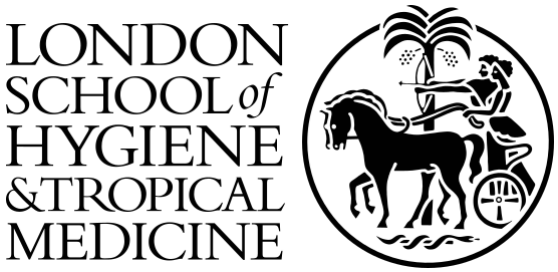


Growth in Infancy in Three Contemporary Cohorts: Selection Bias and Other Methodological Issues

Costanza Pizzi

A thesis submitted to the University of London for the degree of Doctor of Philosophy

London School of Hygiene and Tropical Medicine, 2013

A Nonno Enzo ed Olmo,
per i loro sguardi teneri, che mi mancano molto

Declaration

Statement of Own Work

All students are required to complete the following declaration when submitting their thesis. A shortened version of the School's definition of Plagiarism and Cheating is as follows (the full definition is given in the Research Degrees Handbook):
The following definition of plagiarism will be used:
Plagiarism is the act of presenting the ideas or discoveries of another as ones own. To copy sentences, phrases or even striking expressions without acknowledgement in a manner which may deceive the reader as to the source is plagiarism. Where such copying or close paraphrase has occurred the mere mention of the source in a biography will not be deemed sufficient acknowledgement; in each instance, it must be referred specifically to its source. Verbatim quotations must be directly acknowledged, either in inverted commas or by indenting. (University of Kent)
Plagiarism may include collusion with another student, or the unacknowledged use of a fellow student's work with or without their knowledge and consent. Similarly, the direct copying by students of their own original writings qualifies as plagiarism if the fact that the work has been or is to be presented elsewhere is not clearly stated.
Cheating is similar to plagiarism, but more serious. Cheating means submitting another student's work, knowledge or ideas, while pretending that they are your own, for formal assessment or evaluation.
Supervisors should be consulted if there are any doubts about what is permissible.

Declaration by candidate

I have read and understood the School's definition of plagiarism and cheating given in the Research Degrees Handbook. I declare that this thesis is my own work, and that I have acknowledged all results and quotations from the published or unpublished work of other people.

Signed:

Date: **May 1, 2013**

Full name: **Costanza Pizzi**

Abstract

There is broad recognition that early life growth trajectories are important predictors for the onset of several diseases. This thesis addresses two methodological challenges that arise in life-course studies of infant growth: (i) the bias that may derive from participants' selection in cohort studies, and (ii) the modelling of individual growth trajectories. Data from socio-economically diverse populations were used to address them: the Italian NINFEA web-based birth cohort, the Portuguese GXXI birth cohort, and the Chilean GOCS cohort.

Participants' selection affects all cohorts, but web-based designs are thought to be more affected than traditional ones. The thesis first examines possible selection mechanisms by Monte Carlo simulations and then uses population registry data to assess evidence of selection bias among NINFEA participants. The simulations show that under sensible scenarios there is only weak bias in the effects estimated from a selected sample. Comparisons of NINFEA participants with their source population (via registry data) show that the confounding patterns present in NINFEA differ from those in the source population, revealing that participants' restriction may either increase or decrease the confounding bias in an association of interest.

Studying individual early life growth data requires dealing with the quality of the growth measurements and the nonlinearity of the trajectories. Alternative models are compared in terms of their ability to address these problems while extracting salient features of weight growth. SITAR results to be the most useful model for life-course enquiries. An extension of this model, that includes explanatory variables, is fitted on the three cohorts to study the effect of prenatal exposures on different biologically defined dimensions of the growth process. This reveals some interesting mechanisms.

This thesis contributes to the interpretation of results obtained from cohort studies with restricted participation, and to the implementation of advanced growth models useful for life-course research.

Contents

List of Tables

List of Figures

Acknowledgements

First and foremost I want to thank my supervisor and "mentor" Bianca De Stavola, without whom this thesis would not exist. It is hard to summarise in a few words my endless gratitude, my huge admiration and my great affection for Bianca. I thank her for her invaluable help, her encouragement and her unconditional support and trust. I also very much appreciate her constructive criticism which has always contributed to improve my work and my own skills and understanding. I am also deeply thankful to her for granting me her time notwithstanding all the commitments and hurdles she has faced over the past years. Thank you always, Bianca.

I am extremely grateful to my co-authors and to the members of my advisory panel. Thanks to Lorenzo Richiardi for following me throughout this PhD and particularly for being supportive during the most crucial times. Thanks to Tim Cole who offered me guidance and thoughtful insights, which have enriched my knowledge and strengthened my thesis. Thanks to Isabel Dos Santos Silva who offered me an excellent guidance and for teaching me so much in these years of collaboration; finally thanks to Camila Corvalan and Chris Frost for their help and stimulating opinions. Special and deep thanks go to Franco Merletti who gave me the chance to embark on this PhD, and who welcomed me to his research team in Torino putting his trust in me. I hope I deserved it!

I could have never ended this thesis without the help of all the friends I had beside me along these years. First of all thanks to Pimo, for being the friend of my life and my greatest success. Thanks to the "family of friends" I have in Ancona, for the dinners, the games, the debates that fill up my soul and make me feel a rich person: thanks to Ale, Fede, Alessia, Nico and Zizza. A special thank goes to Giorgio, for housing me so many times, cuddling me and always making me feel at home in

London even when I did not live in London anymore (I will miss our dinners!). Thanks to my friends in Torino who welcome me and have been close to me almost every day during this PhD: thanks to Sere and Ale for our solid friendship and because I owe it all to you if Turin has become my second home. Thanks to Emilio for housing me when I was wandering around! Thanks to Mila, Dani and Enrica, for all the chats, for the support at work, for our friendship that goes beyond the office and to have made the "stanzone" together with Lorenzo, Manu, Anna, Emanuele and Marine a familiar place where I always happily return. A special thanks goes to Rhian for her huge support and for all the laughs we have had about "pumpkins" and other crazy stuff. Thanks to Antonio for his endless (and miraculous!!!) patience in answering all my questions about statistics. Finally thanks to my furthest friends, who always remain close no matter how many km away: thanks to Delia for cheering me up in my hard times and for the hours spent on skype chatting about love and politics; thanks to Flami who's always close no matter how far; and thanks to Francesca for all the emails and for following me all over Europe, even just for 48 hours.

My family is always a fundamental part of my life: without it I would not be here. Thanks to my mum and dad for being there always, for their love, their attention and help. Thanks to Bene, for what we share today and will do forever, and thanks to AnnaSofi for being one of the most beautiful parts of my life. Finally thanks to my extended family which is a priceless asset and source of serenity.

My final thanks is to Francesco. There are very many reasons, but mostly I thank you because you take the worst parts of me with a smile, always letting me feel free to be myself.

Acronyms and Abbreviations

ARR	Associational Risk Ratio
BMI	Body Mass Index
CI	Confidence Interval
CRR	Causal Risk Ratio
DAG	Directed Acyclic Graph
GOCS	Growth and Obesity Cohort Study
GXXI	Generation XXI
HR	Hazard Ratio
ICP	Infancy - Childhood - Puberty
JB	Jenss-Bayley
LBW	Low Birth Weight
MAR	Missing At Random
NINFEA	Nascita e INFanzia: gli Effetti dell'Ambiente
OR	Odds Ratio
PBR	Piedmont Birth Registry
ROR	Relative Odds Ratio
RR	Risk Ratio
SD	Standard Deviation
SEP	Socio Economic Position
SGA	Small for Gestational Age
SITAR	SuperImposition by Translation And Rotation

Preface

This is a research paper style thesis and therefore consists of a series of publications, preceded by some introductory chapters and followed by a conclusion. In total four research papers, one commentary and a book chapter are included. I am the first-author of all the four research papers, and a co-author of the commentary and the book chapter. The commentary, the book chapter and three of the four research papers have already appeared or have been accepted for publication, while the last research paper has been submitted and is currently under review. This Ph.D. addresses two main methodological challenges that may arise in the design and analysis of life course studies of infant growth: the potential selection bias due to selection of cohort study participants and the modelling of individual growth trajectories. As a consequence, the publications included in this thesis focus on these two separated strands of research. The thesis is composed of three parts: 'Introduction' (Part I), 'Selected Publications' (Part II) and 'Discussion' (Part III). Part I consists of three chapters. Chapter 1 gives a brief introduction to the epidemiological motivation for life course studies of growth in infancy, with particular focus on the existing literature on the main exposures influencing growth during infancy. The aims of this thesis are presented in Chapter 2, while in Chapter 3 a description of the three datasets used to address the research questions is provided. Part II includes the selected publications, which are divided into two chapters. Chapter 4 includes the contributions focused on selection bias: these consist of two research papers, the commentary and the book chapter. These publications are preceded by a brief introduction. A summary overview of the alternative models suggested in the anthropometric and statistical literature to describe growth in infancy is provided as an introduction to Chapter 5, where the two research papers dedicated to growth modelling and

the study of its predictors are included. Part III consists of the conclusive Chapter 6, in which final considerations and areas of future work are discussed.

Part I

Introduction

Chapter 1

Background

Pregnancy is a highly susceptible period that can result in unfavourable outcomes for the baby, such as prematurity or low birth weight for gestational age, which in turn may have consequences for postnatal growth and a number of subsequent diseases. The concept that the causes of adult non-communicable diseases can be tracked back to fetal life, known as "Developmental Origins of Health and Disease", is at the core of life course epidemiology. It is built upon the notion that biological, social and environmental factors acting in early life affect the response of an individual to later environments and his/her susceptibility to diseases in adult life (Barker, 1998; Gluckman *et al.*, 2010; Kuh and Ben-Shlomo, 2004). The body of data supporting this hypothesis is now overwhelming, with several studies showing how birth size and early growth influence the onset and development of a wide range of chronic diseases, including cardiovascular and coronary heart diseases (Barker, 1998; Eriksson, 2011), diabetes (Whincup *et al.*, 2008), hypertension (Ben-Shlomo *et al.*, 2008; Gamborg *et al.*, 2009; Huxley *et al.*, 2000; de Jong *et al.*, 2012), obesity (Baird *et al.*, 2005; Monteiro and Victora, 2005; Ong *et al.*, 2000), cancer (De Stavola *et al.*, 2004), cognitive function (Raikkonen *et al.*, 2009; Yang *et al.*, 2011), as well as overall mortality (Crump *et al.*, 2013; Risnes *et al.*, 2011). Moreover there is evidence that the developmental environment induces effects that are transmitted to succeeding generations (Godfrey *et al.*, 2010). For example it has been shown that women exposed to a poor fetal environment, but to rich nutritional conditions later in life have an increased propensity to becoming obese and experience

gestational diabetes (Warner and Ozanne, 2010). These in turn are risk factors for their offspring's obesity risk in adult life (Fall, 2011). Another example of intergenerational effects is that of women exposed to diethylstilbestrol during their fetal life, who have been demonstrated to be at greater risk of preterm delivery as well as other adverse pregnancy outcomes (Hoover *et al.*, 2011). Prematurity in turn is likely to increase the offspring's risk of adverse outcomes in adulthood, including cardiovascular and metabolic diseases (Roggero *et al.*, 2013). Adopting a developmental perspective may therefore guide the implementation of timely interventions to improve the health of the future generations. This will also require the assessment of whether the associations of birth size and early growth with later outcomes are modified by factors that influence intrauterine growth as well as by postnatal factors that influence early life growth.

Understanding the latter, and in particular postnatal catch-up growth, has been given much attention in the last decades. Catch-up growth is defined as that process whereby children small at birth compensate for intrauterine restraint with a rapid postnatal growth, returning within the first two years of life to their genetic trajectory (Tanner, 1986). This type of growth pattern is very important as it has been extensively demonstrated that children exposed to a rapid postnatal weight gain have an increased risk of become obese (Baird *et al.*, 2005; Monteiro and Victora, 2005; Ong and Loos, 2006; Ong *et al.*, 2000; Tzoulaki *et al.*, 2010) and of other related adverse outcomes (van der Gugten *et al.*, 2012; Huxley *et al.*, 2000; Leunissen *et al.*, 2009, 2012). As a consequence there is an increasing interest in evaluating potential predictors of rapid weight gain in infancy (Batista *et al.*, 2012; Mihrshahi *et al.*, 2011; Ong *et al.*, 2000, 2002; Wijlaars *et al.*, 2011) as well as early life determinants of overweight and obesity later in life (Dubois and Girard, 2006; Fall, 2011; Monasta *et al.*, 2010; Ong, 2010). In particular Ong et al (Ong *et al.*, 2000) showed that the factors affecting catch-up growth were those related to intrauterine growth restriction, in particular maternal smoking during pregnancy, parity and parental size.

The majority of the studies that have investigated the relationship between birth weight and size or growth in infancy with later outcomes have defined infant/childhood growth in height or weight in terms of the difference between measurements taken at two fixed time points, limiting therefore the understanding of how particular features of the growth pattern may contribute to the outcome of

interest. Whilst for example the relationship between the adiposity rebound, a concept introduced by Rolland-Cachera et al (Rolland-Cachera *et al.*, 1987) to describe the period between 5 and 7 years of age when body mass index (BMI) begins to increase following a minimum, and later obesity has been studied widely (Ohlsson *et al.*, 2012; Rolland-Cachera, 1998; Williams and Goulding, 2012), similar features of the infant growth curve have been examined only more recently. Silverwood et al (Silverwood *et al.*, 2009b) studied the association between the infant BMI peak, defined as the maximum reached by the BMI trajectory around age 9 months, and BMI later in life, finding that both BMI and age at BMI peak were positively associated with BMI in adulthood. More recently much attention has been given to the role of these and other developmental milestones, such as age at menarche and pubertal growth spurt, in influencing the risk of later diseases (Ohlsson *et al.*, 2012; Ong *et al.*, 2012; Prentice and Viner, 2012; Silverwood *et al.*, 2009b). As a consequence there is also an increasing interest in identifying factors influencing the timing of such milestones (DAloisio *et al.*, 2013; Deardorff *et al.*, 2012; Hui *et al.*, 2012; Ong *et al.*, 2012; Wehkalampi *et al.*, 2011).

Life course models (Ben-Shlomo and Kuh, 2002) have also highlighted the potential importance of risk factors acting in early life. In particular Dietz (Dietz, 1994) recognized the existence of three critical periods for the development of obesity, identifying gestation and early infancy as the first relevant stage. The importance of the first two years of life was also stressed by Cole (Cole, 2000), who showed that the increase in adult height observed from one generation to the next (secular trend) occurs mainly within this period, with height at 2 years best predicting adult height.

The first periods of life, including fetal life, could thus be a window of opportunity for intervention. There is growing consensus that the study of the factors influencing fetal and infant growth, especially within the first 6 months of life (Botton *et al.*, 2008) - the period of fastest growth in the entire life span, and hence most susceptible to adverse condition - is of major importance for the prevention of many later diseases, especially obesity in adolescence and later in adulthood. However, studies of early growth and of its predictors should take into account the complexity of the infant growth process. A review of the main aspects related to the study of growth in infancy is provided below.

1.1 Growth in Infancy

Human growth can be defined as the series of changes to the size, shape and functions of an individual organism which occur over time from conception to maturity (Hauspie, 1998). Its general postnatal growth has been classified in five stages: infancy, childhood, juvenile, adolescence (or puberty) and adulthood. Individual organs, tissues and growth dimensions have been observed to expand at different rates according to these periods (Bogin, 1998). More specifically, postnatal linear growth (that is growth in length/height) can be divided into three phases, infant, childhood and puberty periods, and their features have been modelled by Karlberg in the infancy-childhood-puberty (ICP) model (Karlberg, 1998). It characterises a high growth rate immediately after birth followed by a rapid deceleration until 3 years of age. After that, the velocity slows down until the onset of puberty, when the growth rate starts increasing again. From a biological point of view, linear growth during infancy has been considered as a continuation of the intrauterine growth: in utero this is regulated by placental functions, while after birth it is the outcome of the interplay between food supply and growth hormones (Weaver, 2006). Weight growth generally follows the same pattern, however it is characterised by more fluctuations that may include decreases. As for linear growth, postnatal weight growth is also regulated by nutrition. Most infants with small size at birth experience rapid postnatal weight gain (catch-up growth), and, as discussed above, the latter has been consistently associated with increased risk of subsequent obesity (Baird *et al.*, 2005; Monteiro and Victora, 2005; Ong and Loos, 2006; Tzoulaki *et al.*, 2010). However, low birth weight has also immediate adverse effects. Indeed it has been shown that not inducing catch-up growth in these children, either via improved nutrition or via growth hormone therapy, is associated with an increased risk of postnatal infections and mortality (Ong, 2007). The balance between the disadvantages and the benefits of postnatal catch-up growth has been described in the literature as 'the catch-up dilemma'. Weaver (Weaver, 2006) has distinguished between two types of postnatal rapid growth: (i) catch-up growth, which follows in utero growth restriction and which consists of linear and muscle growth together with fat accumulation, and (ii) accelerated growth, which may occur at any age due to overfeeding, and which implies fat accumulation without linear growth. The former process (catch-up growth) could be beneficial, while accelerated growth could be the mechanism leading to subsequent obesity (Weaver, 2006). In order to

provide valid guidelines for postnatal nutritional strategies directed at children with small size at birth, Ong suggested that "the concept of 'healthy catch-up growth' should be the goal of future research" (Ong, 2007).

When referring to infancy it is important to acknowledge that the term infant derives from the Latin word "infans", meaning "unable to speak" and it is typically applied to babies up to 24 months of age. However, definitions of human infancy vary considerably, and according to the specific context. While in the ICP linear growth model (Karlberg, 1989), the infancy component is assumed to start during fetal life and end at 3 years of age, from a natural science perspective (Bogin, 1998) human infancy is associated with the lactation phase, which ends when the child is weaned from the breast and thus corresponds to a less precise time point.

1.2 Assessment of Growth

The older and simpler method for describing the growth of an individual is anthropometry, that is the measurement of the body. Human growth is indeed usually expressed as the change in any one of many anthropometric dimensions. We now review the main ones, distinguishing between prenatal (fetal) and postnatal growth.

1.2.1 Prenatal growth

As already discussed above, evidence of intrauterine growth is usually assessed at birth, typically through direct measurement of birth weight. Since this is strongly determined by the gestational age at delivery, it is important to analyze and interpret birth weight conditionally on gestational age. Many studies, indeed, define a restricted growth newborn as that infant with a birth weight smaller than 'normal' for its gestational age, i.e. small for gestational age (SGA). The usual approach to adjust birth weight for gestational age is via standardisation of the weight measurements using a reference distribution to generate z-scores or Standard Deviation-scores (SD-scores). Specifically, these are created by subtracting the mean value of a standard distribution corresponding to the gestational age of the newborn from the observed value for that newborn and then dividing this measure by the SD of the reference distribution for that gestational age. This score can be interpreted as a measure

of distance (in term of SDs) between the observed measure and the centre of the reference population. Usually the birth weight distribution is further conditioned on other factors, in particular sex, as it is known that birth weight strongly differs also according to gender. The reference distribution could be internal, when the mean and SD of the observed distribution are used (appropriate for large samples), or external, when a separate dataset is selected to be representative of the population from which the sample originated. SGA infants are then identified as those with SD scores less than the 10th percentile of the standardised birth weight distribution (Goldenberg and Cliver, 1997). Problems may arise when the chosen standard population is inappropriate for the analyzed dataset (see Section 1.2.2).

It should also be acknowledged that failure to achieve accurate measurements of gestational age at birth is frequent and might be due to several factors (Salomon *et al.*, 2010). Gestational age is indeed usually derived by estimating the date of conception using the last menstrual period and assuming a regular 28-day menstrual cycle. However this approach may be limited by memory bias or by irregular menstrual cycle, which would lead to inaccurate dating of the ovulation. An alternative approach is the use of ultrasound examination of anthropometric dimensions such as crown-rump length and biparietal diameter in the first or second trimester of pregnancy. The latter has proven to be more reliable in predicting the date of conception (Salomon *et al.*, 2010).

1.2.2 Postnatal growth

Height and weight are the two most widely used dimensions in growth studies, with height being a composite of linear dimensions and weight a composite of varying tissues, used as an estimate of the total mass of the body (Malina, 1998). Percentage fat, fat mass and fat-free mass are usually indirectly estimated from skinfold thickness measurements (i.e. triceps, biceps, subscapular, abdominal, suprailiac), which are sensitive to the extremes of adiposity, using previously developed prediction equations (Cameron, 1998). When these are not available, adipose tissue in infancy is mostly assessed on the basis of BMI (Cole, 1986), though this index does not distinguish between fat and muscle mass and is influenced by body build and lean tissues (Cole, 2002; Rolland-Cachera *et al.*, 1987).

When studying growth, it is crucial to consider which aspect of the growth process we want to capture: absolute weight and height may describe the size of a subject, but if we are interested in exploring

change in body-shape we need measurements that are independent of size. As regards weight, this can be achieved by calculating either of these weight for height indices:

- relative weight: ratio of the observed persons weight to a reference weight for a person of his height and sex (or other standardising variables).
- indices of the form $\frac{W}{H^p}$, such as the weight-height ratio (p=1), the BMI (p=2), the Ponderal Index (p=3), or the Benn Index (where p is allowed to take a non-integer value) (Benn, 1971), choosing the index with the smallest correlation with height. The p that better meets these criteria in infants and children differs with age and gender. During infancy it has been shown that the optimal value of p increases in the first months of life becoming greater than 2 and then slightly decreases again until about 20 months (Gartside *et al.*, 1984). Further, in order to use the Benn index with infants an optimal value of p could be calculated for different narrow periods; however it has been argued that Benn's assumption of small correlation between index and height holds in adults but may be violated in children (Lazarus *et al.*, 1996).

In order to account for temporal changes, these indices may need to be adjusted for age, using a standard population and creating a relative index. An alternative is the standardisation of the anthropometric measurements to generate z-scores or SD-scores, as already described for birth weight for gestational age. This approach is valid when the variables are normally distributed, if not a transformation should be first used. Alternatively the LMS method proposed by Cole (Cole and Green, 1992), could be applied: this is a general approach for time-dependent variables in which median, variability and skewness of the distribution are all allowed to change over time. When using standardisation, the choice of the reference population is a crucial issue. As for birth weight, this could be internal (appropriate for large sample), or external. Usually sex- and age-specific standards are used. Problems may arise when the chosen standard population is inappropriate for the analyzed dataset: for example when unknown differences in the sex- and age-specific distribution between the reference data and the observed data exist; or when the reference data refer to a different time period with respect to the observed data (Silverwood *et al.*, 2009a). Similar considerations may apply to geographical aspects, or to ethnicity and socio-economical status issues (Hermanussen *et al.*, 2012).

As stated above, anthropometric dimensions needed to investigate growth vary according to the aspect

of the growth process/growth disorder we are interested in. It has been suggested that, during infancy, weight is more useful than length to assess poor growth, as well as to identify children experiencing a catch-up growth (Cole, 2002). In contrast height is considered the most important indicator to assess poor growth beyond infancy (Cole, 2002). Moreover obtaining accurate measures of length in infancy usually requires more sophisticated techniques and better trained personnel than those needed to properly measure weight (Cole, 2002). Due to these reasons in this thesis weight measurements are used to assess growth in infancy. This decision is also driven by the fact that growth data of the children in one of the cohorts available for this Ph.D. are gathered from questionnaires completed by their mothers, and the accuracy of the length measures for this cohort is lower than that of the weight measures (see Chapter 3).

As a consequence growth in weight is the main focus of the following sections.

1.3 Factors Influencing Growth

As stated in the introduction, it has been shown that several exposures occurring prior to and during gestation are correlated with pregnancy outcomes and in particular birth weight. These together with other postnatal exposures have also been found to be associated with growth during infancy. We now review the main ones.

1.3.1 Prenatal growth

Maternal constraint, defined as that process by which a series of maternal and placental non-genetic factors limit fetal growth (presumably as a reflection of limited nutrients supply), acts in all pregnancies. However it has been argued that its effect is greater among those involving mothers who are younger, smaller and of lower parity (Gluckman and Hanson, 2004).

In 1987 the World Health Organization carried out a review of the studies, published in the period 1970-1984, that concerned potential determinants of intrauterine growth or gestational duration, upon which weight at birth depends (Kramer, 1987). According to this review, cigarette-smoking is the largest direct causal factor for birth weight, followed by low caloric intake or low weight gain during pregnancy. The other main determinants that were identified are maternal pre-pregnancy weight,

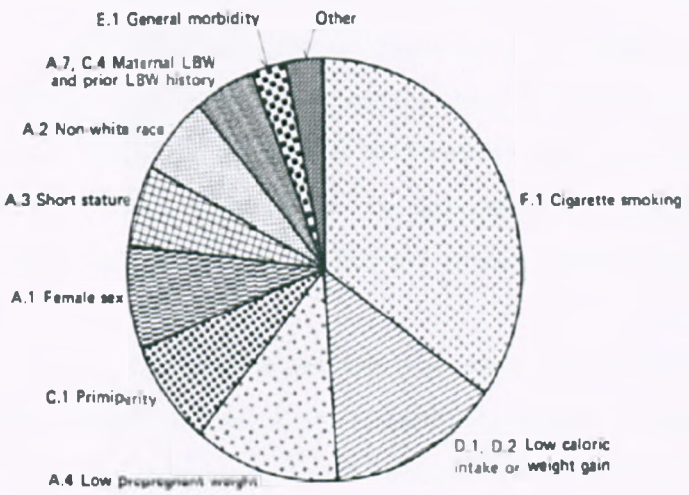


Figure 1.1: Factors with a causal influence on low birth weight in developed countries. From Kramer, 1987 (Kramer, 1987).

parity, infant sex, maternal height, ethnicity, maternal low birth weight and previous low birth weight children (see Figure 1.1). Few years later, Brooke et al (Brooke *et al.*, 1989) confirmed that, among numerous potential factors including alcohol, caffeine, socioeconomic and psychosocial characteristics, smoking during pregnancy has the most important effect on birth weight for gestational age. Several other studies showed that mothers who smoke cigarettes during pregnancy have babies weighing less than those of non-smoker women, and a dose-response relationship was also established (Britton *et al.*, 2013; Erickson and Arbour, 2012; Nobile *et al.*, 2007; Ong *et al.*, 2002; Schell, 1998).

Recent studies confirmed many of the other associations reported by Kramer in 1987, in particular with parental anthropometry - especially from the maternal side - parity, maternal pre-pregnancy weight and pregnancy weight gain, socioeconomic status and infant sex. For example, mother's leg

length in childhood was found to be positively associated with children's birth weight independently of maternal birth weight and adult height (Martin *et al.*, 2004). Knight et al (Knight *et al.*, 2005) showed that paternal height, but not BMI, influences size at birth, in particular length and measures of skeletal growth, while both maternal height and BMI are associated with offspring birth weight and length. A positive association between paternal height and birth weight was also found by Shah et al (Shah, 2010). Similar results were observed by Griffiths et al (Griffiths *et al.*, 2007), who reported that contribution of parental height on offspring birth weight is similar and significant, while maternal weight is more influential than paternal weight. The intergenerational (parental-offspring) correlation in size at birth has also been observed by more recent studies (De Stavola *et al.*, 2011; Selling *et al.*, 2006; Shah, 2010).

A huge number of investigations in the last few years have focused on the role of maternal pre-pregnancy overweight/obesity, maternal weight gain during pregnancy and related adverse conditions, such as gestational diabetes, on fetal growth and birth outcomes. According to the fetal overnutrition hypothesis greater maternal adiposity during pregnancy alters the fetus developmental environment and in particular the transfer of glucose, free fatty acids and amino acids to the fetus, leading to permanent changes in the appetite control and metabolic systems of the offspring. The body of data supporting the association of maternal pre-pregnancy obesity, maternal diabetes and excessive weight gain during pregnancy with newborn macrosomia is now overwhelming (Catalano *et al.*, 2012; Mamun *et al.*, 2011; Metzger *et al.*, 2008; Roland *et al.*, 2012). Moreover there are studies showing that maternal underweight, and low weight gain during pregnancy negatively affect fetal growth, being associated with increased risk of preterm births and SGA infants (Campbell *et al.*, 2012; Dietz *et al.*, 2006; Jeric *et al.*, 2013).

It is also well-documented that mean birth weight increases with parity and that nulliparity increases the risk of SGA children (Campbell *et al.*, 2012; Catalano *et al.*, 2012; Erickson and Arbour, 2012; Nobile *et al.*, 2007; Ong *et al.*, 2002; Roland *et al.*, 2012). Similarly low maternal socioeconomic status, usually proxied by educational level, has been repeatedly associated with adverse pregnancy outcomes (Bouthoorn *et al.*, 2012; Dubois and Girard, 2006; Jansen *et al.*, 2009; Nobile *et al.*, 2007). The latter finding might be partly explained by the fact that women of low socioeconomic status are more likely

to be obese, to smoke during pregnancy and to suffer of hypertension (Jansen *et al.*, 2009).

In addition to the risk factors identified by the 1987 World Health Organization review (Kramer, 1987), other exposures have been related to fetal growth and birth outcomes. For example, the influence of maternal age has been investigated, although the evidence gathered from these studies is conflicting. While a number of researchers showed that young maternal age increases the risk of infant low birth weight or intrauterine growth restriction (Borja and Adair, 2003; Kirchengast and Hartmann, 2003; Malabarey *et al.*, 2012; Strobino *et al.*, 1995), other studies found that women having later age pregnancies (i.e. age $\geq$ 35 years) are more likely to experience adverse pregnancy outcomes (Campbell *et al.*, 2012; Kenny *et al.*, 2013; Koo *et al.*, 2012). Shah et al (Shah, 2010) also found a significant association between advanced paternal age and increased risk of offspring low birth weight. It has been argued that the association observed between young maternal age and low birth size may partly be explained by factors that are proxies for social class differences and behavioural characteristics that underlie early childbearing. Moreover it has been suggested that biological immaturity of the teenager mother may also influence the offspring birth size (Scholl, 1998).

An association between alcohol intake during pregnancy and birth weight has also been established, even if with less support than the one for smoking. Overall, studies found that intake of one or more alcohol drinks per day during pregnancy leads to reduced offspring birth weight and occurrence of preterm delivery (Carter *et al.*, 2007, 2013; Feldman *et al.*, 2012; Goldberg, 1998). Furthermore, gestational hypertension/preeclampsia has been associated with reduced fetal growth and prematurity (Jacquemyn *et al.*, 2006; Spiegler *et al.*, 2013; Xiong *et al.*, 2002). A study which explored factors influencing pre- and postnatal growth identified several exposures associated with placental weight: modifiable factors (parity, smoking during pregnancy), and non modifiable factors (parental height, gender) (Hindmarsh *et al.*, 2008). Placental weight was shown to reflect mainly the effect of other determinants of fetal size, possibly playing a role of mediator in the association between many maternal characteristics and birth weight (Campbell *et al.*, 2012; Ouyang *et al.*, 2012; Roland *et al.*, 2012; Salafia *et al.*, 2008).

Finally in the last decade much attention has been given to the influence of air pollution on pregnancy outcomes, with studies providing some evidence that exposure to pollution during pregnancy increases

the risk of preterm birth, restricted fetal growth and low birth weight (van den Hooven *et al.*, 2012; Laurent *et al.*, 2013; Maisonet *et al.*, 2004; Olsson *et al.*, 2013; Padula *et al.*, 2012).

1.3.2 Postnatal growth

The influence of many of these factors persists after birth thus affecting growth during infancy. As already discussed, children born small are likely to compensate for their experience of intrauterine restraint with a rapid postnatal growth. In a recent twin study, Touwslager et al (Touwslager *et al.*, 2011) showed that the higher the birth weight the slower the postnatal growth rate, although it should be acknowledged that distribution of size and growth pattern of twins differ from those of singleton. In a study investigating the factors influencing catch-up and catch-down growth in height, defined as a change in z-score between child height z-score and birth length z-score greater than 0.67 and lower than -0.67 respectively, it was observed that intrauterine growth restriction (defined on the basis of the ratio between observed birth weight and mean weight for gestational age) in term infants is associated with increased risk of postnatal catch-up growth (Batista *et al.*, 2012). Since smoking during pregnancy has a strong negative effect on birth weight for gestational age (see previous section), it is not surprising that several researchers observed that children of smokers with a low weight at birth are more likely to quickly catch-up and then exceed their expected weight compared to children of non-smokers (Chen *et al.*, 2006; Howe *et al.*, 2012; Power *et al.*, 2003). In 2002, Ong et al (Ong *et al.*, 2002) claimed that infants of smoking mothers with a low birth size complete their catch-up within the first 12 months of life without being heavier or taller than infants of non-smoking mothers. However, the latter result partly conflicts with the evidence, gathered from several other studies, that smoking during pregnancy is a strong risk factor for childhood obesity (Dubois and Girard, 2006; Monasta *et al.*, 2010).
Females remain generally somewhat shorter than males all over infancy and childhood until adolescence, and are also typically a little lighter than boys (Tanner, 1989). Griffiths et al (Griffiths *et al.*, 2007) found that parental size influences both birth weight and infant weight gain with maternal weight having a greater influence on weight at birth than paternal one (as discussed before), while parental height and weight contributing equally to postnatal weight growth. It has been repeatedly observed that maternal height affects postnatal growth rate (Bocca-Tjeertes *et al.*, 2011; Botton *et al.*, 2010;

Lourenco *et al.*, 2012; Power *et al.*, 2003) being positively associated with growth velocity, especially in height. Botton et al (Botton *et al.*, 2010) observed that paternal BMI is instead associated with increased weight growth rate.

When investigating the factors influencing failure to thrive, Blair et al (Blair *et al.*, 2004) found parental height and parity to be respectively negatively and positively associated with slow postnatal weight gain. The effect of parity with postnatal growth, especially in weight, has been consistently observed (Batista *et al.*, 2012; Bulk-Bunschoten *et al.*, 2002; Hui *et al.*, 2010; Ong *et al.*, 2002; Wells *et al.*, 2011), with infants of primiparous pregnancies having faster postnatal growth.

As observed for birth weight, the body of data supporting the association of maternal pre-pregnancy obesity, maternal diabetes and gestational weight gain during pregnancy with growth in weight during infancy is large. Several studies reported a positive influence of these factors, especially maternal obesity, on either weight gain in the first months of life (Deierlein *et al.*, 2011; Regnault *et al.*, 2010), or increasing adiposity tissue (Modi *et al.*, 2011; Vohr *et al.*, 1999). Linabery et al (Linabery *et al.*, 2012) found an effect only at birth and after 1.5 years of life. Moreover, as for smoking, maternal pre-pregnancy obesity has been consistently numbered among the most relevant risk factors of offspring obesity later in childhood (Dubois and Girard, 2006; Monasta *et al.*, 2010). Little is known instead on the influence of gestational hypertension/preeclampsia on postnatal growth. A study evaluating the effect of pregnancy-induced hypertension on weight gain at 28 or 42 days postpartum in a Chinese birth cohort showed that infants of mothers with hypertension are at greater risk of postnatal catch-up growth only if babies experienced intrauterine growth restriction (Baulon *et al.*, 2005).

Much attention has been given to the role of socioeconomic status on postnatal growth in weight and height. While Howe at al (Howe *et al.*, 2011) found that the effect of maternal education on ponderal index is significant only after infancy, another study reported a significant inverse association between high socioeconomic status and weight velocity in the first months of life (Wijlaars *et al.*, 2011). An opposite result – increasing weight velocity during the first year of life at increasing level of maternal education – was instead observed among children belonging to an Hong Kong Chinese birth cohort (Hui *et al.*, 2010). Power et al (Power *et al.*, 2003) observed that infants of lower socioeconomic status are at higher risk of having a low birth weight and a subsequent high BMI in adult life, while Bocca-Tjeertes

et al (Bocca-Tjeertes *et al.*, 2011) found that deprived children are more likely to experience postnatal growth restraint in weight and height. Studies carried out in developing countries reported a positive association between high socioeconomic status and postnatal growth velocity in height (Kang Sim *et al.*, 2012; Lourenco *et al.*, 2012; Matijasevich *et al.*, 2012), while the opposite was observed in a Danish cohort (Silva *et al.*, 2012). Furthermore findings from a UK birth cohort showed that inequalities in height during childhood arise mainly through socioeconomic differential in birth length (Howe *et al.*, 2012). A recent study found that maternal depression reduces growth at 6 months of life, and the authors argued that this effect might be partly due to social deprivation status, and in turn to life-style factors such as smoking status and alcohol consumption (Traviss *et al.*, 2012).

The association between alcohol consumption during pregnancy and postnatal infant growth has been less studied. Carter et al observed that infants of mothers who drank moderately during pregnancy are on average lighter and shorter at birth with differences persisting up to 12 months of life, and with stronger effects for infants of binge-drinking mothers during pregnancy (Carter *et al.*, 2007, 2013).

Finally growth hormones have been suggested to have a minimal effect on postnatal growth, since infant growth up to 6 months is considered as a continuation from prenatal growth which is influenced mainly by nutrition whilst largely growth hormone independent (Mehta *et al.*, 2005). Mathematically (Karlberg, 1987) and clinically (Mehta *et al.*, 2005) it has been observed that infant growth begins to be hormone dependent after 6 months of age. Growth during infancy indeed has been suggested to be the outcome of the interplay between nutrition and the rate of growth programmed in pregnancy, mainly reflecting parental size (Cole, 2000). Breastfeeding is in fact the most relevant factor in the nutrition and growth of the infants (see below) (Binns, 1998).

During the first year of life height velocity is fast and energy needs are consequently high. It is well-documented that breastfed children have an adequate nutritional supply from birth until at least 4-6 months of life with breast-milk alone meeting the nutrient needs for an appropriate growth rate and body composition and better protection from infant morbidity (Binns, 1998; Kramer and Kakuma, 2012). Moreover introduction of supplementary food in the first 6 months of life has been shown not to provide any growth advantage (Dewey, 2001). After 6 months of age neither breast-milk nor infant formula alone provide enough energy for growth and evidence concerning the effect and duration of

breastfeeding on infant growth is conflicting, mostly depending on the population been investigated. A number of studies indeed claimed that children exposed to prolonged and exclusive breastfeeding experience a greater weight gain for the first weeks of life followed by a reduced rate of accretion so that at about 1 year of age breastfed children appear on average shorter and thinner relative to those bottle- or mixed-fed (Beath, 2007; Dewey, 2001; Hediger *et al.*, 2000; Victora *et al.*, 1998). Other studies restricted their evaluation to the first months of life, showing that exclusive breastfeeding is associated with reduced weight growth velocity (Bulk-Bunschoten *et al.*, 2002; de Hoog *et al.*, 2011; Mihrshahi *et al.*, 2011; Regnault *et al.*, 2010; Zhang *et al.*, 2012) or that shorter duration of exclusive breastfeeding is associated with increased growth rate (Betoko *et al.*, 2012). However Dewey (Dewey, 2001) argued that this is the result of infant self-regulation of energy intake rather than of nutritional deficits and other studies reported that infants who were exclusively breastfed for six or more months did not experience a deficit in weight or length gain after 12 months of age (Agostoni *et al.*, 1999; Kramer *et al.*, 2002, 2004). Haschke et al (Haschke and van't Hof, 2000) showed that duration of breastfeeding has an inverse association with length and weight until 24 months but not at 36 months of age, while in a recent study Woo et al (Woo *et al.*, 2013) observed that increasing duration of exclusive breastfeeding only modestly decreases the weight-for-age measure at 1 year. Furthermore, early introduction of solid food, associated with lower rate of ever been breastfed, was found to be a positive risk factor for weight gain during the first year (Kim and Peterson, 2008). Some authors argued that the negative association observed between breastfeeding and subsequent weight gain could be attributable to reverse causation. In a recent study Kramer et al (Kramer *et al.*, 2011) showed that, at least during the first 6 months of life, lower weight-for-age are associated with increased risk of weaning and of discontinuing exclusive breastfeeding in the following month. Based on these findings, the same author argued against a causal protective effect of exclusive breastfeeding on risk of later obesity, which has been reported by several studies (see a review of systematic reviews in (Monasta *et al.*, 2010)), suggesting that infants who grow slowly are satisfied by the breast-milk, while those growing faster request supplementary nutrition due to their increased growth rate (Kramer *et al.*, 2012). Other researchers argued that the negative association between breastfeeding and weight gain could be due to residual confounding mainly due to smoking and socioeconomic status. In particular

it was observed that breastfed infants have greater birth weight (Bulk-Bunschoten *et al.*, 2002), babies from the most affluent families are more likely to be breastfed (Fawzi *et al.*, 1998; Wright *et al.*, 2006) or to be breastfed for longer periods (Betoko *et al.*, 2013), and children of smoking mothers are more likely to be breastfed for a shorter duration (Nafstad *et al.*, 1997).

Finally, a recent twin study compared the genetic and environmental influences on three different dimensions of the weight trajectories in infancy: the size, the velocity and the tempo of growth (i.e. developmental age), where the latter is indicated by the age at peak weight velocity (Johnson *et al.*, 2011). The authors concluded that the size and velocity parameters are primarily driven by genetic factors, while the tempo dimension mainly depends on environmental determinants.

1.4 Final Comments

Fetal and early life growth are known to influence the onset and development of a wide range of chronic diseases later in life. The identification of exposures and conditions occurring during and before pregnancy which affect fetal an infant growth is therefore relevant for a life course perspective. However approaches commonly used to analyse growth data and their association with early exposures often focus on relatively simple aspects of growth, such as differences in size at pre-specified age intervals, and do not provide a comprehensive summary of the patterns of infant growth, limiting therefore the understanding of how prenatal factors relate to different dimensions of the growth process. The timing of developmental milestones has been given increasing interest in life course epidemiology. However, although infancy has been identified as a critical period for the development of diseases later in life, most of the attention has been focussed on studying adiposity rebound in childhood and height growth spurt in puberty. In this thesis I aim to examine features of growth during infancy, using methods that allow the identification of biologically meaningful features of growth trajectories in infancy, within a life course framework. Studying growth during this period should allow to identify early factors affecting infant growth and indicate how these factors affect specific components of the infant growth process (e.g. growth rate, timing of weight/adiposity peak), thus offering insights into the mechanisms governing infant growth and contributing to the understanding of its role in the development of a wide range of later diseases.

