm)	18.19	16.84	16.28	15.88
Standard deviation (cm)	2.66	2.47	1.89	1.99
25th percentile (cm)	16.40	15.15	15.10	14.66
Median (cm)	18.08	16.87	16.19	15.66
75th percentile (cm)	20.00	18.50	17.42	16.83
Minimum (cm)	10.94	10.33	11.30	10.16
Maximum (cm)	25.20	24.09	22.97	23.44

Table 92: Distribution of BMI (kg/m^2) across waves, after data cleaning: Cohort K

	Wave 1	**Wave 2**	**Wave 3**	**Wave 4**
n	451	530	493	494
Mean (cm)	16.30	15.91	15.89	16.33
Standard deviation (cm)	2.31	2.61	2.37	2.85
25th percentile (cm)	14.99	14.14	14.44	14.50
Median (cm)	15.95	15.72	15.54	15.78
75th percentile (cm)	17.34	17.47	16.93	17.35
Minimum (cm)	10.52	9.97	10.49	10.67
Maximum (cm)	25.00	27.21	27.52	31.07

Table 93: Distribution of BMI (kg/m^2) across waves, after data cleaning: both cohorts**.**

	Wave 1	**Wave 2**	**Wave 3**	**Wave 4**
n	996	1,207	1,149	1,170
Mean (cm)	17.33	16.43	16.12	16.07
Standard deviation (cm)	2.68	2.57	2.11	2.40
25th percentile (cm)	15.47	14.61	14.71	14.61
Median (cm)	16395	16.36	15.91	15.71
75th percentile (cm)	18.83	17.99	17.23	17.02
Minimum (cm)	10.52	9.97	10.49	10.16
Maximum (cm)	25.20	27.21	27.52	31.07

Table 94: Distribution of BMI-for-age z-scores across waves, after data cleaning: children under five years of age

	Wave 1	**Wave 2**	**Wave 3**	**Wave 4**
n	971	942	666	652
Mean	0.92	0.56	0.50	0.32
Standard deviation	1.66	1.77	1.36	1.37
25th percentile	-0.07	-0.55	-0.28	-0.46
Median	0.89	0.70	0.54	0.28
75th percentile	1.96	1.74	1.33	1.10
Minimum	-4.84	-4.99	-4.02	-4.97
Maximum	5.00	4.96	4.75	4.47
Low BMI-for-age (z < -2)	40 (4.12%)	79 (8.39%)	27 (4.05%)	26 (3.99%)

	Wave 1	Wave 2	Wave 3	Wave 4
Normal BMI-for-age ($-2 \leq z \leq +2$)	697 (71.78%)	668 (70.91%)	558 (83.78%)	562 (86.20%)
Overweight BMI-for-age ($+2 < z \leq +3$)	126 (12.98%)	133 (14.12%)	59 (8.86%)	41 (6.29%)
Obese BMI-for-age ($z > +3$)	108 (11.12%)	62 (6.58%)	22 (3.30%)	23 (3.53%)

Table 95: Distribution of BMI-for-age z-scores across waves, after data cleaning: children over five years of age

	Wave 1	Wave 2	Wave 3	Wave 4
n	25	265	483	518
Mean	0.36	0.10	0.18	0.24
Standard deviation	1.32	1.68	1.43	1.47
25^{th} percentile	-0.14	-0.94	-0.66	-0.69
Median	0.45	0.27	0.15	0.19
75^{th} percentile	1.46	1.08	0.98	1.02
Minimum	-2.94	-4.67	-4.82	-4.73
Maximum	2.61	4.91	4.88	4.96
Low BMI-for-age ($z < -2$)	2 (8.00%)	28 (10.57%)	26 (5.38%)	26 (5.02%)
Normal BMI-for-age ($-2 \leq z \leq +1$)	15 (60.00%)	164 (61.89%)	338 (69.98%)	361 (69.69%)
Overweight BMI-for-age ($+1 < z \leq +2$)	7 (28.00%)	47 (17.74%)	75 (15.53%)	71 (13.71%)
Obese BMI-for-age ($z > +2$)	1 (4.00%)	26 (9.81%)	44 (9.11%)	60 (11.58%)

Table 96: Distribution of BMI-for-age z-scores across waves, after data cleaning: both cohorts

	Wave 1	Wave 2	Wave 3	Wave 4
n	996	1,207	1,149	1,170
Mean	0.90	0.45	0.36	0.29
Standard deviation	1.66	1.76	1.40	1.42
25^{th} percentile	-0.08	-0.62	-0.5	-0.59
Median	0.89	0.56	0.35	0.24
75^{th} percentile	1.94	1.66	1.24	1.08
Minimum	-4.84	-4.99	-4.82	-4.97
Maximum	5.00	4.96	4.88	4.96
Low BMI-for-age ($z < -2$)	42 (4.22%)	107 (8.86%)	53 (4.61%)	52 (4.44%)
Normal BMI-for-age ($-2 \leq z \leq +2$ for children under 5; $-2 \leq z \leq +1$ for children over 5)	712 (71.49%)	832 (68.93%)	896 (77.98%)	923 (78.89%)
Overweight BMI-for-age ($+2 < z \leq +3$ for children under 5; $+1 < z \leq +2$ for children over 5)	133 (13.35%)	180 (14.91%)	134 (11.66%)	112 (9.57%)

