2010

Individual, behavioural and environmental pathways to adolescent obesity

Paola Teresa Chivers
University of Notre Dame Australia

Follow this and additional works at: http://researchonline.nd.edu.au/theses

Part of the Medicine and Health Sciences Commons

COMMONWEALTH OF AUSTRALIA
Copyright Regulations 1969

WARNING
The material in this communication may be subject to copyright under the Act. Any further copying or communication of this material by you may be the subject of copyright protection under the Act.
Do not remove this notice.

Publication Details

This dissertation/thesis is brought to you by ResearchOnline@ND. It has been accepted for inclusion in Theses by an authorized administrator of ResearchOnline@ND. For more information, please contact researchonline@nd.edu.au.
Chapter Three
Methods

This study is based on data gathered as part of a longitudinal research project, The Western Australian Pregnancy Cohort (Raine) Study. This study design aims to track patterns of behaviour from birth and determine their relative influences on weight status at age 14 years. The study design was determined retrospectively to the data collection.

The methodology for this research is focussed on the statistical analysis of the data, investigating the interrelationships between key variables at each survey point and over time, and weight status at 14 years. This chapter will describe the database from which data are drawn, the sample, measures and their treatment, and the data analysis processes undertaken.

Data

The data for each variable was cleaned by authorised research officers of the Institute for Child Health Research using standardised protocols. Permission to use the data were provided by the Raine Executive Committee, with selected data provided to research teams through the Raine Study de-identified database.

Selection of obesogenic variables from the Raine database involved identifying individual, behavioural and environmental variables linked to obesity from the literature, as well as other potential factors such as motor competence, physical fitness, and timing of transition to solids. Variables selected required review, and if necessary, transformation to an appropriate form for statistical analysis (e.g. height and weight converted to BMI scores, continuous or categorical variables).
Sample

The Raine Study enrolled mothers of 2,979 children *in utero* from antenatal clinics at King Edward Memorial Hospital for Women (KEMH), Perth’s primary specialist obstetric health care facility, and has followed 2,868 live birth children. Women were enrolled into the project over 30 months from May 1989 to November 1991. Enrolment criteria included gestational age of 16-20 weeks, basic proficiency in English for informed consent, expectation of delivery at KEMH, and intention to remain in the state of Western Australia. All mothers gave written informed consent and the study was approved by the institutional ethics committees. The protocol for the original study has been previously reported describing the antenatal (Newnham et al., 1993) and postnatal periods (Joseph-Bowen, de Klerk, Firth, Kendall, Holt, & Sly, 2004; Oddy et al., 1999).

The Raine Study children have been assessed at birth, and ages 1, 2, 3, 6, 8, 10 and 14 years, with testing of 17 years olds completed early in 2010. The Raine Study families are broadly representative of the Western Australian population\(^6\): 10.7% of parents being never married (versus 9.8%), 7.5% children were born <37 weeks (versus 6.9%), with a slight overrepresentation children born <2,500g, 8.6 (versus 6.5%). The Raine cohort is well established and there is frequent contact between enrolled families and study organisers (Li, Kendall, Henderson, Downie, Landsborough, & Oddy, 2008). There has been attrition over time in sample size from each survey wave and among variables (refer to Table 4). Overall retention rates are high for each survey wave (92% at 1-year, 74% at 2-years, 85% at 6-years, 82% at 8- and 10-years, and 79% at 14-years), with enthusiasm amongst participants to provide high quality information. At follow-up age 2 years, funding limitations restricted the number of individuals assessed across physical measures, which impacted upon the sample size for BMI.

\(^6\) Western Australian population was drawn from all births 1989-1992 utilising the Western Australian Maternal and Child Health Research Database at Telethon Institute for Child Health Research (Kendall, 2003).
Table 4

*Raine Study Sample (n) of Participants who Completed All or Part of Each Follow-up*

<table>
<thead>
<tr>
<th>Mean age at Follow-up</th>
<th>Completed</th>
<th>Deferred</th>
<th>Lost</th>
<th>Withdrawn</th>
<th>Deceased</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2,441</td>
<td>204</td>
<td>174</td>
<td>21</td>
<td>28</td>
<td>2,868</td>
</tr>
<tr>
<td>2</td>
<td>1,988</td>
<td>381</td>
<td>418</td>
<td>51</td>
<td>30</td>
<td>2,868</td>
</tr>
<tr>
<td>3</td>
<td>2,280</td>
<td>321</td>
<td>158</td>
<td>79</td>
<td>30</td>
<td>2,868</td>
</tr>
<tr>
<td>6</td>
<td>2,237</td>
<td>339</td>
<td>135</td>
<td>127</td>
<td>30</td>
<td>2,868</td>
</tr>
<tr>
<td>8</td>
<td>2,140</td>
<td>376</td>
<td>124</td>
<td>198</td>
<td>30</td>
<td>2,868</td>
</tr>
<tr>
<td>10</td>
<td>2,047</td>
<td>281</td>
<td>162</td>
<td>348</td>
<td>30</td>
<td>2,868</td>
</tr>
<tr>
<td>14</td>
<td>1,860</td>
<td>357</td>
<td>207</td>
<td>412</td>
<td>32</td>
<td>2,868</td>
</tr>
</tbody>
</table>

*Note.* Completed = participant completed all or part of the follow-up; Deferred = participant did not participate in this follow-up but remained part of the cohort; Lost = participant was unable to be contacted for this follow-up, but remained in the cohort for future contact; Withdrawn = participant has withdrawn from the cohort and no further contact was made. (The Raine Study, 2010)

**Measures and Treatment**

This research draws on the extensive data base of the longitudinal Raine Study cohort. From a background perspective the following summarises the data collection methods used in the Raine Study (Newnham et al., 1993; The Raine Study, 2010).

The Raine Study began in concept in 1988 by Professor John Newnham, receiving funding through the Raine Foundation in 1989. The original goal for the study was to determine the effects of growth before birth and other factors during pregnancy on outcomes through childhood and into adult life (The Raine Study, 2010). Table 5 provides an overview of the data collected at each follow-up.
Table 5