Chapter 2

Aims

This Ph.D. aims at modelling individual infant weight trajectories using three recently established cohort studies: the "Nascita e INFanzia: gli Effetti dell'Ambiente" (NINFEA) birth cohort study, based in Italy, the Geracão XXI (GXXI) birth cohort study, based in Northern Portugal, and the Growth and Obesity Cohort Study (GOCS), based in Chile.

Because NINFEA members are volunteers who participate only via a web-based system, analyzing their data also requires the consideration of selection bias problems. Hence the thesis first examines possible selection mechanisms by Monte Carlo simulations of cohort studies, assessing their impact on the crude estimate of the effect of interest, and then uses population registry data from the same region as NINFEA (i.e. data from the source population) to assess how selection into the web-based NINFEA birth cohort (the sample from the restricted source population) affects the confounding patterns for the associations of interest.

Alternative modelling approaches for the available growth data are then studied in the second part of this PhD. They are compared in term of their ability to validly represent the data and to address issues related to the structure of the measurements (number, irregularity and completeness of the observations) and to the nature (e.g. non-linearity) of the growth trajectories, while identifying biologically meaningful features of growth. Models suggested in the anthropometric and in the statistical literatures are evaluated, with particular focus given to their implementation within the random effects

models framework. The availability of information on several prenatal variables in each cohort is then exploited to study their association with growth trajectories using the SuperImposition by Translation And Rotation (SITAR) model, a shape invariant random effects model that is extended to allow for the inclusion of explanatory variables.

In summary the aims of this Ph.D. are:

1. Examine under which circumstances the selection mechanisms that drive recruitment into cohort studies may induce confounding in the selected sample and therefore may or not lead to biased estimates of effects of interest.

2. Modelling individual infant weight trajectories using three recently established cohort studies.

These are accomplished by addressing the following objectives:

(a) examine sample selection mechanisms for the recruitment of participants into cohort studies and their consequences for the estimate of an effect of interest;

(b) investigate how selection into the web-based NINFEA birth cohort affects the confounding patterns present in the source population;

(c) compare the ability of alternative growth modelling approaches to fit weight trajectories in infancy and to identify biologically meaningful features;

(d) assess whether the results of fitting these models are affected by the type of available data (age range, number and timing of follow-up);

(e) study the prenatal influences of weight trajectories in infancy and compare them across the three cohorts.

Chapter 3

Case Studies

The thesis examines data from three studies: the NINFEA web-based birth cohort study concerning children born in Italy, the GXXI birth cohort established in the Porto region of Northern Portugal and the GOCS cohort based in Chile. All cohorts involve children born in the new millennium.
This chapter describes the datasets, introduces statistical and epidemiological issues specific to these cohorts and provides a summary of the growth measurements available in each cohort. Since this thesis focuses on weight trajectories in infancy, a more detailed description of the weight data available in each cohort is provided. In the last section of this chapter a summary of the cohorts/datasets involved in the analyses of each Research Paper is given.

3.1 NINFEA Study

The NINFEA study is an ongoing web-based birth cohort, co-ordinated by the Cancer Epidemiology Unit of the University of Turin that aims at recruiting pregnant women via the Internet and following up their children.

3.1.1 Background

NINFEA started in July 2005 as a pilot study in the city of Turin, Italy, and since December 2007 has been extended to the whole nation. Recruitment is still on-going and will last until at least a sample size of 7,500 babies is achieved.

As stated in Richiardi et al (Richiardi *et al.*, 2007), "the source population of the NINFEA cohort includes all babies born to pregnant women who have enough knowledge of the Italian language and the use of the Internet to complete the online questionnaires. Members of the cohort are babies born to women in this source population who become aware of the study and volunteer to participate through registration on the study website (www.progettoninfea.it)". Recruitment therefore occurs during pregnancy, when the women complete the first online questionnaire, with participants being self-selected.

The study is advertised both through active and passive methods. Active advertisement is carried out in the Piedmont region (mainly in the city of Turin), and since July 2010 in the Tuscany region, with the following methods: (i) posters at the main hospitals of the city; (ii) leaflets enclosed with the results of laboratory tests and ultrasounds; (iii) leaflets distributed at the pre-delivery classes and (iv) leaflets enclosed with the baby health book. Therefore, this approach is inherently limited to selected geographical areas and targets; in order to enlarge the study to the whole of Italy at the end of the pilot phase the NINFEA study was also advertised online through: (5) links to NINFEA study website posted on the hospitals' websites and on websites dedicated to pregnant women; (6) participation in discussion forums related to pregnancy, and (7) a NINFEA page in facebook. It is also likely that a proportion of participants obtained the information from other members of the cohort.

The cohort is followed-up actively through questionnaires: three long Internet-based questionnaires administered during pregnancy (Q1), at 6 (Q2) and 18 (Q3) months after delivery, and then other short questionnaires, the first of which is administered when the children are 4 years old (Q4) and the second when they are 7 years old (Q5). It is important to notice that the time of administration of the questionnaires could vary within an established age range: for the Q2 between 12-24 months after conception (because contact times were defined from the time of the first questionnaire, when only expected date of delivery was known), for the Q3 between 16-30 months after delivery (or between

24-36 months after conception if date of delivery was not reported at Q2), for the Q4 between 46-59 months after delivery (no Q4 is administered if date of delivery is not gathered at Q2 or Q3) and for the Q5 between 82-94 months after delivery. The Q1 includes questions on several family and maternal characteristics (demographic, lifestyle, environmental and occupational exposures) as well as on the pregnancy and on the maternal reproductive and medical history. Moreover questions about the use of the Internet are included in Q1 (Richiardi *et al.*, 2007). With Q2 and Q3, information on maternal characteristics in the last part of pregnancy, pregnancy outcome, delivery, gestational duration and characteristics of the baby at birth are collected. Questions on feeding type, child diet, sleep pattern, child diseases and medicines, mother-father-child relationship and child development are included as well. The Q4 is a short questionnaire focused on the cognitive, functional and anthropometric development of the children, while the Q5 is focused on the respiratory health of the children. Since September 2008 biological samples of saliva of both the child and the mother are collected.

Because of logistic reasons recruitment and follow-up questionnaires have been adapted since the onset; thus questions slightly differ with year of recruitment, with the final version introduced in November 2008. NINFEA is an on-going study, therefore available data vary with time of data download and analyses reported in each paper included in this thesis have been performed using different versions of the NINFEA database (see section 3.4.1). All the downloaded datasets include both the pilot study and the updated version. The data used to generate Research Paper IV (included in Chapter 5) are described below, as this is the most up to date version of the data used in this thesis. This was downloaded in March 2012 and consists of 4,787 records; records wrongly generated by the system (n=1) as well as records relative to multiple pregnancies (n=151), miscarriages (n=38), babies who died at delivery (n=8), and records relative to women who registered after delivery (67) and for whom the Q2 was not opened, that is for whom conception occurred less that 12 months before the date of the data download (n=572), were excluded. These exclusions lead to a dataset of 3,950 records relative to 3,666 mothers.

Out of these, 816 records were excluded from the growth data analyses because weight measurements for the first 4 years of life (that is the age range analyzed in Research Paper III, see Section 3.4.1) were not available: 268 records were relative to children whose mothers completed the first questionnaires,

	Included N=3,134		Excluded from growth analyses: Lost to follow-up N=268		Not filled Q2 yet N=476		No growth data N=72	
	Mean±*SD* or %	*N*	Mean±*SD* or %	*N*	Mean±*SD* or %	*N*	Mean±*SD* or %	*N*
Maternal characteristics								
Age at registration	33.1 ±4.1	3,134	32.4 ±5.1	268	32.8 ±4.7	476	33.2 ±4.2	72
Pre-pregnancy weight (Kg)	60.9 ±11.4	3,043	60.2 ±10.6	217	61.1 ±10.8	443	59.8 ±10.9	65
Height (cm)	164.7 ±6.2	3,039	164.6 ±6.5	217	164.8 ±6.2	443	163.4 ±5.5	65
Pre-pregnancy BMI (Kg/m2)	22.4 ±3.9	3,031	22.2 ±3.6	217	22.5 ±3.7	443	22.4 ±3.9	65
Italian nationality	96.1%	3,013	95.9%	257	94.3%	449	97.2%	70
Graduated	58.4%	1,809	42.9%	106	50.0%	230	56.3%	40
First pregnancy	73.9%	2,250	57.9%	103	66.4%	300	62.3%	33
Infertility treatment	7.2%	218	5.3%	11	7.1%	31	10.6%	7
Smoking during pregnancy	8.6%	265	12.8%	31	10.3%	47	14.9%	10
Paternal characteristics								
Weight (Kg)	78.6 ±11.4	3,003	78.3 ±11.8	210	78.8 ±10.8	438	79.1 ±15.5	69
Height (cm)	177.7 ±6.8	3,024	177.6 ±6.9	212	178.3 ±6.7	439	177.9 ±6.8	68
BMI (Kg/m2)	24.9 ±3.1	2,992	24.8±3.3	210	24.8 ±3.2	438	24.5 ±2.7	68
Graduated	40.2%	1,219	30.5%	65	31.7%	139	33.3%	23
Italian native speaker	96.7%	2,932	95.3%	201	93.2%	408	94.2%	65

Table 3.1: NINFEA: Baseline characteristics stratified by rate of questionnaires completion.

	Weight growth data			
	Complete (N=1,937)		Partial (N=1,197)	
	Mean±SD or %	N	Mean±SD or %	N
Maternal characteristics				
Age at registration	33.2 ±4.2	1,937	33.1 ±4.2	1,197
Pre-pregnancy weight (Kg)	60.9 ±11.3	1,899	60.9 ±11.6	1,144
Height (cm)	164.6 ±6.0	1,895	165.0 ±6.4	1,144
Pre-pregnancy BMI (Kg/m2)	22.5 ±3.9	1,891	22.4 ±3.9	1,140
Italian nationality	96.6%	1,872	95.3%	1,141
Graduated	58.6%	1,126	58.0%	683
First pregnancy	75.8%	1,456	70.6%	794
Infertility treatment	7.3%	138	7.0%	80
Smoking during pregnancy	8.3%	159	9.1%	106
Paternal characteristics				
Weight (Kg)	78.6 ±11.7	1,891	78.6 ±11.0	1,126
Height (cm)	177.7 ±6.8	1,886	177.8 ±6.9	1,138
BMI (Kg/m2)	24.9 ±3.2	1,868	24.9 ±3.1	1,124
Graduated	38.3%	726	43.2%	493
Italian native speaker	96.6%	1,829	96.8%	1,103

Table 3.2: NINFEA: Baseline characteristics stratified by rate of weight data completion.

but were subsequently lost to follow-up; 476 to children for whom, at the time of the data download, the Q2 had not yet been completed although it could still have been completed; and 72 to children for whom (at least) Q2 was completed but no weight measurements was reported in the questionnaires. Participants with no growth data recorded were automatically excluded from the growth data analyses, while growth analyses carried out on those subjects with incomplete weight measurements (e.g. with weight measures available only for a subset of follow-up occasions) were based on the uninformative missingness (Missing At Random (MAR) assumption (Rubin, 1976)).

In order to check whether subjects included in the analyses are a representative sample of the whole cohort and to assess the validity of the MAR assumption and hence to investigate potential biases arising from the analyses of the subjects with incomplete weight data, baseline characteristics of the study-subjects stratified in the categories discussed above, were compared. Results are reported in Table 3.1 and Table 3.2. Baseline factors known to be potentially associated with birth size and infant growth were investigated; in particular parental size, nationality and educational level, maternal age at registration into the study, smoking and alcohol consumption during pregnancy, parity and use of infertility treatment.

There is some evidence of an association between maternal and paternal educational level and retention into the study, with the smallest proportion of graduated mothers and fathers observed among those lost to follow-up (43% and 30% respectively) and the highest among those who were included in the growth data analyses – see Chapter 5 – (58% of mothers and 40% of fathers). The distribution of nulliparous women follows the same pattern, with the prevalence decreasing from 74% among those included to 58% among those lost to follow-up. The distribution of maternal smoking follows instead the opposite pattern, with the proportion of women who smoked during pregnancy slightly higher among those lost to follow-up (13%) or without any weight measurements (15%). No major differences were highlighted when comparing the characteristics of those who had not completed the Q2 yet with those included in the analyses (Table 3.1). Although there is an indication for potential selection of the participants who remain into the study, mainly driven by socioeconomic factors, this might also be due to the small size of the group of those lost to follow-up.

When comparing those with complete with those with incomplete growth data, the distributions of

the prenatal characteristics appeared to be very similar (Table 3.2).

3.1.2 Growth measurements

All growth measurements of NINFEA children are gathered from the questionnaires that are completed by their mothers. Revisions of the questionnaires were undertaken after enrolment of approximately the first 1,500 participants, thus the available data vary with year of recruitment. In particular, women who enrolled until November 2008 were asked to report the child's weight and length at birth, 3 and 6 months at the time of the second questionnaire, as well as the head circumference at 3 and 6 months (Q2), while at Q3 they were asked to report the weight and length measurements of the baby at 12 and 18 months. When the Q2 and Q3 questionnaires were updated, a new question was introduced regarding the child's weight and height at the actual time of completion of these questionnaires and the source of information used to enter the growth measurements (i.e. whether the mothers used the child health book or they simply recalled the measure). In contrast Q4 was set up from the outset to include questions on anthropometric measures both at 4 years of age and at the time of completion of the questionnaire.

The response for the length/height measurements was lower compared to that for the weight measures. Although this was partly predictable, as measuring and recording length during the first 2 years of life is less common, further analyses highlighted an additional explanation. It appears to be that a number of mothers, when completing the information on their child's length, wrongly entered the head circumference measures instead. This observations, drawn from the data distribution, is supported by the fact that in the "Piedmont children health and growth book", completed whenever a child is examined by the paediatrician, weight and head circumference are recorded up to 2 years of age and it is highly likely that mothers referred to that book to answer and thus reported the head circumference value instead of length.

Table 3.3 shows a summary of the weight measurements available at fixed time points only (0, 3, 6, 12, 18 and 48 months of age) among the 3,134 NINFEA participants included in the growth data analyses, stratified by gender. As expected, newborn males are heavier than females and these differences persist during infancy. The birth weight data are more complete than the measures at 3 and 6 months, all

reported at the time of completion of Q2. We should stress that the 3 months and 12 months growth measures were reported at 6 and 18 months respectively (when the follow-up questionnaires were administered) and thus may be affected by recall bias. The same may apply to birth weight, also collected in Q2 – when the child was about 6 months old, but the extent of this bias is expected to be much lower, as mothers tend to remember correctly their children's birth weight. The extent to which measurements error affects the NINFEA birth data is the focus of the next section.

	Weight (kg)					
Follow-up age	Males (n = 1,585)			Females (n = 1,549)		
(month)	N	Mean	SD	N	Mean	SD
0	1,527	3.33	0.48	1,509	3.18	0.46
3	1,301	5.98	0.82	1,239	5.52	0.72
6	1,230	7.89	0.93	1,198	7.26	0.85
12	891	10.12	1.15	904	9.33	1.07
18	898	11.71	1.26	902	10.87	1.24
48	357	16.89	2.20	340	16.08	2.30

Table 3.3: Weight measurements of NINFEA participants by gender.

3.1.3 Link to the Piedmont birth registry

When establishing the NINFEA study, the major concern raised by funders and ethics committees was the validity of a cohort in which participants are recruited using the internet. The main issue is thus one of **selection bias**, due to the restriction of the source population from which the study-subjects potentially arise. This bias may indeed occur as participants in the NINFEA study differ from the population of pregnant women both because the source population is restricted to internet users and because they volunteer to participate.

Moreover **measurement error** is likely to affect the growth measurements recorded in the NINFEA study, as these are both self-reported and in part collected retrospectively.

In order to address both these issues a probabilistic linkage of the NINFEA dataset downloaded in July 2010 (that is with a previous version of the dataset than the one used for the growth data analyses of this Ph.D.), with the Piedmont Birth Registry (PBR) for the period 2005-2008 was carried out. This was done after exclusions of multiple pregnancies, births occurred after 31 December 2008 and

births that had occurred outside Piedmont. The birth registry recorded information both specific to the parental characteristics and to the pregnancy. In particular data on maternal and paternal age at delivery, educational level and occupation were available together with maternal smoking and alcohol consumption during pregnancy as well as intake of folic acid. Maternal parity, occurrence of previous miscarriages and use of infertility treatment to aid conception, type of delivery, gestational age, birth size (weight, length and head circumference), baby health status at birth and feeding pattern soon after birth were also recorded in the registry data.

These linked data formed the basis of the assessment of the potential selection bias affecting NINFEA. This is the topic of Research Paper II (Pizzi *et al.*, 2012), which is included in Chapter 4.

Validation of NINFEA birth data

The linkage with the birth registry data was also used in order to assess the extent of any measurements error affecting the NINFEA birth data quality. In particular a validation of the birth weight and gestational age measurements was performed. The probabilistic linkage was successful for 1,298 births, but births with gestational age earlier than the 25th week or later than the 44th week, as well as those with an implausible birth weight in either dataset, were further excluded, reducing the linked dataset to 1,295 births. Due to missingness among the NINFEA data, validation of the birth weight variable was carried out on 1,160 subjects, while validation of gestational age on 1,219. Bland-Altman plots (Martin Bland and Altman, 1986) were used to assess agreement between the two data sources (PBR and NINFEA). In particular it was assessed whether there is a systematic bias in the NINFEA measurements and whether the size of the difference between the two measurements is approximately constant throughout the range of measurements.

Birth weight A Bland-Altman plot of birth weight measurements is shown in Figure 3.1. The 95% limits of agreement, given by the mean difference plus or minus 1.96 times the standard deviation of the differences, goes from -189.8 gr to 192.8 gr, with the mean difference being 1.46 gr. The 95% confidence interval of this difference is (-4.2; 7.1) and correlation between the difference and the mean of the two set of measurements was 0.05 (p-value=0.06). These results suggest that there is no systematic bias in the birth weight measures self-reported by the mothers.

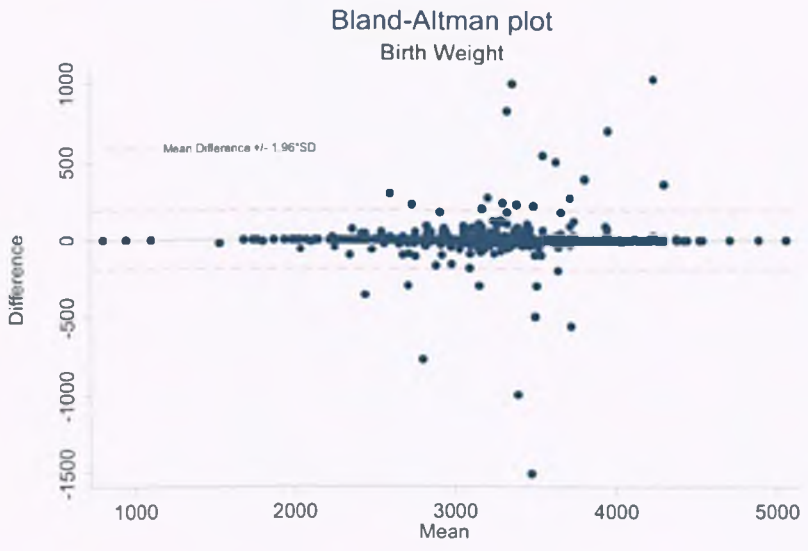


Figure 3.1: Bland-Altman plot of birth weight data collected in NINFEA and in the Piedmont Birth Registry.

Gestational age Figure 3.2 display the Bland-Altman plot for the gestational age at birth in weeks, where each point has been weighted by the absolute frequency. The mean difference is equal to 0.56 (with a 95% confidence interval of (0.51; 0.62)), meaning that on average the gestational age reported in the NINFEA questionnaire is half a week greater than the one recorded in the PBR. The 95% limits of agreement of the Bland-Altman plot (Figure 3.2) range from -1.3 to 2.4 weeks. Correlation between the differences and the mean was 0.07 (p-value=0.01). Overall there is an indication of a light tendency for the women to over-report their gestational age when filling the NINFEA questionnaire, compared to registry data.

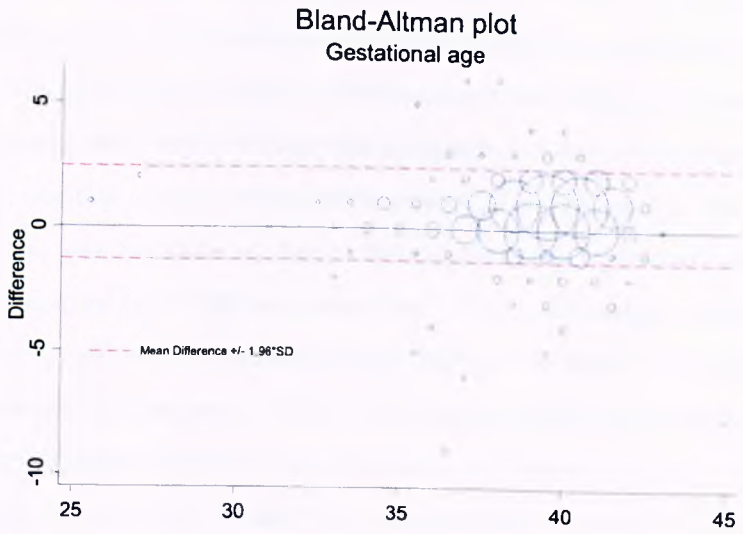


Figure 3.2: Bland-Altman plot of gestational age measures collected in NINFEA and in the Piedmont Birth Registry.

3.2 Geracão XXI Study

The GXXI study is the first Portuguese birth cohort, carried out by the Department of Hygiene and Epidemiology of the University of Porto. It aimed to recruit approximately 10,000 newborn babies and followed them up from birth to young adulthood in order to examine pre- and postnatal growth and development and relate these to health outcomes later in life.

3.2.1 Background

The cohort was established in 2005 in the Porto region of Northern Portugal, the second largest city in Portugal. All live children born to women resident in one of the six Districts included in the Porto region, admitted to one of the five public hospitals in Porto for delivery with a gestational age at birth greater than 24 weeks were eligible to participate. The recruitment period lasted from the end of April 2005 until August 2006. Blood collection was also planned but this was slightly delayed and actually

started in March 2006. A group of 1500 mothers, included in the main cohort, was recruited before delivery, if the first prenatal visit took place within the twelfth gestational week. They were actively followed during their pregnancy with various developmental and dietary data collected at regular times. The remaining women were enrolled during an appointment few days before their due date and, for the majority, baseline questionnaires were completed between 24 and 72 hours after delivery; these included three questionnaires: one for the baby, one for the mother (both completed by the mother), and one for the father (completed by the fathers themselves). With the baseline questionnaires, information on parental socioeconomic and demographic characteristics, reproductive history, disease history, diet and life-style characteristics, pregnancy, delivery and anthropometric history were collected.

Children were actively followed-up through questionnaires planned at 3, 6, 12/15, 24 and 48 months. Due to logistic and financial limitations it was not possible to interview every participant at each follow-up visit. Therefore a restricted time period was allocated for each follow-up occasion (cross-sectional interviews). The 3 months interviews were performed on a small random sample of the cohort by telephone. The 6-15 months evaluations were carried out when mother attended the local reference Health Centre where trained interviewers administered the questionnaires; while for the 2-year follow-up, mothers were offered an incentive of a free dental or eye check to attend this appointment. At this occasion child growth data (weight, height and head circumference), collected prospectively in the parent held health records during each child health professional examination, were gathered for entry into the database. Follow-up children questionnaires recorded information on feeding and diet habits, sleep and crying pattern, health status, environment characteristics and baby development.

The data available for the Ph.D. analyses are those relative to the baseline questionnaires (the child and mother's, not the father's one) for the whole study and those collected at the 2-year questionnaire for about 1,000 participants. In total, 8646 infants were born to mothers in the Porto study during the study period and were recruited, representing approximately 95% of those eligible. Among these, 290 babies from multiple pregnancies were excluded and a further 45 newborns were excluded because of missing mother's baseline questionnaire. The remaining baseline data consist of 8,311 singleton children. At the 2-year follow-up, data for 826 infants were available, of which 794 from singleton pregnancies. Data from both baseline (child and mother) and follow-up questionnaire were available

for 786 subjects (8 subjects with follow-up data but without baseline questionnaire), while for 7,525 participants only baseline characteristics are currently available.

Growth data analyses were therefore carried out in the 786 children who participated to the 2-year follow-up. As already discussed above, women were offered to attend this appointment with an incentive of a free dental or eye check; moreover a restricted time period was allocated for the 2-year follow-up interviews. In order to check if the probability of accepting the appointment was influenced by specific factors, i.e. to check for the potential occurrence of a selection mechanism, or to check if the babies offered to participate in the 2-year follow-up were different from the rest of the cohort, baseline maternal and newborn characteristic of these 2 groups were compared. Results are reported in Table 3.4: according to the baseline factors investigated, no major differences were found and thus it is reasonable to consider the 786 subjects who completed the 2-year questionnaire a representative sample of the whole cohort.

3.2.2 Growth measurements

As already reported in the previous section, at the 2-year follow-up interview, the child growth data, collected prospectively in the parent held health records during each child health professional examination, were gathered for entry into the database. For each subject, weight, height and head circumference values measured at about 1, 2, 4, 6, 9, 12, 15, 18 and 24 months of age were recorded, together with the dates when the measurements were actually taken. The measurements collected at birth, at 12/15 and 24 months, however, differ from the others, as these were measured directly from the interviewers of the GXXI study. So it is important to acknowledge that the source of the latter measurements is different and these are more standardised. Moreover up to 6 additional anthropometric measurements, taken routinely by the health professionals whenever the babies went to the doctor, with the corresponding examination date were entered into the database. Thus, anthropometric data collected were up to 16 time points (at birth and then 9 other fixed occasions plus 6 additional), ones for each of the growth dimensions.

	With 2-year follow-up (n=786)		Without 2-year follow-up (n=7,525)	
	Mean ±SD or %	*N*	*Mean ±SD or %*	*N*
Maternal characteristics				
Age	30.4 ±5.0	785	29.4 ±5.7	7,512
Height (cm)	161.6 ±5.9	670	160.6 ±6.2	5,767
Pre-pregnancy weight (Kg)	62.2 ±11.7	750	61.6 ±11.4	7,321
Pregnancy weight gain (Kg)	13.0 ±5.1	585	12.9 ±5.5	5,853
Birth weight (gr)				
<2500	8.8%	21	7.2%	160
2500-4000	81.9%	194	86.0%	1,923
>4000	9.3%	22	6.8%	151
Years of education >12	28.2 %	217	23.3%	1,737
Have a partner	94.4%	736	93.4%	6,996
Employed	76.5%	530	70.0%	5,093
Partner employed	94.1%	622	93.8%	6,535
No previous pregnancy	50.5%	394	47.7%	3,579
Nulliparous	62.6%	490	59.4%	4,451
Use of infertility treatment	2.5%	19	1.7%	129
Smoking during pregnancy	18.8%	145	22.4%	1,666
Child's characteristics				
Birth weight (gr)	3,185 ±484	758	3,191 ±490	7,394
Birth length (cm)	48.9 ±2.3	724	48.8 ±2.2	7,062
Birth head circumference (cm)	34.3 ±1.5	697	34.3 ±1.5	6,892
Gestational age (weeks)	38.7 ±1.6	738	38.6 ±1.8	7,226
Male gender	50.7 %	385	51.2 %	3,775
Natural childbirth	53.1 %	399	50.2 %	3,733
Malformation at birth	1.5%	10	1.3%	87

Table 3.4: GXXI: Baseline characteristics stratified by 2-year follow-up participation.

Among the 786 children considered, 3 were found to have missing values for all the follow-up growth variables and were therefore excluded from subsequent analyses. Table 3.5 shows descriptive statistics of the weight measures recorded. As shown in Table 3.4, at the time the GXXI data were made available for this Ph.D. gender information was missing for 26 babies.

	Weight (kg)					
Follow-up age	Males (n = 384)			Females (n = 373)		
(month)	N	Mean	SD	N	Mean	SD
0	378	3.23	0.50	370	3.15	0.45
1	366	4.12	0.69	352	3.87	0.59
2	341	5.19	0.75	342	4.82	0.68
4	342	6.76	0.87	338	6.27	0.81
6	333	7.87	0.96	333	7.36	0.89
9	303	9.11	1.15	292	8.54	1.01
12	316	10.09	1.24	309	9.52	1.12
15	265	10.89	1.33	270	10.36	1.20
18	266	11.61	1.43	256	11.05	1.38
24	248	12.83	1.58	236	12.39	1.52

Table 3.5: Weight measurements of GXXI participants by gender.

As expected, the proportion of missing values increases with time in each gender group and therefore mean values may be affected by selection bias. Weight data are available for all the time points for about 33.5% of the females babies and 29.5% of the male babies. As observed for NINFEA participants, newborn males are heavier than females with differences persisting all over the infancy period. Only for 19 participants information on 6 extra growth measurements were recorded, however data on at least one additional visit were available for 378 babies. Among these, about 22% concerns the first 3 months of life, more than 55% covers the first year and only 5% were collected when the children were older than 2 years.

3.3 Growth and Obesity Cohort Study

GOCS is an on-going Chilean cohort aiming to follow the children until the end of puberty to study the association of early growth with children's maturation, adiposity and associated metabolic complications.

3.3.1 Background

The study was initiated in 2006, when all children aged 3.0-4.9 years who attended nursery schools of the National Nursery Schools Council Program in the south area of Santiago, Chile, and who met

the inclusion criteria were invited to participate in the study (Corvalan *et al.*, 2009). The following inclusion criteria were applied to the study: i) single birth with a gestational age of 37-42 weeks and a birth weight between 2500 and 4500 grams; ii) absence of physical and psychological conditions that could severely alter growth (only 6 children excluded due to these conditions). Among the 1,498 children eligible to participate, 1,195 accepted (80%) to take part in the study.

At the time of enrollment, mothers were interviewed to collect information on sociodemographic status, perinatal and feeding history of the child, current physical activity and feeding behaviour of the child, and maternal ginecological and obstetric history. Since 2006 GOCS children and their families have been contacted on a yearly basis and in 2009 a second follow-up of the cohort was initiated. In this occasion, 1,045 children out of the 1,195 belonging to the original cohort (87%) were evaluated. In order to measure metabolic markers, blood samples were collected on a subsample of 300 children at recruitment (children aged 4 years), and then on the whole cohort when children were 7 years old and when children reached the Tanner stage II of maturation (Tanner, 1962). Moreover children body composition was measured on a subsample of 500 children in 2007 (when children were about 6 years old) and annually thereafter via bio-impedance measurements. In 2010-2011 body composition of a subsample of 500 children was also measured using deuterium dilution techniques.

3.3.2 Growth measurements

Anthropometric data from birth up to 36 months of life were retrospectively gathered from child health records. For each subject, weight and height values measured at about 1, 2, 4, 6, 12, 18, 24 and 36 months of age were recorded, together with their actual date at measurements. From 3 years onwards (after recruitment) children were prospectively measured annually. These measurements were taken by a dietitian who visited the nursery school in 2006 and 2007; thereafter measurements were taken at the Institute of Nutrition and Food Technology of the University of Chile. The data available for this Ph.D. include measurements up to 8 years of age (though measurements up to age 10-11 are now available), however, for direct comparison with the other cohorts, only data up to age 4 were used. Among the 1,195 children considered, 41 were found to have missing values for all the 0-4 years follow-up weight measurements and were therefore excluded from the subsequent analyses. Table 3.6

shows descriptive statistics of the 0-4 years weight measures recorded.

	Weight (kg)					
Follow-up age	Males (n = 570)			Females (n = 579)		
(month)	N	Mean	SD	N	Mean	SD
0	569	3.44	0.43	579	3.37	0.39
1	405	4.43	0.59	408	4.26	0.53
2	401	5.57	0.67	399	5.26	0.57
4	383	7.19	0.83	389	6.67	0.74
6	366	8.17	0.88	381	7.64	0.87
12	464	10.16	1.13	473	9.57	1.07
18	413	11.54	1.28	438	11.01	1.23
24	446	12.96	1.52	450	12.44	1.51
36	320	15.28	1.89	364	14.96	1.92
48	458	18.08	2.51	478	17.98	2.66

Table 3.6: Weight measurements of GOCS participants by gender.

As observed for the other cohorts, the proportion of missing values generally increases with time in each gender group, and thus mean values may be affected by selection bias. Newborn males are heavier than females and these differences persist during infancy. Due to its inclusion criteria children included in the GOCS study are slightly heavier than NINFEA and GXXI participants at all ages and in both genders. By age 4 the difference appears to widen considerably, with GOCS males and females weighting on average 18.1 and 18.0 kg respectively (Table 3.6), while the corresponding figures in NINFEA are 16.9 and 16.1 kg (Table 3.3). It should however be noted that GOCS growth data were not measured at exact ages, and those intended to be measured at 4 years (follow-up age of 48 *months* in Table 3.6) were actually measured between 37 and 56 months of age.

3.4 Cohorts Summary

Figure 3.3 summarizes the 0-4 years weight measures available in the three cohorts (including, for NINFEA and GXXI, only fixed time points at which the anthropometric measures were planned to be observed).

An issue to be considered when dealing with longitudinal growth data is the potential effects of missingness, either due to loss to follow-up because of migration, refusal to continue participation or due

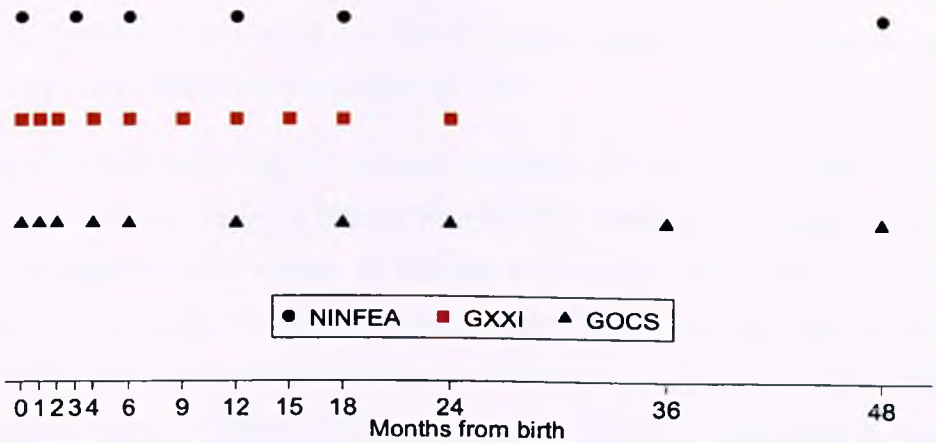


Figure 3.3: Weight measurements available in the three cohort studies.

to 'item non-response', i.e. missing values for some of the variables collected at a given wave. The latter arises whenever the sequence of measures from one study-subject is incomplete at intermediate times. Possible other sources of bias occurring in these cohorts are the effect of the measurement error, which is likely to particularly affect the maternal prenatal GOCS data due to its study design (the questionnaire on prenatal exposures is administered when the children are approximately 3-4 years old), and the potential follow-up selection mechanism, already discussed above.

3.4.1 Analyzed datasets

The GXXI and GOCS cohorts are used only for the growth analyses, and thus contribute to Research Papers III and IV. NINFEA is instead involved both in the analyses focused on selection bias and in the growth modelling analyses. However, as already discussed in this chapter, NINFEA is an on-going study, therefore analyses reported in each Research Paper have been performed using different versions of the NINFEA database. In detail:

- Research Paper I (see Chapter 4) uses simulated data to address objective *(a)* (see Chapter 2), and therefore none of the three cohorts is involved in the analyses.
- Research Paper II (see Chapter 4) focuses on objective *(b)* (see Chapter 2). Data from the

Piedmont Birth Registry were used to address this objective, after these were linked to the NINFEA database downloaded in July 2010. The latter includes 1,105 singleton births occurred within December 2008 in the municipality of Turin.

- Research Paper III (see Chapter 5) addresses objectives *(c)* and *(d)* (see Chapter 2) and involves all the three cohorts. In detail 845 singleton NINFEA children, 783 singleton GXXI children and 1,149 singleton GOCS children are included in the analyses of this paper. In order to assess the sensitivity of the growth models to the age range analyzed (objective *(d)*), for the NINFEA and GOCS cohorts, for which growth data are available up to age 4 and 8 respectively, data up to 4 years of age are included. This is in contrast to GXXI, for which data up to age 2 only are available. The NINFEA dataset used for this paper was downloaded in November 2011 and includes those children who, at the time of the data download, were eligible for completion of the 4-year follow-up questionnaire. These are the children for whom growth measurements are available at fixed time points only (0, 3, 6, 12, 18 and 48 months).

- Research Paper IV (see Chapter 5) investigates the prenatal influences of weight trajectories in infancy (objectives *(e)* in Chapter 2), and therefore involves weight data up to 2 years of age. In particular it includes 2,925 singleton NINFEA children with available data on gestational age at birth and whose mothers were born in Italy, 738 singleton GXXI babies of likely Portuguese origin with known gestational age at birth, and 959 singleton GOCS children of non-indigenous origin with exact gestational age data. The NINFEA dataset used for this paper was downloaded in March 2012 and includes children with at least one growth measure within the first two years of life (this is why the NINFEA sample involved in this paper is much bigger than the one included in the analyses of Research Paper III). Growth data used in the analyses of this paper include both measurements available at fixed time points (0, 3, 6, 12, and 18 months) and at the time of completion of the 6-months and 18-month follow-up questionnaires.

Part II

Selected Publications

Chapter 4

Selection Bias

4.1 Preamble

In epidemiology the term "selection bias" is used to indicate several types of bias, for example: bias resulting from selection of the participants from a restricted source population (i.e. the population from which the sample is drawn) according to specific criteria, bias due to participants' self-selection, bias due to informative censoring in longitudinal studies, or to more general missing data mechanisms, bias deriving from erroneous selection of controls in case-control study (Hernan *et al.*, 2004).

This chapter focuses on the potential bias that may arise from selection of study participants from a restricted source population, with particular attention given to cohort designs. Issue of generalizability of the study findings is beyond the scope of this thesis. The main focus is instead on investigating under which circumstances the selection mechanisms may induce (or reduce) confounding in the study sample and therefore may lead the selected study to be more or less affected by such bias than a study where participants are representative of the source population (see Figure 4.1).

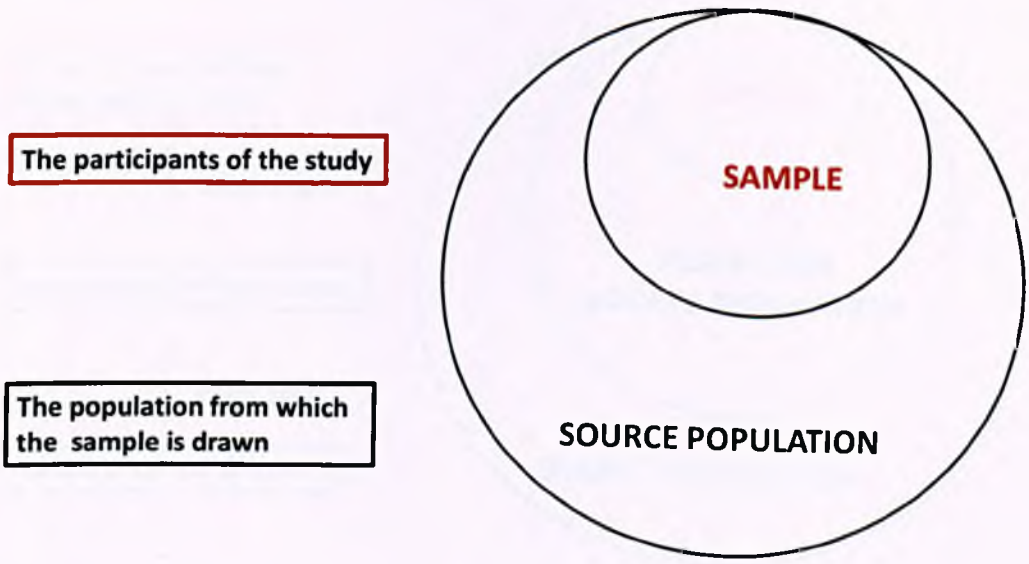


Figure 4.1: Study sample and source population.

Selection of study participants potentially affects all cohort studies, but web-based designs, such as the NINFEA birth cohort used in this thesis, are generally thought to be more prone to than traditional designs. Participants of internet-based studies are required to have access to the internet in order to become aware of the study and volunteer. Hence they are selected both because they are based on a restricted source population – the internet users – (we will refer to this as *intentional non-representativeness*) and because the study-population is a self-selected sample of this restricted source population (*unintentional non-representativeness*) (see Figure 4.2). This is the motivation for investigating the extent and direction of bias that may affect the results of studies like NINFEA. Research Paper I addresses the issue of the quantification of the bias from a theoretical perspective while Research Paper II deals with considerations specific to NINFEA. The Book Chapter addresses methodological issues about internet-based studies and discusses examples of the use of the internet in the context of different types of epidemiological designs.

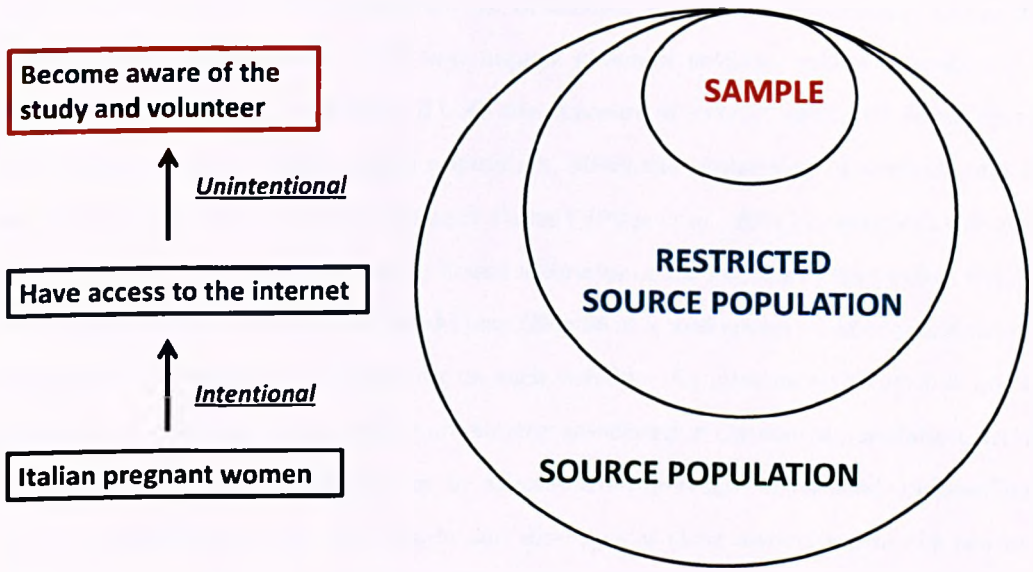


Figure 4.2: Selection into the web-based NINFEA birth cohort.