	Wave 1	Wave 2	Wave 3	Wave 4
Obese BMI-for-age (z > +3 for children under 5; z > +2 for children over 5)	109 (10.94%)	88 (7.29%)	66 (5.74%)	83 (7.09%)

Appendix N: Distribution of birth weight and gestational age in LSIC

Table 97: Distribution of birth weight in LSIC (for both cohorts)

n	1,315
Mean (grams)	3,296.49
Standard deviation (grams)	622.92
25th percentile (grams)	2,948.348
Median (grams)	3,316.892
75th percentile (grams)	3,713.784
Minimum (grams)	907.18
Maximum (grams)	4,819.42
Low birth weight (less than 2,500 grams)	134 (10.19%)
Normal birth weight (between 2,500 and 4,000 grams)	1,034 (78.63%)
High birth weight (more than 4,000 grams)	147 (11.18%)

Table 98: Distribution of gestational age in LSIC (for both cohorts)

n	1,619
Mean (weeks)	39.143
Standard deviation (weeks)	2.184
25th percentile (weeks)	38
Median (weeks)	40
75th percentile (weeks)	40
Minimum (weeks)	32
Maximum (weeks)	42
Pre-term (< 37 weeks)	175 (10.81%)
Term (between 37 and 41 weeks)	1,307 (80.73%)
Post-term (≥ 42 weeks)	137 (8.46%)

Table 99: Distribution of birth weight z-scores in LSIC (for both cohorts)

n	1,304
Mean	-0.224
Standard deviation	1.131
25th percentile	-1.025
Median	-0.288
75th percentile	0.488
Minimum	-2.995
Maximum	2.964
Small for gestational age ($z < -1.28$)	233 (17.87%)
Appropriate for gestational age ($-1.28 \leq z \leq +1.28$)	947 (72.62%)
Large for gestational age ($z > +1.28$)	124 (9.51%)

Bibliography

1. Sanson-Fisher RW, Campbell EM, Perkins JJ, Blunden SV, Davis BB. Indigenous health research: a critical review of outputs over time. Med J Aust. 2006; **184**(10):502-5.
2. Pyett P. Towards reconciliation in Indigenous health research: The responsibilities of the non-Indigenous researcher. Contemp Nurse. 2002; **14**(1):56-65.
3. Grove N, Brough M, Canuto C, Dobson A. Aboriginal and Torres Strait Islander health research and the conduct of longitudinal studies: issues for debate. Aust N Z J Public Health. 2003; **27**(6):637-41.
4. Priest N, Mackean T, Waters E, Davis E, Riggs E. Indigenous child health research: a critical analysis of Australian studies. Aust N Z J Public Health. 2009; **33**(1):55-63.
5. Cole TJ, Flegal KM, Nicholls D, Jackson AA. Body mass index cut offs to define thinness in children and adolescents: international survey. BMJ. 2007; **335**(7612):194.
6. Cole TJ, Bellizzi MC, Flegal KM, Dietz WH. Establishing a standard definition for child overweight and obesity worldwide: international survey. BMJ. 2000; **320**(7244):1240.
7. The Longitudinal Study of Indigenous Children: Key summary report from Wave 3. Australian Government Department of Families, Housing, Community Services and Indigenous Affairs; 2012.
8. The Longitudinal Study of Indigenous Children: An Australian Government initiative - Data User Guide 3.1. Australian Government Department of Families, Housing, Community Services and Indigenous Affairs 2012.
9. Gracey MS. Nutrition-related disorders in Indigenous Australians: how things have changed. Med J Aust. 2007; **186**(1):15-7.
10. Hoy WE, Rees M, Kile E, Mathews JD, Wang Z. A new dimension to the Barker hypothesis: Low birthweight and susceptibility to renal disease. Kidney Int. 1999; **56**:1072-7.
11. Ong KK, Dunger DB. Birth weight, infant growth and insulin resistance. Eur J Endocrinol. 2004; **151**(Suppl 3):U131-U9.
12. Singh GR, Hoy WE. The association between birthweight and current blood pressure: a cross-sectional study in an Australian Aboriginal community. Med J Aust. 2003; **179**:532-5.
13. Hediger ML, Overpeck MD, Kuczmarski RJ, McGlynn A, Maurer KR, Davis WW. Muscularity and fatness of infants and young children born small-or large-for-gestational-age. Pediatrics. 1998; **102**(5):e60.
14. Ong K, Petry C, Emmett P, et al. Insulin sensitivity and secretion in normal children related to size at birth, postnatal growth, and plasma insulin-like growth factor-I levels. Diabetologia. 2004; **47**(6):1064-70.
15. Garnett S, Cowell C, Baur L, et al. Abdominal fat and birth size in healthy prepubertal children. Int J Obes. 2001; **25**(11):1667-73.
16. Kondalsamy-Chenakesavan S, Hoy WE, Wang Z, et al. Anthropometric measurements of Australian Aboriginal adults living in remote areas: Comparison with nationally representative findings. American Journal of Human Biology. 2008; **20**:317-24.