*Summary Table of Raine Study Data Collection (The Raine Study, 2010)*

<table>
<thead>
<tr>
<th>Average Age (Time line)</th>
<th>Data collection - key variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>18-20 weeks gestation</td>
<td>Maternal 108 item questionnaire (SES, lifestyle, medical history, environmental exposures)</td>
</tr>
<tr>
<td></td>
<td>Male partner 30 item questionnaire (physical size, education, occupation, environmental exposures)</td>
</tr>
<tr>
<td>34 weeks gestation</td>
<td>Maternal questionnaire (exposures during 18-34 weeks)</td>
</tr>
<tr>
<td>1-year-olds (1990-)</td>
<td>Caregiver questionnaire (child's health, growth and development since birth; parental questions)</td>
</tr>
<tr>
<td></td>
<td>Physical examination (height, weight, blood pressure, and physical health and development assessment)</td>
</tr>
<tr>
<td>2-year-olds</td>
<td>Caregiver questionnaire (child’s health, growth and development since age one; parental questions)</td>
</tr>
<tr>
<td></td>
<td>Physical examination (height, weight, blood pressure, and physical health and development assessment)</td>
</tr>
<tr>
<td>3-year-olds (-1994)</td>
<td>Caregiver questionnaire (child’s health, growth and development since age two; parental questions)</td>
</tr>
<tr>
<td></td>
<td>Physical examination (height, weight, blood pressure, and physical health and development assessment)</td>
</tr>
<tr>
<td>6-year-olds (1994-1996)</td>
<td>Caregiver questionnaire (child’s health, growth and development since age three; parental questions, family history of asthma, eczema and hay fever)</td>
</tr>
<tr>
<td></td>
<td>Physical examination (height, weight, blood pressure, physical health and development assessment, spirometry and methacholine challenge or bronchodilator response test, skin prick tests for atopy and blood sample tests)</td>
</tr>
<tr>
<td>Average Age (Time line)</td>
<td>Data collection - key variables</td>
</tr>
<tr>
<td>-------------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>8-year-olds (1996-1997)</td>
<td>Caregiver questionnaire (child’s health, growth and development since age five; parental questions, family history of cardiovascular disease, fat intake and physical activity), short fat questionnaire / physical activity. Physical examination (height, weight, blood pressure and blood sample tests for both parent and child), spirometry and blood pressure assessment during a sub-maximal bicycle.</td>
</tr>
<tr>
<td>10-year-olds (1997-2000)</td>
<td>Caregiver questionnaire (child’s home life, leisure activities, schooling, language development, behaviour and general health since age eight); parental questions (health and happiness), parent-teacher questionnaire (child’s mental health, cognitive, language and neuromuscular development) and school report on academic achievement. Family history of cardiovascular disease, fat intake and physical activity), short fat questionnaire / physical activity. Physical examination (height, weight, blood pressure)</td>
</tr>
<tr>
<td>14-year-olds (2003-2006)</td>
<td>Caregiver questionnaire (similar to previous, since age ten); food frequency questionnaire. Child questionnaire (self-perceptions, risk-taking behaviour). Principal – teacher questionnaire (school environment, child participation, academic achievement). Physical examination (height, weight, blood pressure – child and caregiver). Stress test (PWC170), lung function test, bronchial responsiveness test, skin prick test for allergies, urine sample, 7-day pedometer assessment and diary of physical activity, blood tests and postural body images</td>
</tr>
</tbody>
</table>
A methodical review of all variables collected as part of the Raine Study was conducted. Based on the literature, a subset of individual, behavioural and environmental variables was selected as obesogenic variables of interest.

Variables collected as part of the Raine study, have been acquired using standard protocols, by specialist trained research staff at the Institute for Child Health Research, under the guidance of specific academic research teams and institutions. This study involved the use of variables that were already cleaned, scored or categorised and available via a de-identified data base. The variables used in this study are discussed individually below, with any additional transformation to variables detailed. Where possible, protocols used by the Raine Study team for physical assessments are provided.

**Age and time.**

Age and time variables were used for this study as recommended by Singer and Willet (2003). Age was reported in the smallest meaningful unit, in this study, months. Age in months was calculated at each follow-up by computing the difference between the date of physical assessment for that follow-up and the child’s date of birth.

Time was coded according to the average age in years of the cohort at each measurement wave i.e. 0, 1, 2, 3, 6, 8, 10.5, and 14. In this longitudinal sample, “occasional creep – over time” (Singer & Willet, 2003 p.141) was found where the actual ages exceeded that originally planned. In this study the year five follow-up occurred when individuals were an average of 6 years of age, the 10 year follow-up occurred when individuals were an average of 10.5 years of age, and 13 year follow-up occurred when individuals were an average of 14 years of age. This variable forms a time-structured predictor that records time on a scale comparable numerically to the individual’s age, that is, the values indicate the individual’s expected age on each follow-up assessment (Singer & Willet, 2003).
Individual factors.

Broadly, individual factors considered for this study included gestational age, anthropometric measures, adiposity rebound, diet, developmental milestones and motor competence, physical fitness, and puberty. Some factors had multiple indicators.

Gestational age.

Gestational age is the duration of time, measured in weeks, that the foetus has spent developing in utero. This duration is estimated using the date of the mother’s last menstrual period. At birth, the gestational age (gweeks) was used as a marker for pre-term birth exclusion from the sample, with pre term births determined to be a gestational age of less than 37 weeks.

Anthropometric measures.

Anthropometric measures were taken by a small group of extensively trained staff of the Telethon Institute of Child Health Research using scientific protocols, with intra and inter-rater reliability established at the beginning of each follow-up survey. Depending upon anthropometric measurements assessed at follow-up, the sequence of measures taken were Biacromial measurement, waist measurement, hip measurement, height, weight, and arm circumference.

The neonatal examination was conducted between 24 and 72 hours following birth. Length was measured by two people using the Harpenden Neonatometer (Holtain Ltd. Crosswell, United Kingdom) to the nearest 0.1cm. Newborns and infants (follow-up one year) were laid in supine position, with their head held by one person against a curved head plate in mid-line. The other person stretched the legs straight, knees held together, ankle flexed at right angles to the lower leg, moving the mobile plate to rest against the baby’s feet. Weight was measured to the

---

7 This measure was not used in this research study.
nearest 100g, using calibrated hospital scales at birth and a Wedderburn digital chair scales in follow-up one year.

In follow-ups year two and on, where appropriate, children were standing in the anatomical position, palms facing forward. Each area was measured at least twice in sequence with measures within one centimetre. Unless stated, measures were taken at expiration. Height was measured using a Holtain stadiometer (Holtain Ltd. Crosswell, United Kingdom) with shoes off, and heels, bottom and head against the board. The chin was positioned to straighten the neck and the measure taken with a breath intake. Weight was measured with children wearing light clothing (running shorts and singlet top), to the nearest 100g, using a Wedderburn digital chair scales (Wedderburn, Australia).

**Skin-fold measurement.**

Skin-fold measurements were assessed at follow-ups 1, 2, 3, 6 and 10 years. The protocol involved children standing in the anatomical position, palms facing forward. All measurements were taken on the individual’s left side. A skin fold was obtained by pinching, using a slight rolling action of the left thumb and index finger. The skin fold was raised at the marked site with firm pressure and held throughout the measurement. A Holtain skin fold calliper was applied 1cm from finger/thumb, released and after a two second wait, the measurement read. Measurements were taken in sequence (not consecutively), with a minimum of two measures with less than 10% difference. Additional measures were taken as necessary. The sequence for measurement was subscapular, triceps, suprailiac, and mid abdominal with details of protocol detailed at Table 6 (Telethon Institute for Child Health Research, n.d.).
Table 6

*Protocol for Skin-fold Site Identification and Procedure*

<table>
<thead>
<tr>
<th>Skin fold site</th>
<th>Site identification</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subscapular</td>
<td>Site is 1cm below the lower angle of the left scapular. The skin fold runs downward</td>
<td>Pinch left fold immediately beneath the apex of the left scapula, with the arm handing at the subject’s side.</td>
</tr>
<tr>
<td></td>
<td>at an angle of about 45 degrees from the horizontal.</td>
<td></td>
</tr>
<tr>
<td>Triceps</td>
<td>The skin fold is measured on the back of the left upper arm.</td>
<td>Pinch a vertical skin fold on the posterior aspect of the upper arm, half way between the olecranaon and acromium, with the arm hanging by the subject’s side.</td>
</tr>
<tr>
<td>Suprailliac</td>
<td>Leading edge of the iliac crest ... or ...Feel laterally for the iliac crest and</td>
<td>Pinch skin fold at a 45 degree angle immediately above the left anterior superior iliac spine.</td>
</tr>
<tr>
<td></td>
<td>follow along anteriorly.</td>
<td></td>
</tr>
<tr>
<td>Abdominal</td>
<td>5cms to the left of the mid point of the navel.</td>
<td>Pinch vertical skin fold and raise.</td>
</tr>
</tbody>
</table>

*Note.* Adapted from “Skin folds measurement sites,” by Telethon Institute for Child Health Research. (n.d.), *The Western Australian Pregnancy Cohort (Raine) Study* [Protocol]. Perth.