Participation in web-based studies is undoubtedly associated with socioeconomic status, as observed for the NINFEA birth cohort (Pizzi *et al.*, 2012). However conducting cohorts studies in specific subgroups of the source population (intentional selection) is an approach frequently used in epidemiology due to its numerous advantages, such as better follow-up rates, increased study feasibility and sample size as well as potentially better control of confounding. The British Doctor Study is a typical example of such a study design (Doll *et al.*, 2004). Moreover unintentional non-representativeness may occur in every cohort study, even those based on random sampling from the source population, due to refusal to participate or other reasons, such as unsuccessful contact. This is the case of the Danish National Birth Cohort, whose overall participation rate is 31%, with women of low socioeconomic status strongly unrepresented (Jacobsen *et al.*, 2010).

The concern thus should be whether correct scientific inference can be drawn from non-representative samples, and this is the topic of the Commentary included in this chapter. This issue has occasionally prompted rather heated debates, as in the case of the two members of the US National Children's Study Steering Committee, who vacated their position when the sampling design was changed to recruit participants from offices of physicians instead of from the general population (Chi, 2012). The

major criticism usually raised against the use of selected samples is the potential bias in the exposure-outcome association of interest. This may happen in cohort studies – where the outcome occurs after participant selection into the study – if both the exposure of interest and a risk factor for the outcome, which are independent in the general population, affect the probability of selection and thus become associated in the selected sample. Research Paper I (Pizzi *et al.*, 2011) investigates this setting. Under this scenario the restriction will lead to biased estimates of the exposure effect unless the risk factor for the outcome (which in this context would play the role of a confounder) is known and measured so that analyses can be carried out conditioning on such variable. An alternative scenario is one in which the exposure of interest and the risk factor are already associated in the source population. In this situation the analysis of the selected sample may be affected from increased or reduced confounding due to this risk factor depending on the magnitude and direction of their association in the source population and of the associations of both exposure of interest and risk factor with the selection process. As before valid inference is achieved if the confounder is controlled for, and this applies whether or not the cohort is randomly selected from the source population or from a restricted subgroup. A special case of this latter scenario, particularly relevant if the confounder is unknown or unmeasured, is when the association induced by the sample selection process between the exposure and the confounder perfectly compensates the association originally present in the population, so that the estimate of the exposure-outcome relation obtained in the selected sample will be unbiased. This issue is related to violations of what in graph theory is known as *faithfulness* (Glymour, 2006). It follows that it is difficult to predict whether control of confounding can be better achieved in the source population or in its restricted version, as each of them will have its own confounding pattern. Obviously these conclusions cannot be applied to other types of selection, such as informative loss to follow-up, or to other study designs, such as case-control studies, where the outcome may affect selection.

The Research Papers, Commentary and Book Chapter included in this thesis chapter use the terminology of Directed Acyclic Graphs (DAGs) to illustrate the problems potentially arising from selecting participants from a restricted source population or from self-selection of study participants. For this reason the next section gives a brief introduction to DAGs in the context of sample selection in cohort studies.

4.1.1 Directed Acyclic Graphs*

Causal diagrams are a powerful tool to integrate statistical and subject-matter information and explicitly draw assumptions about the causal relations existing among the variables of interest. These graphs allows the researchers to distinguish between causal effects and associations, to identify potential sources of bias in the estimate of the causal effect of interest and to choose an appropriate analytical strategy.

A DAG is a graph made of nodes, representing variables, and edges (arrows) between nodes indicating the direction of the causality. These graph are called "directed" since each edge is a single-headed arrow and "acyclic" since they contain no cycles (no variable causes itself) (Greenland *et al.*, 1999; Pearl, 1995, 2000). A DAG is causal when all common causes of each pair of nodes are included in the graphs (Hernan and Robins, 2012). For example, Figure 4.3*(a)* represents a DAG where E causes D and there are no common causes for these variables, alternatively these other causes would need to be depicted in the graph. A *path* is defined as sequence of edges (pointing in any directions) connecting one variable to another (see Figure 4.3). Two types of path exist: a *directed path* is a path between any two variables where all arrows are single-headed and point 'forwards' (see Figure 4.3*(c)*); a *backdoor path* is a path between any two variables, e.g. E and D, that starts with an arrow pointing to the first (e.g. path E–C–D in Figure 4.3*(b)*). Associations are transmitted along paths, with directed paths being causal and backdoor paths being associational. Nodes in a path can take various names, in particular they can be defined as *parent*, *child*, *ancestor* and *descendant*. In Figure 4.3*(c)*, for example, E is the parent of C, D is the child of C, E is the ancestor of D and D is the descendant of E. A *collider* is a node within a path with at least two parents: for example C is a collider in Figure 4.3*(d)*. Conditioning on a collider or on a child/descendant of a collider, induces an association (spurious association) among its parents. This is the case of Figure 4.4, in which E and D become associated (represented by a dashed line) because of the conditioning (represented by a square around C) on the collider C. More precisely, if E and D are independent in the population (as in Figure 4.4), they will not be necessarily independent within strata of C; while if they are not independent, their association might be altered within strata of C. Finally a path can be *open* or *blocked* (closed). In particular

*I acknowledge the lecture by Rhian Daniel entitled "An introduction to causal inference using causal diagrams" (Module of Advanced Statistical Methods in Epidemiology, LSHTM Masters Programme 2012-2013)

a path is blocked if it contains at least one collider and we do not condition on it nor on any of its descendant (as in Figure 4.3*(d)*), or if we condition on at least one node in the path which is not a collider (e.g. we block the path of Figure 4.3*(b)* conditioning on C). A path is open if it is not blocked, with open paths transmitting associations and blocked paths not transmitting associations.

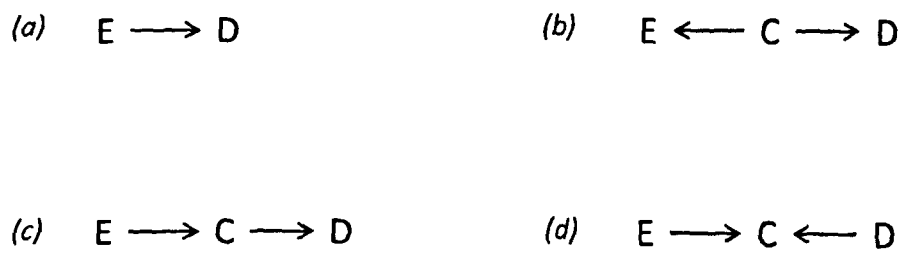


Figure 4.3: Some simple DAGs.

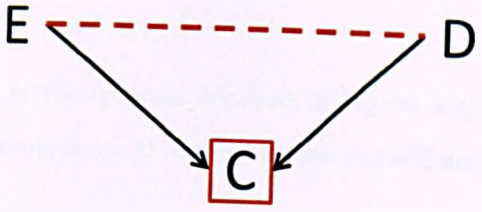


Figure 4.4: The consequence of conditioning on a collider. The dashed line represents an association induced by conditioning on C (represented by a square around C).

It follows that in order to have an unbiased estimate of the relation of interest (causal effect) backdoor paths (i.e. associational paths) have to be blocked, while directed paths (i.e. causal paths) have to be open. Criteria to identify variables to be controlled for in order to obtain an unbiased estimate (that is to avoid confounding for the exposure-outcome association of interest) have been suggested. In 1995 Pearl (Pearl, 1995) stated that there is no confounding if, after removing all edges out of the exposure (E), there are no unblocked backdoor paths from E to D (the outcome). In 1999 Greenland (Greenland *et al.*, 1999) described the relevant algorithm in the epidemiological literature as follows: a set of variables Z, none of which are descendants of the exposure (E) or the outcome (D), is sufficient to control for confounding if, having removed all edges pointing out of E and having linked all pairs of variables that share a child or descendant in Z, there is no unblocked path from E to D which does not pass through Z.

The spurious associations induced by conditioning on a collider is the key element for discussing sample selection in epidemiological studies, and in this context it has been referred to as collider bias. Causation and association are two different concepts, with causal relations, as already stated above, being directed, while associations being the sum of directed and undirected paths. Moreover while sample associations can be directly observable, causal effects cannot. Informally, there is bias when the true causal effect is different from the corresponding estimate. In particular when exposure and outcome are associated in the sample, but the null hypothesis of no causal effect holds, we say that

there is "bias under the null". Under the assumption of no measurement error, the latter can derive from two causal structures:

1. Common causes: this is the scenario depicted in Figure 4.3*(b)*, where the exposure and the outcome share the common cause C. Under this structure E and D will be associated even if not causally related.

2. Conditioning on a common effect: this is the scenario shown in Figure 4.4, where E and D become associated within strata of C.

Most epidemiologists refer to the former bias as *confounding*, while to the latter as *selection bias*. However the bias resulting from conditioning on a common effect (structure 2 above) should be more generally called *collider bias*. As extensively discussed in all the publications included in this chapter, collider bias may induce confounding, and this is the reason why, in the context of cohort studies, the bias induced by the selection process (e.g. restriction of the source population) in the exposure-outcome association of interest has been defined with different terms. While some researchers refer to this as selection bias (Hernan *et al.*, 2004), others defines it as a special case of confounding (Rothman *et al.*, 2008). Those arguing in favour of the term "selection bias" distinguish between the bias that arises from the presence of a common cause of exposure and outcome, i.e. confounding bias (structure 1 above), from the bias due to the confounding induced by conditioning on a collider (structure 2 above), that is by the restriction of the source population. In this framework the use of the term "collider bias" or "collider-stratification bias", which derives from the causal diagrams terminology (Greenland, 2003), is a formal way to overcome the overlap between selection bias and confounding. Clearly both these causal structures result in lack of comparability between the exposed and the unexposed. A simple DAG representing this latter situation (i.e. collider bias inducing confounding in the E-D relation) is shown in Figure 4.5. This DAG is the starting point of Research Paper I and of the following publications.

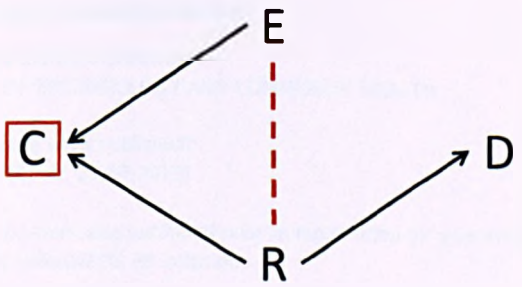


Figure 4.5: Confounding induced by conditioning on a collider. The dashed line represents an association induced by conditioning on C (represented by a square around C).

4.2 Research Paper I

Research paper cover sheet

For a 'research paper' already published

1.1. Where was the work published?
JOURNAL OF EPIDEMIOLOGY AND COMMUNITY HEALTH

1.2. When was the work published?
MAY 2011 (Epub: 29-09-2010)

1.2.1. If the work was published prior to registration for your research degree, give a brief rationale for its inclusion

1.3. Was the work subject to academic peer review? YES

1.4. Have you retained the copyright for the work? NO
I acknowledge permission of the BMJ Publishing Group to include in this thesis the paper published in the Journal of Epidemiology and Community Health (2011), 65(5),407-411.

2. For a 'research paper' prepared for publication but not yet published
2.1. Where is the work intended to be published? _____
2.2. List the paper's authors in the intended authorship order
2.3. Stage of publication – Not yet submitted/Submitted/Undergoing revision from peer reviewers' comments/In press

3. For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper.

Lorenzo Richiardi and Bianca De Stavola gave me substantial advice in designing the study. I designed the study's analytic strategy and performed the simulation analyses, including writing the program code in Stata to run the simulations. I wrote the first draft of the article. All co-authors helped interpret the findings and revised the manuscript, providing useful comments.

Candidate's signature

Supervisor or senior author's signature to confirm role as stated in (3) (Bianca De Stavola)

Sample selection and validity of exposure–disease association estimates in cohort studies

Costanza Pizzi,[1,2] Bianca De Stavola,[2] Franco Merletti,[1] Rino Bellocco,[3,4] Isabel dos Santos Silva,[5] Neil Pearce,[2,6] Lorenzo Richiardi[1]

► An additional table is published online only. To view this file please visit the journal online (http://jech.bmj.com).

[1]Cancer Epidemiology Unit, CeRMS and CPO-Piemonte, University of Turin, Italy
[2]Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK
[3]Department of Statistics, University of Milano Bicocca, Milan, Italy
[4]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
[5]Department of Non-communicable Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, UK
[6]Centre for Public Health Research, Massey University Wellington Campus, New Zealand

Correspondence to
Costanza Pizzi, Via Santena 7, 10126 Torino, Italy;
costanza.pizzi@lshtm.ac.uk

Accepted 29 July 2010
Published Online First 29 September 2010

ABSTRACT

Background Participants in cohort studies are frequently selected from restricted source populations. It has been recognised that such restriction may affect the study validity.

Objectives To assess the bias that may arise when analyses involve data from cohorts based on restricted source populations, an area little studied in quantitative terms.

Methods Monte Carlo simulations were used, based on a setting where the exposure and one risk factor for the outcome, which are not associated in the general population, influence selection into the cohort. All the parameters involved in the simulations (ie, prevalence and effects of exposure and risk factor on both the selection and outcome process, selection prevalence, baseline outcome incidence rate, and sample size) were allowed to vary to reflect real life settings.

Results The simulations show that when the exposure and risk factor are strongly associated with selection (ORs of 4 or 0.25) and the unmeasured risk factor is associated with a disease HR of 4, the bias in the estimated log HR for the exposure–disease association is ±0.15. When these associations decrease to values more commonly seen in epidemiological studies (eg, ORs and HRs of 2 or 0.5), the bias in the log HR drops to just ±0.02.

Conclusions Using a restricted source population for a cohort study will, under a range of sensible scenarios, produce only relatively weak bias in estimates of the exposure–disease associations.

INTRODUCTION

Selection of study subjects from restricted source populations according to prespecified criteria is an approach that is frequently used in cohort studies. The purposes of such restrictions are to enhance study feasibility and to increase the prevalence of exposure or the completeness of follow-up, thereby increasing study validity and precision. Typically this may involve recruiting participants from a subgroup of the general population, rather than sampling directly from the entire general population. Such subgroups may be defined on the basis of occupation, gender, geographical area, birth cohort, etc. The British Doctors' Study[1] and the Nurses' Health Study,[2] occupational cohorts,[3] follow-up of participants in specific events,[4] analyses restricted to specific subgroups of the population, such as non-smokers,[5] ancillary analyses of non-randomised exposures in randomised studies,[6] and follow-up studies of screening attendants[7] are all examples of cohort studies based on restricted samples.

Undoubtedly, restriction of the source population may introduce problems of generalisability of the study findings, but this also applies to studies that are based on the general population (eg, most cardiovascular epidemiology involves cohort studies in specific communities rather than true general population samples). We will therefore not consider issues of generalisability here; rather, our focus is on whether using a restricted source population may affect the validity of the exposure–disease associations.[8 9] In particular, bias will be introduced if a risk factor for disease is not associated with exposure in the general population but is associated with exposure in the study population, as a result of the selection process. Such biases can be represented using directed acyclical graphs (DAGs).[8 10–12] The example depicted in figure 1A represents a population in which there is no association between an exposure (E) and a disease (D); there is another risk factor (R) for the disease, but this is not a source of confounding as it is not associated with the exposure. However, E and R both affect the likelihood of being selected (S=1) into the study. When analyses are restricted to the selected subjects, there is an inherent conditioning on S (as represented by a square around S in figure 1B), which leads to a spurious association between E and R (represented by a dashed line). Under this scenario, even if E has no causal effect on D, the backdoor path E–R–D is opened and the estimated associational RR between the exposure and the disease (ARR_{DE}) may differ from the causal RR (CRR_{DE}). This could, for example, be the situation in a cohort study of the effect of obesity (E) on breast cancer (D) based on breast cancer screening participants (the restricted source population). In this example, obese women (E) are less likely to attend the screening programmes,[13,14] while women with a family history of breast cancer (R) are more likely to participate. Among those who attend screening (ie, conditioning on those with S=1), obesity (E) and family history of breast cancer (R) become positively correlated. In fact an obese woman is more likely to have a family history of breast cancer within the selected sample than in the general population, because otherwise she may not have participated in the screening programme. As a result, family history of breast cancer is a confounder of the obesity–breast cancer association if studied among screening attendees, but is not—or to a lesser extent—in the general population.

This type of bias has been extensively discussed in the causal inference literature from a theoretical point of view.[8 9] Hernan and colleagues' 2004 paper

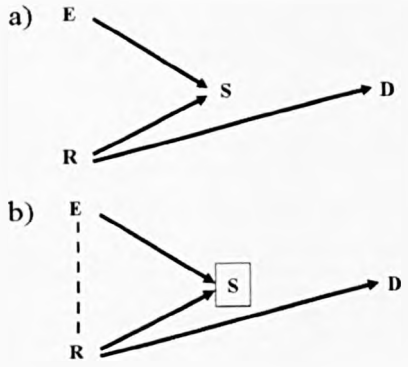


Figure 1 Diagram of a cohort based on a selected sample. (A) In the population the exposure of interest (E) is not associated with the disease of interest (D) that is caused by a risk factor (R). Both E and R affect the probability of being selected (S) as a member of the cohort. (B) The study is carried out in the selected sample, and therefore there is an inherent conditioning on S (a box around a variable means conditioning for that variable) which generated an induced association between E and R (represented by a dashed line).

on selection bias provides the conceptual framework and indicates that if the risk factor associated with the selection process is known and measured, it is possible to adjust for selection bias, whereas if the risk factor is unmeasured, the effect estimates may be biased. However, although the theoretical basis of selection bias is clear, there have been few attempts to quantify the likely strength of such biases. One exception is that of Greenland,[15] who studied the setting of figure 1B with dichotomous exposure and outcome variables, employing methods originally developed to quantify the impact of unmeasured confounding.[16] He calculated the likely maximum strength of the bias in the estimation of the E–D association in the S=1 stratum as a function of the ORs corresponding to the true associations depicted in figure 1B (ie, OR_{SE}, OR_{SR}, OR_{DR}). However, it is not clear how these results apply to cohort studies. Because of the increasing frequency of cohort studies based on selected populations, such as the internet-based birth cohort studies based in Italy (NINFEA cohort) and New Zealand (ELFS),[17] quantifying the potential biases involved in analysing such data is timely and relevant.

Our aim is therefore to study the extent of these biases. We use simulations to mimic a variety of cohort restrictions and disease settings and examine the consequent bias in the estimated exposure hazard (or rate) ratio (HR) of disease. We then discuss these results in terms of whether, and under what circumstances, the resulting selection bias is serious enough to strongly bias the exposure effect estimates. For simplicity, we will assume throughout the paper that there is negligible random variation, that all variables are measured without error, and that there is uninformative censoring.

SAMPLE SELECTION AND DISEASE RISK FACTORS

As previously recognised,[9 18] a fundamental characteristic of selection bias in restricted cohort studies is that the selection process makes a disease risk factor, which may not be associated with the exposure in the general population, become associated with the exposure among the study population and therefore act as a confounder.

Confounders in the general population and risk factors that become confounders in a restricted source population are usually indistinguishable when the study is analysed. Although typically some disease risk factors (ie, potential confounders) are known a priori, it is seldom known whether these are associated with the exposure of interest in the specific population in which the study will be carried out. Both in general population-based and restricted cohorts, therefore, researchers attempt to collect information on all known and suspected important risk factors of the disease in the population that they are studying, regardless of their expectations about whether these are associated with the exposure or not. The example of the association between smoking and socioeconomic position (SEP) illustrates this point well. Depending on the population and the calendar period, SEP can be positively or negatively associated, or not associated at all, with smoking. Researchers aiming to estimate the association between smoking and mortality will always attempt to collect information on SEP and, in most instances, will control for it, irrespective of whether the confounding effect of SEP is due to a real association between SEP and smoking in the general population or a spurious association caused by the sample selection process.

Another possible consequence of the selection mechanism is a change in magnitude, and in extreme cases direction, of the confounding effect of a risk factor. This may occur if the strength of the association between the risk factor and the exposure in the selected sample differs from that originally present in the general population. For example, when two (parent) variables influence a third (child) variable in the same direction, conditioning on the child variable likely leads to a negative association between the parent variables.[8] Thus, if an exposure and a confounder influence the selection process in the same direction, the original association between exposure and confounder will be reduced in the subset of those who participate if they were originally positively associated, or increased if their original association was negative. For example, in many populations smoking and physical exercise are negatively associated. In a hypothetical study restricted to blood donors, who typically have a healthy lifestyle and thus smoke less and exercise more than the average individual in the general population, the sample selection would add a positive association between smoking and physical exercise. Therefore, the original negative association present in the general population would be, if anything, attenuated among blood donors.

In the next section, we use simulations to quantify the likely extent of selection bias arising from the use of restricted cohorts.

QUANTIFICATION OF THE BIAS

Methods

We conducted Monte Carlo simulations of alternative settings corresponding to the scenario of figure 1B to quantify the resulting bias in the estimation of the E–D effect when conditioning on S=1 and not adjusting for R.[19] The generation process of the four variables of figure 1B is described below.

We generated E and R as marginally independent binary variables, with prevalence, respectively P_E and P_R, initially set equal to 0.5 in the source population. They were later allowed to decrease to 0.25 for P_E and to 0.1 for P_R, in order to investigate scenarios more frequently addressed by epidemiologists.

The binary variable S was generated using a logistic regression model with baseline prevalence, P_S, equal to 0.5 and with the ORs for the explanatory binary variables E and R taking values 0.25, 0.33, 0.50, 2, 3 and 4. Specifically, with α_S indicating the log

(odds) of S=1 among the non-exposed, β_{SE} indicating the log (OR) corresponding to exposure E and β_{SR} indicating the log (OR) corresponding to R, the generating model was:

$$logit(S = 1) = \alpha_S + \beta_{SE}E + \beta_{SR}R \quad (1)$$

A more complex model that included an interaction term between E and R was also considered:

$$logit(S = 1) = \alpha_S + \beta_{SE}E + \beta_{SR}R + \beta_{inter}E*R \quad (2)$$

with OR_{inter} corresponding to $\exp(\beta_{inter})$, set at values 0.5 or 2. The interaction term was introduced to examine more realistic selection settings. For example, in the first empirical demonstration of Berkson's bias, Roberts and colleagues found that not only do chronic conditions increase the chance of hospitalisation, but they often also interact more than multiplicatively.[20]

We generated time to the outcome D assuming a constant rate λ—that is, we assumed that time to event followed an exponential distribution.[21] The baseline rate λ_0 was set equal to 0.01, 0.03 or 0.06 events/year, with administrative censoring time set at 5 years. The rate λ was allowed to be affected only by R, with HR_{DR} taking values 0.25, 0.33, 0.50, 2, 3 and 4, while we assumed no E–D association—that is, $HR_{DE}=1$. Specifically, with β_{DE} indicating the log(HR) of D for the exposure E and β_{DR} indicating the log(HR) of D for the risk factor R, the log rate function for D, $\log(\lambda)$, was defined as:

$$log(\lambda) = log\lambda_0 + \beta_{DE}E + \beta_{DR}R \quad (3)$$

with β_{DE} fixed at 0.

We generated a total of 1000 Monte Carlo simulated datasets of 5000 subjects for each combination of the parameters described above. We also used a size of 2500 subjects, increasing the number of simulations (n=2000), to deal with the greater impact of random variation.

In each simulated dataset, we estimated two main parameters in the stratum S=1 (which sample size varies as a consequence of the selected parameters for the selection process): the association between E and R (OR_{ER}) and the association between E and D (HR_{DE}) which is induced by the selection process. The estimate of HR_{DE} was obtained fitting a Cox proportional hazards regression model with no adjustment for R.[22] We then calculated the bias in the E–D association as the difference between zero, that is the true value of β_{DE}, and the logarithm of the estimated HR_{DE}. For each scenario, we summarised the bias, and the estimated values of β_{DE}, in terms of means, SD, and 5th and 95th percentiles.

Results

We first considered the situation with prevalence of E and R both equal to 0.5, $OR_{inter}=1$ (ie, no multiplicative interaction), and $\lambda_0=0.03$ (the 'reference scenario' in table 1). As expected, the size of the bias in the estimation of OR_{DE} depended on: (i) the induced association between the exposure and the risk factor (OR_{ER}), which increased in absolute terms with the absolute size of OR_{SR} and OR_{SE}; and (ii) the magnitude of the association between the risk factor and the disease (HR_{DR}). The largest values of the bias in the log OR were ±0.15 (table 1, 'reference scenario'), which were reached when OR_{SE}, OR_{SR} and HR_{DR} were furthest from the null value (ie, equal to 0.25 or 4). Note that in table 1 the range for $\log(OR_{ER|S=1})$ is not symmetrical because the magnitude of the association induced by the selection between E and R also depends on the prevalence of S in the population (P_s), with the strongest association obtained when $P_s=0.5$. Supplementary table 1 presents the complete results for all combinations of the values of OR_{SE}, OR_{SR} and HR_{DR}. The mean bias decreased from ±0.15 to just ±0.02 when the three ORs/HRs were equal to 2 or 0.5.

Table 1 Bias in the crude estimation of the E–D association by selected values of the data generating parameters; results from 1000 simulations

N	p (E=1)	p (R=1)	Baseline rate of D	Interaction	Mean β_{ER} range*	Mean E–D bias range†
Reference scenario						
5000	0.5	0.5	0.03	No	−0.31; 0.45	−0.15; 0.15
Alternative scenarios						
5000	0.5	0.5	0.03	Yes	−0.98; 0.74	−0.24; 0.27
5000	0.25	0.5	0.03	No	−0.31; 0.45	−0.15; 0.15
5000	0.5	0.5	0.01	No	−0.31; 0.45	−0.15; 0.16
5000	0.5	0.5	0.06	No	−0.31; 0.45	−0.14; 0.15
2500‡	0.5	0.5	0.03	No	−0.32; 0.45	−0.15; 0.15
5000	0.5	0.1	0.03	No	−0.33; 0.45	−0.12; 0.07

*β_{ER} expressed as log(OR).
†β_{DE} expressed as log(HR).
‡Results from 2000 simulations.
E, exposure; R, risk factor; D, disease

When an interaction term between E and R was included in the model generating S, the induced E–R association increased considerably (figure 2), up to a log(OR) of −0.98 (table 1, row 2) when OR_{SE} and OR_{SR} were equal to 0.25 and the OR_{inter} was 0.5. The bias increased accordingly, ranging from −0.24 to 0.27 (table 1, row 2). This situation is equivalent, in terms of induced bias, to those involving very strong marginal associations with selection. It is clear from figure 2 that the impact of the interaction is not the same for all the parameter combinations, as the magnitude of the induced E–R association is strengthened or reduced according to the sign of the interaction term but also to the size of the stratum of subjects exposed to both E and R.

Neither the prevalence of the exposure E (table 1, row 3) nor the baseline rate for the disease D (table 1, rows 4–5) or the sample size (table 1, row 6) affected the extent of the bias. Conversely, the prevalence of R, which becomes a confounder of the E–D association when S=1, had a non-marginal effect. For a given value of the induced E–R association, the bias reached its peak when the prevalence of R among the selected subjects (S=1 stratum) was 0.5. For this reason, when the population prevalence of R was set equal to 0.1 instead of 0.5, the range of the mean bias decreased to (−0.12; 0.07) (table 1, row 7).

DISCUSSION

Conducting cohort studies in a restricted sample of the general population may offer several advantages, including more precise measurement of the exposure, higher exposure prevalence, enhanced feasibility of the study, better control of confounding, increased sample size, higher recruitment rates, and a higher completeness of follow-up. These advantages should be balanced against issues of validity.

In this paper we have shown, via simulations, that the possible bias introduced by restriction of the source population is usually weak when internal comparisons are carried out within the cohort, with a maximum bias in the log(HR) of ±0.15.

These results are in agreement with those of Greenland,[15] who used an analytical approach to quantify the maximum selection bias in settings where the outcome risk is rare so that the analysis of cohort data can be performed using logistic

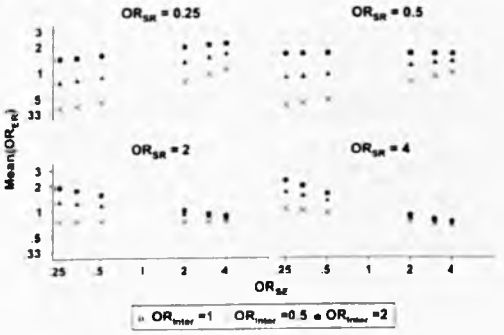


Figure 2 Mean OR of the induced E–R association in the stratum of those selected (S=1) by selected values of the association of the exposure (OR_{SE}) and the risk factor (OR_{SR}) with the selection process and of the E–R interaction (OR_{inter}); results from 1000 simulations.

regression. Our simulations add further insight to these results as we examined a wide range of disease and selection parameters, including exposure and risk factor prevalence, which highlighted their individual role in influencing the extent of the bias. Further, we considered settings where exposure and risk factor interact when influencing the selection process. Some additional points are warranted.

First, the bias is necessarily small when the association between the exposure of interest and the selection process is relatively weak (ie, 0.5<OR<2). In particular, when the exposure-selection OR is equal to 2 or 0.5, while the risk factor-selection OR and the risk factor-disease HR are allowed to take values up to 4 or down to 0.25, the maximum bias in the estimated exposure–disease association is within the ±0.07 range (on the log hazard scale). For example, consider the Million Women Study, a cohort nested within the breast screening programme in the UK.[7] From the study carried out to compare the characteristics of the study participants with the rest of the population (women who attended the screening but did not join the study plus not attendants),[23] the participation OR for current use of hormone replacement therapy, which is the main exposure of interest of the study, was derived. This estimated OR was about 1.6. On the basis of this information it is possible to assume that, in this cohort, the bias introduced by the baseline selection on the estimates of the effect of hormone replacement therapy on the outcome of interest would be negligible.

Second, selection must be associated with one or more unmeasured or unknown disease risk factors in order to introduce bias. However, unknown or unmeasured disease risk factors can introduce bias whether or not the cohort is based on the general population or a restricted source population; in the latter case, the sample selection can either increase or decrease the overall bias, with a magnitude and direction difficult to predict if there are multiple risk factors involved.[24]

Third, we have shown that even when all of the associations involved in the selection and outcome mechanisms are reasonably large (eg, all ORs/HRs of 4.0 or 0.25), the prevalence of the risk factor R is about 50% and there is no adjustment for R, the resulting bias is relatively weak (ie, ±0.15 on the log scale). This is reassuring, as this scenario is rather extreme and very unlikely to occur in practice. Besides, a disease risk factor with a 50% prevalence and a disease HR of 4.0 would have an attributable fraction of 60% and is therefore unlikely not to have been known and measured when a study is planned.

The scenarios considered in our simulations were restricted to binary exposure and binary risk factor and assumed no association between the exposure and the risk factor in the general population. A limitation is that we examined only the case of a single unmeasured determinant of the disease that also influences the selection process. However, we believe it is unlikely that multiple and independent important disease risk factors would affect the sample selection. It is indeed reasonable to consider R as a vector resulting from the combination of a set of correlated risk factors, all moderately associated with S. Finally, we only showed the findings derived from the analyses based on the assumption of a null causal association between the exposure and the outcome of interest; however choosing a true associational value, β_{DL}, different from zero would not modify the simulation results and therefore our conclusions.

We conclude that using a restricted source population for a cohort study will, under a range of sensible scenarios, produce only weak bias in estimates of the exposure–disease associations. On the other hand, the use of such restrictions may increase the response rate and the exposure prevalence, as well as being the only feasible approach in many circumstances.

What we already know on this subject

- Baseline selection of participants in cohort studies may affect the study validity.
- This happens when, because of the selection process, the confounding effect of an unknown or unmeasured disease risk factor is larger in the selected sample than in the general population.

What this study adds

- We conducted Monte Carlo simulations to quantify the likely extent of the selection bias affecting the exposure-disease association, varying all the parameters involved: prevalence and effects of exposure and risk factor on both the selection and outcome process, selection prevalence, baseline incidence rate of the outcome and sample size.
- The maximum bias is relatively weak (±0.15 in the log Hazard Ratio scale). When scenarios typically seen in epidemiological studies were considered the bias in the log Hazard Ratio drops to ±0.02.

Funding The study was conducted within projects partially funded by Compagnia SanPaolo/FIRMS, the Piedmont Region, the Italian Ministry of University and Research (MIUR), the Italian Association for Research on Cancer (AIRC) and the Massey University Research Fund (MURF). The Centre for Public Health Research is supported by a Programme Grant from the Health Research Council of New Zealand.

Competing interests None

Contributors All authors are responsible for the reported research and have made substantial contributions to the conception and design of the study, acquisition of data, analysis and interpretation of data, drafting the article or revising it critically for important intellectual content. All authors have seen and approved the final version of the manuscript.

Provenance and peer review Not commissioned, externally peer reviewed

REFERENCES

1 **Doll R,** Peto R, Boreham J, *et al.* Mortality in relation to smoking: 50 years' observations on male British doctors. *BMJ* 2004;**328**:1519.
2 **Colditz GA,** Hankinson SE. The Nurses' Health Study: lifestyle and health among women. *Nat Rev Cancer* 2005;**5**:388–96.
3 **Magnani C,** Ferrante D, Barone-Adesi F, *et al.* Cancer risk after cessation of asbestos exposure: a cohort study of Italian asbestos cement workers. *Occup Environ Med* 2008;**65**:164–70.
4 **Lagerros YT,** Bellocco R, Adami HO, *et al.* Measures of physical activity and their correlates: the Swedish National March Cohort. *Eur J Epidemiol* 2009;**24**:161–9.
5 **Vineis P,** Airoldi L, Veglia F, *et al.* Environmental tobacco smoke and risk of respiratory cancer and chronic obstructive pulmonary disease in former smokers and never smokers in the EPIC prospective study. *BMJ* 2005;**330**:277.
6 **Beattie MS,** Costantino JP, Cummings SR, *et al.* Endogenous sex hormones, breast cancer risk, and tamoxifen response: an ancillary study in the NSABP Breast Cancer Prevention Trial (P-1). *J Natl Cancer Inst* 2006;**98**:110–15.
7 **Anon.** The Million Women Study: design and characteristics of the study population. The Million Women Study Collaborative Group. *Breast Cancer Res* 1999;**1**:73–80.
8 **Glymour MM.** Using causal diagrams to understand common problems in social epidemiology. In: Jossey-Bass, ed. *Methods in social epidemiology*. San Francisco: Jossey-Bass; 2006.
9 **Hernan MA,** Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology* 2004;**15**:615–25.
10 **Greenland S,** Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology* 1999;**10**:37–48.
11 **Hernan MA,** Hernandez-Diaz S, Werler MM, *et al.* Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am J Epidemiol* 2002;**155**:176–84.
12 **Pearl J.** Causal diagrams for empirical research. *Biometrika* 1995;**82**:669–88.
13 **Pearl J.** *Causality: models, reasoning, and inference*. Cambridge, UK: Cambridge University Press; 2000.
14 **Ferrante JM,** Chen PH, Crabtree BF, *et al.* Cancer screening in women: body mass index and adherence to physician recommendations. *Am J Prev Med* 2007;**32**:525–31.
15 **Greenland S.** Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology* 2003;**14**:300–6.
16 **Yanagawa T.** Case-control studies: assessing the effect of a confounding factor. *Biometrika* 1984;**71**:191–4.
17 **Richiardi L,** Baussano I, Vizzini L, *et al.* Feasibility of recruiting a birth cohort through the Internet: the experience of the NINFEA cohort. *Eur J Epidemiol* 2007;**22**:831–7.
18 **Rothman KJ,** Greenland S, Lash TL. Distinguishing selection bias from confounding. In: Rothman KJ, Greenland S, Lash TL, eds. *Modern epidemiology*. Philadelphia: Lippincott Williams & Wilkins; 2008:136–7.
19 **Burton A,** Altman DG, Royston P, *et al.* The design of simulation studies in medical statistics. *Stat Med* 2006;**25**:4279–92.
20 **Roberts RS,** Spitzer WO, Delmore T, *et al.* An empirical demonstration of Berkson's bias. *J Chronic Dis* 1978;**31**:119–28.
21 **Bender R,** Augustin T, Blettner M. Generating survival times to simulate Cox proportional hazards models. *Stat Med* 2005;**24**:1713–23.
22 **Clayton D,** Hills M. *Statistical models in epidemiology*. Oxford: Oxford University Press; 1993.
23 **Banks E,** Beral V, Cameron R, *et al.* Comparison of various characteristics of women who do and do not attend for breast cancer screening. *Breast Cancer Res* 2002;**4**:R1.
24 **VanderWeele TJ,** Hernan MA, Robins JM. Causal directed acyclic graphs and the direction of unmeasured confounding bias. *Epidemiology* 2008;**19**:720–8.

Supplementary Table 1: Mean and Standard Deviation (SD) of the Bias in the Crude Estimation of HR_{DE} by Selected Values of the OR_{SE}, OR_{SR} and HR_{DR} Parameters: Results From 1,000 Simulations

OR_{SE}	OR_{SR}	HR_{DR}	Mean bias (β_{DE})	SD bias (β_{DE})	p(E=1|S=1)	p(R=1|S=1)	p(D=1|S=1)
0.25	0.25	0.25	-0.049	0.192	0.270	0.270	0.112
0.25	0.25	0.33	-0.038	0.203	0.270	0.270	0.115
0.25	0.25	0.5	-0.033	0.194	0.270	0.270	0.121
0.25	0.25	2	0.053	0.161	0.270	0.270	0.171
0.25	0.25	3	0.073	0.155	0.270	0.270	0.200
0.25	0.25	4	0.100	0.142	0.270	0.270	0.224
0.25	0.5	0.25	-0.040	0.187	0.272	0.388	0.100
0.25	0.5	0.33	-0.034	0.182	0.271	0.388	0.104
0.25	0.5	0.5	-0.012	0.181	0.272	0.388	0.113
0.25	0.5	2	0.034	0.145	0.272	0.389	0.185
0.25	0.5	3	0.054	0.129	0.272	0.388	0.226
0.25	0.5	4	0.060	0.115	0.272	0.388	0.261
0.25	2	0.25	0.089	0.174	0.314	0.588	0.079
0.25	2	0.33	0.070	0.163	0.314	0.588	0.086
0.25	2	0.5	0.044	0.153	0.314	0.588	0.100
0.25	2	2	-0.033	0.101	0.314	0.588	0.209
0.25	2	3	-0.053	0.089	0.314	0.588	0.271
0.25	2	4	-0.060	0.080	0.313	0.588	0.323
0.25	4	0.25	0.150	0.158	0.350	0.650	0.073
0.25	4	0.33	0.114	0.145	0.350	0.650	0.081
0.25	4	0.5	0.074	0.144	0.350	0.650	0.096
0.25	4	2	-0.056	0.090	0.350	0.651	0.217
0.25	4	3	-0.091	0.078	0.349	0.650	0.285
0.25	4	4	-0.109	0.072	0.350	0.650	0.342
0.5	0.25	0.25	-0.033	0.166	0.388	0.272	0.111
0.5	0.25	0.33	-0.032	0.160	0.388	0.272	0.115
0.5	0.25	0.5	-0.015	0.159	0.389	0.271	0.121
0.5	0.25	2	0.033	0.133	0.388	0.272	0.172
0.5	0.25	3	0.048	0.124	0.389	0.272	0.200
0.5	0.25	4	0.056	0.116	0.388	0.272	0.225
0.5	0.5	0.25	-0.027	0.158	0.390	0.391	0.099
0.5	0.5	0.33	-0.028	0.151	0.391	0.391	0.104
0.5	0.5	0.5	-0.020	0.147	0.390	0.390	0.113
0.5	0.5	2	0.019	0.116	0.390	0.390	0.186
0.5	0.5	3	0.026	0.105	0.390	0.391	0.227
0.5	0.5	4	0.036	0.098	0.390	0.390	0.261
0.5	2	0.25	0.041	0.145	0.417	0.583	0.079
0.5	2	0.33	0.037	0.134	0.417	0.584	0.087
0.5	2	0.5	0.023	0.131	0.417	0.583	0.100
0.5	2	2	-0.013	0.087	0.417	0.583	0.209
0.5	2	3	-0.026	0.080	0.416	0.583	0.269
0.5	2	4	-0.033	0.072	0.417	0.583	0.321
0.5	4	0.25	0.071	0.141	0.435	0.637	0.074
0.5	4	0.33	0.059	0.134	0.435	0.638	0.081
0.5	4	0.5	0.041	0.119	0.435	0.638	0.096
0.5	4	2	-0.028	0.080	0.434	0.637	0.215
0.5	4	3	-0.052	0.072	0.435	0.638	0.281
0.5	4	4	-0.059	0.065	0.434	0.638	0.338
2	0.25	0.25	0.050	0.139	0.588	0.314	0.107

2	0.25	0.33	0.041	0.135	0.588	0.313	0.111
2	0.25	0.5	0.026	0.130	0.588	0.314	0.119
2	0.25	2	-0.031	0.107	0.588	0.314	0.177
2	0.25	3	-0.062	0.099	0.588	0.314	0.209
2	0.25	4	-0.077	0.094	0.588	0.314	0.238
2	0.5	0.25	0.031	0.129	0.583	0.417	0.097
2	0.5	0.33	0.030	0.129	0.584	0.417	0.101
2	0.5	0.5	0.022	0.118	0.583	0.417	0.111
2	0.5	2	-0.012	0.092	0.583	0.417	0.189
2	0.5	3	-0.032	0.082	0.583	0.417	0.232
2	0.5	4	-0.042	0.078	0.583	0.416	0.269
2	2	0.25	-0.033	0.125	0.557	0.557	0.082
2	2	0.33	-0.026	0.121	0.557	0.556	0.089
2	2	0.5	-0.021	0.111	0.557	0.556	0.102
2	2	2	0.018	0.079	0.557	0.557	0.206
2	2	3	0.026	0.069	0.558	0.557	0.263
2	2	4	0.031	0.063	0.557	0.557	0.313
2	4	0.25	-0.069	0.119	0.545	0.591	0.079
2	4	0.33	-0.049	0.111	0.544	0.592	0.086
2	4	0.5	-0.029	0.106	0.545	0.591	0.099
2	4	2	0.030	0.076	0.545	0.591	0.210
2	4	3	0.043	0.065	0.545	0.592	0.272
2	4	4	0.049	0.059	0.545	0.591	0.324
4	0.25	0.25	0.101	0.125	0.650	0.350	0.103
4	0.25	0.33	0.089	0.126	0.650	0.350	0.108
4	0.25	0.5	0.060	0.122	0.650	0.350	0.116
4	0.25	2	-0.080	0.100	0.650	0.350	0.181
4	0.25	3	-0.120	0.095	0.650	0.350	0.217
4	0.25	4	-0.149	0.087	0.650	0.350	0.249
4	0.5	0.25	0.059	0.129	0.638	0.435	0.095
4	0.5	0.33	0.049	0.122	0.638	0.435	0.100
4	0.5	0.5	0.034	0.115	0.637	0.435	0.110
4	0.5	2	-0.036	0.089	0.638	0.435	0.192
4	0.5	3	-0.056	0.079	0.637	0.435	0.236
4	0.5	4	-0.074	0.072	0.638	0.434	0.275
4	2	0.25	-0.062	0.117	0.591	0.544	0.084
4	2	0.33	-0.043	0.116	0.591	0.545	0.090
4	2	0.5	-0.031	0.102	0.592	0.545	0.103
4	2	2	0.028	0.075	0.592	0.545	0.205
4	2	3	0.045	0.066	0.591	0.545	0.261
4	2	4	0.055	0.064	0.591	0.544	0.309
4	4	0.25	-0.104	0.114	0.573	0.572	0.080
4	4	0.33	-0.085	0.114	0.572	0.573	0.088
4	4	0.5	-0.054	0.106	0.573	0.572	0.101
4	4	2	0.050	0.074	0.572	0.572	0.208
4	4	3	0.076	0.063	0.572	0.572	0.267
4	4	4	0.089	0.059	0.573	0.573	0.318

4.2.1 Addendum to Research Paper I

Continuous Outcome

The scenarios investigated in Research Paper I were restricted to a binary outcome. Below additional results from settings with continuous instead of a binary outcome are presented.