17. Dobbins TA, Sullivan EA, Roberts CL, Simpson JM. Australian national birthweight percentiles by sex and gestational age, 1998-2007. Med J Aust. 2012; **197**(5):291-4.
18. Lake P. A decade of Aboriginal health research. Aboriginal Health Information Bulletin. 1992; **17**:12-6.
19. Humphery K. Indigenous health and 'Western research'. Discussion Paper: VicHealth Koori Health Research & Community development Unit; 2000.
20. Derrick GE, Hayen A, Chapman S, Haynes AS, Webster BM, Anderson I. A bibliometric analysis of research on Indigenous health in Australia, 1972–2008. Aust N Z J Public Health. **36**(3):269-73.
21. Monasta L, Andersson N, Ledogar RJ, Cockcroft A. Minority health and small numbers epidemiology: a case study of living conditions and the health of children in 5 foreign Roma camps in Italy. Journal Information. 2008; **98**(11).
22. Brough M. Healthy Imaginations: A social history of the epidemiology of Aboriginal and Torres Strait Islander health. Med Anthropol. 2001; **20**(1):65-90.
23. Pyett P, Waples-Crowe P, Van Der Sterren A. Engaging with Aboriginal communities in an urban context: some practical suggestions for public health researchers. Aust N Z J Public Health. 2009; **33**(1):51-4.
24. Nicholson JM, Rempel LA. Australian and New Zealand birth cohort studies: Breadth, quality and contributions. J Paediatr Child Health. 2004; **40**(3):87-95.
25. Curtin L, Feinleib M. Considerations in the design of longitudinal surveys of health. Oxford University Press, New York; 1992. p. 49-67.
26. Young A, Powers J, Wheway V. Working with longitudinal data: Attrition and retention, data quality, measures of change and other analytical issues. International Journal of Multiple Research Approaches. 2007; **1**(2):175-87.
27. Farrington DP. Longitudinal research strategies: Advantages, problems, and prospects. J Am Acad Child Adolesc Psychiatry. 1991; **30**(3):369-74.
28. McKenzie M, Tulsky JP, Long HL, Chesney M, Moss A. Tracking and follow-up of marginalized populations: a review. J Health Care Poor Underserved. 1999; **10**(4):409-29.
29. The Longitudinal Study of Indigenous Children: An Australian Government initiative - Data User Guide 2.0. Australian Government Department of Families, Housing, Community Services and Indigenous Affairs; 2011.
30. Denzin NK, Lincoln YS. The discipline and practice of qualitative research. The Sage handbook of qualitative research: Sage Publications, Incorporated; 2005. p. 1-28.
31. Humphery K. Dirty questions: Indigenous health and 'Western research'. Aust N Z J Public Health. 2001; **25**(3):197-202.
32. Anderson I. Ethics and health research in Aboriginal communities. Ethical intersections: health research, methods and researcher responsibility. 1996:153-64.
33. Holmes W, Stewart P, Garrow A, Anderson I, Thorpe L. Researching Aboriginal health: experience from a study of urban young people's health and well-being. Soc Sci Med. 2002; **54**(8):1267-79.
34. Koolmatrie T. Finding my ground in public health research: lessons from my Grandmother's kitchen. BMC Public Health. **11**(Suppl 5):S2.
35. Sherwood J. Do no harm: decolonising Aboriginal health research. Sydney: University of New South Wales; 2010.
36. Dunbar T, Scrimgeour M. Ethics in Indigenous Research–Connecting with Community. Journal of Bioethical Inquiry. 2006; **3**(3):179-85.
37. Christopher S, Watts V, McCormick AKHG, Young S. Building and maintaining trust in a community-based participatory research partnership. Am J Public Health. 2008; **98**(8):1398.