**Waist girth.**

At age 14 years, waist girth measures were taken. For waist girth, the individual held the measuring tape in place over the umbilicus on the skin in two places. The tape was kept horizontal and firm to the skin. The research assistant moved the tape around the body ensuring the tape was horizontal and in the same plane at the back before recording the measurement in centimetres.

Waist-height Ratio (WHtR) was calculated using the formula waist girth (cm)/height (cm). WHtR cut-offs were selected to represent the cut-offs for BMI as proposed by Kahn and colleagues (2005), namely ≥0.539 obese (BMI ≥30), 0.490-0.539 at risk or overweight (BMI 25-30) and <0.490 (normal) (BMI<25), with the McCarthy and
Ashwell (2006) 0.5 cut-off used for the two category comparison normal weight and overweight.

*Hip girth.*

Hip girth was recorded at age 14 years. Individuals were dressed in light clothing. The research assistant stood to the side of the individual and placed the tape measure around the body, over the widest part of the buttocks. The individual held the tape measure firm and horizontal in two places at the front of their body, while a measurement was taken.

Waist-hip Ratio (WHR) was calculated using the formula waist girth (cm) / hip girth (cm). Age and gender adjusted waist girth cut-offs between normal and overweight were calculated according to those reported in Taylor and colleagues (Taylor, Jones, Williams, & Goulding, 2000).

*Body mass index (BMI).*

At each follow-up BMI was calculated from measured height and weight using the formula weight (kg)/height (m)². At birth and year one, a proxy for BMI was used (weight/length²) (Rolland-Cachera, Deheeger, Akrout, & Bellisle, 1995). At age 14 years, BMI cut-off points equivalent to 25 and 30, adjusted for age and gender, classified the participants as normal weight, overweight or obese as defined by the IOTF criteria (Cole et al., 2000).

*Adiposity rebound.*

Adiposity rebound, defined as the last minimum (nadir) BMI before the continuous increase with age (Rolland-Cachera et al., 1984), was determined in a subset of individuals (n=173) for whom a complete set of BMI data were available. This small sample was due to limited anthropometric measurements taken at the two year follow-up (insufficient funding), but is similar in size to another study (Rolland-Cachera et al., 1995). Adiposity rebound was based on the child’s age in months.
Given the wide range of participant ages in the three and six year follow-ups, data were available for age 3-4 years, and age 5 ½-6 years leaving a small gap between ages 4 and 5½ years without data. This is pertinent to the calculation of adiposity rebound. BMI, and age at nadir were calculated for both the raw BMI and predicted BMI (based on the longitudinal LMM which adjusts for age, gender, weight status and gestational age and is discussed later).

**Puberty.**

Pubertal data were collected at age 14 years and included the Tanner Stages of pubic hair development for males and females, and for females, Tanner Stages of breast development (Tanner, 1962) and the date of first menstruation. The self-assessment of sexual maturation (Tanner, 1962) has been shown to have excellent agreement with physical assessment by medical professionals (Duke et al., 1980).

For girls, menstruation was determined as the age at which they experienced their first menstruation. This was calculated by subtracting their birth date (day/month/year) from the reported date of their first menstruation (day/month/year), with a reported measure in years. Female breast development was reported according to the five stages of development (Tanner, 1962).

During the data collection for the Tanner Stages an error was detected where there were incorrect codes indicated on the Tanner image from which the adolescents made a selection of their pubertal developmental stage. In effect, codes of 1, 2, 3, and 4 that corresponded with images 1, 2, 3, and 4, were instead coded 2, 3, 4 and 5. It was believed that less than a third of the sample was affected by this error (n=334). All affected individual’s data for this variable that were labelled *doubtful* were re-coded for this study as missing. Pubic hair development was reported according to five stages of development (Tanner, 1962).
**Diet.**

Diet variables available for analysis for this study included early infant feeding, self-reported adolescent fruit and vegetable intake at age 14 years, and fat intake at ages 8 years (parent report) and 14 years (self-report).

**Early infant feeding.**

Parent recall to specific questions in follow-ups conducted at one, two and three years collected information pertaining to early infant feeding. Mothers reported the age breast feeding stopped (in months), age milk other than breast milk was introduced (in months), and age they first gave their baby solids (weaning) (in months). Milk other than breast milk typically included infant formula (67.7%), cow (87.7%) or soy milk (20.8%). The age breastfeeding stopped, the age at which milk other than breast milk was introduced, and weaning were determined from the mother’s diary of early feeding milestones, an interview with the study nurse, and survey questions at later follow-ups.

Like Burke and colleagues (2005) who also analysed early feeding data for this cohort, no distinction between exclusive or partial breastfeeding was made. The variables for age breastfeeding stopped, age other milk introduced, and weaning were categorized using a 4-month cut-point (categories ≤4 months and >4 months). This cut-point split the cohort almost equally, was the World Health Organisation recommendation for duration of exclusive breastfeeding at the time of that data being collected (1990-1993), and has been reported by others for this cohort (Burke et al., 2005; Oddy, Li, Landsborough, Kendall, Henderson, & Downie, 2006a).

**Fruit and vegetable intake.**

Adolescents responded to a survey item asking them to report how often they ate fruit (fresh or canned, but excluding dried fruit, juices, fruit bars or frozen desserts), and vegetables (fresh, frozen, canned or salads). Response categories included
rarely / never, 1-2 times a month, 1-2 times a week, 3-5 times a week, and 6 or more times a week.

**Fat intake.**
At follow-up age 8 years, parents reported their child intake on the following identified fat intake items: *food with batter coating, gravy or sauces, add butter to food, fried or roasted veg, eat sausages, bacon; chips/french fries; pastries and cakes; chocolate intake; crisps intake; cream intake; icecream intake; cheese intake; skin on chicken; and fat on meat*. Responses were categorised never (0), less than once a week (1), 1-2 times a week (2), 3-5 times a week (3), and 6 or more times a week (4). A derived fat intake score was computed, as the listwise sum score of these categories (numbers in brackets to categories above), with a score range from 0-56.

At follow-up 14 years, adolescents reported their fat intake. Most items were similar to follow-up eight years, except for *fried food with batter coating, hard cheese intake*, and a rewording and response categories for chicken and meat fat. The question was reworded to: *how much fat on meat do you usually eat, and how much skin on chicken do you usually eat*, with response categories none (0), some (1), and most or all (2). A derived fat intake score was computed as the listwise sum score of these categories (as for follow-up 8 years), with a score range from 0-52.

**Developmental milestones and motor competence.**
Developmental milestones and motor competence was assessed in this study by parent reported developmental milestones (1-3 years), Denver Developmental Screening Test, 2nd Edition (1-3 years) (Frankenburg, Dodds, Archer, Shapiro, & Bresnick, 1992), Infant Monitoring Questionnaire (1-3 years) (Bricker & Squires, 1989), parent perceived motor skills (10 years), and the McCarron Assessment of Neuromuscular Development (10 and 14 years) (McCarron, 1997).
Parent reported developmental milestones.

Questionnaire items at follow-up age 1 year evaluated the timing of developmental milestones (months) based on parent recall. Items included when the child first smiled (weeks), sat up, babbled, crawled, stood, tottered, walked, spoke their first word, showed handedness, and lost their first tooth. Questions on when the child first walked, spoke their first word and showed handedness were repeated at follow-up age 2 years. Data were merged for these items across the 2 years. If data were present at the 1 year follow-up, the year 2 data were ignored; otherwise, if data were missing at the 1 year follow-up, the year 2 data were used.

Denver Developmental Screening Test, 2nd edition (Denver II).