Monte Carlo simulations of alternative settings corresponding to the scenario of Figure 1b of Research Paper I were conducted. Again, E and R were generated as marginally independent binary variables, with prevalence set equal to 0.5 in the source population. The binary variable S was generated using a logistic regression model with baseline prevalence equal to 0.5 and with the ORs for the explanatory binary variables E and R, OR_{SE} and OR_{SR}, each taking values 0.25 and 4. The situation in which there is no association between the exposure and the selection ($OR_{SE} = 1$) was also considered. The continuous outcome D was simulated to have a normal distribution with mean 3 (α_D) to mimic the distribution of the birth weight (in kg) of Caucasian babies. The outcome was allowed to be affected both by R, with β_{DR} taking values 1 and 0.5 and by E, with β_{DE} taking value 0 or 0.5 according to the following model:

$$E(D) = \alpha_D + \beta_{DE}E + \beta_{DR}R + \epsilon \tag{4.1}$$

with ϵ normally distribute with mean 0 and SD equal to 1.

In each simulated dataset, we estimated the marginal association between E and D (β_{DE}) in the stratum S=1, i.e. the subset of the study participants defined by the selection process. This is obtained by fitting a linear regression model of D on E with no adjustment for R. The bias in the E–D association was then calculated as the difference between the true coefficient of E in the marginal regression of D on E in the population (0 or 0.5 in this setting) and the estimated coefficient of E in the marginal regression of D on E in the sample (that is in the stratum S=1). For each scenario, the bias was summarised in terms of means and SD.

The results are shown in Table 4.1. As observed for the binary outcome, the size of the bias depended on the size of OR_{SR} and OR_{SE}, and on the magnitude of the association between the risk factor and the disease (β_{DR}). The largest value of the bias was -0.1 (corresponding to minus 100 grams), which was

reached when OR_{SR}, OR_{SE} and β_{DR} were furthest from the null value. The estimated bias decreased to a maximum of -0.05 when β_{DR} decreased from 1 to 0.5. Table 4.1 also shows that, assuming no association between the exposure and the selection indicator ($OR_{SE} = 1$) and no interaction between E and R in their effect on D, there was no bias, both when the scenario of a null association between the exposure and the outcome of interest was considered ($\beta_{DE} = 0$) and when a true effect of E on D different from the null was considered ($\beta_{DE} = 0.5$).

OR_{SE}	OR_{SR}	β_{DR}	β_{DE}	**Estimated Bias(β_{DE})**	**SD Bias(β_{DE})**
0.25	0.25	0.5	0	**0.03**	0.06
			0.5	**0.03**	0.06
		1	0	**0.06**	0.07
			0.5	**0.06**	0.07
	4	0.5	0	**-0.05**	0.04
			0.5	**-0.05**	0.04
		1	0	**-0.10**	0.05
			0.5	**-0.10**	0.05
1	0.25	0.5	0	**0.00**	0.05
			0.5	**0.00**	0.05
		1	0	**0.00**	0.05
			0.5	**0.00**	0.05
	4	0.5	0	**0.00**	0.04
			0.5	**0.00**	0.04
		1	0	**0.00**	0.04
			0.5	**0.00**	0.04
4	0.25	0.5	0	**-0.05**	0.04
			0.5	**-0.05**	0.04
		1	0	**-0.10**	0.04
			0.5	**-0.10**	0.05
	4	0.5	0	**0.04**	0.03
			0.5	**0.04**	0.03
		1	0	**0.07**	0.04
			0.5	**0.07**	0.04

Table 4.1: Estimated bias in the marginal (with respect to R) estimation of β_{DE} by selected values of the OR_{SE}, OR_{SR} and β_{DR} parameters: results from 1,000 simulated datasets of 5,000 subjects.

Missing data, selection bias and non-collapsibility

As stated in the introduction to this Chapter the term "selection bias" is used to indicate several types of bias, including bias arising from missing data mechanisms. The relationship between the potential bias that may arise from selection of study participants from a restricted source population, as investigated in this thesis, and the issues arising when the aim is to estimate causal effects from incomplete data is of special interest. In particular the paper by Daniel et al (Daniel *et al.*, 2012), which uses causal diagrams to guide analyses in missing data problems, is of relevance to further examine the scenarios considered in Research Paper I and the additional scenarios investigated in this section. In their paper the authors formally extend the backdoor criterion suggested by Pearl (Pearl, 1995) and Greenland (Greenland *et al.*, 1999) to incomplete data settings. In particular they show that in order to identify the causal effect of an exposure on the outcome in the subset with complete data two conditions need to be satisfied (Daniel *et al.*, 2012). Condition 1 is equivalent to the original backdoor criterion (as described in section 4.1.1) but applied to an extended causal diagram. This is the original diagram to which all parents of the exposure of interest and all parents of descendants of the exposure of interest (except for parent of the selection/missingness indicator S) are added. The set $\{S\} \cup \mathcal{Z}$ of variables that satisfy this extended backdoor criterion includes the variables that block any open backdoor path in this extended diagram, where additionally all their common parents are joined by a dotted line, as well as the parents of any of their ancestors. Condition 2 is satisfied if, having removed in the original causal diagram all arrows into the exposure and having linked any pair of variables that are both parents of a variable in $\mathcal{Z}$ or that share a child which is an ancestor of a variable in $\mathcal{Z}$, all paths from S to D not passing through E are blocked (Daniel *et al.*, 2012).

The extended causal diagram corresponding to Figure 1b of Research Paper I is shown in Figure 4.6, where in the latter the unknown parents of E, U_1, have been added. In the setting studied in Research Paper I it is not possible to condition on R as this is an unknown risk factor (that is R cannot be included in the set $\mathcal{Z}$), thus the backdoor path E-R-D is open and Condition 1 is not satisfied. For the same reason the path S-R-D is not blocked and Condition 2 is not satisfied. This setting thus leads, in general, to biased estimates of causal effects.

According to Daniel et al (Daniel *et al.*, 2012) Condition 1 ensures that any association estimated

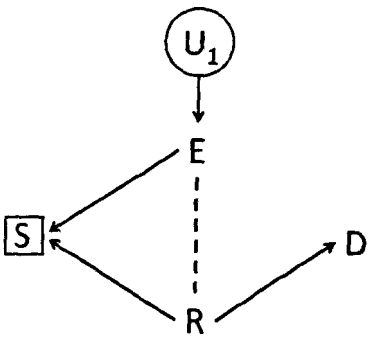


Figure 4.6: Extended causal diagram corresponding to Figure 1b of Research Paper I.

between the exposure and the outcome conditional on $\mathcal{Z}$ is causal, while Condition 2 concerns only "the intercept", i.e. the mean or prevalence of D (at the baseline level of all other variables in the model), depending on whether the outcome is continuous or binary. However following personal correspondence with the first author, it has transpired that this is not correct. Condition 2 is in general needed for an unbiased estimate of all estimands but there are exceptions where Condition 2 is not met but certain causal contrasts can be estimated without bias in the selected population for which S=1 by conditioning only on $\mathcal{Z}$. One such exception was seen in the settings with continuous outcome and a *null effect of* E *on* S reported in Table 4.1. There, Condition 2 is not met and still no evidence of bias in the slope parameter was found. This occurs because – under the assumed parametric model for D that did not include any E-R interactions – the mean outcome difference between exposed and unexposed in the full population, i.e. the marginal mean difference (marginal with respect to R), is the same as the marginal mean difference in the two strata defined by S=0 and S=1 separately. There are two crucial features of this setting which leads to this: the mean difference is a collapsible contrast and there is no E-R interaction. The estimated intercept is always biased, and even the estimated contrast would be biased in any other setting, i.e. when the contrast is noncollapsible (eg an odds ratio or hazard ratio) and/or when there is an interaction between E and R in their effect on D. The association between R and S, and the unavailability of R, means that intercepts and contrasts are marginalised over the distribution of R, which is different in the population and the selected sample

by virtue of the association between R and S. In the special case of a collapsible contrast and no E-R interaction, it does not matter how the marginalisation over R is done. Condition 2 must therefore be met when the estimand of interest is not collapsible and/or there is interaction between E and R in their effect on D. This is the case, for example, when the outcome is binary and interest focuses on the exposure-outcome hazard ratio or odds ratio.

For this reason we revisited the simulations of Research Paper I to consider the scenario of a null association between the exposure and the selection indicator ($OR_{SE} = 1$), both assuming a null association between E and D (see Figure 4.7) and assuming a value different from the null (see Figure 4.8). The extended causal diagrams corresponding to these two scenarios are represented in Figure 4.9 and 4.10 respectively, with U_2 being the unknow parents of the outcome D. According to the algorithm suggested by Daniel et al (*Daniel et al.*, 2012) in both settings Condition 1 is satisfied, but Condition 2 is not.

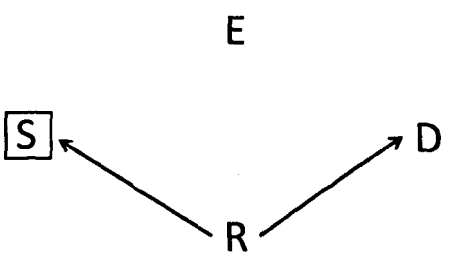


Figure 4.7: Causal diagram of a study conducted in a selected sample: a null association between the exposure of interest and both the selection indicator and the outcome is assumed.

When it was assumed a null association between the exposure and the outcome, as in Research Paper I, the bias resulted to be 0. However this was not the case when a true causal effect different from the null was assumed. In particular, we set OR_{SR} equal to either 0.25 or 4, HR_{DR} - conditional on E -

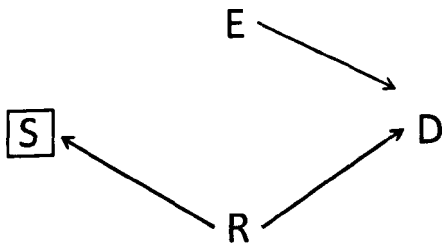


Figure 4.8: Causal diagram of a study conducted in a selected sample: the exposure of interest is associated with the outcome but not with the selection indicator.

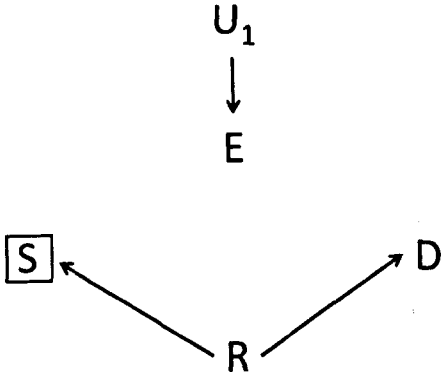


Figure 4.9: Extended causal diagram of a study conducted in a selected sample: a null association between the exposure of interest and both the selection indicator and the outcome is assumed.

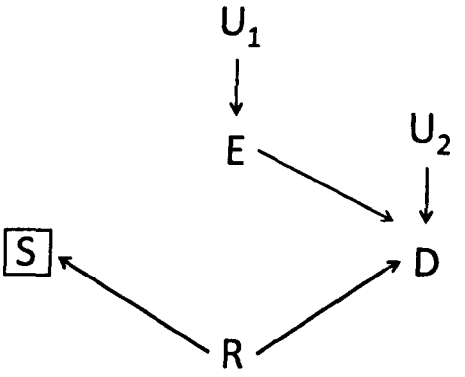


Figure 4.10: Extended causal diagram of a study conducted in a selected sample: the exposure of interest is associated with the outcome but not with the selection indicator.

equal to either 0.25 or 4, HR_{DE} – conditional on R – equal to 2, the baseline rate equal to 0.03, 0.1 and 0.2 events/year and the administrative censoring time at 5 years. The size of the bias was found to increase as the baseline rate increased, reaching a maximum of -0.01 when the baseline rate was equal to 0.03, and a maximum of -0.03 and -0.05 when the baseline rate was equal to 0.1 and 0.2 events/year respectively. The hazard ratio is noncollapsible and therefore the marginal HR_{DE} (with respect to R) is different from the conditional HR_{DE} (conditional on R) even when R is not a confounder, with the difference increasing as the prevalence of the outcome increases. Informally the marginal HR is equal to the conditional HR integrated over the distribution of R. The conditional HR is the same in the source population and in the selected sample but the distribution of R is different because R affects the selection (see Figure 4.10). It follows that when the conditional HR is integrated over the distribution of R to get the marginal HR the resulting effect is different in the source pupoulation and in the sample. However, under the null hypothesis (no effect of E on D) the conditional HR (on the log scale) is 0 for all values of R, both in the population and in the sample. Therefore even if averaging over two different distributions of R the marginal HR (on the log scale) with respect to R will also be 0. This is the reason why the setting depicted in Figures 4.7 leads to unbiased estimates of the causal effect, even when the parameter of interest is noncollapsible.

In summary if the parameter of interest is a collapsible contrast (e.g. a mean or a risk difference), and

if, conditional on R, is the same at every level of R (thus assuming no interaction between E and R in their effect on D), then only the intercept will be biased if Condition 2 fails and Condition 1 holds. In contrast whenever the parameter of interest, when calculated conditional on R, depends on R (i.e. E and R interact in their effect on D), or whenever the effect of interest is noncollapsible, such as an HR (or OR), and is different from the null, then the estimator of the marginal effect (with respect to R) in the selected sample will be unbiased for the marginal effect (with respect to R) in the population only if, in addition, Condition 2 holds.

4.3 Research Paper II

Research paper cover sheet

1. **For a 'research paper' already published**
   1.1. Where was the work published?
   JOURNAL OF EPIDEMIOLOGY AND COMMUNITY HEALTH

   1.2. When was the work published?
   NOVEMBER 2012 (Epub: 06-12-2011)

      1.2.1. If the work was published prior to registration for your research degree, give a brief rationale for its inclusion

   1.3. Was the work subject to academic peer review? YES

   1.4. Have you retained the copyright for the work? NO
   I acknowledge permission of the BMJ Publishing Group to include in this thesis the paper published in the Journal of Epidemiology and Community Health (2012), 66(11), 976-81.

2. For a 'research paper' prepared for publication but not yet published
   2.1. Where is the work intended to be published? _____
   2.2. List the paper's authors in the intended authorship order
   2.3. Stage of publication – Not yet submitted/Submitted/Undergoing revision from peer reviewers' comments/In press

3. For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper.

   Lorenzo Richiardi gave me substantial advice in designing the study. Together with Fulvio Lazzarato I performed the linkage between the NINFEA dataset and the Birth Registry database. I performed the statistical analyses and wrote the first draft of the article. All co-authors helped interpret the findings and revised the manuscript, providing useful comments.

Candidate's signature

Supervisor or senior author's signature to confirm role as stated in (3) (Bianca De Stavola)

Selection bias and patterns of confounding in cohort studies: the case of the NINFEA web-based birth cohort

Costanza Pizzi,[1,2] Bianca L De Stavola,[2] Neil Pearce,[2,3] Fulvio Lazzarato,[1] Paola Ghiotti,[4] Franco Merletti,[1] Lorenzo Richiardi[1]

[1]Cancer Epidemiology Unit, CeRMS and CPO-Piemonte, University of Turin, Turin, Italy
[2]Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK
[3]Centre for Public Health Research, Massey University, Wellington, New Zealand
[4]Department of Health, Piedmont Region, Turin, Italy

Correspondence to
Costanza Pizzi, Cancer Epidemiology Unit, CeRMS and CPO-Piemonte, University of Turin, Via Santena 7, Turin 10126, Italy.
costanza.pizzi@lshtm.ac.uk

Accepted 10 November 2011
Published Online First 12 June 2011

ABSTRACT

Background Several studies have examined the effects of sample selection on the exposure–outcome association estimates in cohort studies, but the reasons why this selection may induce bias have not been fully explored.

Aims To investigate how sample selection of the web-based NINFEA birth cohort may change the confounding patterns present in the source population.

Methods The characteristics of the NINFEA participants (n=1105) were compared with those of the wider source population—the Piedmont Birth Registry (PBR)—(n=36 092), and the association of two exposures (parity and educational level) with two outcomes (low birth weight and birth by caesarean section), while controlling for other risk factors, was studied. Specifically the associations among measured risk factors within each dataset were examined and the exposure–outcome estimates compared in terms of relative ORs.

Results The associations of educational level with the other risk factors (alcohol consumption, folic acid intake, maternal age, pregnancy weight gain, previous miscarriages) partly differed between PBR and NINFEA. This was not observed for parity. Overall, the exposure–outcome estimates derived from NINFEA only differed moderately from those obtained in PBR, with relative ORs ranging between 0.74 and 1.03.

Conclusions Sample selection in cohort studies may alter the confounding patterns originally present in the general population. However, this does not necessarily introduce selection bias in the exposure–outcome estimates, as sample selection may reduce some of the residual confounding present in the general population

INTRODUCTION

Cohort studies are frequently conducted in selected populations, with the study subjects either self-selected or selected according to some prespecified criteria. The consequence of this selection process on the internal validity of the exposure–outcome associations has been defined as selection bias,[1] or a special case of confounding,[2] and have been extensively discussed in the literature. In particular, the use of self-selected or restricted populations for cohort studies may introduce bias if the selection mechanism alters the original (population-level) associations between the exposure(s) of interest and the other risk factor(s) for the outcome, therefore changing the strength and, possibly, the direction of confounding.[1 2]

Several papers have compared the characteristics of participants in cohort studies with those of non-participants to assess the representativeness of the study sample.[3–11] Two of these have also evaluated the potential effects of the selection process on the exposure–outcome estimates of interest within the context of birth cohorts.[4 5] However, none of these earlier studies have specifically explored the mechanisms through which bias can be induced by the sample selection process.

In this paper, we aim to examine these mechanisms, focusing on comparisons of the confounding patterns (ie, the associations among all outcome risk factors, including the exposure of interest) for the associations of interest in the general population and in the selected sample.

We will use data from the NINFEA study, an established ongoing web-based birth cohort, in which study subjects enrol through the internet.[12] It started as a pilot study in the city of Turin (capital of the Piedmont Region, Italy) but it has been extended, since December 2007, to the rest of the country. Recruitment occurs during pregnancy, when the women are informed about the study and may choose to register on the study website. The existence of the study is advertised at hospitals and family clinics of selected areas and through websites of interest for pregnant women. Undoubtedly the NINFEA participants are a selected sample of the source population, with participation strongly associated with socioeconomic factors.[12]

We decided to investigate the impact of the changes in the confounding patterns due to the sample selection on the exposure–outcome associational estimates observed in NINFEA, by comparing the characteristics of the study participants with those of the wider source population (the Piedmont Birth Registry, PBR). Our specific objectives were: (1) to explore the frequency of selected variables in PBR and NINFEA; (2) to compare the associations between the exposure of interest and the other risk factors available in the sample and in the source population; (3) to formally compare the exposure–disease association estimates obtained in PBR and in NINFEA; and (4) to examine alternative potential mechanisms leading to these results using directed acyclic graphs (DAGs).[13–15]

MATERIAL AND METHODS

Data

We used the PBR data for the period 2005–08, which includes 145 885 pregnancy records, created by midwives at the time of the delivery.[16]

Compulsory computerised birth registration was established in the whole of Italy in 2001. In Piedmont it is of particularly high quality and completeness.[17] Records from PBR were linked to those from the NINFEA study (data downloaded in July 2010). The latter included 1547 singleton pregnancies after exclusion of births occurring after December 2008 and those occurring outside Piedmont. The date of birth of the mother, the date of birth and the sex of the child, and the hospital and ward where the delivery occurred were used as (deterministic) linkage key variables. The linkage was successful for 1298 (84%) births. We further excluded births with gestational age earlier than the 25th or later than the 44th week, as well as those with implausible birth weights in the PBR data. This led to reducing the original dataset to 145 496 records, 1295 of which were linked to NINFEA. Since most of the NINFEA births occurred in the city of Turin and most of their parents were Italian, we further restricted the analyses to children born in Turin from Italian parents to avoid strata with sparse data. Thus the final PBR dataset was substantially reduced (n=36 092), unlike the NINFEA dataset (n=1105).

The PBR holds information on maternal and child/delivery characteristics. In particular, data on parents' age, educational level and occupation were available, together with maternal smoking and alcohol consumption during pregnancy as well as pregnancy weight gain and intake of folic acid. Information on the reproductive history of the mother (ie, parity, previous miscarriages and use of infertility treatment) and information on reproductive outcomes (ie, type, gestational age, birth size) is also recorded in the PBR.

Statistical analysis

In order to assess the impact of selection on the estimate of an effect of interest, we examined two outcomes: low birth weight for gestational age (LBW, defined as birth weight lower than the 20th percentile of the internally gestational age-standardised distribution) and birth by caesarean section.

We then selected the main potential risk factors recorded in the PBR for these two outcomes and examined their prevalence ORs of participation in NINFEA. Self-selection or cohort restriction may introduce non-negligible bias in cohort studies when the exposure–selection OR is ≥2 or ≤0.5.[18] In our study, only parity and maternal education satisfied this criterion. Thus, we chose these two main predictors of participation as the exposures of interest and studied them in association with the two outcomes, while we treated the other variables—maternal age, weight gain during pregnancy, consumption of folic acid, alcohol during pregnancy and history of previous miscarriages—as potential confounders. Maternal smoking and use of infertility treatment were not considered further as potential confounders because of their low population prevalence.

The analyses involved: (1) estimating the Prevalence ORs (POR); (2) investigating the effects on the exposure–potential confounder associations of restricting the analyses to the selected sample; and (3) for each outcome separately, formally comparing the estimated exposure–outcome associations obtained in the selected sample and in the original population.

For simplicity, all continuous variables were dichotomised. Namely, low pregnancy weight gain was defined as a weight gain lower than the 20th percentile in PBR (10 kg); young maternal age as lower than the PBR median (33 years); and maternal parity as nulliparous (ie, no previous live births) versus parous. Logistic regression analyses were performed to study associations, leading to estimated ORs and corresponding 95% CIs. For the two outcomes of interest, both the crude and the fully adjusted (by all potential confounders) ORs for the two exposures were estimated in each dataset. Their formal comparison was performed in terms of relative ORs—that is, the ratios of the NINFEA-based OR over the PBR-based OR, with CIs obtained as in Nohr *et al*.[5] We focused on ORs in line with previous publications on self-selection bias[4 5] although results were substantially unchanged when based on risk ratios.

RESULTS

Of 36 092 delivery records included in the PBR dataset, 1105 were participants of the NINFEA cohort—that is, a participation proportion of 3.1%.

Predictors of participation

The two exposures of interest (maternal education and parity) and most of the seven other potential selected risk factors were associated with participation in NINFEA. Low parity, high educational level and non-smoking during pregnancy were the strongest predictors in both crude and mutually adjusted analyses (table 1). There was some evidence of effect modification between maternal education and age (p<0.001), with the OR of participation for high education level increasing from 1.9 among older women (>32 years) to 3.4 among younger ones (≤32 years). When this interaction was included in the model the adjusted ORs for the other factors did not change (data not shown).

Table 1 Frequency distribution of potential risk factors and crude and adjusted ORs of participation into the NINFEA cohort study

	Frequency distribution (%)		Effect on participation (3.1%)		
Potential risk factors	PBR	NINFEA	Crude OR	Adjusted OR*	95% CI
Parity (1+)	43.5	21.2	0.34	0.32	0.27 to 0.37
Maternal education (graduate)	18.1	34.1	2.42	2.30	2.02 to 2.62
Smoking during pregnancy	8.9	3.7	0.39	0.43	0.31 to 0.59
Alcohol during pregnancy	21.5	27.4	1.39	1.32	1.15 to 1.51
Infertility treatment	2.0	3.8	2.00	1.31	0.95 to 1.81
Folic acid intake	80.3	85.8	1.50	1.34	1.13 to 1.60
Maternal age (>32 years)	53.7	58.4	1.22	1.30	1.14 to 1.48
Previous miscarriages	17.5	17.7	1.02	1.10	0.93 to 1.29
Low pregnancy weight gain (<10 kg)	18.6	19.6	1.07	1.10	0.95 to 1.29

*Adjusted for all the other variables in the table.

Associations between exposures and risk factors

Parity

There is a substantial difference in the distribution of parity across the two datasets: about 45% of PBR records involved women with parity greater than 0, compared with about 20% in the NINFEA study (table 1).

In the PBR population, parity was associated with almost all the potential confounders, although ORs were greater than 1.5 only for maternal age and history of previous miscarriages (table 2). Findings in NINFEA were generally similar, although the association between maternal age and parity was slightly stronger (OR=3.17; 95% CI 2.27 to 4.45 in NINFEA vs OR=2.45; 95% CI 2.35 to 2.56 in PBR) (table 2).

Maternal education

Maternal education also has a different distribution in the two datasets, with more educated women contributing to NINFEA (table 1). In PBR, maternal education was strongly associated with maternal age (OR=2.09, 95% CI 1.97 to 2.21) while in NINFEA the association with maternal age was weaker (OR=1.22, 95% CI 0.95 to 1.57), and that with folic acid intake was stronger (table 3).

Associations between exposures and outcomes

Caesarean section

The upper left side of table 4 reports the crude ORs of caesarean section for parity and maternal education, obtained in PBR and NINFEA. The estimates obtained from NINFEA are closer to the null value than those obtained from PBR, as reflected by relative ORs below 1.0. When the ORs were adjusted for all other potential risk factors, the estimates from PBR and NINFEA were both reduced to a similar extent, leading to substantially unchanged relative ORs (table 4).

Low birth weight for gestational age

The crude ORs of LBW for parity were reasonably similar when estimated in PBR and NINFEA (relative OR=0.79; table 4). When these estimates were adjusted for potential confounders, their relative sizes did not change markedly (relative OR=0.74, 95% CI 0.49 to 1.13). However, residual confounding should not be discounted, as information on a number of known risk factors for LBW was not available, and it is therefore hard to predict whether the adjusted OR estimate of 0.43 found in NINFEA is more or less affected by residual confounding than that of 0.58 found in PBR.

High level of education was found to be inversely associated with LBW. The ORs estimated in PBR and in NINFEA were equal, leading to crude and adjusted relative ORs of 1.0 (table 4).

Table 2 Crude associations (OR and 95% CI) of parity (parous vs nulliparous) with other potential risk factors in the Piedmont Birth Registry (PBR) population and among the NINFEA participants

	Parity (1+ vs 0)	
Potential risk factors	PBR OR (95% CI)	NINFEA OR (95% CI)
Alcohol during pregnancy	0.92 (0.87 to 0.97)	1.08 (0.78 to 1.49)
Folic acid intake	0.77 (0.74 to 0.82)	0.89 (0.59 to 1.33)
Maternal education (graduate)	1.11 (1.05 to 1.17)	1.24 (0.92 to 1.68)
Maternal age (>32)	2.45 (2.35 to 2.56)	3.17 (2.27 to 4.45)
Low pregnancy weight gain (<10 kg)	1.33 (1.26 to 1.40)	1.26 (0.89 to 1.79)
Previous miscarriages	1.67 (1.58 to 1.76)	1.59 (1.12 to 2.26)

Table 3 Crude associations (OR and 95% CI) of maternal education with other potential risk factors in the Piedmont Birth Registry (PBR) population and among the NINFEA participants

	Maternal education (graduate vs non-graduate)	
Potential risk factors	PBR OR (95% CI)	NINFEA OR (95% CI)
Alcohol during pregnancy	1.09 (1.02 to 1.16)	1.03 (0.78 to 1.36)
Folic acid intake	1.01 (0.95 to 1.08)	1.44 (0.99 to 2.10)
Parity (1+)	1.11 (1.05 to 1.17)	1.24 (0.92 to 1.68)
Maternal age (>32 years)	2.09 (1.97 to 2.21)	1.22 (0.95 to 1.57)
Low pregnancy weight gain (<10 kg)	1.01 (0.94 to 1.08)	1.02 (0.75 to 1.40)
Previous miscarriages	0.95 (0.89 to 1.02)	0.85 (0.61 to 1.18)

PBR, Piedmont Birth Registry.

DAGs illustrating the effect of changes in the confounding pattern due to sample selection on bias

In a web-based cohort study, such as NINFEA, the sample selection process is driven by two main mechanisms: (1) the restriction of the source population to internet users; and (2) the decision to participate. For simplicity we will focus here only on the latter mechanism; however, this line of reasoning can be generalised to any other selection mechanism.

Conditioning on volunteers may either induce or attenuate bias depending on the population-level relationships between the exposure of interest and the other risk factors. The two main scenarios that may arise are illustrated, using DAGs, in figures 1 and 2, where E indicates the exposure of interest, R is a risk factor for the disease of interest D, U_1 and U_2 are other unmeasured variables, and S is an indicator of selection into the sample. For simplicity, we assume no causal effect of the exposure on the outcome, no interaction between E and R in their effect on D, and that no mediator on the E–D pathway can influence selection. In the first scenario (figure 1), E and R are *independent* in the general population and both affect the likelihood of being selected (through volunteering), either directly (figure 1A) or as proxy of some other factors U_1 and U_2 (figure 1B). Restricting on S induces an association between E and R, making R a confounder in the subset of the participants. Thus selection is likely to induce bias in the exposure–disease association unless the back-door path E–R–D is blocked, for example by adjustment. In our data, this may be the case of maternal education (E) and folic acid intake (R), which are both associated with participation (table 1). These variables are independent in the PBR population (OR=1.01 in table 3), but become associated among the NINFEA participants (OR=1.44 in table 3).

In the second scenario (figure 2), E and R still affect the probability of participation, but now they share a common cause U (figure 2A), implying that R is already a confounder in the general population. Under this scenario the selection process can either induce or attenuate the bias, depending both on the strength and direction of the E–R association present in the general population and of the equivalent associations within the restricted sample defined by S. Although there are a number of exceptions,[19] typically when two variables influence a third one in the opposite (*same*) direction, conditioning on the latter leads to a positive (*negative*) spurious association between the first two variables.[13 18] It follows that when E and R are, for example, positively associated in the general population, selection is likely to attenuate the bias in an unadjusted estimate of the E–D association, if E and R have a qualitatively similar effect on the probability of volunteering. The opposite would happen if they

Table 4 Crude and fully adjusted ORs for the effect of parity and maternal education on caesarean section and LBW in the birth registry population and in the NINFEA cohort, together with the corresponding relative ORs

Outcome	Crude OR (95% CI)			Fully adjusted OR (95% CI)		
Risk factors	PBR	NINFEA	ROR	PBR	NINFEA	ROR
Caesarean section						
Parity	1.68 (1.60 to 1.76)	1.34 (0.92 to 1.93)	0.80 (0.55 to 1.15)	1.51 (1.44 to 1.59)*	1.22 (0.84 to 1.78)*	0.81 (0.56 to 1.18)*
Maternal education	1.15 (1.08 to 1.23)	1.0 (0.71 to 1.39)	0.86 (0.62 to 1.20)	1.07 (1.01 to 1.14)†	0.97 (0.69 to 1.35)†	0.90 (0.65 to 1.25)†
LBW						
Parity	0.59 (0.56 to 0.63)	0.47 (0.31 to 0.71)	0.79 (0.53 to 1.19)	0.58 (0.55 to 0.61)*	0.43 (0.28 to 0.66)*	0.74 (0.49 to 1.13)*
Maternal education	0.86 (0.80 to 0.92)	0.86 (0.64 to 1.17)	1.0 (0.75 to 1.35)	0.86 (0.80 to 0.92)†	0.89 (0.65 to 1.21)†	1.03 (0.76 to 1.39)†

*Adjusted for maternal age, occurrence of previous miscarriages, folic acid intake, alcohol consumption during pregnancy, maternal pregnancy weight gain and maternal education.
†Adjusted for maternal age, occurrence of previous miscarriages, folic acid intake, alcohol consumption during pregnancy, maternal pregnancy weight gain and parity.
LBW, low birth weight; PBR, Piedmont Birth Registry; ROR, relative OR.

had a qualitatively opposite effect on selection. We observed this pattern in our data. Let us consider the case of parity and maternal age, which are, respectively, negatively (OR=0.32) and positively (OR=1.30) associated with participation in NINFEA (table 1) while they are positively associated in the source population (OR=2.45; table 2). When analyses are restricted to the NINFEA data, the OR for their association increases to 3.17 (table 2). In contrast, the OR for the association between maternal education and maternal age, which are both positively associated with the probability of being selected (table 1)—decreased from 2.09 in PBR to 1.22 in NINFEA (table 3). As a consequence the OR for maternal education and caesarean section, for example, estimated adjusting for all other known risk factors except for maternal age, shows a smaller degree of residual confounding than that estimated from PBR. In particular the estimate derived from PBR increases from 1.07 (95% CI 1.01 to 1.14; table 4) when fully adjusted, to 1.14 (95% CI 1.08 to 1.22) when adjusted for all measured risk factors except for maternal age. The corresponding figures from NINFEA are 0.97 (95% CI 0.69 to 1.35; table 4) and 0.98 (95% CI 0.70 to 1.37), which are almost identical.

A special and relevant case is when the association in the population between E and R is due to a common unmeasured cause U, which also influences the probability of volunteering (figure 2B). Under this scenario, conditioning on S may imply a partial conditioning on U and a consequent attenuation of the confounding bias due to E–U–R–D. For example, if socioeconomic condition (U) is a determinant of smoking (E) and drinking (R) in pregnancy as well as of volunteering, restricting the analyses to the selected sample is equivalent to adjusting for a proxy of socioeconomic status. As a consequence the association between smoking and drinking would be attenuated in the selected sample.

DISCUSSION

In this paper we have investigated whether the patterns of confounding present in NINFEA differed from the ones observed in the whole PBR population, and assessed the extent of selection bias affecting the NINFEA study, by comparing the exposure–outcome associational estimates derived from these two datasets.

Undoubtedly the NINFEA participants differ substantially from the general population. Consistent with what has been found in other birth cohort studies, participating mothers are more likely to take folic acid during pregnancy, to be nulliparous, to be less likely to smoke,[4 5] and are on average older at delivery and with a higher educational level[3] than in the general population.

Several papers have examined the baseline characteristics of participants in cohort studies in comparison with those of non-participants to assess the representativeness of the study sample,[3–11] but none of them has specifically explored the mechanisms through which the extent of bias in the exposure–outcome estimates may be related to changes in the confounding patterns for the associations of interest. We have

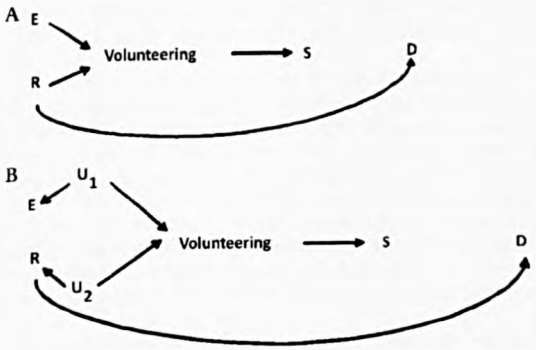


Figure 1 Diagram of a cohort where subjects volunteer to participate. (A) In the population the exposure of interest (E) is independent of the disease risk factor (R) and both affect the likelihood of being selected into the sample (S), through volunteering. (B) (E) is independent of the disease risk factor (R) and both affect the likelihood of S (through volunteering) as a proxy of some other factors (U).

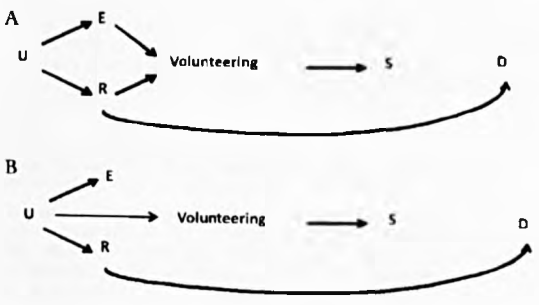


Figure 2 Diagram of a cohort where subjects volunteer to participate. (A) In the population the exposure of interest (E) and the disease risk factor (R) are associated between each other as they share a common cause of some other factors (U), and they both affect the probability of participation (S), through volunteering. (B) In the population E and R share a common cause U, and U also affects the probability of volunteering.

focused on the associations between the exposures of interest and the other risk factors for which we had information both in the general population, and in the selected sample. The extent of the bias affecting the exposure–disease association estimates in NINFEA was then assessed in the light of the potential effect of changes in the confounding patterns due to the sample selection.

Our findings indicate that the main variables to be controlled to minimise bias identified for the general population may not be the most important or relevant ones for analyses involving a cohort study of a selected subgroup of the general population. For example, our data showed that the estimated associations between maternal education and the other potential risk factors in NINFEA differed from those in PBR (table 3). Similar reasoning would apply to the role of unknown confounders, and therefore it is not possible to predict whether estimates based on a selected cohort would be more or less biased than those based on the equivalent population-based cohort. In fact, consistent results between the source population and the selected cohorts would argue in favour of similar confounding patterns or absence of confounding. Analogously, if results differ, this may imply that either residual confounding is an important issue for the exposure–disease association of interest or that the distribution of some unknown modifiers of this association differs between the selected and the population-based cohort. Obviously, these two scenarios are not mutually exclusive.

In order to understand whether the estimates derived from the selected sample are more or less valid than those obtained in the general population, it would be necessary to distinguish between the scenarios depicted in figures 1 and 2. In other words, it would be necessary to know whether the exposure (E) and the potential unmeasured confounder (R) are already associated in the general population or become associated in the selected sample. In some cases, expert opinion could be invoked to assess the likelihood of one scenario over the other. However, there is little published data on E–R associations in different populations but associations among risk factors are very likely.[20]

The findings presented here are based on a relative small study size, as our analyses of the NINFEA cohort include only 1105 subjects: thus, the precision of some estimates was low. In particular the numbers were too small to explore the effects of some important variables, such as maternal smoking during pregnancy and use of infertility treatment. Another limitation of this study concerns the lack of data on important risk factors and the consequent potential effect of residual confounding which may partly explain the observed differences in the estimated effects.

Taking this into consideration, our study nevertheless suggests that the estimates derived from NINFEA do not differ considerably from those obtained from PBR, with relative ORs ranging between 0.74 and 1.03. In agreement with previous studies investigating the effect of non-participation in birth cohort studies,[4 5] we have shown that even in a web-based birth cohort, selection does not induce substantial bias in the exposure–outcome associations we investigated. It is however important to consider the expected magnitude of these effects of interest when evaluating potential biases: when this is really small, even a moderate bias becomes relevant.

It should also be noted that this study only concerns baseline selection in cohort studies, which is by definition independent of the outcome. Our results cannot be extrapolated to the selection arising because of informative drop-outs, or to other epidemiological study designs, such as cross-sectional studies or case–control studies, where the disease/outcome status may affect participation.

What is already known on this subject

- Several papers have examined the baseline characteristics of participants in cohort studies in comparison with those of non-participants and some have also evaluated the potential effects of the selection process on the exposure–outcome estimates of interest.
- However none of these studies have specifically explored the mechanisms through which bias can be induced by the sample selection process.

What this study adds

- By comparing the confounding patterns and the exposure–outcome associations between the general population and the NINFEA internet-based birth cohort study sample, we found that possible differences in the estimates of the exposure–outcome associations between a selected and a population-based cohort reflect changes in the confounding patterns due to the sample selection process.
- Therefore, unless all relevant confounders in the two cohorts are known and measured, it is not possible to predict whether estimates based on a cohort selected at baseline would be more or less biased than those based on the equivalent population-based cohort, as sample selection might also reduce the confounding already present in the general population.

In conclusion, possible differences in the estimates of the exposure–outcome associations between a selected and a population-based cohort reflect changes in the confounding patterns due to the sample selection process. Therefore, unless all relevant confounders in the two cohorts are known and measured, it is not possible to predict whether estimates based on a cohort selected at baseline would be more or less biased than those based on the equivalent population-based cohort, as sample selection might also reduce the confounding already present in the general population.

Funding This work was supported by Compagnia SanPaolo/FIRMS, the Piedmont Region, the Italian Ministry of University and Research (MIUR), the Italian Association for Research on Cancer (AIRC) and the Massey University Research Fund (MURF). The Centre for Public Health Research is supported by a Programme Grant from the Health Research Council of New Zealand

Competing interests None

Ethics approval The study was approved by the Ethical Committee of the San Giovanni Battista Hospital—A.S.O. C.T.O./C.R.F./Maria Adelaide, Turin, Italy

Contributors All authors are responsible for the reported research and have made substantial contributions to the conception and design of the study, acquisition of data, analysis and interpretation of data, drafting the article or revising it critically for important intellectual content. All authors have seen and approved the final version of the manuscript.