38. Sayers SM, Mackerras D, Singh G, Reid A. In an Aboriginal birth cohort, only child size and not birth size predicts insulin and glucose concentrations in childhood. Diabetes Researach and Clinical Practice. 2004; **65**:151-7.
39. Sayers S, Mott S, Singh G. Fetal growth restriction and 18-year growth and nutritional status: Aboriginal Birth Cohort 1987-2007. American Journal of Human Biology. 2011; **23**:417-9.
40. Sayers S, Singh G, Mackerras D, et al. Australian Aboriginal Birth Cohort study: follow-up processes at 20 years. BMC International Health and Human Rights. 2009; **9**(23).
41. Anderson W. The cultivation of whiteness: Science, health and racial destiny in Australia: Basic Books (AZ); 2003.
42. Gracey M. Historical, cultural, political, and social influences on dietary patterns and nutrition in Australian Aboriginal children. Am J Clin Nutr. 2000; **72**(suppl):1361S-7S.
43. Rasmussen M, Guo X, Wang Y, et al. An Aboriginal Australian genome reveals separate human dispersals into Asia. Science. 2011; **334**(6052):94-8.
44. Brown T, Townsend GC. Adolescent growth in height of Australian Aboriginals analysed by the Preece-Baines function: a longitudinal study. Ann Hum Biol. 1982; **9**(6):495-505.
45. Headline indicators for children's health, development and wellbeing 2011. Canberra: Australian Institute of Health and Welfare; 2011.
46. Propert DN, Edmonds R, Parsons PA. Birth weights and growth rates up to one year for full-blood and mixed-blood Australian Aboriginal children. J Paediatr Child Health. 1968; **4**(2):134-43.
47. Abbie AA. A preliminary survey of the growth pattern of central Australian Aboriginal males. Oceania. 1961; **31**(3):215-21.
48. Dugdale AE, Muller M, Alsop-Shields L. Patterns of weight growth in Aboriginal children on Queensland communities. J Paediatr Child Health. 1994; **30**(1):55-8.
49. Dugdale AE, Musgrave A, Streatfield K. The changing growth of Aboriginal children. Journal of Paediatric Child Health. 1990; **26**:192-296.
50. Maxwell GE, R. B. Nutritional state of Australian aboriginal children Am J Clin Nutr. 1969; **22**(6):716-25.
51. Rousham EK, Gracey M. Persistent growth faltering among Aboriginal infants and young children in north-west Australia: a retrospective study from 1969 to 1993. Acta Paediatrics. 1997; **86**:46-50.
52. Dugdale A, Lovell S. The effect of early growth on distribution of subcutaneous fat in aboriginal children. Ecol Food Nutr. 1981; **11**(1):53-5.
53. Gracey M, Sullivan H. Growth of Aboriginal infants in the first year of life in remote communities in north-west Australia. Ann Hum Biol. 1988; **15**(5):375-82.
54. Ruben AR. Undernutrition and obesity in Indigenous children: Epidemiology, prevention, and treatment. Pediatr Clin North Am. 2009; **56**:1285-302.
55. Li SQ, Guthridge S, d'Espaignet E, Paterson B. From infancy to young adulthood: health status in the Northern Territory, 2006. In: Services NTDoHaC, editor.: Government Printer of the Northern Territory; 2007.
56. Wang Z, Hoy W, McDonald S. Body Mass Index in Aboriginal Australians in remote communities. Aust N Z J Public Health. 2000; **24**(6):570-9.
57. Cunningham J, Mackerras D. Overweight and obesity: Indigenous Australians 1994. Occasional Paper. 1998.
58. Schultz R. Prevalences of overweight and obesity among children in remote Aboriginal communities in central Australia. Rural Remote Health. **12**(1872).

59. Mackerras DEM, Reid A, Sayers SM, Singh GR, Bucens IK, Flynn KA. Growth and morbidity in children in the Aboriginal Birth Cohort Study: the urban-remote differential. Med J Aust. 2003; **178**(2):56-60.
60. Sellers EAC, Singh GR, Sayers SM. Large waist but low body mass index: the metabolic syndrome in Australian Aboriginal children. The Journal of Pediatrics. 2008; **153**(2):222-7.
61. Rutishauser IH, McKay H. Anthropometric status and body composition in aboriginal women of the Kimberley region. Med J Aust. 1986; **23**(144):S8-S10.
62. World Health Organization. Diet, nutrition and the prevention of chronic diseases. World Health Organization, Joint and FAO Expert Consultation: 1991.
63. Leonard D, McDermott R, O'Dea K, et al. Obesity, diabetes and associated cardiovascular risk factors among Torres Strait Islander people. Aust N Z J Public Health. 2002; **26**(2):142-9.
64. Piers L, Rowley KG, Soares M, O'Dea K. Relation of adiposity and body fat distribution to body mass index in Australians of Aboriginal and European ancestry. Eur J Clin Nutr. 2003; **57**(8):956-63.
65. Hennenberg M, Shilitz A, Lambert KM. Assessment of the growth of children and the physical status of adults in two aboriginal communities in South Australia. American Journal of Human Biology. 2001; **13**:603-11.
66. Abbie A. Morphological variation in the adult Australian Aboriginal. The origin of the Australians. 1976; (6):211.
67. Heath DL, Panaretto KS. Nutrition status of primary school children in Townsville. Aust J Rural Health. 2005; **13**(5):282-9.
68. Magarey AM, Daniels LA, Boulton TJC. Prevalence of overweight and obesity in Australian children and adolescents: reassessment of 1985 and 1995 data against new standard international definitions. Med J Aust. 2001; **174**(11):561-4.
69. Vos DT. The burden of disease and injury in Aboriginal and Torres Strait Islander peoples 2003. Centre for Burden of Disease and Cost-Effectiveness, School of Population Health, University of Queensland: 2007.
70. Gracey M, Bridge E, Martin D, et al. An Aboriginal-driven program to prevent, control and manage nutrition-related "lifestyle" diseases including diabetes. Asia Pac J Clin Nutr. 2006; **15**(2):178-88.
71. Gracey M, King M. Indigenous health part 1: determinants and disease patterns. The Lancet. 2009; **374**:65-75.
72. Reyes L, Manalich R. Long-term consequences of low birth weight. Kidney Int. 2005; **68**:S107-S11.
73. Hoy WE, Mathews JD, McCredie DA, et al. The multidimensional nature of renal disease: Rates and associations of albuminuria in an Australian Aboriginal community. Kidney Int. 1998; **54**:1296-304.
74. Zimmet P, Alberti K, George MM, et al. The metabolic syndrome in children and adolescents - an IDF consensus report. Pediatric Diabetes. 2007; **8**(5):299-306.
75. Kim S, Popkin BM. Commentary: Understanding the epidemiology of overweight and obesity - a real global public health concern. Int J Epidemiol. 2006; **35**:60-7.
76. Burns J, Thomson N. Review of nutrition and growth among Indigenous peoples. Australian Indigenous Health Info Net; 2008 [23/01/2012]; Available from: http://www.healthinfonet.ecu.edu.au/health-risks/nutrition/reviews/our-review.
77. Back N, Cohen IR, Kritchevsky D, Lajtha A, Paoletti R, editors. Early nutrition and its later consequences: New opportunities. Berlin: Springer; 2005.
78. Plagemann A. A matter of insulin: developmental programming of body weight regulation. J Matern Fetal Neonatal Med. 2008; **21**(3):143-8.