The Denver II assessment is used to screen for possible developmental problems in pre-school children (Frankenburg et al., 1992). In this study it was conducted at follow-up ages 1, 2 and 3 years (Newnham, Doherty, Kendall, Zubrick, Landau, & Stanley, 2004). It reported on gross motor, fine motor, personal social, language, and overall development (age 3 years only). Categories included uncooperative, fail, suspect, adequate, and normal. The validity and reliability of the measure is unclear, although the revised version (Denver II) was re-standardized and is considered to be a quick screening tool (Glascoe & Frankenburg, 2002). The test-retest reliability (.90) and inter-rater reliability (.99) is excellent. Validity of the tool though is weak, and tends to result in a high rate of false positives (Naar-King, Ellis, & Frey, 2004).

Infant Monitoring Questionnaire (IMQ).

The IMQ is a series of parent completed questionnaire about motor, communication, social/personal and adaptive development (Bricker & Squires, 1989) and screens for developmental competence (Houck, 1999). Validity and reliability of the IMQ is good. It has good agreement with standardized tests (86% - 91%), low under- (6%) and over- (6%) screening rates, high inter-observer reliability (87%) and test-retest agreement (91%) (Bricker & Squires, 1989), and more recently
was found as an accurate screening measure for identifying Australian infants (aged 0-18 months) at risk of developmental delay (Dixon, Badawi, French, & Kurinczuk, 2009).

Parent responses were obtained at follow-ups 1, 2 and 3 years (Newnham et al., 2004). For this study, only items related to gross motor (1, 2 and 3 years), fine motor (1, 2 and 3 years), and overall development (2 and 3 years) were analysed. Response categories for gross and fine motor development were not yet, sometimes and yes. Overall development questions were a no or yes response.

**Parent perceived child motor skills.**
At follow-up age 10 years, parents rated their child’s motor skills in locomotion (running, jumping, hopping, skipping, dodging); object control (throwing, catching, kicking, striking); balance (biking, balancing); and overall coordination and development progress. Categories for responses were poor, below average, average, above average and excellent. For most items, sample size for response poor was less than 10, hence category poor was merged with below average for analyses to create four categories.

**McCarron Assessment of Neuromuscular Development (MAND).**
The MAND (McCarron, 1997) was used to assess motor competence at age 10 and 14 years. The test involved five fine motor tasks (beads in box, beads on rod, finger tapping, nut and bolt and rod slide) and five gross motor tasks (hand strength, finger-nose-finger, jumping, heal-toe walk, standing on one foot. The test has been shown to have good validity and reliability (McCarron, 1997). Evidence of the content, construct, predictive and concurrent validity of the test is provided by McCarron (1997). Tan, Parker and Larkin (2001) provided further evidence of the concurrent validity of the MAND. Performance rankings for the MAND were highly correlated with the Bruininks-Oseretisky Test of Motor Proficiency-Short Form (r =
and Movement ABC (.86) in a sample of 69 children aged between 4 and 10 years.

Based on absolute quantitative and qualitative test performance, raw scores were awarded. These raw scores were adjusted for age, then summed to yield an overall scaled score for fine motor competence, gross motor competence, and total motor competence. The scaled total score was converted to a Neuromuscular Development Index (NDI) score, which is normally distributed with a mean of 100 and a standard deviation of 15 (Hands et al., 2008). In addition, a three category variable was derived, low (≤85), medium (86-114) and high (≥115) motor competence based on the McCarron (1997) NDI cut-points of ≤85 below normal performance, and 85-115 normal performance.

Physical fitness.
Objective measures of physical fitness included aerobic fitness, muscle endurance, muscle strength, and flexibility at age 14 years.

Aerobic fitness.
Aerobic fitness was represented by the physical working capacity 170 (PWC 170) Monark cycle ergonometer test at follow-up 14 years (Hands et al., 2008). Although this test was also conducted at the 8-year-old follow-up, essential information for the calculation of the PWC 170 was missing, and unavailable for this study.

Aerobic fitness is an index of oxygen delivery and oxidative mechanisms of muscles during exercise (Armstrong, Welsman, Nevill, & Kirby, 1999), with PWC 170 providing an estimate of maximum oxygen consumption (VO₂ Max). At present, within the literature, both absolute and relative PWC 170 (Physical Working Capacity at an extrapolated heart rate of 170 beats per minute) measures are reported to be suitable in adolescents (American College of Sports Medicine, 2009;
Armstrong et al., 1999; Cleland et al., 2009; Hands et al., 2008), with either methodology used in respect to studies of obesity (Cleland et al., 2009; Hands et al., 2008). In adolescents the predictive ability of VO2 Max from PWC 170 is good (r=.71 and .70 for females and males respectively (Rowland, Rambusch, Staab, Unnithan, & Siconolfi, 1993); and r=.84 (Boreham, Paliczka, & Nichols, 1990), although there is wide variability (Rowland et al., 1993). An evaluation by Armstrong and colleagues (1999) demonstrated that it was an inappropriate assumption that scaling using body mass or other mass components controls adequately for body size differences in children and adolescents, with the relationship between VO2max and PWC 170 weakened (r=.65 and r=.48 for females and males respectively) when adjusted for body weight (Rowland et al., 1993). Age, gender and pubertal development are instead strong confounders to the measure of aerobic fitness (Armstrong et al., 1999). Further it is argued that excess body weight does not necessarily mean maximum oxygen consumption is reduced, and that VO2 max should be considered separately (Goran et al., 2000).

Adjustment of PWC 170 for body mass was not made, with this raw measure considered an absolute measure. This decision was based on considerations of mixed adjustment usage within the literature, evidence to suggest that adjustment for body mass does not adequately control for body size differences in children and adolescents (Armstrong et al., 1999), and that the outcome measure for this study is BMI, with an adjustment for body mass considered to be made twice.

**Muscle endurance.**

Abdominal muscle endurance was evaluated using a curl-ups test. Participants lay with their back and head on a mat, knees bent to 90 degrees and feet flat on the floor. The arms were outstretched, grasping a pencil in the fists and resting on the thighs. The examiner placed a ruler over the peak of the knees to act as a target. No pressure was placed on participants’ feet or knees. At a cadence of 20 curls per minute, set by a metronome, participants repeatedly curled up to touch their fingers on the ruler and returned to the starting position. The test ended after 60
curls or if the following occurred for two consecutive curls: curls were not in time with the beat, the grip on the pencil was released, one or both soles of the feet left the floor, one or both fists did not touch the ruler, the head did not make contact with the mat, or the arms/elbows were bent. The number of completed curls was counted (Hands et al., 2008).

**Muscle strength.**

Upper body strength was measured by a chest pass test. Participants sat on the floor with their back, buttocks, shoulder and head flat against a wall, and their legs extended with the feet together. The examiner placed a hoop on top of participants’ toes. Participants held a basketball with elbows touching the wall, and the ball touching their chest, and then used a two-handed chest to pass the ball through the top of the hoop. Distance was measured from the wall to the landing point and the better of two trials was recorded (Hands et al., 2008).

**Flexibility.**

Two tests of flexibility were conducted: hip flexibility (sit and reach), and shoulder flexibility (shoulder stretch). The Figure-Finder Flex tester was used to measure hip flexibility. To measure the right leg, with shoes removed, the participants sat on the floor and extended their right leg so that their right foot was placed flat against a 30cm high box placed against a wall. The opposite knee was bent, with the sole against the medial border of the extended knee. With palms prone and touching, participants reached as far forward as possible, holding for 2–3 seconds, with no jerkiness, unevenness of hands or bending of knees. The meter rule used for measurement projected from the box toward participants, with the 0 cm mark nearest to the participant and the 50cm mark at the front edge of the box. The 23cm mark was located above the toes. The better of two trials was recorded. This was then repeated on the left leg and finally with both feet placed against the box together (Hands et al., 2008).
The shoulder stretch test involved participants reaching their right hand over and behind their right shoulder, and their left hand behind their back toward the right shoulder, attempting to touch their hands together. The stretch was graded as able if participants could touch at least fingertips together. This was repeated on the opposite side (Hands et al., 2008).