Provenance and peer review Not commissioned; externally peer reviewed

REFERENCES

1. **Hernan MA**, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology* 2004;**15**:615–25.
2. **Rothman KJ**, Greenland S, Lash TL. *Distinguishing Selection Bias from Confounding. Modern Epidemiology*. Philadelphia: Lippincott Williams & Wilkins, 2008:136–7.

3 **Clarisse B,** Nikasinovic L, Poinsard R, *et al*. The Paris prospective birth cohort study: which design and who participates? *Eur J Epidemiol* 2007;**22**:203–10.
4 **Nilsen RM,** Vollset SE, Gjessing HK, *et al*. Self-selection and bias in a large prospective pregnancy cohort in Norway. *Paediatr Perinat Epidemiol* 2009;**23**:597–608.
5. **Nohr EA,** Frydenberg M, Henriksen TB, *et al*. Does low participation in cohort studies induce bias? *Epidemiology* 2006;**17**:413–18.
6 **Goldberg M,** Chastang JF, Leclerc A, *et al*. Socioeconomic, demographic, occupational, and health factors associated with participation in a long-term epidemiologic survey: a prospective study of the French GAZEL cohort and its target population. *Am J Epidemiol* 2001;**154**:373–84.
7 **Heilbrun LK,** Ross PD, Wasnich RD, *et al*. Characteristics of respondents and nonrespondents in a prospective study of osteoporosis. *J Clin Epidemiol* 1991;**44**:233–9.
8. **Jackson R,** Chambless LE, Yang K, *et al*. Differences between respondents and nonrespondents in a multicenter community-based study vary by gender ethnicity. The Atherosclerosis Risk in Communities (ARIC) Study Investigators. *J Clin Epidemiol* 1996;**49**:1441–6.
9. **Macera CA,** Jackson KL, Davis DR, *et al*. Patterns of non-response to a mail survey. *J Clin Epidemiol* 1990;**43**:1427–30.
10. **Shahar E,** Folsom AR, Jackson R. The effect of nonresponse on prevalence estimates for a referent population: insights from a population-based cohort study. Atherosclerosis Risk in Communities (ARIC) Study Investigators. *Ann Epidemiol* 1996;**6**:498–506.
11. **Walker M,** Shaper AG, Cook DG. Non-participation and mortality in a prospective study of cardiovascular disease. *J Epidemiol Community Health* 1987;**41**:295–9.
12 **Richiardi L,** Baussano I, Vizzini L, *et al*. Feasibility of recruiting a birth cohort through the Internet: the experience of the NINFEA cohort. *Eur J Epidemiol* 2007;**22**:831–7.
13 **Glymour MM.** Using causal diagrams to understand common problems in social epidemiology. In: Oakes JM, Kaufman JS, eds. *Methods in Social Epidemiology*. San Francisco: Jossey-Bass, 2006.
14. **Greenland S,** Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology* 1999;**10**:37–48.
15. **Pearl J.** *Causality: Models, Reasoning, and Inference*. Cambridge, UK: Cambridge University Press, 2000.
16. **Assessorato della sanità Regione Piemonte.** *Nascere in Piemonte—Executive Summary*. Torino, 2007. www.regione.piemonte.it/sanita/program_sanita/dip_materno_inf/dwd/nas_piem_sum.pdf
17 **Ministero del Lavoro della Salute e delle Politiche Sociali—Settore Salute.** *Certificato di Assistenza al Parto (CeDAP). Analisi dell'evento Nascita—Anno*. Roma, 2006. www.salute.gov.it/imgs/C_17_pubblicazioni_1120_allegato.pdf
18. **Pizzi C,** De Stavola B, Merletti F, *et al*. Sample selection and validity of exposure–disease association estimates in cohort studies. *J Epidemiol Community Health* 2011;**65**:407–11.
19 **VanderWeele TJ,** Robins JM. Directed acyclic graphs, sufficient causes, and the properties of conditioning on a common effect. *Am J Epidemiol* 2007;**166**:1096–104.
20 **Smith GD,** Lawlor DA, Harbord R, *et al*. Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. *PLoS Med* 2007;**4**:e352.

4.4 Book Chapter

Book chapter cover sheet

For a 'research paper' already published

1.1. Where was the work published?
1.2. When was the work published?
1.2.1. If the work was published prior to registration for your research degree, give a brief rationale for its inclusion

---

1.3. Was the work subject to academic peer review?
1.4. Have you retained the copyright for the work?
If yes, attach evidence of retention
If no, or if the work is being included in its published format, attach evidence of permission from copyright holder (publisher or other author) to include work

2. For a 'research paper' prepared for publication but not yet published

2.1. Where is the work intended to be published?
Book title: "Handbook of Epidemiology, 2nd edition"; Editors: "Ahrens Wolfgang, Pigeot Iris"; Publisher: "Springer"

2.2. List the paper's authors in the intended authorship order
Lorenzo Richiardi, Costanza Pizzi, and Daniela Paolotti

2.3. Stage of publication – Not yet submitted/Submitted/Undergoing revision from peer reviewers' comments/In press
IN PRESS
(I acknowledge Springer Science+Business Media for kind permission of including in this PhD thesis the Chapter entitled "Internet-Based Epidemiology", which is part of the Springer book "Handbook of Epidemiology, 2nd edition", whose Editors are Ahrens Wolfgang and Pigeot Iris)

3. For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)

I wrote the first draft of Section 2 of the chapter entitled "Selection and Selection Bias in Internet-Based Studies". I gave substantial contribution to the other sections, revising the whole chapter and providing several comments, most of which have been included in the final version of the chapter.

Candidate's signature

Supervisor or senior author's signature to confirm role as stated in (3) (Lorenzo Richiardi)

Internet-based epidemiology

Lorenzo Richiardi[1], Costanza Pizzi[1,2], Daniela Paolotti[3]

1 Cancer Epidemiology Unit, Department of Medical Sciences, University of Turin and CPO-Piemonte, Italy

2 Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK

3 Computational Epidemiology Lab, ISI Foundation, Turin, Italy

1. INTRODUCTION

Currently almost 2 billion persons worldwide, i.e. 30% of the world population, have access to the internet (Internet World Stats). These numbers are accurate as of June 30, 2010 and will soon be outdated: access has grown by 445% from 2000. The proportion of users varies by country, being 77% in North America and 58% in Europe, with notable peaks in the Nordic countries, where Denmark and Finland are above 85% and Norway and Sweden are above 90%. Figure 1 provides a global view on the proportion of users and growth in use in the last ten years.

The proportion of users is not homogenously distributed in the population. In 2007, for example, according to Eurostat data, in all EU countries, with no exceptions, the proportion of internet users was higher among men than women and among people aged 16 to 24 than older persons (United Nations Economic Commission for Europe -UNECE- 2011). It is reasonable, however, to expect that sooner or later almost everybody in the world will have access to the internet, with no marked country, age or sex differences. This is a very attractive prospect for epidemiologists and human researchers in general, who already recognized the possibility of using the internet to conduct field studies in the early 1990s.

It is perhaps surprising that the internet has been initially used mainly to conduct surveys rather than longitudinal studies or interventions, although the latter are less vulnerable to selection bias. The first medical surveys, such as those of patients with inflammatory bowel diseases (Hilsden et al. 1999; Soetikno et al. 1997) or diabetes (Baehring et al. 1997), have been published at the end of the 1990s.

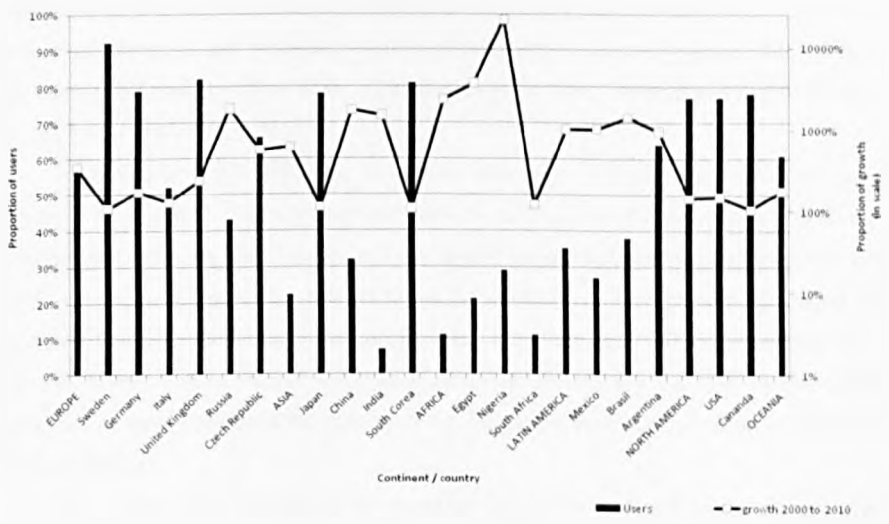


Figure 1. Proportion of internet users on June 30, 2010 and growth from 2000 to 2010 in selected countries. (Internet World Stats 2011)

These studies were pioneered by surveys carried out by psychologists and sociologists (Berrens et al. 2003; Buchanan and Smith 1999; Kraut et al. 2004; Skitka and Sargis 2006). In 1997, Kushi and colleagues reported the launch of a pilot study for an internet-based cohort on diet and breast cancer (Kushi et al. 1997) and a small number of web-based birth cohort studies have been conducted in the last ten years (Hercberg et al. 2010; Mikkelsen et al. 2009; Richiardi et al. 2007; Treadwell et al. 1999; Turner et al. 2009). However, most of the internet-based medical studies are currently intervention trials. The internet was first suggested as a tool to manage all aspects of the trial, including randomization and data acquisition (Kelly and Oldham 1997; Pepine et al. 1998) but, in the last ten years, it has often been used for the purpose of recruiting study participants (McAlindon et al. 2003; Wang and Etter 2004).

The idea of using the internet in empirical research in general, and epidemiological research in particular, often receives skeptical reactions. Typical concerns include problems related to lack of exposure heterogeneity in the study participants, phantom participations, duplicate records by the same participant, low data quality, and confidentiality issues (Gosling et al. 2004; Skitka and Sargis 2006; van Gelder et al. 2010). In fact, most of these problems have limited implications or can be solved technically. For example, the lack of heterogeneity does not apply to most of the exposures,

as in many countries more than half of the population has access to the internet. Similarly, various methods have been developed to identify duplicate entries, to enhance data quality and to ensure data security and confidentiality, including the use of encrypted connections, registration through individual username and password, the use of screening questions to detect duplicates and checks for implausible answers (Baer et al. 2002; Bowen et al. 2008; Dillman and Smyth 2007; Gosling et al. 2004; van Gelder et al. 2010).

As for any study based on volunteers, the main critical issue related to internet-based research is, however, the representativeness of participants for the study population and the likelihood of selection bias. This is currently largely debated. At an international conference held in 2008, for example, one of the authors heard the editor of an epidemiological journal saying that, because of concerns about self-selection, he was a priori against publishing results of internet-based surveys. Issues of selection and selection bias will be extensively discussed in this chapter both in an ad-hoc section and within the context of the discussion about the use of the internet in each type of study design.

In contrast to its limitations, the use of the internet in epidemiological research offers several advantages, including decreased costs, simplified logistics, rapidity, flexibility, the possibility to tailor the questionnaire to the participants' characteristics, instantaneous checks to identify inconsistencies as well as to reduce errors resulting from data entry (Baer et al. 2002; Dillman and Smyth 2007; van Gelder et al. 2010). Moreover, on-line studies have fewer constraints than traditional studies, both from a geographical and a temporal point of view: they can reach distant or "hidden" populations as well as they can recruit continuously for several years. Some other advantages are specific to the different study designs and will be discussed further in the corresponding sections of this chapter. However, a common feature of internet-based studies, often overlooked, is the active involvement and empowerment of the study subjects (Rhodes et al. 2003). They can give feedback over the whole duration of the project and receive (and comment on) information about the study results. For example, it is not uncommon that, within the framework of an internet-based study, researchers keep constant contact with the study participants using social networks.

In several studies, the internet has been used to submit on-line questionnaires to a pre-specified population, including members of a "traditional" cohort (Ekman et al. 2006; Russell et al. 2010), members of internet panels (Silver et al. 2002; West et al. 2006) or members of a mailing list (Ruddy et al. 2010). In all these studies, the internet was not used to directly recruit participants, whereas it was used as a tool to deliver a questionnaire to a pre-selected population. We will not discuss advantages and limitations of on-line questionnaires and technicalities on how to prepare

them in this chapter (about these issues see for example references (Baer et al. 2002; Dillman et al. 2009; Ekman et al. 2007; Kongsved et al. 2007; Russell et al. 2010; Schleyer and Forrest 2000; van Gelder et al. 2010). Rather, we will focus on the internet as a method to recruit study participants and its influence on study design and validity. Typically, internet-based studies use on-line questionnaires but this is not a necessary feature. Indeed, researchers may have a direct contact with participants via the internet, for example, to complete a telephone interview or to obtain biological samples (Etter et al. 2005; Richiardi et al. 2007).

In this chapter we will address general methodological issues about internet-based studies and discuss examples of the use of the internet in the context of different types of epidemiological designs, from surveys to randomized studies. Debates about the use of the internet in epidemiological research may be affected by preconceived opinions either in favor or against it. Being involved in internet-based research, we cannot be objective but we will aim at discussing the different issues impartially.

2. SELECTION AND SELECTION BIAS IN INTERNET-BASED STUDIES

Participants in an internet-based study are doubtless a selected population, regardless of the epidemiological design chosen to carry out the study. This is due to two main reasons: i) the source population is restricted to internet users (either all internet users or users of specific websites), and ii) participation of subjects should be voluntary. What does instead depend on the study design is the mechanism through which bias may be induced by the sample selection process. We will illustrate these mechanisms in different study designs also using directed acyclic graphs (DAG) (see Chap. Directed Acyclic Graphs of this handbook.) indicating E as the exposure of interest, R as a risk factor (or a set of risk factors) for the outcome of interest D, U_i as other unknown/unmeasured variables and S as the indicator of selection into the sample. In all figures, a square around S will indicate conditioning on study participation.

In surveys aiming at estimating the prevalence of a disease, there is selection bias if the disease status or any determinant of the disease is associated with the selection probability. This is very likely to happen in internet-based studies. If determinants of selection are known, it is possible to apply weights and obtain a valid estimate of the prevalence. However, in most situations it is very difficult to obtain a good picture of the determinants involved in the selection in a study based on volunteers recruited through the internet.

In studies aiming at estimating associations, in general there is selection bias if the probability of selection depends on both the exposure (E) and the outcome (D) of interest (Hernán

et al. 2004). This is illustrated in Figure 2. For example, in an internet-based cross-sectional study on asthma, the probability of volunteering to take part into the study could be associated with having asthma (D) as well as with living in a heavily polluted area (E). Under these circumstances, because of the conditioning on selection, an internet-based study would most likely find an association between air pollution and asthma even in the case of a lack of a true causal association between these two variables. This type of selection bias can be large.

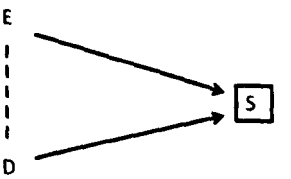


Figure 2. Diagram of a study where selection of the study subjects (S) depends on both the exposure (E) and the outcome (D) of interest. The dashed line represents an association induced by conditioning on S (represented by a square around S).

2.1 Longitudinal studies

In longitudinal studies where the outcome of interest occurs after being selected into the study sample, the mechanism that we have just described does not apply but selection bias may still occur (Glymour 2006; Hernan et al. 2004). In particular, if the likelihood of participation in the study depends on both exposure (E) and a disease risk factor (R) for the outcome of interest (D), and these are independent of each other in the general population, the selection process induces an association between E and R; R becomes a confounder of the E-D association and thus, if we cannot adjust for it in the analyses, the estimate of the association is biased (Figure 3). We (Pizzi et al. 2011) and others (Greenland 2003) showed that the potential bias induced by this mechanism is usually moderate. For example, assuming no effect of the exposure on the incidence of the disease (true rate ratio = 1.0), the estimated exposure-disease rate ratio is 0.95 even if i) the odds ratio of selection associated with the exposure is 2.0, ii) the odds ratio of selection associated with the risk factor is 4.0 and iii) the risk factor increases the risk of the disease by 4-fold (Pizzi et al. 2011). The ideal situation is a randomized study in which, thanks to randomization which occurs after selection of study subjects, the exposure is not associated with selection in the study. Under this situation, assuming compliance to the treatment assigned, there is no possibility for selection bias even when some risk factors (R) are strong determinants of the selection.

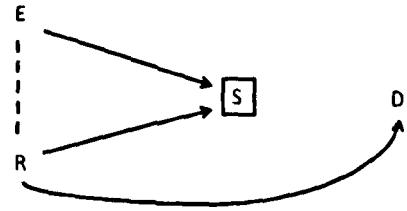


Figure 3: Diagram of a cohort study based on a selected sample. In the population the exposure of interest (E) is independent of the risk factor (R) and is not associated with the outcome of interest (D). Both E and R affect the likelihood of being selected as member of the study (S). In the selected sample (i.e. conditioning on S - represented by a square around S), E and R become associated (indicated by a dashed line).

It should be noted that, when the exposure and the disease risk factor are associated in the general population, the selection mechanism will alter the confounding effect of the risk factor, either increasing or decreasing it, according to the strength and direction of the association between E and R (in the general population) and also of the associations between E and R with S.

Let us now focus on the specific case of an internet-based *cohort study* and the effect of restriction on internet users, by discussing some hypothetical examples.

(i) In the first example, the exposure of interest (E) and the risk factor (R) are independent in the general population and both are associated, either directly (Figure 4a) or indirectly (Figure 4b), with the likelihood of being an internet user and thus of being selected into the sample.

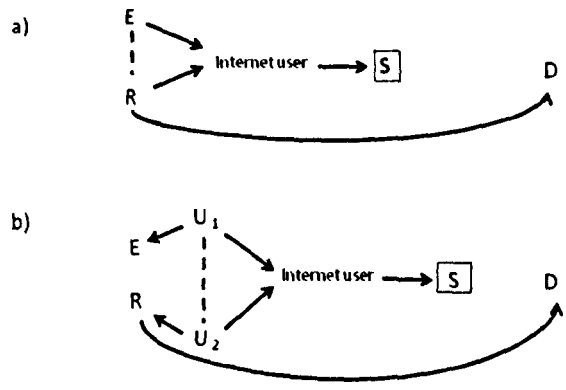


Figure 4: Diagram of an internet-based cohort study. a) In the population, the exposure (E) and the disease risk factor (R) are independent, and E is not associated with the outcome of interest (D). Both E and R affect the likelihood of being an internet user and thus of being selected as member of the study (S). In the selected

sample (i.e. conditioning on S - represented by a square around S), E and R become associated (indicated by a dashed line). b) Here, E and R affect the likelihood of being an internet user as proxy of other factors (U_1 and U_2).

In Figure 4a, for example, socioeconomic status (E) as well as year of birth (R), which are assumed to be independent in the general population, could affect the probability of being an Internet user. In Figure 4b, socio-economic status (U_1) could be a cause of being an internet user and of being a smoker (E), while year of birth (U_2) could affect both the likelihood of being an internet user and height (R). In these scenarios, the restriction on internet users induces a spurious association between E and R, and, therefore, year of birth becomes associated with socio-economic status (Figure 4a) and smoking becomes associated with height (Figure 4b).

(ii) In a second example, the exposure of interest (E) and the disease risk factor (R) are already associated in the general population as they share a common cause (U_3). This scenario is illustrated in Figure 5.

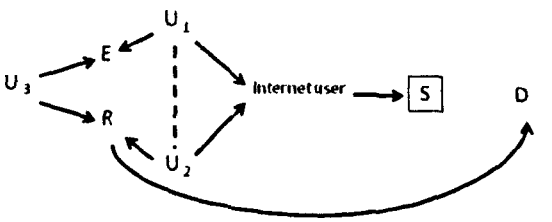


Figure 5: Diagram of an internet-based cohort study. In the population the exposure (E) and the disease risk factor (R) are associated as they share a common cause U_3, and both E and R are associated with the likelihood of being an internet user. In the selected sample (i.e. conditioning on S - represented by a square around S), U_1 and U_2 become associated (indicated by a dashed line) thus altering the original association between E and R.

This is in fact an extension of the scenario depicted in Figure 4b, where an additional factor (U_3), for example place of birth, could affect both smoking (E) and height (R). Under this scenario it is hard to predict whether the restriction of the source population to internet users would increase or decrease the bias, as this depends on the strength and direction of the E-R association existing in the general population and on the strength and direction of the spurious E-R association induced by the sample selection process (as a consequence of the association induced between U_1 and U_2).

(iii) A further example of interest is when E and R are associated in the general population because they share a common cause which, in turn, is a determinant of being an internet user. This is shown in Figure 6 that summarizes the case of a study in which socio-economic status (U_3) is a cause of being an internet user as well as of being a smoker (E) and taking regular exercise (R).

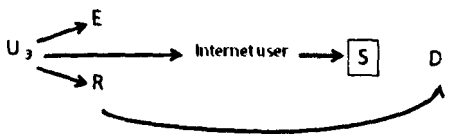


Figure 6: Diagram of an internet-based cohort study. In the population the exposure (E) and the disease risk factor (R) are associated as they share a common cause U_3, which also affects the likelihood of being an internet user. In the sample, the condition on S (represented by a square around S) implies a partial condition on U_3, thus attenuating the association between E and R.

Under this scenario, restricting the study to internet users implies a partial conditioning on U_3 and therefore a likely attenuation of the E-R association among the study participants compared with the general population. This means that the restriction would diminish the confounding effect of R and thus the corresponding bias induced in the exposure-outcome association.

In these examples we only considered the restriction to internet users as a potential source of selection. Similar considerations can be made for the second source of selection that is volunteering. Our examples demonstrate, however, that the effect of selection on the exposure-disease association in longitudinal studies is difficult to predict and can induce or attenuate confounding bias present in the general population (Pizzi et al. 2012). However, even when it induces bias, its magnitude is expected to be small as discussed above.

If the aim of the study is to estimate the incidence of the disease, bias is more likely to occur. The disease status itself cannot affect the likelihood of selection (as it may occur in surveys) but any disease risk factor (R in Figures 3 to 6) that is associated with being an internet user or volunteering would bias the estimate of the incidence.

2.2 Case-control studies

Case-control studies can be restricted to a specific population and still reveal valid associations (apart from some of the considerations just mentioned for cohort studies). Thus, the restriction to cases who are internet users (or, for example, who seek information about their disease

on the internet) is not expected to introduce a large bias by itself. The concern is, however, on the effect of this restriction on the control selection, as control subjects should be representative of the source population from which the cases originated. When cases and controls arise from two different source populations, there are clearly factors with different distributions across the two groups, which, when associated with the exposure of interest, lead to cases and controls being no longer comparable.

Even when the principle of the same source population is met, selection bias may occur if the exposure is associated, either directly or through other factors, with the probability of selection, i.e. the sample fraction of controls (and cases when applicable) is not constant across exposure levels. In this situation, the causal structure becomes similar to that depicted in Figure 2 or Figure 7, both implying selection bias. Figure 7, for example, could depict a scenario in which cases with a disease of interest (D) are compared with controls originating from the same source population but having another disease (C). If the disease (C) is associated with the exposure of interest (E) (in this scenario through the common cause U_1) then there is selection bias (Hernán et al. 2004).

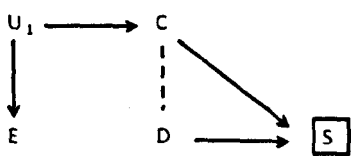


Figure 7. Diagram of a case-control study in which cases have the disease of interest (D) and controls have a control disease (C). Both C and D affect the probability of selection in the study (S) and C is also associated with the exposure of interest (E), as they share a common cause U_1. In the selected sample (i.e., conditioning on S - represented by a square around S), C and D become associated (indicated by a dashed line) thus altering the original association between E and D.

Since participants in an internet-based case-control study are self-selected, it is not unlikely that some of the determinants of the exposure or the exposure itself are associated with volunteering and thus with selection. It should be noted, however, that the causal structures summarized in Figure 2 and 7 do not always introduce bias. Indeed, there is no bias if the ratio of the selection probabilities of exposed and unexposed cases is the same as the ratio of the corresponding selection probabilities in exposed and unexposed controls. Under this special scenario, E and D are still determinants of the selection but the association remains to be valid.

3. INTERNET-BASED RECRUITMENT

The use of the internet allows recruitment of study participants from large populations with a decrease in cost and time. Internet recruitment is similar to "traditional" studies in that it requires the specification of the source population. For example the source population of an internet-based study aiming at surveying cancer patients residing in Italy is defined by those internet users who reside in Italy and have cancer. However, in internet-based studies there is no available list for random sampling from the source population and participants are self-selected volunteers. Thus, apart from defining the source population by means of eligibility criteria, researchers have little possibility to influence the selection of study participants. Internet-based recruitment typically goes through two processes: i) people in the source population need to become aware of the existence of the study and ii) they need to agree to participate.

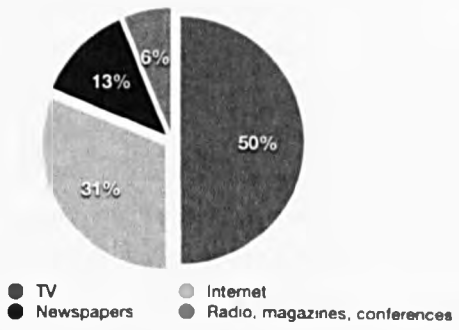


Figure 8. Distribution on how participants to the Influweb study got to know about the existence of the study (based on about 2000 respondents). Influweb, Italy, 2008-2009 (Influenzanet 2011)

The study has to be advertised in order to efficiently reach all members of the source population. The advertisement of a study of individuals with a specific disease or exposure would have to involve websites targeting that disease/exposure, while a study targeting the general population would be more efficiently advertised by means of a publicity campaign involving television, radio, and newspaper coverage and word of mouth. The internet-based Nutrinet-Santé study, for example, which aims at recruiting a large cohort of individuals from the French general population, advertises the existence of the study through a multimedia campaign, different websites, and professional health channels (Hercberg et al. 2010). On the basis of our experience in the Italian Influweb study (an internet-based monitoring project for influenza surveillance in Italy), when the target is the general population, television seems to be the most effective means of communication,

especially if the advertisement takes place in Scientific Communication programs, while articles on the front page of online newspapers seem to be very efficient in terms of visits to the website but not in terms of new volunteers (Figure 8) (Paolotti et al. 2010).

The Danish Pregnancy Planning Study, an internet-based cohort study of women of reproductive age, was advertised through a pop-up advertisement in a national health-related website as well as with press releases to reach the media (Mikkelsen et al. 2009). An internet-based study of determinants of gout attacks was advertised using a Google advertisement linked to the search term "gout" (Zhang et al. 2007). Over a period of 11 months, the advertisement was displayed 866,703 times, 6.6% of which led to a visit of the study website.

An often overlooked issue is that an integrated approach to advertise an internet-based study aiming at recruiting for a long period implies intense and continuous efforts. Conversely, if the study is only advertised via selected websites, the recruitment process is less demanding, although the recruitment rate will be lower. Thus, to reach large sample sizes, an integrated approach is generally necessary. Furthermore, the methods used to advertise the existence of the study will impact on the characteristics of the study participants and the probability of selection bias. If a mother-child cohort, for example, is advertised only during antenatal courses, there will be an oversampling of nulliparous women who are more likely to attend these courses. An integrated approach could reduce this source of selection and spread the information about the study to the whole source population. A recent study compared three methods of advertisement for a survey of young adult smokers, namely: (1) advertisement on a single general website (Craiglist 2011); (2) an internet campaign to target social networks and lifestyle-based websites; (3) an invitation sent to members of an internet panel (Ramo et al. 2010). The internet campaign yielded the largest number of participants but it was less cost-effective (about 43 US$ per completed survey) and was the method associated with the largest proportion of incomplete surveys. Roughly spoken, the three samples differed in terms of age, sex, ethnicity, education attained, nicotine dependence and recent use of marijuana and cigars. Fewer or no differences were found in other variables, including alcohol use and smoking prevalence.

Once the members of the source population are aware of the study, they should access the study website and volunteer to participate. This is obviously a key issue. During the pandemic season, from October 2009 to March 2010, the Influweb website (Influenzanet 2011), for example, was visited 90,000 times and roughly 3,000 persons participated. The success of volunteering depends on several elements: the study website should not only induce participation but the study topic should also be of interest among the source population. For example, influenza during the H1N1 pandemic was a topic prone to gain the interest of the public and to reach the media. In other

words, internet-based recruitment is more efficient when the study is about a topic of general interest (Paolotti et al. 2010; Tilston et al. 2010), or targets a strongly motivated population, such as smokers trying to quit smoking (Civljak et al. 2010) or pregnant women (Richiardi et al. 2007). Even when an integrated approach for advertisement is used and the population is motivated to participate in the study, the participation proportion is not going to be high compared to traditional studies. For example, we used several means to advertise the existence of the NINFEA (Nascita ed INFanzia: gli Eetti dell'Ambiente) study (NINFEA 2011), a birth cohort study, to the population of the city of Turin, Italy. We have recently estimated that about 3-4% of the total number of pregnant women present in the population (excluding those born outside Italy) participated in the study (Pizzi et al. 2012). Considering that about 60% of the pregnant women have access to the internet, and thus belong to the source population, this proportion translates into a participation proportion of 5-7%.

Some internet-based surveys collect anonymous data but many internet-based studies require registration and collect demographic information. This is an obvious requirement for follow-up studies in which subjects should be identified and be re-contacted but it is also used in surveys, for example to limit the problem of duplicate or phantom participations. Researchers undertaking a web-based study should therefore check that their platform is compliant with privacy regulations in the country where the study is ongoing. Upon registration, users have to be informed what these requirements are and, as in any epidemiological study collecting non-anonymous data through questionnaires, they should sign an informed consent form. In many studies an online informed consent is used to decrease costs and organizational efforts and to enhance participation (the alternative being a mailed hard copy of the informed consent to be signed by participants). For example, in the NINFEA study women provide online consent when they register and complete the study questionnaires, while they provide an additional written consent if they also donate a saliva sample.

The use of the internet to collect non-anonymous data may raise concerns about safety and confidentiality issues. Researchers typically use technical solutions, such as encryption, firewalls, HTTPS (Hypertext Transfer Protocol Secure) protocol, etc, which provide the same level of safety as traditional epidemiological studies, as, for example, mail surveys (Baer et al. 2002; Kraut et al. 2004). It is possible that potential participants are reluctant to participate in an internet-based questionnaire fearing that their data could be accessed from people from outside the study. To address this issue, within the NINFEA study, we interviewed a small number (n = 37) of women who were both aware of the study and had access to the internet but did not participate in the study

(Richiardi et al. 2007). No woman reported that she did not participate for fear of revealing personal information, while the most common reason of non-participation was lack of interest in the study.

4. STUDY DESIGNS

The internet has been used in the context of most of the classical epidemiological studies, including cross-sectional (see Chap. Descriptive Studies of this handbook.), cohort studies (see Chap. Cohort Studies of this handbook.), interventions (see Chap. Intervention Trials of this handbook.), and case-control studies (see Chap. Case-Control Studies of this handbook.). Basic characteristics of these study types are discussed in detail in the chapters listed above.

4.1 Cross-sectional studies

Cross-sectional studies or surveys are typically carried out to measure prevalence and are particularly vulnerable to selection bias. If participants have a different prevalence of the disease of interest compared to non-participants, and it is not possible to apply weights to counterbalance this difference, results of the study will be difficult to interpret. This drawback makes the use of the internet in surveys very problematic.

Although characteristics of internet users can be investigated and may sometimes be known for certain populations, determinants of self-selection in a specific study are almost unknown. It is possible to obtain a rough estimate of the number of individuals who visit the study website to estimate a sort of response proportion (number of participants out of the number of visitors). However, this proportion is relatively useless as the number of visitors differs from the number of subjects who became aware of the study. Furthermore, the response proportion would give little information on the amount of bias in the estimate of the disease prevalence, as the key issue is whether volunteering is associated or not with the probability of having the characteristic of interest.

Having this strong limitation in mind, internet-based surveys can still be useful for a number of reasons, including to conduct qualitative studies in which population representativeness is less relevant; to rapidly obtain information to generate hypotheses or develop a study protocol; to reach hidden populations; to identify patients with rare diseases.

Simmons and colleagues carried out an internet-based cross-sectional study between 2001 and 2002 to obtain information on possible precipitating factors of multiple sclerosis (MS) (Simmons et al. 2004). The aim was to generate hypotheses and select potential precipitating factors

to be investigated in a cohort study of MS patients. An anonymous questionnaire in English was posted on the MS Australia and the MS International Federation websites for a period of about 10 months. About 2,500 self-selected patients from 60 countries in total, mainly from USA, Australian and UK, completed the questionnaire. They reported factors that in their opinion were improving their condition or worsening their MS symptoms.

Behavioral research or studies on HIV and sexually transmitted diseases amongst men who have sex with men (MSM) are most often conducted on convenience samples, as MSM is a hard-to-reach population (a so-called "hidden" population). Therefore, internet-based surveys using anonymous questionnaires are becoming increasingly common in this field (Elford et al. 2009; Evans et al. 2007; Hirshfield et al. 2004; Rosser et al. 2009). A study conducted in London in 2002 and 2003 recruited about 4,000 participants from HIV-positive patients attending outpatient clinics (12%), men seeking an HIV-test (10%), men using gyms in central London (35%) and internet users, the latter either via chat rooms or the websites of gaydar (2011) and gay.com (2011) (43%). The samples had different socio-economic and behavioral characteristics and most likely none of them was representative of the whole MSM population.

A number of studies aimed at comparing characteristics of participants in an internet-based cross-sectional study with study participants based on a representative sample of the population (Andersson et al. 2002; Etter and Perneger 2001; Evans et al. 2007; Klovning et al. 2009; Marcus et al. 2009; Miller et al. 2010; Ross et al. 2005). Unsurprisingly, participants recruited via the internet had different characteristics and disease prevalence, most often in an unpredictable direction and magnitude. This reinforces the concept that cross-sectional studies can only be conducted over the internet if their aim does not require a population representative sample.

4.2 Cohort studies

In 1997, an editorial on the use of the internet in epidemiology suggested the possibility of conducting a cohort study of internet users, defining this study population as an epidemiologists' dream coming true (Rothman et al. 1997). Authors were briefly listing some of the most evident advantages, such as fast enrolment of a large sample, prolonged contact with cohort members, inexpensive and efficient follow-up, as well as some possible problems, underlying the risk of marauder and phantom users, and issues of information validity. According to a response letter to this editorial, Kushi and colleagues were already piloting at that time the feasibility of an internet-based cohort study (Kushi et al. 1997).

Indeed, the use of the internet to recruit a cohort is very attractive as, in longitudinal studies, a representative sample is not a necessary requirement to get valid associational estimates. The epidemiological dream, however, has infrequently materialized in the last decade (Hercberg et al. 2010; Mikkelsen et al. 2009; Richiardi et al. 2007; Treadwell et al. 1999; Turner et al. 2009).

In 2005 we started the NINFEA study, which is an internet-based mother-child cohort carried out in Italy (NINFEA 2011). A parallel study was started two years later in New Zealand (The Early Life Factors Study of Childhood Diseases). We will therefore use the NINFEA cohort as an example to illustrate advantages and limitations of using the internet to conduct a cohort study.

NINFEA is a multi-purpose cohort aiming at investigating the effects of certain exposures during pre-natal and early post-natal life on infant, child and adult health (Richiardi et al. 2007). It enrolls pregnant women in order to follow-up their children for at least 18 years. Members of the cohort are children born to women who are internet users, become aware of the study and volunteer to participate. At any time during the pregnancy, they can register through the project website (www.progettoninfea.it) and complete the first questionnaire that lasts about 30 minutes. They are asked to complete two other 30-minute long questionnaires at 6 and 18 months after delivery. Long-term follow-up involves linkage with available population registries and periodical very short on-line questionnaires focusing on specific outcomes (e.g. cognitive development, respiratory diseases, etc.).

We advertise the existence of the study using both "active" and "passive" methods. Active methods involve the collaboration of health personnel to distribute leaflets and/or to introduce the study to pregnant women when they reach hospitals or family clinics for reasons related to the pregnancy. Therefore, this approach is inherently limited to selected geographical areas and targets a (roughly) pre-specified catchment population. Currently, the NINFEA study is actively advertised in the city of Turin (900,000 inhabitants), in the Tuscany Region (4,000,000 inhabitants) and, with a lower intensity, in the Piedmont Region (4,000,000 inhabitants including those living in Turin). One of the potential advantages of active recruitment that we have not yet explored is the involvement of specific populations characterized by high levels of specific exposure or diseases of interest. Let us suppose, for example, that in a community living around a large industrial area there are concerns on the possible reproductive effects of industrial emissions. It would be hard to quickly set up a "traditional" mother-child cohort in this population, especially if there are no research infrastructures already available in the area. It would, however, be possible to actively advertise the existence of the internet-based cohort in that population in order to recruit a sufficient number of pregnant women. The online questionnaires could be modified accordingly to incorporate questions about the exposure of interest in the area.

Passive recruitment includes methods that do not involve the health personnel, including the internet and the media. So far we have not launched a media campaign to advertise the NINFEA study, while we use the internet in various ways: links to our study website posted on the hospitals' websites and on websites dedicated to pregnant women, participation in discussion forums related to pregnancy, and a NINFEA page in Facebook. This type of passive recruitment is not entirely automatic, as forums change constantly as well as they become more or less popular among internet users. This implies constant monitoring of the accesses and the need to routinely post reminder messages. Furthermore, there should be bilateral interaction with the users of the forum to keep the discussion lively and attract new participants.

Doubtless, participants in the NINFEA cohort are strongly selected. When compared with the general population we found that NINFEA participants have a higher socio-economic status, a lower parity, are less frequently non-Italian citizens, and smoke less but have a higher alcohol consumption during pregnancy (Pizzi et al. 2012; Richiardi et al. 2007).

As mentioned above, participating women should complete 30-minute follow-up questionnaires at 6 and 18 months after delivery and shorter questionnaires thereafter. The use of the internet makes the follow-up rather efficient. In the NINFEA study we collect information via email, landline telephone, cell phone and postal address at the time of the registration. When it is time to complete a follow-up questionnaire we email the women asking to access the website and complete the questionnaire. Non-responders are additionally contacted first by email and then by telephone and regular mail. Currently, about 60-65% of the women reply after email contacts, while remaining women have to be contacted using traditional approaches. Overall, the final response to the second and third questionnaires is about 85-90%.

During the first five years of the study we have learned some valuable lessons regarding the follow-up. First, it is fundamental that contact information is obtained through mandatory questions at the time of the registration. This allows a much higher follow-up completeness at the cost of a small baseline dropout of participants who are not willing to reveal this type of information. Indeed, in cohort studies baseline selection is a much smaller problem than incomplete follow-up and the initial contact strategy should aim at assembling a cohort whose members guarantee high long-term participation. Second, although many authors have concerns about phantom participants in internet-based studies, in the NINFEA cohort this was a minor problem. Some people registered to the website to further understand about the questionnaires, but if public information about the project is clear enough we believe that phantom participants and registration from non-eligible individuals are not important issues. For example, in our cohort participation should occur before delivery. Indeed, so far, nobody participated after delivery, but we have been contacted by women asking if it is

possible to participate after the baby was born. This suggests that the information on the website was clear enough to prevent registration after delivery. The possibility of duplicate registrations of the same participant is a more relevant issue. Although it is not difficult to identify them at the time of the statistical analyses, using key-variables based on the available demographic information, duplicate registrations can make the follow-up procedures more complex. Let us take the example of a pregnant woman who registers twice with two different dates for the last menstruation (because of typos or because the pregnancy was re-dated between the two registrations): what date should be considered for the follow-up? When should the woman be re-contacted? Most likely this person will be contacted twice. It is possible to introduce checks in order to limit the number of duplicate registrations and facilitate the follow-up procedures but, in our experience, it is not possible to avoid them completely. A third issue is change of email address. Pregnant women and families with small children quite frequently change job and/or home. It is therefore important to keep frequent contact with the participants to give them updates about the study as well as to check contact information. If the email address has changed it is possible to contact the woman by telephone or letter and ask her to update her information. Indeed "traditional" cohort studies are affected by the same problem, and having the participants' email address and a population restricted to internet users only helps in obtaining a high follow-up participation proportion.

A potential limitation of the internet-based recruitment is that there is no direct contact with the participants and thus, exposure information is self-reported and it is more difficult to obtain biological samples. In the NINFEA cohort, we collect samples of saliva from the mothers and the children using self-collection kits sent by regular mail. About 60% of the members of the cohort, so far, agreed to donate a sample. Successful collection of biological samples has also been achieved in other internet-based studies, such as collection of saliva samples from subjects enrolled through a smoking cessation website (Etter et al. 2009). In an ongoing nationwide French cohort study aiming at recruiting half a million individuals (Nutri-Santeé cohort), participants can donate a blood sample by visiting local sample collection centers (Hercberg et al. 2010).

Currently the NINFEA study recruits and follows up about 25 subjects per week, employing overall (including IT experts and research assistants) less than three persons-years per year. Its internet-based design offers two main advantages, namely efficiency and flexibility, which are obtained at the cost of population representativeness. Flexibility is an often overlooked characteristic. The cohort will be able to recruit for an indefinite period and its population coverage has changed and will change over time. Other Italian regions will be able to start active recruitment in the future and it will be possible to adapt the questionnaire to other populations. Indeed, a parallel study has been launched in 2007 in New Zealand (The Early Life Factors Study of Childhood

Diseases 2011) and other countries may join in the future. Furthermore, compared to traditional studies, it is easier to add, delete or modify questions to all participants or selected subgroups, both to improve the questionnaire and to assess exposures previously uncovered. For example, during the H1N1 pandemic in the winter of 2010 we added some specific questions about vaccinations that were not planned in 2005. Based on our experience and other internet-based cohort studies, we believe that this methodology will be increasingly used in the next years, especially when the target of the cohorts will concern highly motivated people (e.g. pregnant women), or populations difficult to reach or difficult to follow-up (e.g. short-term migrants) or when researchers will aim at fast recruitment of large samples.