79. Forsen T, Eriksson J, Tuomilehto J, Reunanen A, Osmond C, Barker D. The fetal and childhood growth of persons who develop Type 2 diabetes. Ann Intern Med. 2000; **133**(3):176-82.
80. Harder T, Rodekamp E, Schellong K, Dudenhausen JW, Plagemann A. Birth Weight and Subsequent Risk of Type 2 Diabetes: A Meta-Analysis. Am J Epidemiol. 2007 April 15, 2007; **165**(8):849-57.
81. Eriksson JG, Forsen TJ, Osmond C, Barker DJP. Pathways of infant and childhood growth that lead to type 2 diabetes. Diabetes Care. 2003; **26**(11):3006-10.
82. Shannon C. Acculturation: Aboriginal and Torres Strait Islander nutrition. Asia Pac J Clin Nutr. 2002; **11**(Suppl):S576-S8.
83. Humphrey M, Holzheimer D. Fetal growth charts for Aboriginal fetuses. Aust N Z J Obstet Gynaecol. 2000; **40**(4):388-93.
84. Rousham EK, Gracey M. Factors affecting birthweight of rural Australian Aborigines. Ann Hum Biol. 2002; **29**(4):363-72.
85. Rousham EK, Gracey M. Seasonality of low birthweight in Indigenous Australians: an increase in pre-term birth or intrauterine growth retardation? Aust N Z J Public Health. 1998; **22**(6):669-72.
86. Shaw J. Epidemiology of childhood type 2 diabetes and obesity. Pediatric Diabetes. 2007; **8**:7-15.
87. Wahlqvist ML. Dietary fat and the prevention of chronic disease. Asia Pac J Clin Nutr. 2005; **14**(4):313.
88. McDermott R. Ethics, epidemiology and the thrifty gene: biological determinism as a health hazard. Soc Sci Med. 1998; **47**(9):1189-95.
89. Roberts CL, Lancaster PAL. Australian national birthweight percentiles by gestational age. Med J Aust. 1999; **170**:114-8.
90. Figueras F, Gardosi J. Should we customize fetal growth standards? Fetal Diagn Ther. 2009; **25**(3):297-303.
91. Sayers S, Powers J. Risk factors for Aboriginal low birthweight, intrauterine growth retardation and preterm birth in the Darwin Health Region. Aust N Z J Public Health. 1997; **21**(5):524-30.
92. Smith RM, Smith PA, McKinnon M, Gracey M. Birthweights and growth of infants in five Aboriginal communities. Aust N Z J Public Health. 2000; **24**(2):124-35.
93. Panaretto KS, Lee HM, Mitchell MR, et al. Impact of a collaborative shared antenatal care program for urban Indigenous women: a prospective cohort study. Med J Aust. 2005; **182**(10):514-9.
94. Raio L, Ghezzi F, Naro ED, et al. Perinatal outcome of fetuses with a birth weight greater than 4500 g: an analysis of 3356 cases. European Journal of Obstetrics & Gynecology and Reproductive Biology. 2003; **109**(2):160-5.
95. Oral E, Cagdas A, Gezer A, Kaleli S, Aydinli K, Ocer F. Perinatal and maternal outcomes of fetal macrosomia. European Journal of Obstetrics & Gynecology and Reproductive Biology. 2001; **99**(2):167-71.
96. Bromwich P. Big babies. British Medical Journal (Clinical research ed.). 1986; **293**(6559):1387.
97. Sayers S, Singh G, Mott S, McDonnell J, Hoy W. Relationships between birthweight and biomarkers of chronic disease in childhood: Aboriginal Birth Cohort Study 1987-2001. Paediatr Perinat Epidemiol. 2009; **23**(6):548-56.
98. Beltrand J, Verkauskiene R, Nicolescu R, et al. Adaptive changes in neonatal hormonal and metabolic profiles induced by fetal growth restriction. J Clin Endocrinol Metab. 2008; **93**(10):4027-32.
99. Beltrand J, Nicolescu R, Kaguelidou F, et al. Catch-up growth following fetal growth restriction promotes rapid restoration of fat mass but without metabolic consequences at one year of age. PLoS ONE. 2009; **4**(4):e5343.