**Behavioural factors.**

Broadly, behavioural factors considered for this study included physical activity behaviours, sedentary behaviour, self concept, and parent reported progress. For some factors there were multiple indicators.

**Physical activity behaviours.**

Physical activity behaviours were represented by frequency of visits to the park in early childhood (1-3 years), parent reported activity levels in late childhood (6, 8 and 10 years), physical activity level (14 years), and attitudes and values to physical activity.

**Frequency of visits to a park or playground.**

At follow-ups one, two and three years, parents reported the frequency of visits to the park or playground with their children. Category response frequencies included never, seldom (<1/month), occasionally (<1/week), often (>1/week) and every day. This type of survey item has been shown to have moderate reliability and validity (Veitch, Salmon, & Ball, 2009).

**Parent reported activity levels at ages 6, 8 and 10 years.**

Organised sport participation was assessed at follow-up ages 6 and 8 years with a yes or no response to participation. Organised sport was defined in the survey as “at school or with a club”. Similarly organised activity at age 6 years was assessed with a yes or no response to participation and was defined in the survey as “music, dancing, kindygym, other clubs”. At age 8 years, activity level responses were
categorised as *sedentary or little*, *slightly active* and *active*. Activity was defined in the survey as:

- **sedentary or little**: gets very little exercise, ... spends most of (their) time sitting, watching TV, or reading; ...
- **slightly active**: gets some exercise, ... spends more time in active play than reading or watching TV; ...
- **active**: is involved in an organised activity 2 or 3 times per week or walks/runs 2km or more per day”.

At age 10 years, questionnaire items asked to state the activity involved in at each week day, the time spent in minutes, and whether the activity was in the local community (yes or no). Activities reported were then coded into activity levels *no*, *light*, *moderate* or *vigorous* based on the Raine Protocol for Physical Activity Categories (Appendix B). These protocols were based on those used in the WA CAPANS study (Hands et al., 2004).

From this several variables were computed:

- **Vigorous activity during the week** – count of all reported vigorous activity during the week Monday to Friday.
- **Moderate activity during the week** – count of all reported moderate activity during the week Monday to Friday.
- **Light activity during the week** – count of all reported light activity during the week Monday to Friday.
- **No activity during the week** – count of all reported no activity during the week Monday to Friday.

The same four variables were computed for weekend activity levels. That is, for Saturday and Sunday a total count was computed for vigorous, moderate, light and no activity. A total activity count was computed, total vigorous, total moderate and total light activity for the whole week, that is, Monday to Sunday. Activity in the local community was calculated for weekday (Monday to Friday), weekend
(Saturday and Sunday), and total (Monday to Sunday) activity levels. In addition, a total activity time was computed, a sum of activity time in minutes from Monday to Sunday.

**Physical activity level.**

Actual physical activity level was measured by pedometer step count at age 14 years. The step count measure demonstrates convergent (Tudor-Locke, Williams, Reis, & Pluto, 2002) and construct (Tudor-Locke, Williams, Reis, & Pluto, 2004b) validity, with the protocol established as reliable (Hands et al., 2008). Physical activity levels determined by pedometer have good intra-individual variability (Tudor-Locke et al., 2004a) and do not appear to be affected by participant reactivity (Rowe, Mahar, Raedeke & Lore, 2004).

Pedometer steps were recorded over a seven day period, with adolescents wearing the Yamax Digiwalker SW200 pedometers (Yamasa Tokei Keiki Co. Ltd., Tokyo, Japan) on the right hip. Step counts were recorded daily via an electronic or paper diary. Daily step counts below 1,000 or above 44,000 were removed (Hands et al., 2008).

Continuous forms of the variables investigated included a weekday average step count (mean daily step count per weekday) and weekend average step count (mean daily step count per weekend). However the mean daily step count determined for those participants recording feasible step counts for a minimum of 4 days, including at least one weekend day was the primary measure of interest for modelling analyses (mean daily step count per week). This method has been shown to be strongly correlated with 7-day averages (Ridley, Olds, Hands, Larkin, & Parker, 2009). Gender specific categories determined by tertiles, mean step count cut-point, median step count cut-point, and a 12,000 step count cut-point were also investigated.
Physical activity attitudes and values.

At age 14 years the questionnaire included multi-item questions in areas of encouragement and support to exercise (11 items); physical activity and physical education (13 items); effects of physical activity (15 items); importance of physical activity (15 items); excuses for not increasing physical activity in the future (15 items); and other physical activity related perceptions (6 items). Responses were according to a Likert scale, differing dependent upon question. For effects and importance of physical activity the Likert categories extremely unlikely and very unlikely were combined into one category, as well as extremely likely and very likely, due to small sample sizes, resulting in a five category variable. Likewise, excuses for not increasing physical activity, the Likert categories’, applies strongly and applies very strongly were combined into one category, resulting in a four category variable.

A derived score was calculated for variables asking about encouragement and support to exercise (11 items, score 0-4), physical activity and physical education (13 items, score 0-3); effects of physical activity (15 items, score 0-6); importance of physical activity (15 items, score 0-6); excuses for not increasing physical activity in the future (15 items, score 0-4). The derived score was based on the mean of the sum of each variable’s item. The mean was computed for each using listwise selection (that is, no missing data points). Listwise assessment found the number of cases with missing data points for these variables were very few, with a paired sample t-test reporting no significant differences between the listwise derived mean score versus the mean score. The listwise derived score was used for further analysis.

Sedentary behaviour.

Screen time was used as a proxy for sedentary behaviour. Questionnaire responses have moderate reliability with validity results suggesting under-reporting (Clark, Sugiyama, Healy, Salmon, Dunstan, & Owen, 2009). Measures included usual time spent watching television, categorised as none, <3 hrs/wk, up to 1 hr/day (3-7
hrs/wk), 1-2 hrs/day (7-14 hrs/wk), 2-3 hrs/day (14-21 hrs/wk), and >3 hrs/day (>21 hrs/wk) at ages 6, 8 and 10 years. At 10 years the question changed slightly to include “TV and/or playing computer games”. At follow-up 14 years, the question changed further. Categories were none, up to 1 hr/day (3-6 hrs/wk), 1-2 hrs/day (7-13 hrs/wk), 2-3 hrs/day (14-21 hrs/wk), and >4 hrs/day (>21 hrs/wk). The wording changed to watching “TV or videos”. An additional question on usual time spent “playing video games or computer games, use the internet or chat online” was added with the same categories for responses.

A dichotomous television variable was created using the Australian National Television Watching guidelines of no more than 2 hours per day (Australian Government Department of Health and Ageing, 2004), with a 2 hours cut-point (i.e. the first three categories at ages 6, 8 and 10 years and the first two categories at 14 years, were collapsed into one category). At 14 years, a total screen time variable was created combining time spent in both television and computer use with categories none, up to 1 hr/day, 1-2 hrs/day, 3-4 hrs/day (14-21 hrs/wk), and >4hrs/day.