4.3 Case-control studies

In case-control studies, cases and controls should be selected from the same source population, independently of their exposure status. As discussed before in this chapter, this is difficult to achieve using internet-based recruitment. Thus, it is not surprising that we could not find any example of a study in which both cases and controls were selected using the internet. A recent study describes the selection of a control sample to be used for genetic analyses: about 4,500 subjects were selected in the USA both among a group of internet panelists and using ad-hoc internet recruitment (Sanders et al. 2010). However, study cases were traditionally selected and the study focused on genetic variants, which are unlikely to be strongly associated with self-selection in the study.

Case and control selection could be very problematic using the internet but not impossible. We can imagine different approaches, which are described using a hypothetical example of a case-control study on celiac disease in Italy.

In a first approach, cases could be selected via a link posted on a specific website, for example the website of the Italian Celiac Association. All potential cases visiting the website would be informed about the study and a small proportion of them would volunteer to participate. The advertising efforts would have to last until a sufficient number of patients is enrolled in the study. The source population for this study would include all subjects that, if diagnosed with celiac disease, would search information on the internet. Then controls could be recruited by posting information about the study in a number of disease-specific websites, say the websites of the Italian Associations for asthma, chronic bowel disease and type 1 diabetes. Obviously, these diseases would have to be unrelated with the exposure of interest. Again, control participants would be self-selected volunteers and the participation proportion would be expected to be very low. The

association estimates obtained from an internet-based case-control study of this type would be valid had the determinants of self-selection been similar among cases and controls or had these determinants been unrelated with the exposure status. Unfortunately these conditions are both rather strong and difficult to check in the data.

A second approach for an internet-based case-control study would begin with the definition of the source population. For the hypothetical internet-based study on celiac disease determinants in Italy, a broad definition of the source population would include all Italian internet users. Recruitment of case patients would then involve the website of the Italian Celiac Association, any other website of potential interest for celiac disease patients as well as media campaigns. Case participants would again be strongly selected. Controls would have to be selected from the same source population, namely from internet users. Recruitment would thus involve media campaigns and links posted in various websites, not necessarily of interest for celiac patients.

A stricter definition of the source population could improve this study design. For example, the source population could include users of a specific website, say the website of a national newspaper. Everybody accessing the website would be invited to participate in a study, and case patients would include volunteers with celiac disease while control subjects would include all the other volunteers. It could be possible also to adopt a two-stage approach in which, in the first stage, a generic health internet-based survey is offered to persons accessing the website and, at the second stage, cases with celiac disease and a sample of controls without celiac disease are further interviewed.

Irrespectively of whether a broader or a stricter definition for the source population is used, case and controls would again be strongly selected and the critical issue would be whether the determinants of the selection are associated with the exposure of interest and if they differ between cases and controls.

This brief description of hypothetical internet-based case-control studies emphasizes their vulnerability to selection bias. Careful methodological work and empirical testing is still needed before an internet-based case-control study can actually be conducted. Apart from selection bias, however, internet-based case-control studies would have other important limitations. Firstly, recruitment of incident cases through the internet seems even more problematic and most likely an internet-based case-control study would involve prevalent cases, with the well-known corresponding limitations. Secondly, even if a study manages to involve cases and controls from the same source population without selection bias, the use of the internet introduces selection among cases. It is hard to predict if case participants would have a more or less serious disease but the selection would most likely introduce problems of generalizability. Furthermore it would be

difficult to distinguish between factors causing the disease and factors affecting its severity or the patients' overall health status.

A *case-crossover design (see Chap. Modern Epidemiological Study Designs of this handbook)* which does not involve the selection of control subjects could be a more sensible option for an internet-based study. An internet-based case-crossover study on gout (Online Gout Study 2011) has indeed been launched in 2003 (Zhang et al. 2007). The existence of this study was advertised using Google links. Gout patients were invited to register and asked to complete a control-period questionnaire every three months investigating risk factors for gout attack in the preceding two days. Moreover they were asked to complete an "attack questionnaire" if they were experiencing a gout attack. Participants were also provided a hard copy of the attack questionnaire in case they could not access the internet during the gout attack but in fact this option has been rarely used during the study.

As any other case-crossover study, this study aimed at evaluating trigger factors and acute effects (Hunter et al. 2006; Zhang et al. 2006). Exposure information was collected prospectively, because patients were recruited before having the actual gout attack, thus limiting the possibility of selection bias. It is also possible to imagine a different design, in which patients participate when they have an event, say a gout attack, and complete the exposure information for the control period/s retrospectively. Under this scenario, there is possibility of selection bias if their likelihood of accessing the study website and participating in the study depends on some of the trigger factors under investigation. For example a patient could experience a gout attack after having used a diuretic. He or she might suspect that the diuretic was the cause of the gout attack and use the internet to check for this hypothesis. Then, he or she accesses the study website, in which there might also be some general information about the disease, and decide to participate. Obviously, this would induce selection bias.

4.4 Intervention studies

For a number of reasons internet recruitment fits very well with intervention studies, namely: 1) randomization is likely to cancel bias due to self-selection, 2) the randomization can be easily managed centrally through the website, also stratifying for a number of variables and patients' characteristics, and 3) pragmatic and explanatory trials become very similar in design and conduct. Indeed, internet recruitment is the most correct setting in which to test tailored interventions offered via the internet to unselected patients/populations. These types of interventions are becoming more and more frequent. We have carried out an admittedly cursory

PubMed search restricted to clinical trials using "internet-based" and "intervention" as the keywords. It revealed a clear trend in increasing number of papers with time (Figure 9).

Internet-based trials in medicine have been conducted starting from the end of the 1990s / beginning of 2000s. In 2000, for example, McAlidon and colleagues started online recruitment for a feasibility study of an internet-based clinical trial of glucosamine vs. placebo in patients with osteoarthritis of the knee (McAlindon et al. 2003). The main outcome was knee pain, assessed using a validated online self-completion questionnaire 12 weeks after intervention. Volunteers completed an eligibility screening questionnaire and, if applicable, they were asked to send a signed hard copy of the informed consent and copies of medical records. Upon confirmation of the osteoarthritis of the knee, they were randomized into the treatment or placebo group. To those subjects included in the treatment group, the pills were mailed every second week. Although this study involved internet based recruitment, the intervention to be tested was not an online-intervention (as it involved mailed pills) and relied on having access to the actual medical records to confirm the diagnosis and obtain detailed information on the disease. These aspects may decrease the possible methodological advantages of using the internet.

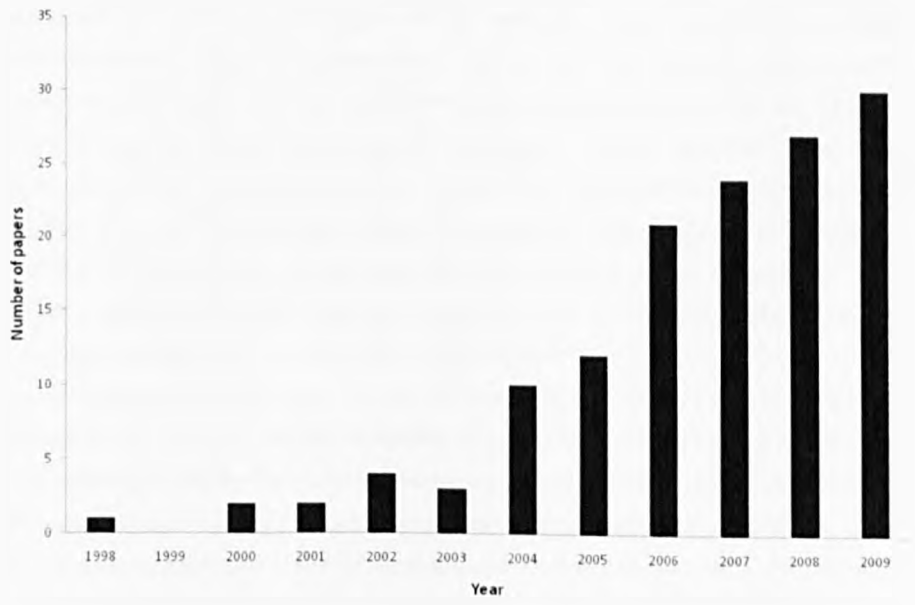


Figure 9. Number of papers identified in PubMed using "internet-based" and "intervention" as the keywords and restricting the search to clinical trials (search carried out in November 2010)

Studies testing online primary care interventions for health risk behaviors are more common, including those on diet and nutrition, physical activity, smoking habit and alcohol consumption (Civljak et al. 2010; Portnoy et al. 2008; Vandelanotte et al. 2007; Webb et al. 2010). A recent Cochrane review on internet-based interventions for smoking cessation identified 20 studies published until June 2010 (Civljak et al. 2010). Most of these studies used an internet-based recruitment and all of them were published after 2004. Some of the trials compared an internet intervention with either no intervention at all or an offline intervention; others compared different types of internet interventions. In general, the study designs were very heterogeneous and many studies had a relatively high proportion of drop-outs. Authors concluded that the evidence of long-term benefits for programs delivered only by the internet (as compared with offline interventions), is very limited, while there is some evidence that tailored (i.e. interventions specifically designed to meet the characteristics of a target individual or group) internet interventions are more effective than non-tailored internet interventions.

Rabius and colleagues advertised an intervention trial targeting smokers who wanted to quit smoking on the website of the American Cancer Society (Rabius et al. 2008). Potential participants were asked to complete a baseline questionnaire and provide informed consent. If eligible, they were randomized to receive access to one of five tailored interactive websites providing interventions for smoking cessation or a more static page set up at the American Cancer Society website serving as control treatment. The main outcome (successful abstinence for the last 30 days) was assessed 13 months after randomization, first emailing a survey questionnaire and then contacting by phone those who did not answer. Out of almost 6,500 individuals enrolled in the study, only 38% answered the follow-up questionnaire. Analyses were first conducted among the respondents only and then re-conducted assuming that non-respondents did not quit smoking: in no case there was a significant difference in smoking cessation across the intervention arms. The high drop-out rate between follow-ups, however, limited the study power.

Low completeness of follow-up was also observed in a trial set up on the Stop-tabac.ch website (Stop-tabac.ch 2011) to compare an original and a modified version of an online tailored program for smoking cessation (Etter 2005). The main outcome (smoking abstinence in the last seven days) was assessed via email 11 weeks after randomization, using up to three reminders. Almost 12,000 subjects were randomized in the study, with a response proportion at the follow-up questionnaire of 35%. In the analyses, non-respondents were assumed to still smoke. The original version of the program was found to be more effective, with 1 additional quitter every 26 participants.

Although many internet-based randomized intervention studies have a high proportion of loss to follow-up, we do not think that this is a characteristic inherent to the study design. It is possible to increase participation by obtaining more detailed contact information at baseline (telephone number, cell phone number, address, second email) and by recruiting a more committed and motivated population, for example by having 1-2 follow-up questionnaires before randomization. Furthermore, in a pragmatic setting, a certain proportion of loss to follow-up should be expected.

Randomized trials recruiting participants through the internet are feasible and, thanks to randomization, do not suffer from selection bias more than if a traditional approach of recruitment is used. Generalizability problems, namely the restriction to self-selected internet users, are minor when the intervention involves online programs. There are, however, some limitations. Firstly, it is difficult to have direct contact with participants and to include medical exams in the protocol, thus limiting the use of this study design in clinical settings. Secondly, in many studies, attrition to follow-up questionnaires was low and thus there is a need of methodological improvements to limit drop-outs and increase motivation of the participants. Thirdly, internet-based recruitment may lose its efficiency when the aim of the trial is to test an offline intervention.

4.5 Surveillance

The main aim of surveillance is to monitor trends in the rate of disease occurrence, both to gain insight in the current situation of an established disease and to detect outbreaks of emerging diseases.

In this paragraph we will not go into detail in describing the aspects of surveillance, instead, we will concentrate on surveillance conducted by means of the internet as a way to collect data from sources not easily accessible by traditional surveillance.

The widespread diffusion of computers and of the internet has provided a tool capable of the earliest possible detection of epidemics, giving allowance to a timely and maximally effective public health response worldwide. Surveillance data, as well as behavioral data, social contacts and risk perception can be collected by exploiting new Information and Communication Technology (ICT) techniques and methodologies to better understand the spread of infectious diseases (for example by collecting information on behavioral data to understand human immunodeficiency virus transmission) and obtain real-time data, which are crucial to rapidly identify public health emergencies, understand global trends, feed realistic data-driven models to assess the impact on the population, and optimize the allocation of resources.

Existing traditional disease surveillance systems have limitations. For example, in the case of influenza-like-illness (ILI), monitoring methods rely on sentinel networks of physicians, laboratory scientists, public health professionals and epidemiologists. Although they may mirror influenza activity, they cannot be implemented as real-time surveillance tools: either they only record proxy measures of influenza, or they contain unavoidable time delays between incidence and reporting. Traditional schemes require individuals to access health services and rely on the propensity of individuals to consult. Age-stratified rates of physician consultation may vary widely with different health care and health insurance systems. For non-severe diseases especially, only a minor (and unknown) fraction of all infected individuals sees a doctor, and frequently after a considerable delay, when a complication has occurred or in case a doctor's certificate is required. A web-based platform can be used to detect cases from those individuals who are less prone to consult a doctor when sick but agree to fill in a brief web survey about their symptoms. Moreover, traditional monitoring schemes typically lack uniform standards for clinical definitions that vary considerably between countries and even between reporters (EISN – European Influenza Surveillance Network). By using web surveys, standard platforms can be used across different countries to collect uniform data without major economic efforts.

In the following, we will focus on the case of influenza-like-illnesses (ILI) for which internet-based technologies have been successfully used. These internet surveillance systems for ILI have been implemented in Belgium and Netherlands in 2003 (under the name "Der Grote Griepmeting" – the Great Influenza Survey), in Portugal in 2005 ("Gripenet"), in Italy in 2007 ("Influweb"), in the UK in 2009 ("Flusurvey") and in Sweden in 2010 ("Influensakoll") with the aim of measuring influenza activity and collecting important public health information in real-time (for example, during the 2009 H1N1 pandemic, the web platform detected the peak activity of the influenza more than one week in advance with respect to the general practitioners (GPs). Platforms are now grouped under the umbrella of Influenzanet, forming a network of platforms to measure ILI in the community at a European level.

For each platform, registration of participants takes place through the web page (see flowchart in Figure 10). Upon registration, following provision with a password protected account, participants are asked to complete a baseline questionnaire with questions about age, sex, household size and composition, occupation, location of home and workplace, membership of a risk group, etc. Participants are also able to create accounts on behalf of their family/household enabling the entry of data from elderly people or children. Each week, participants are asked to complete a symptoms questionnaire as to whether or not they had symptoms in the previous week. To maintain

participants' interest and remind them to complete the questionnaire, a newsletter containing influenza facts and news is sent each week. The platforms, called Internet Monitoring Systems (IMS), are updated on a daily basis during the whole influenza season with items including the estimated incidence and the spatial distribution of cases.

Results from the Belgian, Dutch, and Portuguese surveys have been analyzed under the name "Great Influenza Survey (GIS)" (Friesema et al. 2009). The estimated seasonal influenza incidence curves given by these systems were highly correlated with those obtained through the traditional surveillance method. These analyses also offer the encouraging indication that the internet-based approaches can detect increased influenza activity more rapidly than surveillance-based reports by general practitioners (Marquet et al. 2006). Moreover, it is possible to estimate delays between the onset and the consultation dates and to detect changes in contact patterns and general behavior (Friesema et al. 2009).

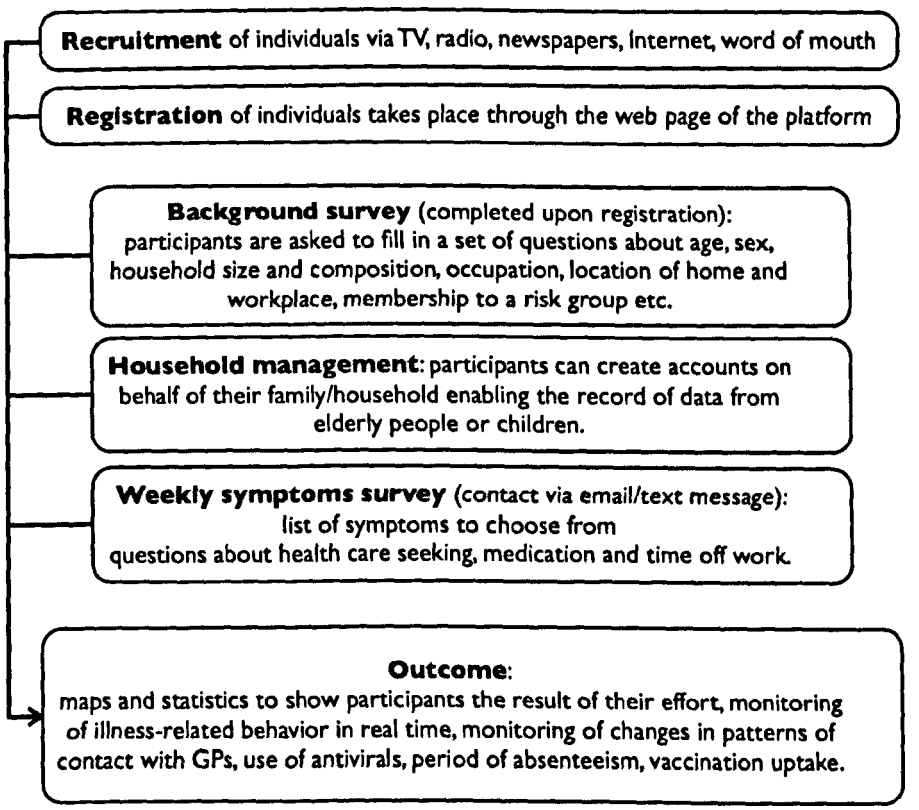


Figure 10. Diagram of recruitment and follow-up in the Influweb study

During the 2009 H1N1 pandemic, IMSs have proved to be valuable tools in gaining an insight into the evolution of the pandemic in real-time. In particular, the web-based system was launched in the UK during the first pandemic wave and went on collecting data until the end of the pandemic. Participants answering the symptoms questionnaires during the pandemic were asked more accurate follow-up questions about healthcare-seeking behavior, the delay of consultation with respect to the onset of symptoms; if they took time off work and for how long; if they took antiviral medication; if they were willing to be vaccinated against H1N1; etc (Tilston et al. 2010).

Since participants would select their symptoms from a pre-specified list, it was also possible to test different definitions of ILI, and to compare the resulting incidence with the one estimated by the Health Protection Agency (HPA). During the pandemic, to get a clearer picture of the epidemic evolution, the HPA asked for random testing of patients accessing different health care settings, which allowed evaluation of the true number of cases and thus adjustment of the estimates, by means of a method that was expensive and induced further delays in the data stream. The IMS to monitor ILI in the community was a direct and timely alternative, providing an incidence curve with a timing of the peak being close to the adjusted HPA case estimates. In making comparisons between web-based system estimates and HPA case numbers, results suggest that trends can be captured by the IMS even more reliably than standard GP-based systems, even though it remains unclear how accurate they are for estimating the absolute level of incidence (Tilston et al. 2010).

The UK web-based surveillance platform that ran continuously from July 2009 to March 2010 (i.e during both the first –summer– and the second –Autumn– pandemic wave in England in 2009) has also detected changes in healthcare seeking behavior between the two waves (Brooks-Pollock et al. 2011). These behavioral modifications, due to changing scientific information, media coverage and public anxiety, affected official case estimates. The web-based platform was able to detect a decrease from 43% to 32% in the percentage of individuals with ILI symptoms who sought medical attention from start to the end of the epidemic. Adjusting official numbers accordingly, it was possible to estimate that there were 1.1 million symptomatic cases in England, over 40% more cases than previously estimated, and that the autumn epidemic wave was 45% bigger than previously thought.

A further aspect for which surveillance is crucial is the development of accurate prediction models. When an outbreak occurs, usually short- and long-term predictions are based on observed data (provided by GP consultations or death/hospitalization records) combined with mathematical models updated as more data arise. For example, Bageuelin and colleagues (2010) carried out a real time assessment of the effectiveness and cost-effectiveness of alternative influenza

A/H1N1v vaccination strategies. To generate plausible autumn scenarios under different vaccination options, they fitted a transmission dynamic model using the estimated number of cases determined by means of a web-based surveillance platform, calls to the UK National Pandemic Flu Service and GP calls and consultations. In this specific case, data collected by means of the web platform were used to estimate the proportion of ILI cases calling the GP during the influenza A/H1N1v in UK.

One of the important limitations that remain related to internet-based surveillance is the lack of medical or laboratory confirmation of diagnosis. Data collected by means of a web platform will never be able to replace the virological analysis or clinical diagnosis carried out by GP surveillance. In direct contact with the patient, the GP can exclude other diseases than ILI and the virological analysis can give further information about the subtype of influenza virus. A possibility to overcome the latter limitation could be to send self-sampling kits to a selected subset of the internet-based system participants. An attempt in this direction has already been made (Cooper et al 2008) and this possibility should be explored further. Another possible limitation of internet-based surveillance is that participants are not representative of the general population. This issue has been addressed by re-weighting the sample according to the age and sex distribution in the general population (Tilston et al. 2010), although it is never possible to exclude that participation in an internet-based system is positively or negatively associated with the risk of ILI.

In conclusion, internet-based surveillance has the potential to capture a wider range of ILI cases than traditional surveillance, as well to track changes in health care attendance patterns in real-time. Even though internet-based surveillance has limitations and cannot replace traditional GP-based surveillance, it can provide an important support to enable the collection of valuable additional information, both in ordinary surveillance and during public health crises when the sentinel GPs surveillance and Public Health systems are under stress. While ILI has been used in the early deployment of the system, in subsequent years these IMSs will consider other diseases and infections.

5. WEB 2.0

Very recently, the internet has started offering new possibilities for epidemiological research based on the so-called web 2.0, which refers to the active generation of contents by the internet users through various means including online communities, web searches, social networks, etc. (Lee 2010). For example, these new communication and information habits over the internet may be used to quantify outbreaks of specific diseases (Eysenbach 2002, 2009). A well-known example is

the use of Google searches to obtain real-time estimates of influenza-like-illness (ILI) in the United States (Ginsberg et al. 2009): Ginsberg and colleagues developed a method to identify automatically ILI – related web search queries and use them to estimate ILI weakly percentages; these estimates had a correlation above 0.90 with data obtained from the US Surveillance Network of the CDC (Centre for Disease Control and Prevention).

The potential sources of information available in the web 2.0 are growing fast. As examples, data on disease outbreaks can be obtained from chat rooms, blogs, press release or Facebook (Brownstein et al. 2009, 2010; Eysenbach 2009), while exposure data, say on air pollution, can be obtained continuously from sensors worn by self-selected volunteers, and automatically transferred in real-time for model analyses (Mobile Environmental Sensing System Across Grid Environments, Message 2008; Mobile Air Quality Monitoring Network, Institute for Software Integrated Systems 2011). There are however obvious limitations in using these sources of data, including issues of generalizability, bias and exposure and outcome measurement. Currently, areas of epidemiological research which may actually benefit from the use of web 2.0 remain to be explored and identified.

6. CONCLUSIONS

Fifteen years ago, the internet was advocated as a promising tool for epidemiological research (Rothman et al. 1997). Since then, some internet-based studies have been conducted, including a number of surveys and intervention trials along with few cohort studies. Nevertheless, the use of the internet in epidemiological research is still very limited. Van Gelder and colleagues have recently reviewed analytic epidemiological papers published in 2008-2009 in four top general medical journals (the British Medical Journal, The Journal of the American Medical Association, the Lancet and The New England Journal of Medicine) and in three top epidemiological journals (American Journal of Epidemiology, Epidemiology and International Journal of Epidemiology), finding that only about 1% of the scrutinized 2,094 studies used internet-based questionnaires (van Gelder et al. 2010). Since online questionnaires can be also used in studies which do not recruit online, the actual number of internet-based studies is smaller. For this chapter, we have scrutinized papers published between September 21, 2009 and September 20, 2010 in eleven leading epidemiological journals (Environmental Health Perspective, American Journal of Epidemiology, Epidemiology, International Journal of Epidemiology, American Journal of Public Health, American Journal of Preventive Medicine, European Journal of Epidemiology, Preventive Medicine, Journal of Epidemiology and Community Health, Journal of Clinical Epidemiology,

Annals of Epidemiology). We have identified only two papers, both methodological, concerning an internet-based cohort (Huybrechts et al. 2010) and an internet-based survey (Klovning et al. 2009).

These data demonstrate that the use of the internet has not yet routinely entered the epidemiological practice. Moreover, most publications of internet-based studies deal with feasibility or proof-of-concept studies.

Evidence on the validity and reliability of web-based questionnaires has started accumulating (Apovian et al. 2010; Brigham et al. 2009; Donker et al. 2009; Miller et al. 2002; Rankin et al. 2008; Touvier et al. 2010, 2011; West et al. 2006) and some studies have successfully used mixed methods involving both web-based and mailed questionnaires for the follow-up of traditionally recruited cohorts (Ekman et al. 2006; Russell et al. 2010). However, methodological research on the use of the internet to recruit unspecified populations in the context of the various types of study designs is still in its infancy. In detail, we need to better understand when an internet-based survey can be carried out and if it is possible to tackle some of its inherent problems of selection bias; we need to further investigate the impact of baseline selection in internet-based cohort studies, as well as evaluate the determinants of completeness of follow-up; we should refine our ability to advertise the existence of internet-based studies to the relevant source population; we should develop methods to improve attrition in internet-based randomized studies; we should understand if it is possible to conduct internet-based case-control studies; and we should understand when surveillance can (or should) be carried out via the internet.

In our opinion the short-, medium-term agenda for internet-based research applied to epidemiology should prioritize innovative field work and methodological studies on the acquired data. In other words, the debate on whether internet-based research is valid or not should overcome a-priori formed opinions and become more evidence-based where this chapter has indicated that randomized trials, cohort studies and surveillance may be successfully carried out using the internet.

Acknowledgments

This work was partially supported by the Compagnia San Paolo and the Piedmont Region. We thank Prof. Neil Pearce for valuable comments and Jose' Lobo for valuable help.

REFERENCES

Andersson G, Lindvall N, Hursti T, Carlbring P (2002) Hypersensitivity to sound (hyperacusis): a prevalence study conducted via the Internet and post. Int J Audiol 41:545-554

Apovian CM, Murphy MC, Cullum-Dugan D, Lin PH, Gilbert KM, Coffman G, Jenkins M, Bakun P, Tucker KL, Moore TJ (2010) Validation of a web-based dietary questionnaire designed for the DASH (dietary approaches to stop hypertension) diet: the DASH online questionnaire. Public Health Nutr 13:615-622

Baehring TU, Schulze H, Bornstein SR, Scherbaum WA (1997) Using the World Wide Web – a new approach to risk identification of diabetes mellitus. Int J Med Inform 46:31-39

Baer A, Saroiu S, Koutsky LA (2002) Obtaining sensitive data through the Web: an example of design and methods. Epidemiology 13:640-645

Baguelin M, Hoek AJ, Jit M, Flasche S, White PJ, Edmunds WJ (2010) Vaccination against pandemic influenza A/H1N1v in England: a real-time economic evaluation. Vaccine 28:2370-2384

Berrens RP, Jenkins-Smith H, Silva C, Weimer DL (2003) The advent of internet surveys for political research: a comparison of telephone and internet samples. Political Analysis 11:1-22

Bowen AM, Williams ML, Daniel CM, Clayton S (2008) Internet based HIV prevention research targeting rural MSM: feasibility, acceptability, and preliminary efficacy. J Behav Med 31463-477

Brigham J, Lessov-Schlaggar CN, Javitz HS, Krasnow RE, McElroy M, Swan GE (2009) Test-retest reliability of web-based retrospective self-report of tobacco exposure and risk. J Med Internet Res 11:e35

Brooks-Pollock E, Tilston N, Edmunds WJ, Eames KT (2011) Using an online survey of healthcare-seeking behaviour to estimate the magnitude and severity of the 2009 H1N1v influenza epidemic in England. BMC Infect Dis 11:68

Brownstein JS, Freifeld CC, Madoff LC (2009) Digital disease detection--harnessing the Web for public health surveillance. N Engl J Med 360:2153-2155, 2157

Brownstein JS, Freifeld CC, Chan EH, Keller M, Sonricker AL, Mekaru SR, Buckeridge DL (2010) Information technology and global surveillance of cases of 2009 H1N1 influenza. N Engl J Med 362:1731-1735

Buchanan T, Smith JL (1999) Using the Internet for psychological research: personality testing on the World Wide Web. Br J Psychol 90 (Pt 1):125-144

Civljak M, Sheikh A, Stead LF, Car J (2010) Internet-based interventions for smoking cessation. Cochrane Database Syst Rev:CD007078

Cooper DL, Smith GE, Chinemana F, Joseph C, Loveridge P, Sebastionpillai P, Gerard E, Zambon M (2008) Linking syndromic surveillance with virological self-sampling. Epidemiol Infect 136:222–224

Craiglist (2011) www.craiglist.com. Accessed on May 9, 2011

Dillman DA, Smyth JD (2007) Design effects in the transition to web-based surveys. Am J Prev Med 32:S90-96

Dillman DA, Smyth JD, Christian LM (2009) Internet, mail, and mixed-mode surveys. the taiolred design method. John Wiley & Sons, Inc., Haboken (NJ)

Donker T, van Straten A, Marks I, Cuijpers P (2009) A brief Web-based screening questionnaire for common mental disorders: development and validation. J Med Internet Res 11:e19

Ekman A, Dickman PW, Klint A, Weiderpass E, Litton JE (2006) Feasibility of using web-based questionnaires in large population-based epidemiological studies. Eur J Epidemiol 21:103-111

Ekman A, Klint A, Dickman PW, Adami HO, Litton JE (2007) Optimizing the design of web-based questionnaires – experience from a population-based study among 50,000 women. Eur J Epidemiol 22:293-300

Elford J, Jeannin A, Spencer B, Gervasoni JP, van de Laar MJ, Dubois-Arber F (2009) HIV and STI behavioural surveillance among men who have sex with men in Europe. Euro Surveill 14: pii: 19414

Etter JF (2005) Comparing the efficacy of two Internet-based, computer-tailored smoking cessation programs: a randomized trial. J Med Internet Res 7:e2

Etter JF, Perneger TV (2001) A comparison of cigarette smokers recruited through the Internet or by mail. Int J Epidemiol 30:521-525

Etter JF, Neidhart E, Bertrand S, Malafosse A, Bertrand D (2005) Collecting saliva by mail for genetic and cotinine analyses in participants recruited through the Internet. Eur J Epidemiol 20:833-838

Etter JF, Hoda JC, Perroud N, Munafo M, Buresi C, Duret C, Neidhart E, Malafosse A, Bertrand D (2009) Association of genes coding for the alpha-4, alpha-5, beta-2 and beta-3 subunits of nicotinic receptors with cigarette smoking and nicotine dependence. Addict Behav 34:772-775

Evans AR, Wiggins RD, Mercer CH, Bolding GJ, Elford J (2007) Men who have sex with men in Great Britain: comparison of a self-selected internet sample with a national probability sample. Sex Transm Infect 83:200-205; discussion 205

Eysenbach G (2002) Infodemiology: The epidemiology of (mis)information. Am J Med 113:763-765

Eysenbach G (2009) Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 11:e11

Friesema IH, Koppeschaar CE, Donker GA, Dijkstra F, van Noort SP, Smallenburg R, van der Hoek W, van der Sande MA (2009) Internet-based monitoring of influenza-like illness in the general population: experience of five influenza seasons in The Netherlands. Vaccine 27:6353-6357

Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L (2009) Detecting influenza epidemics using search engine query data. Nature 457:1012-1014

Glymour MM (2006) Using causal diagrams to understand common problems in social epidemiology. In: Oakes JM, Kaufman JS (eds) Methods in social epidemiology. Jossey-Bass, San Francisco, pp 387-422

Gosling SD, Vazire S, Srivastava S, John OP (2004) Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. Am Psychol 59:93-104

Greenland S (2003) Quantifying biases in causal models: classical confounding vs collider-stratification bias. Epidemiology 14:300-306

Hercberg S, Castetbon K, Czernichow S, Malon A, Mejean C, Kesse E, Touvier M, Galan P (2010) The Nutrinet-Sante Study: a web-based prospective study on the relationship between nutrition and health and determinants of dietary patterns and nutritional status. BMC Public Health 10:242

Hernán MA, Hernández-Diaz S, Robins JM (2004) A structural approach to selection bias. Epidemiology 15:615-625

Hilsden RJ, Meddings JB, Verhoef MJ (1999) Complementary and alternative medicine use by patients with inflammatory bowel disease: An Internet survey. Can J Gastroenterol 13:327-332

Hirshfield S, Remien RH, Humberstone M, Walavalkar I, Chiasson MA (2004) Substance use and high-risk sex among men who have sex with men: a national online study in the USA. AIDS Care 16:1036-1047

gaydar (2011) http://www.gaydar.co.uk/. Accessed on May 9, 2011

gay.com (2011) http://www.gay.com/. Accessed on May 9, 2011

Hunter DJ, York M, Chaisson CE, Woods R, Niu J, Zhang Y (2006) Recent diuretic use and the risk of recurrent gout attacks: the online case-crossover gout study. J Rheumatol 33:1341-1345

Huybrechts KF, Mikkelsen EM, Christensen T, Riis AH, Hatch EE, Wise LA, Sorensen HT, Rothman KJ (2010) A successful implementation of e-epidemiology: the Danish pregnancy planning study 'Snart-Gravid'. Eur J Epidemiol 25:297-304

Internet World Stats (2011) http://www.internetworldstats.com. Accessed on May, 9 2011

Influenzanet (2011) www.influweb.it. Accessed on May, 9 2011

Kelly MA, Oldham J (1997) The Internet and randomised controlled trials. Int J Med Inform 47:91-99

Klovning A, Sandvik H, Hunskaar S (2009) Web-based survey attracted age-biased sample with more severe illness than paper-based survey. J Clin Epidemiol 62:1068-1074

Kongsved SM, Basnov M, Holm-Christensen K, Hjollund NH (2007) Response rate and completeness of questionnaires: a randomized study of Internet versus paper-and-pencil versions. J Med Internet Res 9:e25

Kraut R, Olson J, Banaji M, Bruckman A, Cohen J, Couper M (2004) Psychological research online: report of Board of Scientific Affairs' Advisory Group on the Conduct of Research on the Internet. Am Psychol 59:105-117

Kushi LH, Finnegan J, Martinson B, Rightmyer J, Vachon C, Yochum L (1997) Epidemiology and the Internet. Epidemiology 8:689-690

Lee BK (2010) Epidemiologic research and Web 2.0 – the user-driven Web. Epidemiology 21:760-763

Marcus U, Schmidt AJ, Hamouda O, Bochow M (2009) Estimating the regional distribution of men who have sex with men (MSM) based on Internet surveys. BMC Public Health 9:180

Marquet RL, Bartelds AI, van Noort SP, Koppeschaar CE, Paget J, Schellevis FG, van der Zee J (2006) Internet-based monitoring of influenza-like illness (ILI) in the general population of the Netherlands during the 2003-2004 influenza season. BMC Public Health 6: 242

McAlindon T, Formica M, Kabbara K, LaValley M, Lehmer M (2003) Conducting clinical trials over the internet: feasibility study. BMJ 327:484-487

Mikkelsen EM, Hatch EE, Wise LA, Rothman KJ, Riis A, Sorensen HT (2009) Cohort profile: the Danish Web-based Pregnancy Planning Study – 'Snart-Gravid'. Int J Epidemiol 38:938-943

Miller ET, Neal DJ, Roberts LJ, Baer JS, Cressler SO, Metrik J, Marlatt GA (2002) Test-retest reliability of alcohol measures: is there a difference between internet-based assessment and traditional methods? Psychol Addict Behav 16:56-63

Miller PG, Johnston J, Dunn M, Fry CL, Degenhardt L (2010) Comparing probability and non-probability sampling methods in Ecstasy research: implications for the internet as a research tool. Subst Use Misuse 45:437-450

Institute for Software Integrated Systems (2011) Mobile Air Quality Monitoring Network. http://www.isis.vanderbilt.edu/projects/maqumon. Accessed on May 9, 2011

Message (2008) Mobile Environmental Sensing System Across Grid Environments. http://bioinf.ncl.ac.uk/message/. Accessed on May 9, 2011

NINFEA project (2011) www.progettoninfea.it. Accessed on May, 9 2011

Online Gout Study (2011) https://dcc2.bumc.bu.edu/goutstudy/. Accessed on May 9, 2011

Paolotti D, Giannini C, Colizza A, Vespignani A (2010) Internet-based monitoring system for Influenza-like illness: H1N1 surveillance in Italy. Proceedings of Ehealt, Casablanca

Pepine CJ, Handberg-Thurmond E, Marks RG, Conlon M, Cooper-DeHoff R, Volkers P, Zellig P (1998) Rationale and design of the International Verapamil SR/Trandolapril Study (INVEST): an Internet-based randomized trial in coronary artery disease patients with hypertension. J Am Coll Cardiol 32:1228-1237

Pizzi C, De Stavola B, Merletti F, Bellocco R, Dos Santos Silva I, Pearce N, Richiardi L (2011) Sample selection and validity of exposure-disease association estimates in cohort studies. J Epidemiol Community Health 65:407-411

Pizzi C, De Stavola B, Pearce N, Lazzarato F, Ghiotti P, Merletti F, Richiardi L (2012) Selection bias and patterns of confounding in cohort studies: the case of the NINFEA web-based birth cohort. J Epidemiol Community Health 66:976-81

Portnoy DB, Scott-Sheldon LA, Johnson BT, Carey MP (2008) Computer-delivered interventions for health promotion and behavioral risk reduction: a meta-analysis of 75 randomized controlled trials, 1988-2007. Prev Med 47:3-16

Rabius V, Pike KJ, Wiatrek D, McAlister AL (2008) Comparing internet assistance for smoking cessation: 13-month follow-up of a six-arm randomized controlled trial. J Med Internet Res 10:e45

Ramo DE, Hall SM, Prochaska JJ (2010) Reaching young adult smokers through the internet: comparison of three recruitment mechanisms. Nicotine Tob Res 12:768-775

Rankin KM, Rauscher GH, McCarthy B, Erdal S, Lada P, Il'yasova D, Davis F (2008) Comparing the reliability of responses to telephone-administered versus self-administered Web-based surveys in a case-control study of adult malignant brain cancer. Cancer Epidemiol Biomarkers Prev 17:2639-2646

Rhodes SD, Bowie DA, Hergenrather KC (2003) Collecting behavioural data using the world wide web: considerations for researchers. J Epidemiol Community Health 57:68-73

Richiardi L, Baussano I, Vizzini L, Douwes J, Pearce N, Merletti F (2007) Feasibility of recruiting a birth cohort through the Internet: the experience of the NINFEA cohort. Eur J Epidemiol 22:831-837

Ross MW, Mansson SA, Daneback K, Cooper A, Tikkanen R (2005) Biases in internet sexual health samples: comparison of an internet sexuality survey and a national sexual health survey in Sweden. Soc Sci Med 61:245-252

Rosser BR, Oakes JM, Horvath KJ, Konstan JA, Danilenko GP, Peterson JL (2009) HIV sexual risk behavior by men who use the Internet to seek sex with men: results of the Men's INTernet Sex Study-II (MINTS-II). AIDS Behav 13:488-498

Rothman KJ, Cann CI, Walker AM (1997) Epidemiology and the internet. Epidemiology 8:123-125

Ruddy KJ, Gelber S, Shin J, Garber JE, Rosenberg R, Przypysny M, Partridge AH (2010) Genetic testing in young women with breast cancer: results from a Web-based survey. Ann Oncol 21:741-747

Russell CW, Boggs DA, Palmer JR, Rosenberg L (2010) Use of a web-based questionnaire in the Black Women's Health Study. Am J Epidemiol 172:1286-1291

Sanders AR, Levinson DF, Duan J, Dennis JM, Li R, Kendler KS, Rice JP, Shi J, Mowry BJ, Amin F, Silverman JM, Buccola NG, Byerley WF, Black DW, Freedman R, Cloninger CR, Gejman PV (2010) The Internet-based MGS2 control sample: self report of mental illness. Am J Psychiatry 167:854-865

Schleyer TK, Forrest JL (2000) Methods for the design and administration of web-based surveys. J Am Med Inform Assoc 7:416-425

Silver RC, Holman EA, McIntosh DN, Poulin M, Gil-Rivas V (2002) Nationwide longitudinal study of psychological responses to September 11. JAMA 288:1235-1244

Simmons RD, Ponsonby AL, van der Mei IA, Sheridan P (2004) What affects your MS? Responses to an anonymous, Internet-based epidemiological survey. Mult Scler 10:202-211

Skitka LJ, Sargis EG (2006) The internet as psychological laboratory. Annu Rev Psychol 57:529-555

Soetikno RM, Mrad R, Pao V, Lenert LA (1997) Quality-of-life research on the Internet: feasibility and potential biases in patients with ulcerative colitis. J Am Med Inform Assoc 4:426-435

Stop-tabac.ch (2011) http://www.stop-tabac.ch/fra/. Accessed on May 9, 2011

The Early Life Factors Study of Childhood Diseases (2011) http://www.elfs.org.nz/. Accessed on May, 9 2011

Tilston NL, Eames KT, Paolotti D, Ealden T, Edmunds WJ (2010) Internet-based surveillance of Influenza-like-illness in the UK during the 2009 H1N1 influenza pandemic. BMC Public Health 10:650

Touvier M, Mejean C, Kesse-Guyot E, Pollet C, Malon A, Castetbon K, Hercberg S (2010) Comparison between web-based and paper versions of a self-administered anthropometric questionnaire. Eur J Epidemiol 25:287-296

Touvier M, Kesse-Guyot E, Mejean C, Pollet C, Malon A, Castetbon K, Hercberg S (2011) Comparison between an interactive web-based self-administered 24 h dietary record and an interview by a dietitian for large-scale epidemiological studies. Br J Nutr 105:1055-1064

Treadwell JR, Soetikno RM, Lenert LA (1999) Feasibility of quality-of-life research on the Internet: a follow-up study. Qual Life Res 8:743-747

Turner C, Bain C, Schluter PJ, Yorkston E, Bogossian F, McClure R, Huntington A (2009) Cohort profile: The Nurses and Midwives e-Cohort Study--a novel electronic longitudinal study. Int J Epidemiol 38:53-60

United Nations Economic Commission for Europe (UNECE) (2011) http://w3.unece.org/pxweb/dialog/varval.asp?ma=02_GEICT_InternetUse_r&path=../DATABASE/STAT/30-GE/09-Science_ICT&lang=1. Accessed on May, 9 2011

van Gelder MM, Bretveld RW, Roeleveld N (2010) Web-based questionnaires: the future in epidemiology? Am J Epidemiol 172:1292-1298

Vandelanotte C, Spathonis KM, Eakin EG, Owen N (2007) Website-delivered physical activity interventions a review of the literature. Am J Prev Med 33:54-64

Wang J, Etter JF (2004) Administering an effective health intervention for smoking cessation online: the international users of Stop-Tabac. Prev Med 39:962-968

Webb TL, Joseph J, Yardley L, Michie S (2010) Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 12:e4

West R, Gilsenan A, Coste F, Zhou X, Brouard R, Nonnemaker J, Curry SJ, Sullivan SD (2006) The ATTEMPT cohort: a multi-national longitudinal study of predictors, patterns and consequences of smoking cessation; introduction and evaluation of internet recruitment and data collection methods. Addiction 101:1352-1361

Zhang Y, Woods R, Chaisson CE, Neogi T, Niu J, McAlindon TE, Hunter D (2006) Alcohol consumption as a trigger of recurrent gout attacks. Am J Med 119:800 e813-808

Zhang Y, Chaisson CE, McAlindon T, Woods R, Hunter DJ, Niu J, Neogi T, Felson DT (2007) The online case-crossover study is a novel approach to study triggers for recurrent disease flares. J Clin Epidemiol 60:50-55

4.5 Commentary

Commentary cover sheet

For a 'research paper' already published

1.1. Where was the work published?
1.2. When was the work published?
1.2.1. If the work was published prior to registration for your research degree, give a brief rationale for its inclusion

---

1.3. Was the work subject to academic peer review?
1.4. Have you retained the copyright for the work?
If yes, attach evidence of retention
If no, or if the work is being included in its published format, attach evidence of permission from copyright holder (publisher or other author) to include work

2. For a 'research paper' prepared for publication but not yet published

2.1. Where is the work intended to be published?
INTERNATIONAL JOURNAL OF EPIDEMIOLOGY

2.2. List the paper's authors in the intended authorship order
Lorenzo Richiardi, Costanza Pizzi, Neil Pearce

2.3. Stage of publication – Not yet submitted/Submitted/Undergoing revision from peer reviewers' comments/In press
ACCEPTED, AWAITING PROOF
(I acknowledge the Oxford University Press for permission to include the commentary entitled "Representativeness is not necessary and often should be avoided", accepted for publication in the International Journal of Epidemiology, in this PhD thesis)

3. For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper.

Lorenzo Richiardi led the study design and I provided advice on it. Lorenzo Richiardi wrote the first draft of the article. I provided several comments to the first draft and wrote sections of the revised version.