100. Jaddoe VWV, Witteman JCM. Hypotheses on the fetal origins of adult diseases: contributions of epidemiological studies. Eur J Epidemiol. 2006; **21**:91-102.
101. Golding J, Pembrey M, Jones R. ALSPAC—the Avon Longitudinal Study of Parents and Children: I. Study methodology. Paediatr Perinat Epidemiol. 2001; **15**(1):74-87.
102. Garnett S, Cowell C, Bradford D, et al. Effects of gender, body composition and birth size on IGF-I in 7-and 8-year-old children. Hormone Research in Paediatrics. 1999; **52**(5):221-9.
103. Casey PH, Kraemer HC, Bernbaum J, Yogman MW, Sells JC. Growth status and growth rates of a varied sample of low birth weight, preterm infants: a longitudinal cohort from birth to three years of age. The Journal of Pediatrics. 1991; **119**(4):599-605.
104. Sayers S, Mackerras D, Singh G. Update on the Aboriginal Birth Cohort Study. Aboriginal and Islander Health Worker Journal. 2004; **28**(2):6.
105. Stettler N, Zemel BS, Kumanyika S, Stallings VA. Infant weight gain and childhood overweight status in a multicenter, cohort study. Pediatrics. 2002; **109**(2):194.
106. Ong KKL, Ahmed ML, Emmett PM, Preece MA, Dunger DB. Association between postnatal catch-up growth and obesity in childhood: prospective cohort study. BMJ. 2000; **320**(7240):967-71.
107. Department of Families Housing Community Services and Indigenous Affairs. Footprints in Time: The Longitudinal Study of Indigenous Children - Key Summary Report from Wave 1. Canberra: Department of Families, Housing, Community Services and Indigenous Affairs2009.
108. The Longitudinal Study of Indigenous Children: An Australian Government initiative - Data User Guide, Release 3.0.: Australian Government Department of Families, Housing, Community Services and Indigenous Affairs; 2011.
109. 2006/07 Directions for Footprints in Time: A Longitudinal Study of Indigenous Children. Developing Research Options: Department of Families, Housing, Community Services and Indigenous Affiars; 2007.
110. . Proceedings of a symposium on Aboriginal early childhood: "Improving life chance of Aboriginal children: Birth to 8 years". Aboriginal & Torres Strait Islander early childhood issues in the Adelaide metropolitan region; 2006; Adelaide, South Australia: Australian Government Department of Families, Housing, Community Services and Indigenous Affiars.
111. Penman R. Occasional Paper No. 16: Aboriginal and Torres Strait Islander views on research in their communities. Australian Government Department of Families, Housing, Community Services and Indigenous Affiars; 2006.
112. Values and ethics: guidelines for ethical conduct in Aboriginal and Torres Strait Islander research. Canberra, Australia: National Health and Medical Research Council; 2003.
113. Department of Health and Aged Care: Information and Research Branch. Measuring Remoteness: Accessibility/Remoteness Index of Australia (ARIA). Occasional Papers: New Series Number 142001. p. 1-25.
114. Silburn S. Strengthening the capacity of Aboriginal Children, families and communities: Telethon Institute for Child Health Research; 2006.
115. The survey - objectives, design and process: The Western Australian Aboriginal Child Health Survey - The Health of Aboriginal People and Young People.
116. Pink B. Australian Statistical Geography Standard (ASGS): Volume 2 - Indigenous Structure. Australian Bureau of Statistics; 2011.
117. World Health Organization. Physical status: the use and interpretation of anthropometry. WHO Technical Report Series. Geneva1995.