**Self concept.**

Self concept was represented by the Self Perception Profile for adolescents (SPPA) (Harter, 1988) and was used to help evaluate the relationship between the individual’s personal beliefs and subsequent behaviour, particularly related to physical activity behaviour. All nine domains were included for analysis. These were global self worth, and eight specific domains: physical appearance, athletic competence, scholastic competence, social acceptance, behavioural conduct, romantic appeal, close friendship and job competence. Each domain was scored on a continuous scale of one to four. SPPA has been shown to have good reliability and factor validity (Harter, 1988). Recently SPPA has been shown to be representative of the Australian population, reporting good factor validity and reliability in the Raine Cohort sample (Rose, Larkin, & Hands, 2010).
**Parent reported academic, social and behavioural progress.**

Parent questionnaire responses at age 10 years evaluated parent perceived academic performance categorised as *poor, below average, average, very good,* or *excellent.* A 5-point Likert scale (*very satisfied, satisfied, neither, dissatisfied,* and *very dissatisfied*) was reported for parent satisfaction with child’s progress in learning skills, social and behaviour.

**Environmental factors.**

Broadly, environmental factors considered for this study included socioeconomic status, parental influences, and the built environment. For all factors there were multiple indicators.

**Socioeconomic status (SES).**

SES data were collected via questionnaire at birth and each follow-up. SES data included Socioeconomic Index For Area, ethnicity, income, parent education, father’s occupation, employment and life stress.

**Socioeconomic Index For Area (SEIFA).**

The Australian Bureau of Statistics (ABS) (Australian Bureau of Statistics, 2007a) reported four indexes in its 2001 census data which related to socioeconomic aspects of geographic areas. The index of advantage/disadvantage is a new index that reports the advantage (high scores) to disadvantage (low scores) continuum. This index was considered the most appropriate for use with weight status in this study, used as an indicator of community SES.

The Index of Relative Socioeconomic Advantage / Disadvantage is an Australian wide index encompassing both rural and metropolitan areas. Areas with a relatively high proportion of people on high income and involved in skilled work will report high scores. While areas with a high proportion of people on low income
and involved in unskilled work will report low scores. This index is intended as a summary of area characteristics only (Australian Bureau of Statistics, 2007a).

The child’s postcode at the 14 year follow-up was used to determine the corresponding SEIFA score from the ABS tables (Australian Bureau of Statistics, 2007a). The SEIFA index was investigated as a continuous variable, and also categorised into low (≤25 centile), medium (>25–<75 centile) and high (≥75 centile).

**Parent ethnicity.**

Parent ethnicity included survey response categories other, Aboriginal, Polynesian, Vietnamese, Chinese, Indian and Caucasian. For both males and females, non-Caucasians represented only 9% of the sample. Hence a dichotomous ethnicity variable Caucasian and non-Caucasian was used.

**Income.**

Family income was used as an indicator of individual family wealth, but did not consider levels of debt, and was collected at birth and at each follow-up. Parents reported their family income bracket in survey response. Income brackets differed over the eight follow-ups and are summarised in Table 7. However, a low income cut-point was used as previously reported (Kozyrskyj, Kendall, Jacoby, Sly, & Zubrick, 2009). Cut-points for low income at each follow-up were: Birth and age 1 year - <$24,000, age 2 years - <$27,000, age 3 and 6 years - <$26,000, and age 8, 10 and 14 years - <$30,000.

**Parent education.**

Mother’s and father’s highest level of education responses were categorised as: none, trade certificate or apprenticeship, professional registration (non-degree), college diploma or degree, University degree, and other. The reported education level at the birth survey was used for this study.
Father’s occupation.

Father’s occupation response were categorised as: managerial, professional, par-professional, trade, clerical, sales, plant operator, and labourer. The reported father’s occupation at the birth survey was used for this study.

Table 7

Family Income Brackets across Follow-ups

<table>
<thead>
<tr>
<th>Income (AUS)</th>
<th>Birth</th>
<th>Ages 1,2,3 &amp; 6</th>
<th>Ages 8, 10 &amp; 14</th>
<th>Study Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less 7,000</td>
<td>1</td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Less 8,000</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>7K-11,999</td>
<td>2</td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>8K-13,999</td>
<td>1</td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>8K-15,999</td>
<td></td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>12K-23,999</td>
<td>3</td>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>14K-26,999</td>
<td>2</td>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>16K-24,999</td>
<td>2</td>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>24K-35,999</td>
<td>4</td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>27K-40,999</td>
<td>3</td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>25K-29,999</td>
<td></td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>&gt;36K</td>
<td>5</td>
<td></td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>&gt;41K</td>
<td></td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>30K-34,999</td>
<td></td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>35K-39,999</td>
<td></td>
<td>5</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>40K-49,999</td>
<td></td>
<td>6</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>50K-59,999</td>
<td></td>
<td>7</td>
<td>6 (or 5)</td>
<td></td>
</tr>
<tr>
<td>60K-69,999</td>
<td></td>
<td>8</td>
<td>7 (or 5)</td>
<td></td>
</tr>
<tr>
<td>&gt;70K</td>
<td></td>
<td>9</td>
<td>8 (or 5)</td>
<td></td>
</tr>
</tbody>
</table>

Note. K=1,000. Less than $7,000 and less than $8,000 had very small sample sizes so were merged with the next category.
**Employment.**

Family minimum work hours were investigated in terms of whether having a non-working or part-time parent at home was related with the child’s weight status. The minimum hours were calculated using the minimum hours of either the mother or partner. At ages 1, 2 and 3 years there were no partner hours recorded, so those data represent the mother only. When the father’s data were missing, then mother’s data were used. When mother’s data were missing, then data were coded missing. Minimum work hours were categorised as zero hours, 1-29 hours and 30 or more hours.

Family employment assessed separately whether the mother and partner were working using a yes or no response at birth and ages 1, 2, 3, 6 and 8 years. At 10 years the question changed to a no and multi-option yes (work for payment, unpaid work or other unpaid work) response. For this analysis the multi-yes response was converted to a simple yes response. The variable was converted into three categories which investigated differences between both mother and partner not working, either mother or partner working, and both mother and partner working. Mother’s work was a yes or no response. Mother’s work hours was the nominated hours of work per week reported.

**Life stress.**

Life stress item responses (yes or no) were collected at 18 and 34 weeks gestation, and all follow-up years. Parents were asked whether any of the following stress events had happened to them in the past year. The stress items included pregnancy problems, death of a close relative, death of a close friend, separation or divorce, marital problems, problems with children, own job loss, partner’s job loss, money problems, residential move, and other. A count of life stress events was computed for individuals with all 11 items answered, along with a dichotomous variable, <3 stress events and ≥3 stress events.
**Parental influences.**

Parental influences were reflected through parent weight and BMI prior to and across follow-ups, maternal smoking during pregnancy, as well as parenting styles in later childhood.

**Parent weight.**

Mother’s and father’s birth weight were reported by parent recall at the initial survey. Self reported height and weight was collected by survey for the parents, Mother’s pre-pregnancy and at their child’s age 6, 8 and 14 year follow-up; and father’s at child’s birth and child’s age 8 year follow-up. BMI was calculated for each using self-reported height and weight and the formula weight (kg)/height (m)² (Cole et al., 2000).

**Maternal smoking during pregnancy.**

Mother’s smoking behaviour during pregnancy was assessed by a yes or no response.

**Parenting styles.**

At the 10 year follow-up parents completed the Parent Scale (O’Leary, 2003). The parenting scale score was derived from the individual item scores reported according to the following scoring instructions (O’Leary, 2007).