Candidate's signature

Supervisor or senior author's signature to confirm role as stated in (3) (Lorenzo Richiardi)

Representativeness is usually not necessary and often should be avoided

Lorenzo Richiardi[1], Costanza Pizzi[1,2], Neil Pearce[3,4]

1. Cancer Epidemiology Unit, Department of Medical Sciences, University of Turin and CPO-Piemonte, Italy.
2. Centre for Statistical Methodology, London School of Hygiene and Tropical Medicine, London, United Kingdom
3. Departments of Medical Statistics and Non-communicable Disease Epidemiology, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, United Kingdom
4. Centre for Public Health Research, Massey University, Wellington, New Zealand

Word count: text: 3243; Figures: 2; references: 9

We agree with Rothman and colleagues that scientific inference in epidemiology does not require representativeness of the general population or target population in order to be valid. This is an important message and we welcome Rothman and colleagues' paper which has clearly expressed this position [1].

On the other hand, perhaps Rothman and colleagues go too far in arguing that representativeness should be avoided as a matter of principle, and we consider that there are some situations where representativeness is the most sensible approach. For example, it would be rare for researchers to only study one age-group, and to then attempt to extrapolate their findings to other age-groups, if sufficient numbers and funding were available to also sample adequate numbers from these other age-groups.

In our experience, there are three usual reasons for deliberately opting for non-representativeness in a study design ('intentional' non-representativeness): (i) practical reasons, e.g. it may be most practical to restrict a study to those who have a telephone; (ii) to minimise bias, e.g. by restricting a study to a particular population subgroup (as in the British doctors study [2]) so that there is less likelihood of lifestyle differences between exposed and non-exposed within that group; and (iii) in order to focus on one or more population subgroups, e.g. if we wish to compare exposure-outcome estimates in different ethnic groups [3]. In the first instance, representativeness is not necessary and would usually not improve the feasibility of the study; in the latter two situations it should specifically be avoided.

In addition, non-representativeness may also be 'unintentional', e.g. in longitudinal studies because of low baseline response rates or the recruitment of volunteers rather than a formal sample of a defined population. Such unintentional selection may occur both in studies involving random population samples and in those involving non-representative samples. In this paper we will focus

mainly on the issues involved in intentional non-representativeness, but will also consider issues of unintentional non-representativeness. In this latter situation, the potential for bias may be relatively greater. In particular there is potential for large bias if the outcome of interest or its early signs affect the probability of baseline selection. We will argue however, that, provided that the outcome does not affect selection, situations of intentional and non-intentional non-representativeness are generally similar in terms of validity. Furthermore, baseline self-selection is likely to create a group of more motivated persons in longitudinal studies, which may result in a better response to follow-up and, thus, in decreased selection bias. Thus, the possibility of bias from lack of representativeness needs to be balanced against the likelihood of bias from poor response to follow-up in a more representative sample. For example, most researchers, if given the choice, would opt to base a study on 50% of the population and then achieve good follow-up rates, rather than to start with a representative sample and then only achieve 50% follow-up.

We should also note that in some instances the aim of an epidemiological study is primarily descriptive, e.g. to estimate the prevalence of a condition such as asthma in the general population [4], and in these studies representativeness is necessary to obtain valid estimates. Furthermore, such studies often are not completely descriptive. For example, prognostic research is population and time specific, but the identification of a cause of disease progression may add information on the understanding of a biological phenomenon.

We will focus on 'analytical' studies which aim to estimate a particular exposure-disease association, while appropriately controlling for confounding and avoiding other biases. In this situation, we agree that representativeness is not a goal *per se,* but rather needs to be justified in the context of the particular study. For example, in a clinical trial where we want to understand the efficacy of a treatment for a disease, a random sample is clearly not needed and in many ways it can

be inappropriate. Typically we restrict the initial studies to high risk patients or to patients expected to have a high compliance with the assigned treatment and follow-up.

We have been involved in discussions on representativeness a number of times since 2005, when we started an internet-based birth cohort in Italy (NINFEA cohort, www.progettoninfea.it) [5], followed by a similar study in New Zealand (ELFS cohort, www.elfs.org.nz). Internet-based recruitment has advantages in terms of feasibility, costs and possibilities of reaching traditionally understudied populations. However, this approach is often criticized on the basis of its consequent lack of representativeness of the general population. Internet-based recruitment selects participants who have access to the internet, become aware of the existence of the study, and volunteer to participate. Thus, it is based on a restricted source population and the study population is a self-selected sample of the source population (i.e. non-representativeness is both intentional and unintentional).

In this commentary we describe these criticisms and argue, in line with Rothman and colleagues, that restricting a study to a subgroup of the general population does not usually hamper scientific inference, and may often enhance it. We focus on infant cohort studies, but the same arguments generally apply to the corresponding case-control and cross-sectional studies based on the same restricted populations. We focus on the main two arguments which we have received against using non-representative populations in internet-based birth cohorts: (i) lack of heterogeneity; and (ii) the potential for bias. We also consider a third potential criticism relating to selection on a mediating variable.

Criticism 1: Non-representative cohorts lack heterogeneity.

One major criticism of the use of non-representative samples is the resulting lack of heterogeneity, with regards to exposures, potential effect modifiers, or both. Although it is true that restriction may decrease the range of exposure levels and the magnitude of the contrasts, in fact, we argue that using non-representative samples may often enhance study power to assess main effects and effect modification. To study a rare exposure, for example, either we assemble a very large cohort or we do 'smart selection' of its members. For example, in an internet-based birth cohort study, in which members are characterized by a high socioeconomic status, women having their first pregnancy after their forties are overrepresented. When high maternal age is the exposure of interest, an internet-based birth cohort becomes more efficient than a birth cohort which is representative of the general population. Similarly using non-representative samples may enhance our ability to assess heterogeneity with regards to potential effect modifiers, e.g. by ensuring that there are adequate numbers in each of the ethnic groups to be considered if we suspect or are interested in potential modification by ethnicity.

These arguments refer to issues of study efficiency, but lack of heterogeneity among study participants may be an advantage with regards to controlling confounding. Ideally, the best study in terms of scientific validity would be a design involving large heterogeneity in the exposure and complete homogeneity in all other characteristics (provided we did not wish to investigate effect modification and/or the effects of varying population contexts).

Of course it should be acknowledged that lack of heterogeneity is not always an advantage, particularly when there is important effect modification. It can happen that exposure has strong effects in one population subgroup and weaker or non-existent effects in another. If a study is based on the latter subgroup, then the effects of exposure will not be identified. However, once again, to explore such effect modification usually requires non-representative samples, e.g. by studying equal numbers in each age-group, gender or ethnic group.

Unless we are explicitly interested in, or have a priori reason for, investigating heterogeneity, generalizability is a matter of scientific inference rather than representativeness. There are many situations in which such generalisability is relatively straightforward. Smoking causes lung cancer in every population in which it has been studied, and there was no bias, and considerable practical advantages, to restricting one of the key early studies to British doctors [6]. Similarly, smoking presumably causes lung cancer in those with or without a telephone, those who have registered to vote and those who have not, and in those who use and those who do not use the internet. With rare exceptions, such restrictions may greatly enhance study practicality and thereby response rates and power, and have little or no effect on validity or generalisability.

Criticism 2: If the exposure of interest is associated with the probability of selection, the exposure-outcome associations estimated in a non-representative cohort may be biased.

The second major criticism of the use of non-representative samples is the possibility of introducing selection bias. When conducting a cohort study on a selected population, it is likely that there are factors that are associated with selection and are also determinants of the disease of interest. For example, in a cohort study restricted to British doctors, familial history of early mortality from cardiovascular diseases may affect both the lifetime probability of cardiovascular diseases and the decision of becoming a doctor. As with other risk factors, the exposure of interest may also be associated with the probability of selection: for example socioeconomic status may affect both smoking habits and grades at high school (and therefore the probability of being admitted to a Medical School). If both the exposure and another risk factor for the disease of interest are associated with the probability of selection, baseline restriction can introduce bias in the exposure-outcome association. This is a type of collider bias that has been discussed extensively in the epidemiological literature, including by us in the context of internet-based cohorts [7, 8]. Fortunately,

the amount of bias that is expected to be introduced by this phenomenon is small unless all of the associations involved in generating the bias are very strong. Assuming that all relative risks involved are of 2.0, the bias, in logarithmic scale, will be of 0.02 (i.e. a RR for the exposure-outcome association of 1.02, when the true relative risk is 1.00), while assuming that all RRs are of 4.0 it will be of 0.15 (i.e. a RR of 1.16, when the true RR is 1.00) [7].

However, the exposure of interest is almost always associated with some disease risk factors in the general population, whether or not we study a restricted subpopulation. Indeed each general population, at a given point in time, will have its specific confounding pattern. There is no reason to assume that confounding patterns for, say, the association of smoking with cardiovascular disease in London, UK, in 2012 is the same of that present in Turin, Italy, in 2012: we could, for example, expect that in London smoking is associated with drinking beer while in Turin it is associated with drinking red wine. The confounding pattern in the selected cohort may differ from that of the corresponding general population, but we cannot predict whether the amount of confounding will be greater, similar or smaller. The bottom line is that each population, including a selected study population, has its own confounding pattern. Valid scientific inference is achieved if the confounders are controlled for, and there is no reason to believe that control of confounding can be more easily achieved in a population-based cohort than in a restricted cohort. Indeed, we can intentionally restrict the cohort to decrease confounding bias. For example, if we are not able to precisely measure the amount of alcohol consumption in the general population, and we know that alcohol is a relevant confounder of the association of interest, we can restrict the study to non-drinkers and occasional drinkers.

In a recent paper, we compared, for selected exposures and outcomes of interest, the confounding pattern of the NINFEA internet-based cohort with that present in the corresponding general

population, showing that the overall confounding was not larger, but was qualitatively different, than that present in the general population [8].

As mentioned above, it is not impossible to devise situations in which selection bias could occur due to restriction (i.e. non-representativeness), e.g. when an exposure and an unmeasured risk factor for the disease are independent in the general population but both are associated with the probability of selection. Our argument is not that such bias is impossible, but rather that restricted studies are often likely to be less affected by confounding. Also, any small likelihood of bias from using non-representative samples needs to be balanced against the likelihood of bias if attempts to use random representative samples result in low response rates at follow-up and/or a greater likelihood of information bias. The British doctors study is a relevant example once again, in which the non-representative sample has likely induced better follow-up and greater validity of the smoking information gathered. To insist on doing the study in a random general population sample would have had little or no benefit, and considerable disadvantages in terms of logistics and study validity.

Criticism 3: If an intermediate variable in the causal pathway from the exposure to the outcome is associated with the selection, exposure-outcome associations estimated in a non-representative cohort may be biased.

We would therefore argue that the main reasons for opposing the use of non-representative samples – lack of heterogeneity and the potential for introducing selection bias and/or confounding – are rarely valid, and are generally outweighed by the benefits of this approach, although of course this conclusion is highly hypothesis- and study-dependent. In the rest of this paper, we will consider an issue which has been less debated, namely the situation in which an intermediate variable (a mediator, that is a variable that is on the pathway from the exposure to the outcome) is associated with the probability of selection.

In most circumstances, baseline selection in cohort studies takes place *before* the intermediate variable is manifest. For example in the British doctors study it could be assumed that members of the cohort became doctors before the occurrence of manifest mediators of the effect of the exposure (smoking) on the outcomes of interest. Analogously, in an internet-based birth cohort, having access to the internet likely occurs before pregnancy and, thus, before most of the possible intermediate variables may become manifest. Within this framework, if there is a variable affecting both the intermediate variable and the probability of selection, the use of a non-representative sample could alter the exposure-mediator confounding pattern. This situation is illustrated in Figure 1 using directed acyclic graphs. Figure 1a shows a non-representative cohort in which selection introduces exposure-mediator confounding that was not present in the general population; Figure 1b shows the case of a representative cohort in which there is already exposure-mediator confounding; in Figure 1c a non-representative cohort study is conducted in the same population of Figure 1b; in Figure 1d the exposure-mediator confounder also affects the probability of selection. An example of the scenarios depicted in Figures 1b and 1d would be the effect of pre-pregnancy BMI (E) on pre-term delivery (O), in which gestational hypertension is a possible mediator (M). Socio-economic class (C) would be an exposure-mediator confounder, assuming that it affects both pre-pregnancy BMI and gestational hypertension but, in a simplified scenario, it is not a determinant of pre-term delivery otherwise. In a study restricted to internet users, socioeconomic status would also affect selection (S) (as in Figure 1d) and thus, restriction would be likely to decrease exposure-mediator confounding due to socioeconomic status.

In summary, some of the scenarios described in Figure 1 increase the overall exposure-mediator confounding, while others decrease it. We consider that there is no reason to expect that non-representative cohorts tend to have a larger exposure-mediator confounding than representative cohorts, although we can always plan the selection in order to decrease exposure-mediator confounding. We should acknowledge that a confounder of the exposure-mediator association is

often treated as a confounder of the exposure-outcome association, especially when quantifying the role of the mediator is not the focus of the study. In this context, scenarios described in Figure 1 become very similar to those described in the previous section (Criticism 2).

It is possible that baseline selection occurs *after* an intermediate variable becomes manifest. This may typically happen both in representative and in non-representative cohorts when there is unintentional non-representativeness. In a study involving internet-based recruitment, the fact that participants are volunteers that should need to first be aware of the existence of the study may enhance this potential problem. If the intermediate variable has a direct effect on the selection, a number of different scenarios may occur. The simplest scenario is that described in Figure 2, in which there is only a direct effect from the mediator to the selection. According to the causal relationship described in this figure (in which there are no other factors affecting the selection), the effect of the exposure on the outcome of interest would be attenuated. It should be considered, however, that typically the decision to participate in the study depends on a large number of factors and the selection process is poorly predicted by a single intermediate variable. Thus, the situation described in Figure 2 in most instances should introduce a negligible or modest bias in the estimate of the exposure-outcome association. The example of the effect of maternal pre-pregnancy BMI on preterm delivery in which gestational hypertension is an intermediate factor may be used to illustrate also the situation in which selection is directly affected by an intermediate variable. In particular, the decision of pregnant women to participate in the study could depend on whether they have gestational hypertension or not.

The relationship between intermediate variables and selection can become much more complex than described above [9]: for example, the selection could be affected both by the intermediate variable and by the participant's reaction to the intermediate factor. For example, in the hypothetical study of the effect of maternal pre-pregnancy BMI on the risk of pre-term delivery, where gestational

hypertension acts a mediator, we would have to consider that women are usually monitored during the remaining part of pregnancy and may be prescribed blood pressure medications. Participation in the cohort could be affected both by the gestational hypertension and by the consequent activities, e.g. those taking medications, being more or less likely to volunteer to participate in the study.

The interplay between intermediate variables and selection, as well as the natural history of disease, will have to be fully explored in a future work. However, it should be emphasized that, regarding selection, the issue can be solved by taking into account the temporal relationships between the variables under study and, thus, by enrolling the participants before the intermediate variable or its early signs could become manifest. In a birth cohort study involving enrolment during the first trimester of pregnancy, for example, selection cannot be directly affected by intermediate variables acting later in pregnancy or at birth.

Conclusions

In conclusion, we agree with Rothman and colleagues that scientific inference does not require representativeness, and often explicitly requires that study samples should not be representative. Overall, representativeness can be harmful or beneficial depending on the study question and context. There is no reason to embrace representativeness *per se*, as often restriction can enhance the practicality of a study and/or the validity of the scientific inferences. We acknowledge that further work is needed to fully understand some specific situations, in particular when an intermediate variable directly affects baseline selection. However, leaving aside this specific issue, we consider that the view that studies based on representative samples are clearly better than those based on restricted samples is untenable. Rather, although it is perhaps too strong to argue that representativeness should always be avoided, it is usually not necessary, and often should be avoided.

References

1. Rothman KJ, Gallacher J, Hatch EE. Why Representativeness Should be Avoided. *Int J Epidemiol* 2013.

2. Doll R, Hill AB. The mortality of doctors in relation to their smoking habits: a preliminary report. *Bri Med J* 1954:1451-5.

3. Kolonel LN, Henderson BE, Hankin JH, *et al.* A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. *Am J Epidemiol* 2000;151:346-57.

4. Asher MI, Keil U, Anderson HR, *et al.* International Study of Asthma and Allergies in Childhood (ISAAC): rationale and methods. *Eur Respir J* 1995;8:483-91.

5. Richiardi L, Baussano I, Vizzini L, Douwes J, Pearce N, Merletti F. Feasibility of recruiting a birth cohort through the Internet: the experience of the NINFEA cohort. *Eur J Epidemiol* 2007;22:831-7.

6. Doll R, Hill AB. Mortality in Relation to Smoking: Ten Years' Observations of British Doctors. *Br Med J* 1964;1:1399-410.

7. Pizzi C, De Stavola B, Merletti F, *et al.* Sample selection and validity of exposure-disease association estimates in cohort studies. *J Epidemiol Community Health* 2012;65:407-11.

8. Pizzi C, De Stavola BL, Pearce N, *et al.* Selection bias and patterns of confounding in cohort studies: the case of the NINFEA web-based birth cohort. *J Epidemiol Community Health* 2012.

9. Howards PP, Schisterman EF, Poole C, Kaufman JS, Weinberg CR. "Toward a clearer definition of confounding" revisited with directed acyclic graphs. *Am J Epidemiol* 2012;176:506-11.

Figure 1. Effect of selection in a cohort study in which a mediator (M) of the effect of the exposure (E) on the outcome (D) becomes manifest after the selection has occurred. **a)** Non-representative cohort in which the mediator (M) and the selection (S) are affected by a common cause (R), and the exposure (E) is also associated with selection; **b)** Representative cohort in which there is exposure (E) – mediator (M) confounding from C; **c)** Scenarios described in Figure1a and Figure 1b coexist; **d)** Same as Figure 1b, but the confounder (C) also affects selection.

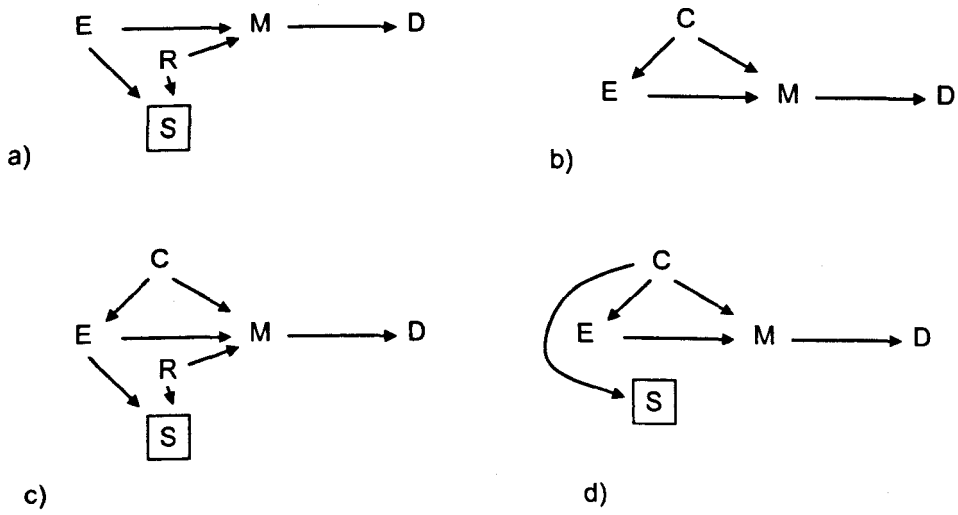
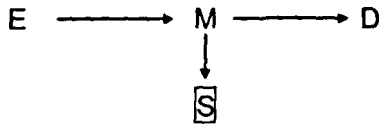


Figure 2. Selection of cohort participants (S) is affected by the mediator (M) of the exposure (E)-outcome (D) association.



Chapter 5

Growth Modelling

5.1 Preamble

The aim of modelling growth is to summarise individual data, made of repeated anthropometric measurements taken on a subject at different ages, into a reduced number of parameters, without significant loss of information. This process should allow the comparison of growth patterns between individuals or populations and, when the growth function is differentiable, to identify particular features of the growth curve such as turning points (e.g. timing of BMI peak during the first year of life) or velocity and acceleration at any time point within the observed age range. The target is thus to summarise the individual growth measurements taken at varying ages into a smooth curve, which could represent the underlying growth trajectory more closely that the raw data themselves. The optimal degree of smoothness, that is the one that provides a satisfactory balance between goodness of fit and both model parsimony and parameters interpretability, thus needs to be found. Further, it is desirable that the fitting procedures used to estimate these models are robust to missing values and measurement error and to irregularity of the ages at which observations are made, which are common issues in longitudinal growth studies (Berkey, 1982).

Several models have been proposed in the literature to describe human growth, either specific to an anthropometric measurement or to a range of ages. Hereafter we will refer to this class of models as

"Biological models". More general statistical models have also been applied to growth data to describe their patterns and extract particular features. These include in particular polynomials and fractional polynomials models (Royston and Altman, 1994), random effect models (Goldstein, 2010), splines methods (Ruppert *et al.*, 2003) and latent growth models (Jung and Wickrama, 2008). Hereafter we will refer to these class of model as "Statistical models". Some of these models will be briefly reviewed here.

5.1.1 Biological models

Among the models specifically developed to describe human growth, many are concerned with at least the first 2 years of life, for example the Jenss-Bayley (Jenss and Bayley, 1937), Count (Count, 1943), Berkey-Reed (Berkey and Reed, 1987), infancy-childhood-puberty (ICP) Karlberg (Karlberg, 1987) and Jolicoeur-Pontier-Pernin-Sempé (Jolicoeur *et al.*, 1988) models, with the last two describing growth in height from birth through to final adult height. On the other hand, the Jenss-Bayley, Count and Berkey-Reed models have been proposed for both height and weight, but covering only the first 6/8 years of life.

All of these models were conceived as models to be fitted separately on each individual. Extensions for the joint modelling of multiple individuals are discussed later. Also these models were originally conceived for un untransformed anthropometric measures. Formal specifications of these models as fitted separately on each subject are given below, with y_{ij} being the anthropometric measurement made for subject j, $j = 1,\ldots,J$, at times t_{ij}, $i = 1,\ldots,n_j$, the α_js being the unknown subject-specific parameters and ϵ_{ij}s being the error terms assumed to be normally distributed with mean 0 and variance $\sigma^2_{\epsilon j}$:

The Jenss-Bayley model

$$y_{ij} = \alpha_{1j} + \alpha_{2j}t_{ij} - e^{\alpha_{3j}+\alpha_{4j}t_{ij}} + \epsilon_{ij} \tag{5.1}$$

where y_{ij} is height or weight at age t_{ij}, with t ranging from birth to 6/8 years. The model accounts for the rapid decelerating growth usually observed after birth via the exponential component, which after infancy becomes negligible.

The Count model

$$y_{ij} = \alpha_{1j} + \alpha_{2j} t_{ij} + \alpha_{3j} \log t_{ij} + \epsilon_{ij} \tag{5.2}$$

This model has been used to describe height and weight as well as some skull dimensions in early childhood. Count presented also a model which considers height throughout the whole period of growth, adding an additional component for the time 6 to 11 years and one for the last phase of growth (Count, 1943).

The Berkey-Reed model

$$y_{ij} = \alpha_{1j} + \alpha_{2j} t_{ij} + \alpha_{3j} \log t_{ij} + \alpha_{4j} \left(\frac{1}{t_{ij}} \right) + \epsilon_{ij} \tag{5.3}$$

$$y_{ij} = \alpha_{1j} + \alpha_{2j} t_{ij} + \alpha_{3j} \log t_{ij} + \alpha_{4j} \left(\frac{1}{t_{ij}} \right) + \alpha_{5j} \left(\frac{1}{t_{ij}^2} \right) + \epsilon_{ij} \tag{5.4}$$

These two versions of the Berkey-Reed model are an extension of the Count model with the addition of one and two decelerating terms respectively, which, in turn, allow for one or more inflection points in the growth curve. Here y can be either height or weight as well as head circumference.

The Kalberg-infancy component model

$$y_{ij} = \alpha_{1j} + \alpha_{2j}(1 - e^{-\alpha_{3j} t_{ij}}) + \epsilon_{ij} \tag{5.5}$$

This is the infancy component of the ICP model, already mentioned above, which consists of a decelerating function lasting until about 3-4 years of age. It is an exponential function, where y is height.

The Jolicoeur-Pontier-Pernin-Sempé model

$$y_{ij} = A\left(1 - \frac{1}{1 + (\frac{t'_{ij}}{\alpha_{1j}})^{\alpha_{2j}} + (\frac{t'_{ij}}{\alpha_{3j}})^{\alpha_{4j}} + (\frac{t'_{ij}}{\alpha_{5j}})^{\alpha_{6j}}}\right) + \epsilon_{ij} \tag{5.6}$$

here t'_{ij} represents the 'total age' of subject j from conception (that is age t_i plus the duration of the pregnancy), y is height and A is adult height, with the model describing its growth from birth until maturity.

Other anthropometric models have been proposed in the literature, for example the Preece-Baines model (Preece and Baines, 1978) used to describe growth in height from age two years to maturity, but they are not presented here because they deal with human growth after the infancy period and therefore are not relevant for this Ph.D. The Jenss-Bayley model (5.1) and the 4-parameters version of the Berkey-Reed model (5.3), which were specifically developed to model weight in the first years of life (as well as height), will be reviewed in detail in Research Paper III.

5.1.2 Statistical models

As already stated, the model described above (defined as "Biological models") were specifically developed to describe human growth in a given anthropometric dimension for a specific age range. These models imply a basic functional form of the growth curve with their parameters usually allowing some functional interpretation (Hauspie and Molinari, 2004). In contrast, general statistical models do not postulate a particular form of the growth curve. Furthermore because the repeated observations for an individual within the sample are not mutually independent inferential problems arise when all data are analyzed together. Data such as these are called hierarchical. These also arise for example with data on patients clustered within hospitals, children within schools, siblings within families etc. Growth data collected on cohorts of children are therefore a special case of hierarchical data (called longitudinal data) where the clusters are the individuals who are followed over time. A particular feature that distinguishes this type of data from standard clustered data is that the strength of dependency within each cluster is influenced by the time interval elapsed between observations, that is the within cluster correlation structure is driven by time. There are alternative approaches to deal with the correlated

structure of observations such as these: fitting models that explicitly acknowledge the hierarchical structure of the data or using a robust approach to the estimation of the precision of the estimates obtained ignoring it. Only the first approach will be considered in this thesis.

The hierarchical models suitable for handling these data differ according to whether the data can be classified as *balanced* or *unbalanced*, a concept that concerns the study design. When analysing repeated growth measurements data can be defined as balanced if the time points at which the anthropometric measures are planned to be observed are the same for each study-subject. Conversely, when the repeated observations are not taken at the same time for each individual the resulting longitudinal growth data are unbalanced. Unbalanced growth data can be analysed using *random effects models*, also called mixed effects or multilevel models (Goldstein, 2010), under the assumption that the irregularity of the observations is not guided by the actual values of the anthropometric variable. An introduction to this class of models and to generalization of this approach particularly relevant for this Ph.D. is provided below.

Random effects models

Random effects models explicitly incorporate the hierarchical structure of the data into the modelling process, i.e. the correlations among the observations on the same child. They also allow to investigate the influence of predictors on the growth trajectories. Using a random effects approach all subjects with at least one growth measure can theoretically be included in the analyses, as this method borrows information across individuals to estimate individual parameters.

Formally, let y_{ij} be the anthropometric measurement made for subject j, $j = 1, \ldots, J$, at times t_{ij}, $i = 1, \ldots, n_j$ and $\sum_{j=1}^{J} n_j = N$ (that is N is the total number of measurements), where the subjects represent the clusters (level-2) and the time points represent the elementary units (level-1). Note that if data are balanced $n_j = n$ and $t_{ij} = t_i$ for every j. Also note that the anthropometric measure y could be transformed (e.g. log-transformed) to meet the distributional assumptions of the model.

The simplest random effects model in this class is called the **random intercept** linear model in which the variation in individual trajectories is modelled via random intercepts. When there are no

explanatory variables apart from time, this is given by

$$y_{ij} = (\mu + u_{0j}) + \beta t_{ij} + e_{ij} \tag{5.7}$$

where u_{0j} represent random variables with mean 0 and variance σ^2_{u0} (the between-subjects variance) and e_{ij} also are random variables with mean 0 and variance σ^2_e (the within-subjects variance). It is generally assumed that $u_{0j} \sim N(0, \sigma^2_{u0})$ and $e_{ij} \sim N(0, \sigma^2_e)$, independent of each other and of the explanatory variables (in this case t_{ij}). Observations from two different subjects are thus independent, while a within-subject correlation is allowed from this model. Given that

$$Cov(y_{1j}, y_{2j}) = Var(u_{0j}, u_{0j}) = \sigma^2_{u0} \tag{5.8}$$

and

$$Var(y_{1j}) = Var(y_{2j}) = \sigma^2_{u0} + \sigma^2_e \tag{5.9}$$

the within-subject correlation (also called *intra-class correlation (ICC)*) is given by:

$$\rho = \frac{\sigma^2_{u0}}{\sigma^2_{u0} + \sigma^2_e} \tag{5.10}$$

representing the proportion of total variance explained by the variation between subjects.

In the linear model (5.7) the parameters μ and β are referred to as fixed effects, while the u_{0j} as random effects. The marginal regression, given by $E(y_{ij}|t_{ij}) = \mu + \beta t_{ij}$, represents the population average relationship between the anthropometric dimension and time. According to this model, the individual intercepts are allowed to vary around a mean value but the slope, β (i.e. the effect of time), is forced to be the same for every subject.

A more appropriate model for longitudinal growth data is the **random intercept and slope** model (also known more generally as *random coefficients* model), which is a natural extension of the previous one, allowing both the slopes and the intercepts to be subject-specific:

$$y_{ij} = (\mu + u_{0j}) + (\beta + u_{1j}) t_{ij} + e_{ij} \tag{5.11}$$

where here both u_{0j} and u_{1j} are considered as random effects. These variables have the following assumptions:

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N(0, \Sigma_u), \quad \Sigma_u = \begin{bmatrix} \sigma^2_{u0} & \\ \sigma_{u01} & \sigma^2_{u1} \end{bmatrix} \tag{5.12}$$

$$e_{ij} \sim N(0, \sigma^2_e)$$

It follows that now the level-2 residuals, the u's, are allowed to be correlated. As before the level-2 residuals are assumed to be independent of the level-1 residuals, e_{ij}, and all the residuals are assumed to be independent of the model's explanatory variables (t_{ij} in this case). It follows that, as before, measurements from different subjects are still uncorrelated, but for observations within the same subject:

$$Cov(y_{1j}, y_{2j}) = Cov(u_{0j} + u_{1j}t_{1j} + e_{1j}, u_{0j} + u_{1j}t_{2j} + e_{2j}) = \sigma^2_{u0} + \sigma^2_{u1}t_{1j}t_{2j} + \sigma_{u01}(t_{1j} + t_{2j}) \tag{5.13}$$

Their covariance thus depends on the times of measurement, that is a non constant covariance between observations in the same individual is allowed. Moreover the model implies that the variance varies with time too and it is given by:

$$Var(y_{ij}) = \sigma^2_{u0} + \sigma^2_{u1}t^2_{ij} + 2\sigma_{u01}(t_{ij}) \tag{5.14}$$

It should be noticed that, if the time interval elapsing between measurements on a subject is very short, the assumption of independence of the within-individual residuals (level-1) may not hold and thus the error structure in longitudinal data may be more complex than this (Goldstein *et al.*, 1994). Finally random effects models can be extended to allow for additional predictors to have random effects, to allow for more than 2 hierarchical levels or to allow for nonlinear relationships (for example adding a quadratic term for the time variable). In general random effects models have a common function of time (i.e. linear or quadratic) to describe the average growth pattern of all the study-subjects from which each departs according to the distribution of the level-2 errors. However different subjects may

follow very different growth trajectories over time, and growth patterns are often more complex than those described by linear or curvilinear functions. Although random slopes for both the linear and the quadratic term can be included to allow for subjects with different growth trajectories, imposing such rigid algebraic form to the growth curves may result in a poor fit to the data. An alternative is to consider more flexible random effects models, such as random effects specification of the biological parametric models described above, or to use spline methods, which can also be implemented within a random effects model framework. A brief introduction to these approaches is given below.

Random effects specification of biological models

As discussed above, the models specifically developed to described human growth were conceived to be fitted separately on each child, with population estimates of the model's parameters derived by summarizing the individual ones (using their means and standard deviations for example). However, this approach suffers from some limitations. In particular individual curves can be fitted only to those subjects for whom a minimum number of measurements is available, depending on the number of parameters to be estimated. Moreover if growth measurements for each subject are available only for a limited number of occasions the variability of the resulting parameter estimates could be very large. In order to overcome these problems distributional assumptions for the subject-specific parameters could be added to define an overall model for all children and therefore extending them to have a random effects model specification. As stated above, using a random effects approach all subjects with at least one growth measure can theoretically be included in the analyses. As a consequence the variability of the random effects is expected to be much lower than that of the corresponding fixed effects, obtained by fitting individual curves. Moreover allowing to fit one single model for all children instead of fitting one model for each child separately this approach significantly reduces the computational burden. For example, using the same notation as in equation (5.1), the child-specific parameters of the Jenss-

Bayley model can be specified as follow:

$$\begin{aligned} \alpha_{1j} &= \alpha_{10} + \alpha_{11j} \\ \alpha_{2j} &= \alpha_{20} + \alpha_{21j} \\ \alpha_{3j} &= \alpha_{30} + \alpha_{31j} \\ \alpha_{4j} &= \alpha_{40} + \alpha_{41j} \end{aligned} \tag{5.15}$$

where α_{10} represents the fixed effect and α_{11j} the child-specific random effect for the parameter α_{1j}, with similar definition for the components of α_{2j}, α_{3j} and α_{4j}. As observed for the random intercept and slope model, described in equations (5.11) and (5.12), the child-specific random effects α_{11j}, α_{21j}, α_{31j} and α_{41j} would be assumed to be drawn from a multivariate normal distribution with mean **0** and covariance matrix $\mathbf{\Phi}$ and are assumed to be independent of the level-1 errors (ϵ_{ij} in equation (5.1)). Similar specification could be applied to the other biological models described above.
This is the approach used in Research Paper III for the Jenss-Bayley and the 4-parameters version of the Berkey-Reed model.

Spline models

Splines methods are used to handle non-linear relationships between an outcome variable and an explanatory variable, with the objective of achieving a good balance between a good fit to the data and a smoothed curve (Green and Silverman, 1994; Ruppert *et al.*, 2003). In order to apply these methods it is important to clarify which are the different types of splines available and what are the alternative options each of them offers. In this context we are concerned with the relationship between an anthropometric variable of interest y_{ij} for subject j, $j = 1, \ldots, J$, and age t_{ij}, $i = 1, \ldots, n_j$.
Generally, the problem is to define a function $g(t)$, the so called spline function, such as

$$y_{ij} = g(t_{ij}) + \epsilon_{ij} \tag{5.16}$$

Formally, a spline is a piecewise-polynomial function of the explanatory variable (e.g. age) on an

interval $[a, b]$ composed of K subintervals, with the order of the polynomial defining the degree of the spline p and with the subintervals joined at points called *knots*. Usually the values of this function and the first $p-1$ derivatives are continuous at the knots. This property guarantees the smoothness of the spline curve but also imposes some constraints in its flexibility.

Let $g(t_{ij})$ be the smooth curve to be used when fitting a model for the growth data for subject j, splines are defined by selecting a *basis*, that is a set of known functions that determine a family of transformations to be applied to the original variable (t_{ij}) (Schimek, 2009). Let $B_s(t_{ij})$ be the sth basis function then $g(t_{ij})$ can be represented by

$$g(t_{ij}) = \sum_{s=1}^{S} \beta_s B_s(t_{ij}) \tag{5.17}$$

where S is the basis dimension and the β_s are the unknown coefficients to be estimated. The basis dimension depends on the number of knots K, the type of basis function and the degree p of the spline. Three basic choices need to be made when applying splines (Ruppert *et al.*, 2003): *(i)* the spline model; *(ii)* the basis function; and *(iii)* the estimation method.

The spline model

(i) The spline model is specified in term of the degree of the polynomial used – with cubic splines being the most frequent choice –, the number of knots and their location, and the imposition of the boundary constraints. The latter refers to the constraint that the spline is linear in the tails beyond the boundary knots. Splines with this constraint are named "natural splines". Automatic procedure to select the knots are available, and often knots are selected to be placed at equally-spaced quantiles of the observed distribution of the explanatory variable, in this case age. Choice of the number of knots, rather than their location, has been shown to be crucial to the fit (Ruppert *et al.*, 2003; Schimek, 2009). An alternative is to place the knots at specific points, where significant changes in the curve of interest are expected.

The basis function

(ii) The simplest basis is that generated applying the truncated power function. Defining $(t-\kappa)_+$ as the function which takes value 0 to the left of κ and value $(t-\kappa)$ from κ onwards, the *truncated power*

basis for a spline with degree p and K knots is:

$$1, t, t^2 ..., t^p, (t - \kappa_1)_+^p, ..., (t - \kappa_K)_+^p \tag{5.18}$$

The corresponding pth degree spline model is:

$$g(t_{ij}) = \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + ...\beta_p t_{ij}^p + \sum_{k=1}^{K} \beta_k (t_{ij} - \kappa_k)_+^p \tag{5.19}$$

These bases have the drawback of not being orthogonal, often leading to numerical instability/less accuracy of the fit, especially when the number of knots is large and no or small penalty is applied. The most common alternative is the *B-spline basis* (for which there is not a straightforward equation), which has more stable numerical properties. Other alternatives, such as radial bases, exist (Ruppert *et al.*, 2003).

The estimation method

(iii) Two broader classes of splines exist: fully-parametric splines, which parameters are estimated via standard approaches (e.g. ordinary least squares or maximum likelihood), and *penalized splines*, which use a penalized likelihood approach. With the former the degree of smoothness of the curve is entirely specified through the number and position of the knots, making these choices crucial to the fit. Penalized splines are a more flexible approach which overcomes this limitation, as it provides automatic procedures to select the degree of the smoothness. In order to deal with fits to the data which are too rough, which is usually due to including too many knots, either the number of knots has to be reduced or all the knots are retained but their influence is constrained. This is the general concept that lies behind the so called penalized splines (Ruppert *et al.*, 2003). The aim is to avoid unconstrained estimations of the coefficients of the K knots which could lead to a wiggly fit: instead of removing some knots, which is equivalent to set to 0 their corresponding coefficients while leaving the remaining ones unconstrained, all coefficients of the spline basis are shrunken towards zero. A roughness penalty term is thus introduced – with the amount of smoothing controlled by a smoothing parameter – to penalize fits which are too rough. A general definition of penalized splines is $\widehat{\beta}^T B(t_{ij})$,

where $\widehat{\beta}$ is the minimiser of

$$\sum_{i=1}^{n} \left\{ y_{ij} - \beta^T B(t_{ij}) \right\}^2 + \alpha \beta^T D \beta \tag{5.20}$$

with $\alpha > 0$ being the *smoothing parameter*, $\alpha\beta^T D\beta$ representing the *roughness penalty* and D, a symmetric positive semidefinite matrix, the penalty matrix (Ruppert *et al.*, 2003). Once the basis function and the penalty have been chosen, the penalty matrix D is automatically determined. The main advantage is that, with this spline, knots selection problem can be overcome, as automatic procedures to select the smoothing parameter exist. It has been shown that, provided the knots cover all the range of the observation points, their number and location does not affect much the result, while the choice of the smoothing parameter α is more relevant (Gurrin *et al.*, 2005; Ruppert *et al.*, 2003).