118. Gorstein J, Sullivan K, Yip R, et al. Issues in the assessment of nutritional status using anthropometry. Bull World Health Organ. 1994; **72**(2):273-83.
119. World Health Organization. WHO Child Growth Standards, Methods and development: Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age. Geneva2006.
120. Duggan M. Anthropometry as a tool for measuring malnutrition: impact of the new WHO growth standards and reference. Annals of Tropical Paediatrics: International Child Health. **30**(1):1-17.
121. Pietrobelli A, Faith MS, Allison DB, Gallagher D, Chiumello G, Heymsfield SB. Body mass index as a measure of adiposity among children and adolescents: A validation study. J Pediatr. 1998; **132**(2).
122. Dietz W, Robinson T. Use of the body mass index (BMI) as a measure of overweight in children and adolescents. The Journal of Pediatrics. 1998; **132**(2):191.
123. Reilly J, Dorosty A, Emmett P. Identification of the obese child: adequacy of the body mass index for clinical practice and epidemiology. International journal of obesity and related metabolic disorders: journal of the International Association for the Study of Obesity. 2000; **24**(12):1623.
124. Lazarus R, Baur L, Webb K, Blyth F, Gliksman M. Recommended body mass index cutoff values for overweight screening programmes in Australian children and adolescents: Comparisons with North American values. J Paediatr Child Health. 1995; **31**(2):143-7.
125. Dietz WH, Bellizzi MC. Introduction: the use of body mass index to assess obesity in children. The American Journal of Clinical Nutrition. 1999; **70**(1):123S-5S.
126. Gracey M. Nutrition of Australian Aboriginal infants and children. Journal of Paediatric Child Health. 1991; **27**:259-71.
127. Himes JH, Dietz WH. Guidelines for overweight in adolescent preventive services: recommendations from an expert committee. The Expert Committee on Clinical Guidelines for Overweight in Adolescent Preventive Services. The American Journal of Clinical Nutrition. 1994; **59**(2):307.
128. De Onis M, Lobstein T. Defining obesity risk status in the general childhood population: Which cut-offs should we use? Int J Pediatr Obes. 2010; **5**(6):458-60.
129. Child Growth Charts in the Northern Territory: Discussion Paper. Considerations of the 2006 WHO growth standards: Discussion and recommendations. Northern Territory: Department of Health and Community Services; 2008.
130. de Onis M, Onyango A, Borghi E, Siyam A, Blossner M, Lutter C. Worldwide implementation of the WHO Child Growth Standards. Public Health Nutr. 2012:1-8.
131. Flegal KM, Tabak CJ, Ogden CL. Overweight in children: definitions and interpretation. Health Educ Res. 2006; **21**(6):755-60.
132. Growth reference 5-19 years. World Health Organization; 2011; Available from: http://www.who.int/growthref/who2007_bmi_for_age/en/.
133. de Onis M, Lobstein T. Defining obesity risk status in the general childhood population: Which cut-offs should we use? Int J Pediatr Obes. **5**(6):458-60.
134. Gardosi J, Figueras F, Clausson B, Francis A. The customised growth potential: an international research tool to study the epidemiology of fetal growth. Paediatr Perinat Epidemiol. 2010; **25**(1):2-10.
135. Willows ND, Sanou D, Bell RC. Assessment of Canadian Cree infants' birth size using the WHO Child Growth Standards. American Journal of Human Biology. 2011; **23**(1):126-31.
136. Delbaere I, Vansteelandt S, De Bacquer D, et al. Should we adjust for gestational age when analysing birth weights? The use of z-scores revisited. Hum Reprod. 2007 August 1, 2007; **22**(8):2080-3.

137. Arbuckle T, Wilkins R, Sherman GJ. Birth weight percentiles by gestational age in Canada. Obstet Gynecol. 1993; **81**(1):39-48.
138. Mikolajczyk RT, Zhang J, Betran AP, et al. A global reference for fetal-weight and birthweight percentiles. Lancet. 2011; **377**(9780):1855-61.
139. Hadlock FP, Harrist RB, Martinez-Poyer J. In utero analysis of fetal growth: a sonographic weight standard. Radiology. 1991; **181**(1):129-33.
140. Kierans W, Joseph K, Luo ZC, Platt R, Wilkins R, Kramer M. Does one size fit all? The case for ethnic-specific standards of fetal growth. BMC Pregnancy and Childbirth. 2008; **8**(1).
141. Lu MC, Halfon N. Racial and ethnic disparities in birth outcomes: a life-course perspective. Matern Child Health J. 2003; **7**(1):13-30.
142. Groom KM, Poppe KK, North RA, McCowan LME. Small-for-gestational-age infants classified by customized or population birthweight centiles: impact of gestational age at delivery. Am J Obstet Gynecol. 2007; **197**(3):239.e1-.e5.
143. Verkauskiene R, Figueras F, Deghmoun S, Chevenne D, Gardosi J, Levy-Marchal M. Birth weight and long-term metabolic outcomes: does the definition of smallness matter? Hormone Research in Paediatrics. 2008; **70**(5):309-15.
144. Davidoff MJ, Dias T, Damus K, et al. Changes in the gestational age distribution among US singleton births: impact on rates of late preterm birth, 1992 to 2002. Semin Perinatol. 2006; **30**(1):8-15.
145. Judd CM, Smith ER, Kidder LH. Questionnaires and Interviews: Asking Questions Effectively. Research methods in social relations. 6th ed: Hardcourt Brace Jovanovich Publishers; 1991. p. 228-65.
146. Marshall M. The key informant technique. Fam Pract. 1996; **13**(1):92-7.
147. Chino M, DeBruyn L. Building true capacity: Indigenous models for Indigenous communities. Journal Information. 2006; **96**(4).
148. Davidson-Hunt IJ, O'Flaherty RM. Researchers, indigenous peoples, and place-based learning communities. Society and Natural Resources. 2007; **20**(4):291-305.
149. World Health Organization. WHO Anthro (version 3.2.2 January 2011) and macros. 2012 [cited 2012 October 1]; Available from: http://www.who.int/childgrowth/software/en/.
150. Blossner M, Siyam A, Borghi E, Onyango A, de Onis M. WHO AnthroPlus for personal computers manual: software for assessing growth of the world's children and adolescents. Geneva, Switzerland: Department of Nutrition for Health and Development; 2009.
151. World Health Organization. Growth reference data for 5-19 years. 2012 [cited 2012 October 1]; Available from: http://www.who.int/growthref/en/.
152. Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. Stat Med. 2006; **11**(10):1305-19.
153. Global Database on Child Growth and Malnutrition. World Health Organization; [11/12/2012]; Available from: http://www.who.int/nutgrowthdb/software/Differences_NCHS_WHO.pdf.
154. Noel PH, Copeland LA, Perrin RA, et al. VHA Corporate Data Warehouse height and weight data: opportunities and challenges for health services research. J Rehabil Res Dev. **47**:739-1489.
155. Ananth CV. Menstrual versus clinical estimate of gestational age dating in the United States: temporal trends and variability in indices of perinatal outcomes. Paediatr Perinat Epidemiol. 2007; **21**:22-30.
156. Growing up in Australia: study questionnaires. Commonwealth of Australia; 2012 [02/02/2012]; Available from: http://www.growingupinaustralia.gov.au/studyqns/index.html.