- Each item receives a 1-7 score, where seven is the "ineffective" end of the item. (i.e. items 2, 3, 6, 9, 10, 13, 14, 17, 19, 20, 23, 26, 27, 30 score seven on the left, the others on the right).
- Total score is the average of all items (listwise).
- Factor score, average the responses on the items on that factor (listwise):
  - Laxness: 7, 8, 12, 15, 16, 19, 20, 21, 24, 26, 30 (11 items).
  - Over-reactivity: 3, 6, 9, 10, 14, 17, 18, 22, 25, 28 (10 items).
At age 14 years, adolescents rated their own parents’ parenting, during the last 6 months, selecting categories never, sometimes, often and very often for 11 items. Each item was analysed separately and included parents smile at me, parents forget rules, parent praise me, parents nag me about little things, parents use rules only when it suits them, parents tell me they appreciate me, parents threaten more than do, parents speak of good things I do, parents enforce or don’t depending upon mood, parents threaten or hit me, and parents proud of me.

**Built environment.**
Proxies for the built environment related indirectly to obesity through opportunities to be physically active. They included home environment factors in the early years, and adolescent school environment factors.

**Home environment.**
The influence of the home environment on opportunities to be active was investigated by factors home swimming pool, outdoor play space, and proximity to the park or playground. At ages 1, 2, and 3 years, parents reported via a yes or no response whether they had a backyard swimming pool, home garden (outdoor play space), and whether they lived close to a park or playground.

**School facilities.**
At age 14 years, school principals responded (yes or no) to whether they had the following facilities: playing fields or ovals, gymnasium, swimming pool and or playing courts (tennis/basketball). This questionnaire was issued to principals on an individual basis (i.e. linked to individual cases), and hence there were duplicated responses across a period of 4 years. Duplicated entries were deleted, with the most recent date of assessment entry retained. Combinations of the various
facilities at the schools are presented in Table 8. Each combination reflects the number of schools with only that combination.

Table 8

*Descriptive Statistics for School Facilities Combination*

<table>
<thead>
<tr>
<th>Facility Combination</th>
<th>Present</th>
<th>Percent (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oval only</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>Gymnasium only</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Swimming pool only</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Court (Tennis, Basketball) only</td>
<td>6</td>
<td>2.7</td>
</tr>
<tr>
<td>Oval &amp; Gymnasium</td>
<td>5</td>
<td>2.3</td>
</tr>
<tr>
<td>Oval &amp; Court</td>
<td>41</td>
<td>18.5</td>
</tr>
<tr>
<td>Gymnasium &amp; Court</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>Oval, Gymnasium &amp; Court</td>
<td>96</td>
<td>43.2</td>
</tr>
<tr>
<td>Oval, Gymnasium &amp; Pool</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>Oval, Court &amp; Pool</td>
<td>8</td>
<td>3.6</td>
</tr>
<tr>
<td>All 4 Facilities</td>
<td>63</td>
<td>28.4</td>
</tr>
</tbody>
</table>

Based on this information, a positive school environment for sports facilities was determined to be high (three or more facility combinations), medium (two facility combinations) and low (one facility). The rationale was that the more physical activity choices available, the higher the likelihood for participation in physical activity by students at and around school times.

Total school facilities included the number of school ovals, council or shire facilities, and other facilities used for physical activity. For the same school, the number of facilities reported may be different from each date the questionnaire was completed. Some possible explanations may be that the school building expansion may reduce the number of ovals, sporting activities requiring council or other facilities may change from term to term. For this variable, duplicate school cases were treated as separate cases and not removed. Percentiles depicted for the total
number of school facilities were used to create three categories low (≤2 or ≤25\textsuperscript{th} percentile), medium (2-4 or 25\textsuperscript{th}-75\textsuperscript{th} percentile) and high (>4 or >75\textsuperscript{th} percentile).

\textit{School representative sports.}

Question 11 and 12 of the school questionnaire evaluated if schools participated in the representative sports program\textsuperscript{8}, and if so, which sports. All responses were yes or no. The categories for representative sport included netball, athletics, swimming, soccer, tennis, rugby, AFL (Australian Football League) and other. All but six school respondents participated in representative sports. A total score for the number of representative sports offered was calculated for each school with a minimum of one and a maximum of eight.

Representative sport data were skewed towards maximum influence with many schools offering six or more options. However the range in the number of representative sports offered was varied enough to warrant further investigation of any influences. Based on data percentiles (low <5\textsuperscript{th} percentile, high >50\textsuperscript{th} percentile), representative sport was categorised according to a positive environment for fostering after school physical activity. The rationale was that representative sport typically entails participation by students outside of their normal school hours. It was also assumed that within a metropolitan area, most schools would have the opportunity to participate in representative athletics and swimming. Therefore, representative sport was categorised as minimal influence: ≤3, moderate influence: 4-6, and maximum influence: >6 sports offered.

\textit{Physical education classes.}

The variable minutes allocated to physical education per week was not used in this study due to the lack of variance between schools, averaged across years 8-10. On average, schools ranged from category two (46-60 minutes) to category five (more

\textsuperscript{8} Representative sports program – sports in which children can represent their school in organised weekly competitions.
than 90 minutes). A few schools reported category one (31-45 minutes) for an individual year, but this was countered by extra hours in the other 2 years. In summary, all schools with data on physical education per week provided a positive environment for physical education.

**Data Analysis**

Generally, the number of variables used was dependent on sample size power, missing data, time points available, relevance to the statistical test, and requirements of the statistical technique being used. As mentioned previously sample size varies across follow-ups (Table 4), as well as for each individual variable.

This research study involved basic descriptive statistics and two key data analysis techniques, longitudinal data analysis using linear mixed modelling (LMM) and relationship modelling using structural equation modelling (SEM). All statistical processes were conducted using statistical software SPSS for Windows (Release 17.0.0. 2008).

**Descriptive statistics.**

Continuous variables were described statistically according to sample size \((n)\), mean \((M)\) and standard deviations \((SD)\). Gender differences between males and females were assessed using independent t-tests \((t)\) with a reported probability \((p)\). Weight status group comparisons were assessed using ANOVA \((F)\) and Dunnett T3 post hoc tests for IOTF weight status groups (normal weight, overweight and obese), determined at 14 years.

Categorical variables were described statistically according to sample size \((n)\). Gender differences between males and females, and weight status group differences were assessed using Chi-square \((\chi^2)\) and reported probability \((p)\).
Comparison of adiposity field measures. Several proxies for adiposity were available at age 14 years. The inter-changeability of these proxies were examined to guide a decision on the most appropriate measure to use considering the longitudinal structure of this research. Similarities among adiposity measures were investigated using the Bland-Altman method (Bland & Altman, 1986) using z-scores, an alternative regression approach (Bland, 2004; Bland & Altman, 2003), and the traditional Pearson Product Moment Correlation. The latter has been criticised as an inappropriate statistical analysis of similarity (Bland, 2004; Bland & Altman, 2003). Differences in weight status categorisation were examined using the International Obesity Task Force (IOTF) BMI cut-offs (Cole et al., 2000), WHtR cut-offs (Kahn et al., 2005; McCarthy & Ashwell, 2006), and waist girth cut-offs (Taylor et al., 2000).

Pearson Product Moment Correlation was calculated between each of the adiposity measures taken at follow-up age 14 years. The categorisations of weight status using BMI, WHtR, and waist girth were compared using a Kappa Chi-square test. Correlations are good at measuring the strength of a relationship, but not the amount of agreement (Bland & Altman, 1986). The Bland-Altman method (Bland & Altman, 1986) was used to assess the limits of agreement, with raw scores converted to z-scores prior to analysis and the resulting confidence intervals converted back into associated raw score units. The technique allows two methods to be compared when neither provides an unequivocally correct measure, and the degree of agreement can be assessed. The Bland regression method (Bland, 2004) was used to determine 95% prediction limits at the mean.

---

Linear mixed modelling (LMM)\textsuperscript{10}.

Multilevel models are becoming popular in health and social sciences for the analysis of longitudinal data and provide an opportunity to investigate within-person and between-person change over time. The multilevel model for change was used to separate differences between individuals at baseline and changes over time within individuals. The model is a mathematical representation of population behaviour (Garson, 2008; Singer & Willet, 2003), which enables correct modelling of correlated errors for repeated, continuous and correlated observations (Garson, 2008), and has an underlying assumption that data are missing at random (West, Welch, & Galecki, 2007). LMM has the advantage over repeated measures ANOVA in that they are more flexible in fitting and testing covariance structures, permit individuals to have missing data points, and allow the inclusion of time-varying factors as well as the time measurement (West et al., 2007).

Model selection is performed using information criteria. Here the \textit{goodness of fit} is based on a log-likelihood statistic. It allows the selection of a subset of interrelated predictors that best captures the effect of a single underlying construct (Singer & Willet, 2003; West et al., 2007). Predominantly the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are used. The AIC parameter is based on the number of model parameters, while the BIC includes the number of model parameters plus sample size (Singer & Willet, 2003).

Investigations using LMM required the dataset to be transformed into person-period data in which each individual had multiple data entries corresponding to each measurement occasion (Singer & Willet, 2003). A LMM was used to model the trajectory of BMI over time from birth to 14 years. Fixed and random effects, interactions and covariance structure were all investigated in the determination of the final model, with residual diagnostics performed. Time was used as a repeated measure. Best model fit was determined based on Akaike’s Information Criterion

(AIC), with a lower result indicating a better fit. BMI was treated as the dependent variable with effects of age, gender, gestational age and weight status (determined at age 14 years) investigated. There was no treatment for missing data, which were assessed as missing at random (MAR) (Chivers, Hands, Parker, Beilin, Kendall, & Bulsara, 2009).

The final LMM was used to test each obesogenic variable. Covariate only, full interactions and subsets were investigated consecutively, with model fit comparisons based on AIC. There was no treatment of missing data (MAR). However, for many variables, measurement occasions were limited to one or less than the complete eight measurement occasions, hence complete missingness for particular measurement occasions was present. For variables not collected at each of the eight measurement occasions, the one measurement occasion value was replicated for each of the other measurement occasions for each individual (that is, no change over time for that covariate.

**Structural equation modelling (SEM).**

SEM is a statistically powerful and flexible tool which enables the exploration of interrelationships among variables (Gao, Thompson, Xiong, & Miller, 2006; Garson, 2007; Thomas & Nelson, 1990). It is able to assess both directional and non-directional relationships (Gao et al., 2006). It allows for overall model testing, use of multiple dependents, investigation of mediating variables, testing across multiple between subject groups. It is also able to handle auto correlated errors across longitudinal data points, non-normal data, and incomplete data. Modelling also accounts for interactions, nonlinearities, correlations, and measurement errors. SEM has been explained as an extension of general linear modelling and incorporates and integrates path and factor analysis (Garson, 2007). SEM is becoming a popular choice in multivariate methods as it allows for mixed model analysis, especially where, traditionally, different levels of measurement have been problematic (Schumacker & Lomax, 2004).
The goal of SEM is to determine how well the theoretical model is supported by the data. It is a way of testing hypotheses or can be confirmatory in approach, although there is some element of exploration in the process (Dragan & Akhtar-Danesh, 2007; Garson, 2007; Schumacker & Lomax, 2004). It takes into account group level (people within groupings) as well as individual processes (Rasbash, 2006). For causation-effect relationships to be established, there must be logical preceding order with preceding variables affecting following variables in direct or indirect ways; existence of covariance or correlation between variables; control for other causes or influences; and measurement on at least an interval level. Structural models in longitudinal data may be able to show changes in the latent variables over time, while modelling both individual and group changes (Schumacker & Lomax, 2004).

In this study however, a purely exploratory approach was taken, using add on SPSS module AMOS version 17 (SPSS for Windows, Release 17.0.0. 2008). At each follow-up, possible interrelationships were explored. The basic hypothesised model at each follow-up had all variables having pathways to BMI only. Normality was examined for each variable and below the recommended limits (skewness <3; kurtosis <10) required for SEM (Kline, 2005). Due to the presence of missing data, a series mean replacement of missing values was conducted in SPSS to create a complete dataset for single-step model building (adding pathways) (Boomsma, 2000; Garson, 2007). This permitted model building to be explored by the addition of theoretically sound pathways based on modification indices (Byrne, 2001; Garson, 2007; Kline, 2005). The original dataset (with missing values) was then used for single step model trimming (removing pathways) (Boomsma, 2000; Garson, 2007), with a maximum likelihood estimation procedure (Byrne, 2001). Each model was assessed separately for males and females and additional significant pathways included. Each final model included all significant pathways for both sexes with gender differences assessed using multi-group analysis (Byrne, 2001).
These procedural steps taken are summarised below:

- Model specification. All variables have a direct pathway to BMI.
- Single-step model building, estimation and testing. Modification indices used to add pathways.
  - Data set contains series mean replacement of missing values.
  - Original data set with missing values.
- Estimation and testing for each gender. Model modification.
  - Original data set with missing values.
- Final model - multi-group analysis for gender.
  - Original data set with missing values.

**Ethics**

Initial data collection and all follow-ups of the Western Australian Pregnancy Cohort (Raine Study) have been approved by the Human Ethics Committee at King Edward Memorial Hospital and Princess Margaret Hospital for Children, Perth, Western Australia. Parents had been informed, prior to agreeing to participate, that the study would be ongoing over a number of years and would involve the investigation of a number of undisclosed child health outcomes. Consent to participate in the study was obtained from the mother of each child at enrolment and a parent at each subsequent follow-up (Kendall, 2003; Newnham et al., 1993).

Data collection has been conducted by approved research personnel of the Telethon Institute of Child Health, under the approved ethical guidelines and procedures established for the Raine Study.

All personal information pertaining to this study is kept secure and confidentially as part of the Telethon Institute of Child Health Research processes and procedures. Copies of completed questionnaires and examination data collected are identified only by study number and are stored separately from personal information in a
restricted access storage facility at the Telethon Institute of Child Health Research. Access to this information was not required for this study. Electronic stored data is also subject de-identified with restricted access by approved personnel with password access. Electronic personal information is stored separately from the data, again with restricted access and password security (Kendall, 2003). This research only accessed subject de-identified data for statistical analysis.

Research study approval and permission to access de-identified data were granted by the Raine Study Executive Committee. An expedited application for ethics approval was made to the University of Notre Dame Australia Human Research Ethics Committee and approved by both the School of Health Science Research Committee and the University of Notre Dame Human Research Ethics Committee. Copies of relevant ethic approval documents are provided at Appendix C.

Summary

In summary, the methodology for this study describes a longitudinal research approach. The research design identified key obesogenic variables of interest, examined their relationship with weight status at 14 years, and identified differences between males and females. Tests of similarity for adiposity measures were used to guide the decision on best proxy for adiposity for use in this study (Chivers, Hands, Parker, Beilin, Kendall, & Bulsara, in press). A model of BMI trajectory was developed using LMM, to investigate the influences of obesogenic variables on adolescent weight status (Chivers et al., 2009) and SEM procedures were designed to explore the interrelationships among identified obesogenic variables at each follow-up.
Eight years

"... That all children are created whole, endowed with innate intelligence, with dignity and wonder, worthy of respect. The embodiment of life, liberty and happiness, children are original blessings, here to learn their own song ...

... To recognize the early years as the foundation of life, and to cherish the contribution of young children to human evolution ..."

By Raffi