157. Bavdekar A, Vaidya UV, Bhave SA, Pandit AN. Catch up growth and its determinants in low birth weight babies: A study using Z scores. Indian Pediatr. 1994; **31**:1483-90.
158. Li C, Goran MI, Kaur H, Nollen N, Ahluwalia JS. Developmental trajectories of overweight during childhood: role of early life factors. Obesity. 2012; **15**(3):760-71.
159. Al Mamun A, Lawlor DA, Alati R, O'Callaghan MJ, Williams GM, Najman JM. Does maternal smoking during pregnancy have a direct effect on future offspring obesity? Evidence from a prospective birth cohort study. Am J Epidemiol. 2006; **164**(4):317-25.
160. Widerøe M, Vik T, Jacobsen G, Bakketeig LS. Does maternal smoking during pregnancy cause childhood overweight? Paediatr Perinat Epidemiol. 2003; **17**(2):171-9.
161. Mizutani T, Suzuki K, Kondo N, Yamagata Z. Association of maternal lifestyles including smoking during pregnancy with childhood obesity. Obesity. 2012; **15**(12):3133-9.
162. La Merrill M, Stein CR, Landrigan P, Engel SM, Savitz DA. Prepregnancy body mass index, smoking during pregnancy, and infant birth weight. Ann Epidemiol. 2011; **21**(6):413-20.
163. Twisk JWR. Applied longitudinal data analysis for epidemiology: a practical guide: Cambridge University Press; 2003.
164. Rabe-Hesketh S, Skrondal A. Multilevel and longitudinal modeling using Stata: STATA press; 2008.
165. Kirkwood BR, Sterne JAC. Essential medical statistics: Wiley-Blackwell; 2003.
166. Salsberry PJ, Reagan PB. Dynamics of early childhood overweight. Pediatrics. 2005; **116**(6):1329-38.
167. Lamb MM, Dabelea D, Yin X, et al. Early-life predictors of higher body mass index in healthy children. Ann Nutr Metab. **56**(1):16-22.
168. McCarthy A, Hughes R, Tilling K, Davies D, Davey Smith G, Ben-Shlomo Y. Birth weight; postnatal, infant, and childhood growth; and obesity in young adulthood: evidence from the Barry Caerphilly Growth Study. The American Journal of Clinical Nutrition. 2007 October 2007; **86**(4):907-13.
169. Abell TD. Low birth weight, intrauterine growth-retarded, and pre-term infants. Human Nature. 1992; **3**(4):335-78.
170. Thomas C, Hypponen E, Power C. Prenatal Exposures and Glucose Metabolism in Adulthood Are effects mediated through birth weight and adiposity? Diabetes Care. 2007; **30**(4):918-24.
171. Wen X, Triche EW, Hogan JW, Shenassa ED, Buka SL. Prenatal factors for childhood blood pressure mediated by intrauterine and/or childhood growth? Pediatrics. 2011; **127**(3):e713-e21.
172. Waterland RA, Garza C. Potential mechanisms of metabolic imprinting that lead to chronic disease. The American Journal of Clinical Nutrition. 1999; **69**(2):179-97.
173. Arenz S, Rückerl R, Koletzko B, Von Kries R. Breast-feeding and childhood obesity—a systematic review. Int J Obes. 2004; **28**(10):1247-56.
174. Biddle N. An Exploratory Analysis of the Longitudinal Survey of Indigenous Children: Australian National University, Centre for Aboriginal Economic Policy Research; 2011.
175. Soothill P, Bobrow C, Holmes R. Small for gestational age is not a diagnosis. Ultrasound Obstet Gynecol. 1999; **13**(4):225-8.