A particular type of spline is the *smoothing spline*. This is the spline obtained by minimizing the *penalized sum of squares*

$$\sum_{i=1}^{n} \{ y_{ij} - \widehat{g}(t_{ij}) \}^2 + \delta \int_a^b \widehat{g''}(t)^2 dt \tag{5.21}$$

where the roughness of the fitted curve is quantified by a penalty of the integrated squared second derivative of the curve $g(t)$ while the goodness of fit is measured by the residual sum of squares. Therefore, similarly to the penalised spline, a roughness penalty approach is used with δ being the smoothing parameter used to compromise between a good fit and a a smoothed curve: as δ gets smaller $g(t_{ij})$ will tend to follow the curves of the observed data more closely. It has been shown that the minimiser of this function is a natural cubic spline with knots placed at each discrete value of the explanatory variable. Penalized splines are however usually preferred to smoothing splines, as a much lower number of knots is needed to obtain the same amount of smoothness while strongly reducing the computational burden (Ruppert *et al.*, 2003; Schimek, 2009).

A particularly useful application of these spline techniques is their implementation within a random effects model framework (Ruppert *et al.*, 2003). Splines can be extended with the inclusion of subject-specific random effects, which capture the departure of a given individual's curve from the average

(population-level) pattern. As discussed above, this will allow to define an overall model for all subjects.

Shape-invariant random effects model

Among the applications of spline methods within a random effects model framework, of relevance for this Ph.D., is the shape-invariant model with random effects introduced by Beath (Beath, 2007) to describe infant weight growth and by Cole (Cole *et al.*, 2010) to analyse pubertal growth in height. Cole used a slightly modified parametrization, which is the one adopted here, and named the model SuperImposition by Translation And Rotation (SITAR). In this model a common spline function for all subjects is modified by shifting and scaling the two axes to adapt the common function to the individual curves (more details are included in Research Papers III and IV). In these applications a natural cubic spline with B-spline basis is used to fit the data by using a non-linear random effects model, with the parameters of the spline treated as fixed effects whereas the coefficients of the shape invariant model are treated as random. Let, again, y_{ij} be the anthropometric measurement made for subject j, $j = 1, \ldots, J$, at times t_{ij}, $i = 1, \ldots, n_j$, then the SITAR model is specified as:

$$y_{ij} = \alpha_j + h\left(\frac{t_{ij} - \beta_j}{e^{-\gamma_j}}\right) + \epsilon_{ij} \tag{5.22}$$

where $h(z)$ is the natural cubic spline curve of the growth variable regressed on z (the transformed age) and $\alpha_j = \alpha_0 + \alpha_{1j}$ (size) is a subject-specific coefficient with α_0 representing the fixed effect and α_{1j} the random effect, with similar definitions for the other two parameters β_j (tempo) and γ_j (velocity). It should be acknowledged that shape-invariant modelling has been applied before to summarize mechanisms that are common to most human being. In particular, shape-invariant models using the logistic or the Gompertz models as shape functions have been used to study the pubertal growth spurt (Hauspie and Molinari, 2004).

In Research Paper III SITAR is described and its fitting and interpretability discussed in comparison with the Jenss-Bayley and Berkey-Reed models, while in Research Paper IV SITAR is extended to include multiple time-fixed explanatory variables and used to investigate the association between several maternal prenatal exposures and size, velocity and tempo of infant growth of children belonging to the three cohorts available for this Ph.D.

5.2 Research Paper III

Research paper cover sheet

For a 'research paper' already published

1.1. Where was the work published?
1.2. When was the work published?
1.2.1. If the work was published prior to registration for your research degree, give a brief rationale for its inclusion
1.3. Was the work subject to academic peer review?
1.4. Have you retained the copyright for the work?
If yes, attach evidence of retention
If no, or if the work is being included in its published format, attach evidence of permission from copyright holder (publisher or other author) to include work

2. For a 'research paper' prepared for publication but not yet published

2.1. Where is the work intended to be published?
JOURNAL OF THE ROYAL STATISTICAL SOCIETY SERIES A

2.2. List the paper's authors in the intended authorship order
Costanza Pizzi, Tim J Cole, Camila Corvalan, Isabel dos Santos Silva, Lorenzo Richiardi, Bianca L De Stavola

2.3. Stage of publication – Not yet submitted/Submitted/Undergoing revision from peer reviewers' comments/In press
IN PRESS
(I acknowledge John Wiley & Sons for permission to include the article entitled "On modelling early life weight trajectories", accepted for publication in the Journal of the Royal Statistical Society, in this PhD thesis)

3. For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper.

I led the study design. Bianca De Stavola and Tim Cole provided advice on the study design. Tim Cole provided the program code in R used to fit the SITAR model. I wrote the program code in R for the other two models and provided several comments to the code written by Tim Cole, many of which have been implemented in the final version of the program. I performed the statistical analyses and wrote the first draft of the article. All co-authors helped interpret the findings and revised the manuscript, providing useful comments.

Candidate's signature

Supervisor or senior author's signature to confirm role as stated in (3) (Bianca De Stavola)

On modelling early life weight trajectories

Costanza Pizzi*[1,2], Tim J. Cole[3], Camila Corvalan[4], Isabel dos Santos Silva[5], Lorenzo Richiardi[1], and Bianca L. De Stavola[2]

[1] *Cancer Epidemiology Unit, Department of Medical Sciences, University of Turin and CPO-Piemonte, Italy*
[2] *Centre for Statistical Methodology, London School of Hygiene and Tropical Medicine, UK*
[3] *MRC Centre of Epidemiology for Child Health, UCL Institute of Child Health, London, UK*
[4] *Institute of Nutrition and Food Technology, University of Chile, Chile*
[5] *Non-communicable Disease Epidemiology Department, London School of Hygiene and Tropical Medicine, UK*

Abstract

There is broad recognition that early life growth trajectories can contribute to the study of the onset and development of several health outcomes. In this paper we review the random effects specifications of two models purposely developed to describe anthropometric data and a shape invariant random effects model recently proposed in the statistical literature. They are compared in terms of their ability to extract salient and biologically meaningful features of growth in infancy and also to validly represent the data. We discuss advantages and limitations in choosing and interpreting each of the models using longitudinal weight data taken from 0 to 4 years from three contemporary birth cohorts.

Keywords: Growth curve; Jenss-Bayley; Random effects; Reed; SITAR; Splines; Weight.

1 Introduction

Much interest in modelling growth data comes from research in life course epidemiology, Barker (1998); Huxley et al. (2000); Kuh and Ben-Shlomo (2004); Baird et al. (2005), where summary growth parameters are used as explanatory variables for the onset of a later outcome. In such settings analyses consist of two stages: the first focussed on modelling the growth data, and the second aimed at relating parameters from the growth model, e.g. the age at peak weight velocity (APWV), to distal health outcomes, e.g. cardiovascular disease later in life. This field of research has been enriched by several new European (CHICOS (2010)) and worldwide (Richter et al. (2011); Brown et al. (2007)) birth cohorts established to study early life influences on the onset and development of a wide range of later diseases. As a consequence, a large body of longitudinal growth data, collected in different populations and periods

*Address for correspondence: Costanza Pizzi, University of Turin, Cancer Epidemiology Unit, Via Santena 7, 10126, Turin, Italy. E-mail: costanza.pizzi@lshtm.ac.uk

and with different approaches and levels of detail, is available. There is therefore an increasing need for understanding and applying modelling strategies that properly exploit these data, also taking into account both the pattern and the quality of the growth measurements.

The traditional approach of analysing anthropometric data consists of finding a parametric curve which is suitable for the variable and the age range of interest, while fitting it separately to each child. For example the *Jenss-Bayley* (JB) model was suggested in 1937, Jenss and Bayley (1937), to describe weight or height growth from birth to age six years. It is a fully parametric model that includes a linear and a nonlinear (exponential) part. Another model widely used with early life height and weight, is the four, or five, parameter *Reed* model, Berkey and Reed (1987), where different parametrizations can be used to model postnatal growth deceleration. Typically these and other anthropometric models are either specific to a particular growth measurement or age range and therefore are not widely applicable, and, most importantly, their parameters do not have an intrinsic biological interpretation. Moreover their fitting of separate individual curves is inefficient. An alternative approach that partly overcomes the latter problem consists of fitting whatever model is selected not separately to each child but simultaneously, using random effects (*i.e.* mixed) models, Verbeke and Malenberghs (2000). Within this framework child-specific parameters are assumed to be drawn from some distribution. Applications include a random effects specification of the JB model used to study the height trajectories of children with and without Turner syndrome, van Dommelen et al. (2005), and two specifications of the Reed model used to analyze growth in weight during the first years of life in an Ethiopian and a Finnish birth cohort, Asefa et al. (1996); Tzoulaki et al. (2010). General specifications of random effects models have also been used to analyse growth data, Verbeke and Malenberghs (2000). Linear and higher degree polynomial functions of time are specified within this framework, dos Santos Silva et al. (2002); Yang et al. (2011), with random effects generally included only for some of the parameters because of identification constraints, Goldstein (2011). However these models can only produce a limited range of possible shapes, which are not usually sufficient to describe infant growth. Furthermore growth patterns may not all cluster around a single population average polynomial curve, therefore leading to poor fit to the data.

For this reason the implementation of spline-based models within the mixed effects framework has raised considerable interest because of their flexibility in fitting different anthropometric variables over different age ranges and the possibility to achieve a satisfactory balance between goodness of fit and smoothness of the growth curve by choosing the number of knots, Ruppert et al. (2003). Among these models, the shape invariant random effects model proposed by Beath in 2007 is particularly relevant, Beath (2007). It was originally used to model infant growth in weight, Beath (2007), but has been used by Cole to analyse height growth during puberty, Cole et al. (2010), and more recently by Hui et al. (Hui et al. (2010)) and Johnson et al. (Johnson et al. (2011)) to study factors influencing infant growth. In this model a common spline function is modified by shifting of the two axes and scaling of the x-axis to fit the individual data. The scaling and shifting identify how each child departs from the common spline function and are captured by parameters which have a direct biological

interpretation, Beath (2007); Cole et al. (2010), unlike any of the methods mentioned above. The application of the shape invariant random effects model however requires fairly advanced computing skills. Its interpretation also requires careful understanding of the parameterization.

The aim of this paper is to compare the two models purposely developed for anthropometric data, JB and Reed, both specified within a random effects framework, with the shape invariant random effects model in terms of their ability to extract salient and biologically meaningful features of growth in infancy and also in their ability to validly represent the data. Comparisons will be carried out using longitudinal weight data from three recent birth cohort studies. The information available from these cohorts differs in term of age range and number and timing of follow-up observations (regular/irregular), and therefore illustrates a variety of settings likely to be encountered in practice.

2 Data

We have analysed data from three contemporary birth cohorts: the Southern European web-based "Nascita e INFanzia gli Effetti dell'Ambiente" (NINFEA) and "Geracão XXI" (GXXI) birth cohort studies, and the "Growth and Obesity Chilean Cohort Study" (GOCS).

2.1 Nascita e INFanzia gli Effetti dell'Ambiente

The NINFEA study is an Italian web-based cohort study which started in 2005 and aimed to recruit pregnant women via the Internet and follow up their children. The study is still ongoing and targets pregnant women, who have access to the Internet and sufficient knowledge of the Italian language to understand the questionnaires. They are recruited via a wide range of advertising campaigns (for more details see Richiardi et al. (2007)). Registration is carried out at the study website (www.progettoninfea.it) where participants complete the first questionnaire (Q1). The cohort is then actively followed-up via other online questionnaires administered at around 6 (Q2), 18 (Q3) and 48 (Q4) months of age of the child.

Revisions of the questionnaires were undertaken after enrollment of the first 1,500 participants, thus the available data vary with year of recruitment. In particular, women who enrolled until November 2008 were asked to report the child's weight at birth, 3 and 6 months at the time of the second questionnaire (Q2), while at Q3 they were asked to report the weight of the baby at 12 and 18 months. When the Q2 and Q3 questionnaires were updated, a new question was introduced regarding the child's weight at the actual time of completion of these questionnaires. In contrast Q4 was set up from the outset to include questions on anthropometric measures both at 4 years of age and at the time of completion of the questionnaire.

Because of these variations in questionnaire design the data used for these analyses include all singleton children who, at the time of the data download (November 2011), were eligible for completion of Q4. These are the children for whom growth

measurements are available at fixed time points only (0, 3, 6, 12, 18 and 48 months). This sample includes 845 children.

2.2 Geracão XXI

The GXXI study is the first Portuguese birth cohort, established in 2005 in the Porto region. All live children born of women resident in one of the six regional Districts, admitted in one of the five hospitals of Porto for delivery with a gestational age at birth greater than 24 weeks, were eligible to participate. The recruitment period lasted from the end of April 2005 until August 2006. Women were enrolled during an appointment a few days before their due date and, for the majority, baseline questionnaires were completed between 24 and 72 hours after delivery. In total the baseline data consist of 8,311 singleton children. Children were actively followed-up through interviewer-administered questionnaires planned at 3, 6, 12/15 and 24 months. Due to logistic and financial limitations it was not possible to interview every participant at each follow-up visit. Therefore a restricted time period was allocated for each follow-up occasion. At the 2 years follow-up visit growth data from the child's health records card, with measures obtained prospectively by health professional, were gathered from the parents for entry into the database.

The data used for this analyses consist of the information from the baseline and the 2-years follow-up, available for 783 singleton babies. These include anthropometric measures reported in the child's health record and referring to measures taken at about 1, 2, 4, 6, 9, 12, 15, 18 and 24 months of age, together with the actual dates of measurement. The values at 24 months were obtained directly by the interviewers. Up to 6 additional measurements and dates reported in the health records were also entered into the database. There are therefore up to 16 weight measurements per child, taken from birth to (about) 24 months.

2.3 Growth and Obesity Chilean Cohort Study

GOCS is an ongoing Chilean cohort aiming to study the association of early growth with children's maturation, adiposity and associated metabolic complications. The study was initiated in 2006 when all children aged 2.6-4 years attending public nursery schools in six counties of Santiago were invited to participate if they: (i) were singleton births with a gestational age at birth between 37 and 42 weeks, (ii) had a birth weight between 2500 and 4500 grams, and (iii) had no physical and psychological conditions that could affect growth (6 children excluded due to these conditions). Among the 1,498 children eligible to participate, 1,195 accepted (80%). For almost all subjects, weight and height measurements from birth up to 36 months of life were retrospectively gathered from health records, and they were prospectively measured every year after recruitment by a dietitian who visited the nursery school.

For these analyses we only used growth data up to around age 4 years for direct comparison with the NINFEA cohort. This leads to including a maximum of 11 measurements per child. After exclusion of subjects with missing growth data, the sample used in these analyses consists of 1,149 children.

3 Methods

In this paper we compare two classes of anthropometric random effects models, the JB model and the four-parameter Reed model, and the shape invariant random effect model. Below we define these models and the specific issues arising from fitting them to the data.

3.1 The Jenss-Bayley model

The original specification

The JB model, Jenss and Bayley (1937), has been widely applied (see van Dommelen et al. (2005); Deming and Washburn (1963); Manwani and Agarwal (1973); Berkey (1982)) to describe childhood growth during the first six years of life in term of both height and weight. According to this model, the observed growth variable $y_i(t)$ of child i at age t, for $i = 1, ..., n$ and $t = 1, ..., T_i$, can be expressed as

$$y_i(t) = c_i + d_i t - e^{a_i + b_i t} + \varepsilon_{it} \tag{1}$$

where a_i, b_i, c_i and d_i are unknown parameters to be estimated separately for each child and ε_{it} is the error term at age t specific to child i, that is assumed to be normally distributed with mean zero and variance σ_i^2. Equation (1) defines $y_i(t)$ as a negatively accelerated exponential curve in t, whose asymptote is a positive straight line. Hence the model accounts for the rapid decelerating growth rate usually observed during infancy via the exponential component $e^{a_i + b_i t}$, while c_i and d_i represent the intercept and the slope of the asymptote, respectively. From equation (1) it follows that the predicted size at birth for child i is $(c_i - e^{a_i})$.

The random effects specification

Instead of modelling each child separately, we can add some distributional assumptions for the child-specific parameters, and to their relation with the residual errors ε_{it}, in order to define an overall model for all children. Using the same notation as in equation (1) we now specify the child-specific parameters as

$$\begin{aligned} a_i &= a_0 + a_{1i} \\ b_i &= b_0 + b_{1i} \\ c_i &= c_0 + c_{1i} \\ d_i &= d_0 + d_{1i} \end{aligned}$$

where a_0 represents the fixed effect and a_{1i} the child-specific random effect, with similar definitions for the components of b_i, c_i and d_i. We also assume that a_{1i}, b_{1i}, c_{1i} and d_{1i} are drawn from a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix Φ and that the errors ε_{it} are independent normally distributed random variables with mean zero and constant variance σ^2, which are independent of the

child-specific random effects. The curve implied by the model when all random effects are set to 0 (the "population-level curve", Pinheiro and Bates (2000)) is therefore $c_0 + d_0 t - e^{a_0 + b_0 t}$. Note that this model does not allow for an inflection point (maximum/minimum in the weight velocity curve), and therefore no APWV, because its second derivative with respect to t is an exponential function in t.

3.2 The Reed model

The original specification

The Reed model was suggested by Berkey and Reed (Berkey and Reed (1987)), as an extension to an earlier model suggested by Count (Count (1943)). It has two versions, with 4 and 5 parameters, respectively. For comparison with the JB model we focus on the 4-parameter version, which has been shown not to be inferior in term of goodness of fit to the 5-parameter one, Simondon et al. (1992). The 4-parameter specification of the Reed model is

$$y_i(t) = a'_i + b'_i t + c'_i ln(t) + \frac{d'_i}{t} + \varepsilon'_{it} \tag{2}$$

where again, $y_i(t)$ is the ith child's growth variable at age t, and a'_i, b'_i, c'_i and d'_i are the child-specific parameters and ε'_{it} is the child- and age-specific error term, assumed to be normal with mean zero and variance σ'^2_i. This specification, unlike that for the JB model, is linear in its parameters and can accommodate one inflection point. From equation (2) it follows that c'_i and d'_i are deceleration terms moderating the original increase captured by b'_i.

The random effects specification

As for the JB, we can specify the Reed model using a random effects approach. Again we define the child-specific parameters as

$$\begin{aligned} a'_i &= a'_0 + a'_{1i} \\ b'_i &= b'_0 + b'_{1i} \\ c'_i &= c'_0 + c'_{1i} \\ d'_i &= d'_0 + d'_{1i} \end{aligned}$$

with, a'_0 representing the fixed effect and a'_{1i} the child-specific random effects, and similarly for the other parameters. The parameters a'_{1i}, b'_{1i}, c'_{1i} and d'_{1i} are assumed to be drawn from a multivariate normal distribution with mean **0** and covariance matrix Φ'. As before the error terms ε'_{it} are assumed to be independent normally distributed random variables with mean zero, constant variance σ'^2, and to be independent of the random effects. This model can have an inflection point and therefore the APWV can be derived by differentiating the weight velocity curve (first derivative of the fitted curve) and setting it to zero.

3.3 The shape invariant random effects model

The shape invariant random effects model was introduced by Beath (Beath (2007)) to describe infant weight growth and by Cole (Cole et al. (2010)) to analyse pubertal growth in height. Cole used a slightly modified parametrization, which is the one adopted here, and named the model "SuperImposition by Translation And Rotation" (SITAR). It is a model where a common spline function for all subjects is modified by shifting the two axes and scaling the x-axis in order to adapt it to the individual trajectories. This model is not specific to an age range or to an anthropometric dimension. The use of a spline function allows this model to deal naturally with non-linearity and therefore to identify inflection points. In this application a natural cubic spline with a B-spline basis matrix is used to fit the data using a non linear random effects model, where the coefficients of the spline function are treated as fixed effects while the parameters of the shape invariant model are treated as random. Formally, let $y_i(t)$ be again the growth dimension of the ith child at age t, then the SITAR model is specified as:

$$y_i(t) = \alpha_i + h\left(\frac{t-\beta_i}{e^{-\gamma_i}}\right) + \eta_{it} \tag{3}$$

where $h(z)$ is the natural cubic spline curve of the growth variable regressed on z (the transformed age) and $\alpha_i = \alpha_0 + \alpha_{1i}$ is a subject-specific coefficient with α_0 representing the fixed effect and α_{1i} the random effect, with similar definitions for the other two parameters. The three random effects α_{1i}, β_{1i} and γ_{1i} are assumed to be drawn from a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix Λ. As before the error terms η_{it} are assumed to be independent normally distributed random variables with mean zero, constant variance τ^2, and to be independent of the random effects. Weight velocity curves can easily be derived by differentiating the spline curve, and, from these, the APWV can be derived.

The specific random effects α_{1i}, β_{1i} and γ_{1i} for subject i correspond to the shift parameter for the y-axis (measure), and the shift and the scale parameter for the x-axis (age), respectively. Cole (Cole et al. (2010)) refers to these three parameters as *size*, *tempo* and *velocity*, where the effect of the first two lead to a translation (in measure and age), while that of γ to a rotation. *Size* is expressed in the units of the y variable, *tempo* in the units of the t variable, while *velocity* is a multiplier, and therefore is scale-free. Therefore, from a biological perspective, when analysing weight in infancy α_{1i} captures differences in size (greater for heavier children), β_{1i} the timing of growth (negative for babies with an earlier APWV) and γ_{1i} the growth rate (positive for children with steeper growth). When these subject-specific random effects are set equal to 0, again the "population-level curve" is obtained.

3.4 Growth models for weight in early life

We fitted the random effects specifications of the JB and Reed models and of the SITAR model to the weight data from the three cohorts described above, initially separately by sex. We used all the available data (for ages 0-4 years in NINFEA

and GOCS, and for ages 0-2 years in GXXI) as well as the NINFEA and GOCS data restricted to ages 0-2 years for comparison with GXXI. In each set of analyses all subjects with at least one observation were included under the assumption that missing visits occurred at random, Rubin (1976).

The JB and Reed models were both originally conceived for an untransformed growth measure. Despite this, natural logarithmic transformations of the observed weight measures were considered when fitting all models and results compared with those obtained when weight was not transformed.

A transformation of the time scale was needed for the Reed model because our data include measures at birth, i.e. at time 0. One option, suggested by Berkey and Reed themselves, is to replace age since birth (in months) t, with t^* where $t^* = \frac{t+9}{9}$, so that $t^* = 0$ at around conception and $t^* = 1$ at birth. With this transformation the size at birth for child i predicted by the Reed model is $a'_i + b'_i + d'_i$. The same transformation was also considered when fitting the SITAR model when investigating its best specification for the data.

The SITAR model was fitted using a natural cubic spline with B-spline basis matrix, placing the internal knots of the spline $h(t)$ at the quantiles of the age distribution, as in Cole at al. (Cole et al. (2010)), according to the number of degrees of freedom specified. Degrees of freedom were chosen according to the richness of available weight measurements over the age time scale. Because of identification issues, alternative constraints on the values of the fixed effects β_0 and γ_0 were considered. In particular we considered either fixing both β_0 and γ_0, or just β_0 or just γ_0 to be equal to zero. To aid interpretation, models were also refitted with the males and females combined.

Estimation of all models was carried out by maximum likelihood estimation implemented in the `nlme` function in R, Lindstrom and Bates (1990). The cubic spline function with B-spline basis matrix was fitted using the `ns` function also in R.

3.5 Comparison of alternative models

Specifications of the SITAR model with different weight and age scales first, and then with alternative constraints on the fixed and random effects, were fitted on the data from all three cohorts, and then compared in terms of Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Plots of predicted population-level curves, and of individual growth and growth velocity curves for a selection of children were also examined to assess their closeness with the observed data. Specifications of the random effect JB model obtained with/without log-transformation of the weight scale were compared using the same criteria; likewise for the random effects Reed model. When the weight scale was transformed, adjusted deviance was calculated, Box and Cox (1964), and corrected AIC and BIC derived.

For each dataset the goodness of fit of the best specifications of each model were then compared again using the same criteria. Moreover, child-specific APWVs were estimated, when possible, and their averages compared with published data, Tzoulaki et al. (2010); Johnson et al. (2011).

For every model and every dataset, the distribution of the estimated error terms was examined graphically using normal quantile plots.

Only a selection of the comparisons and of the plots described above are reported below.

4 Results

4.1 Data description

Table 1 summarizes the baseline characteristics of the children included in the analyses, by cohort. It also includes descriptors of weight and weight change at selected ages. GXXI is the richest cohort in terms of weight measurements gathered, with a median of 10 observations per child within the first 2 years of life, compared to medians of 8 in GOCS and of 5 in NINFEA where growth measures were reported at exactly 0, 3, 6, 12, 18 and 48 months. Mean/median birth weight, as well as weight at two and six months, are slightly higher in GOCS compared to the other two cohorts possibly because of its exclusion of preterm births (gestational age < 37 weeks). This is also reflected by the average weight change in the first months of life, which is again slightly higher in GOCS (1.0 kg/month) than in GXXI (0.9 kg/month) and NINFEA (0.8 kg/month, in the first three months). The lower rates in the European cohorts are probably due to pre-term children not having yet completed their catch-up period by 2/3 months, as opposed to the term children of GOCS. Indeed the rates in weight change are identical across the three cohorts when evaluated from birth to six months. Mean weight and corresponding SD at 18 months are similar across the three cohorts, while GOCS children at 4 years are on average slightly heavier than those in NINFEA (we have no GXXI measures at 4 years). All of these comparisons are however based on different subsets of children at each time point because not all children in each cohort have measures at all times considered.

Plots of weight trajectories for a random selection of children (dark lines for males and lighter lines for females), superimposed over the scatter of data for their full cohort, are shown in Figure 1 (where the cohort differences in terms of frequency and timing of the observations is also evidenced). Each child selected from each cohort represents one of six strata, defined by gender and birth weight category (low/middle/high, defined with cut-off points at 2.5 and 3.8 kg). These plots clearly show how weight growth is faster during the first 3-6 months of life, and then starts to decelerate, especially after the first year of birth. As expected males are generally slightly heavier than females in each birth weight category. This figure highlights specific growth profiles: (i) children with a low weight at birth who remain small (relatively to the other same-sex children); (ii) children with a low weight at birth who experience a high postnatal rate of growth (e.g the low birth weight male in GXXI); (iii) those with a high birth weight who remain constantly heavier compared to the others (e.g the high birth weight male in GOCS); (iv) those with an average weight at each observational time point (e.g the medium birth weight male in GXXI); and finally (v) children with a high birth weight who experience a greater deceleration after the first months of life and thus return to an average weight (e.g the high birth weight male in NINFEA).

	GXXI (n=783)					GOCS (n=1,149)					NINFEA (n=845)				
	N	*Range*	*Median*	*Mean*	*SD*	*N*	*Range*	*Median*	*Mean*	*SD*	*N*	*Range*	*Median*	*Mean*	*SD*
Age range		**0-2 years**					**0-4 years**					**0-4 years**			
Number of measures per child	783	1-16	10	9.4	1.9	1,149	2-11	8	7.6	1.9	845	1-6	5	5.0	1.2
Gestational age (weeks)	748	28-43	39.5	39.2	1.6	959	37-42	39.5	39.6	1.3	837	26-43	39.9	39.7	1.7
Preterm births (<37 weeks)	61	8.2%	-	-	-	0	0.0%	-	-	-	40	4.8%	-	-	-
Female	373	49.3%	-	-	-	579	50.4%	-	-	-	399	47.3%	-	-	-
Birth weight (kg)	756	0.6-4.7	3.2	3.2	0.5	1,148	2.5-4.8	3.4	3.4	0.4	793	0.8-4.8	3.2	3.2	0.5
Weight at 2/3 months[†] (kg)	573	2.5-8.1	5.0	5.0	0.7	800	3.4-7.9	5.4	5.4	0.6	662	2.3-8.7	5.7	5.7	0.8
Weight at 6 months (kg)	522	4.2-12.8	7.6	7.6	0.9	747	5.5-11.2	7.8	7.9	0.9	664	4.6-11.5	7.6	7.6	1.0
Growth rate 0-2/3 months[†] (kg/month)	553	0.3-1.8	0.9	0.9	0.2	799	0.1-2.0	1.0	1.0	0.3	658	0.1-1.8	0.8	0.8	0.2
Growth rate 0-6 months (kg/month)	506	0.4-1.4	0.7	0.7	0.1	746	0.3-1.3	0.7	0.7	0.1	661	0.3-1.4	0.7	0.7	0.1
Weight at 18 months[‡] (kg)	291	7.9-15.7	11.3	11.2	1.3	729	7.5-16.2	11.1	11.3	1.3	752	8.0-16.0	11.2	11.4	1.3
Weight at 48 months[§] (kg)	-	-	-	-	-	136	12.7-31.3	17.2	17.5	2.6	620	11.0-30.0	16.0	16.5	2.3

[†] *Weight data around 2 months are available for GOCS and GXXI, while weight data at 3 months are available for NINFEA*

[‡] *For GOCS and GXXI studies, children with weight measures collected between 17.5 and 18.5 months are included*

[§] *For GOCS study, children with weight measures collected between 47 and 49 months are included*

Table 1: Baseline characteristics and weight and weight changes at selected ages, by cohort

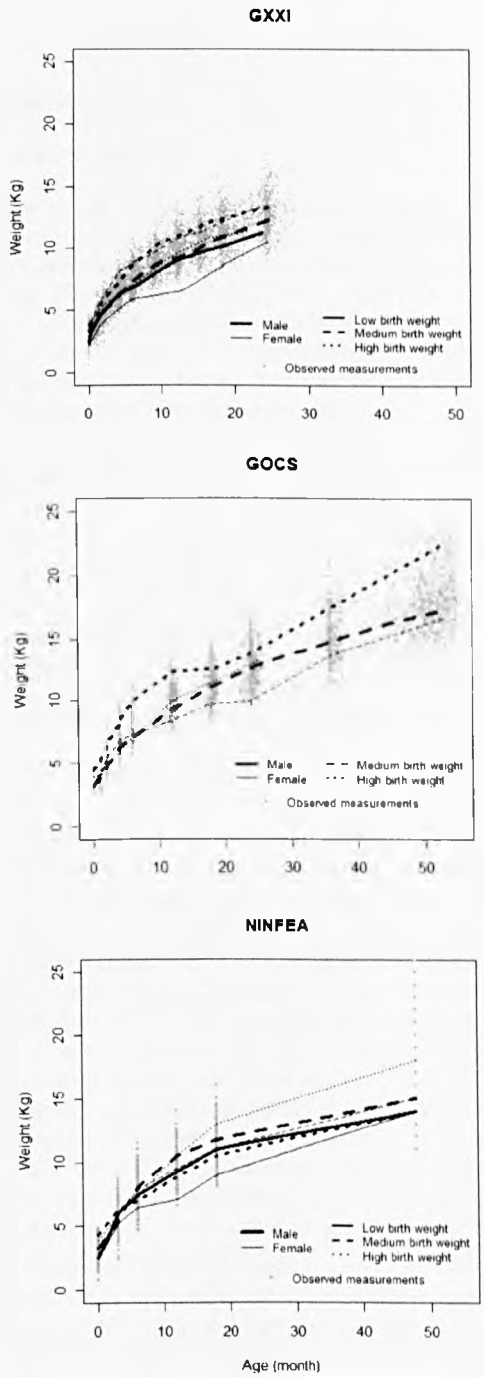


Figure 1: Observed weight growth curves of a stratified random selection of children by cohort, superimposed on the observed data (by gender and birth weight category)

4.2 Growth modelling

Jenss-Bayley and Reed models

Log-transforming the weight scale did not generally improve the fit of the two anthropometric models according to the fit criteria. Therefore only results obtained on the original scale are reported in Table 2, separately by age, gender and cohort. The JB model failed when fitted on the NINFEA 0-2 years data; this is probably due to the small number of measurement times available (see Figure 1), the restricted age range analysed (since the JB model was suggested to describe growth between birth and six years) and the measurement error that may be affecting the NINFEA recalled measurements more than those from the other cohorts.

The estimates of the JB fixed effects were quite stable across cohorts, age and gender, with $\hat{a}_0$ and $\hat{c}_0$ slightly lower in females than males. These estimated parameters were also slightly lower in both GOCS males and females than in the other cohorts, as were the estimates of b_0. The corresponding SDs of the random effects were similar across cohorts and gender, and lower for the 0-4 years data compared to the 0-2.

In agreement with the JB model, the Reed fixed effect estimates differed in GOCS, with all $\hat{a}'_0$, $\hat{b}'_0$, $\hat{c}'_0$ and $\hat{d}'_0$ considerably larger in absolute terms compared to NINFEA and GXXI. The random effects SDs for this model were extremely large because of their high correlations (see below). However they were lower when estimated on the 0-4 years data, as were the random effects correlations.

SITAR models

As described in the methods section, we compared alternative specifications of the SITAR model, by allowing different scales for the dependent variable, weight, and the age time scale, different numbers of knots for the spline function, and by including constraints on the values of the fixed effects.

Models with log-transformed weight always performed better in terms of AIC and BIC than those with untransformed weight; therefore only results from the former are reported. Instead models with log-transformed age performed better only in some settings and hence both sets of results (models with age and log-trasformed age) are reported. Four internal knots were used when analysing the GXXI and GOCS studies and three when analysing the NINFEA cohort, because of its fewer time occasions. Results were not affected when the number of internal knots for GXXI and GOCS was increased to 5 or 6. Alternative constraints on the fixed effects were examined to address identification issues. Among the various combinations fixing either β_0 and γ_0 (Model 1), or just β_0 (Model 2), to be equal to zero led to the best fitting models.

The results are shown in Table 3. According to AIC and BIC, Model 2 performed slightly better than Model 1 in most of the combinations; moreover the estimates of the APWV were more stable when obtained from Model 2. As regards the choice of the age scale, the models fitted best on the untransformed age scale in most cases.

The estimated SDs of the three random effects α_{1i}, β_{1i} and γ_{1i}, did not vary substantially across genders, nor across cohorts and age ranges with the exception of

		Jenss-Bayley								Reed							
		Fixed effects				Random effects (SD)				Fixed effects				Random effects (SD)			
		a_0	b_0	c_0	d_0	a_{1i}	b_{1i}	c_{1i}	d_{1i}	a'_0	b'_0	c'_0	d'_0	a'_{1i}	b'_{1i}	c'_{1i}	d'_{1i}
0-2 years																	
GXXI	Male	1.70	-0.19	8.88	0.17	0.37	0.06	2.08	0.07	12.52	0.85	-0.08†	-10.25	9.78	3.65	14.81	13.08
	Female	1.59	-0.17	8.19	0.18	0.42	0.07	1.84	0.08	9.72	0.79	1.33†	-7.49	9.48	3.84	14.74	12.89
GOCS	Male	1.52	-0.27	8.11	0.20	0.36	0.11	1.81	0.07	19.80	3.29	-10.47	-19.74	11.19	4.10	16.57	14.92
	Female	1.45	-0.22	7.70	0.19	0.36	0.08	1.61	0.07	16.87	3.20	-8.96	-16.77	10.66	3.86	15.58	14.12
NINFEA	Male									10.95	0.66†	1.48†	-8.28	13.85	5.79	22.28	19.22
	Female									6.76	-0.41†	5.87	-3.22	11.80	5.55	20.57	16.90
0-4 years																	
GOCS	Male	1.55	-0.25	8.18	0.20	0.26	0.07	1.35	0.05	18.38	2.69	-8.15	-17.71	6.65	1.52	7.54	7.86
	Female	1.40	-0.24	7.49	0.21	0.29	0.08	1.27	0.05	15.40	2.63	-6.66	-14.71	6.63	1.55	7.56	7.88
NINFEA	Male	1.76	-0.17	9.33	0.16	0.32	0.06	1.94	0.07	11.39	0.91	0.61†	-8.97	8.34	2.02	9.98	10.04
	Female	1.50	-0.18	7.98	0.17	0.33	0.05	1.58	0.06	9.24	0.96	1.08†	-7.08	6.21	1.63	7.66	7.46

† *Not statistically significant at the 5% level*

Table 2: Estimated fixed effects and standard deviations of the random effects of the Jenss-Bayley and Reed models fitted on the original weight scale and stratified by cohort, gender and age range

Scales used for y and x			log(Kg) and month						log(Kg) and log(t)					
			Random effects (SD)			APWV	AIC	BIC	Random effects (SD)			APWV	AIC	BIC
			α	δ	γ	*(month)*			α	δ	γ	*(month)*		
								MALES						
	0-2 years													
GXXI		Model 1	0.18	0.72	0.32	1.53	**5332**	**5407**	0.23	0.09	0.29	1.16	5503	5577
		Model 2	0.18	0.71	0.32	1.65	5333	5411	0.22	0.09	0.28	1.50	5479	5560
GOCS		Model 1	0.17	0.63	0.35	0.78	6120	6193	0.19	0.08	0.22	0.73	6277	6350
		Model 2	0.16	0.60	0.30	1.29	6102	6182	0.12	0.06	0.15	1.10	**6054**	**6134**
NINFEA		Model 1	0.14	0.70	0.26	2.58	4117	4178	0.22	0.10	0.31	1.51	4129	4190
		Model 2	0.15	0.73	0.30	1.98	**4044**	**4110**	0.16	0.08	0.21	1.97	4203	4270
	0-4 years													
GOCS		Model 1	0.18	0.70	0.34	1.98	9306	9383	0.13	0.06	0.10	1.01	9051	9127
		Model 2	0.15	0.64	0.26	1.10	**9031**	**9114**	0.11	0.06	0.10	0.95	9039	9122
NINFEA		Model 1	0.15	0.72	0.31	2.64	5398	5460	0.22	0.11	0.25	1.17	5394	5456
		Model 2	0.17	0.77	0.36	1.83	**5304**	**5372**	0.12	0.07	0.11	1.75	5458	5526
								FEMALES						
	0-2 years													
GXXI		Model 1	0.18	0.72	0.35	1.62	**4716**	**4790**	0.24	0.10	0.29	1.29	1833	1907
		Model 2	0.18	0.71	0.34	1.90	1717	1797	0.23	0.10	0.31	1.62	1794	1874
GOCS		Model 1	0.19	0.71	0.37	0.72	6029	6103	0.24	0.10	0.25	0.55	6065	6140
		Model 2	0.17	0.70	0.32	1.29	**5813**	**5893**	0.13	0.07	0.15	1.28	5936	6017
NINFEA		Model 1	0.14	0.71	0.24	2.61	3372	3432	0.17	0.08	0.25	1.66	**3322**	**3382**
		Model 2	0.15	0.77	0.28	2.04	3360	3425	0.15	0.07	0.23	1.96	3375	3111
	0-4 years													
GOCS		Model 1	0.18	0.71	0.30	1.87	9252	9320	0.12	0.06	0.09	0.93	8963	9040
		Model 2	0.13	0.60	0.20	0.91	9075	9159	0.13	0.06	0.09	0.80	**8957**	**9040**
NINFEA		Model 1	0.17	0.80	0.31	2.61	1627	1680	0.22	0.11	0.22	1.11	1660	1722
		Model 2	0.19	0.89	0.34	2.11	**4536**	**4603**	0.23	0.11	0.21	0.98	1684	1751

SD = Standard Deviation; APWV = Average age at peak weight velocity, AIC = Akaike information criterion, BIC = Bayesian information criterion

Model 1: $\phi_0 = 0$ and $\gamma_0 = 0$; Model 2: $\phi_0 \neq 0$

Within each stratum defined by study, gender and age range the lowest AIC and BIC are reported in bold

§ $t = (month + 9)/9$

Table 3: Results from alternative specifications of the SITAR model stratified by cohort, gender and age range

the estimated SD of β_{1i}, that was generally lower in GOCS than in the other cohorts, probably because of its inclusion criteria (Table 3).

4.3 Models comparison

Goodness of fit

Results from the specification of the SITAR model which best fitted each dataset were compared with those obtained from the JB and Reed models in term of AIC and BIC (Table 4). The two information criteria were always in agreement and show that the Reed model fit the data best, especially among females. The JB model gave the worse fit and failed when fitted on 0-2 years NINFEA.

Examinations of the normal quantile plots of the residual errors estimated for each model showed that the assumption of normality appeared to be better satisfied by the SITAR model than the two anthropometric models (data not shown). This was seen for each dataset, although even the SITAR estimated residuals showed some kurtosis (also reported by Beath at al. (Beath (2007))).

	Jenss-Bayley		Reed		SITAR	
	AIC	*BIC*	*AIC*	*BIC*	*AIC*	*BIC*
			MALES			
0-2 years						
GXXI †	5406	5499	5217	5310	5332	5407
GOCS ‡	6047	6139	5889	5981	6054	6134
NINFEA§			3869	3952	4044	4110
0-4 years						
GOCS §	8949	9045	8813	8908	9031	9114
NINFEA§	5484	5570	5143	5228	5304	5372
			FEMALES			
0-2 years						
GXXI †	4929	5022	4592	4685	4716	4790
GOCS §	5914	6007	5681	5774	5813	5893
NINFEA§§			3253	3335	3322	3382
0-4 years						
GOCS ‡	9124	9220	8846	8942	8957	9040
NINFEA§	4569	4654	4486	4571	4536	4603

AIC= Akaike information criterion; BIC= Bayesian information criterion

† *SITAR: Model 1, log(Kg) and month scales*

‡ *SITAR: Model 2, log(Kg) and log(t) scales*

§ *SITAR: Model 2, log(Kg) and month scales*

§§ *SITAR: Model 1, log(Kg) and log(t) scales*

Table 4: Goodness of fit of the best specification of the three models by cohort, gender and age range

Population-level predicted curves

The population-level predicted curves from the best specification of the three models fitted on the GXXI data (0-2 years) are shown in Figure 2, stratified by gender. Note that in all figures curves predicted by the SITAR models fitted on the log-weight scale have been back-transformed so that weights are shown on the same scale for all the three models. As expected the predicted curves for males lie above those for the females, regardless of the models used, with the difference increasing from birth till about 1 year and becoming approximately constant after that. The curve predicted by the JB model differs from those of the other two models, especially during the first months. In contrast the curves predicted by the Reed and the SITAR models overlap up to about 12-15 months before diverging slightly. These differences do not necessarily reflect differences in goodness of fit, but rather in the shape of their respective predicted curves when all random effects are set at zero. For this reason the curves predicted by the three models for the randomly selected GXXI females of Figure 1 are compared in Figure 3. They show how the Reed and SITAR model most closely interpolate the original data in line with the results from the information criteria.

Further comparisons were made across the three cohorts. Figure 4 shows the predicted population curves and predicted weight velocities obtained by the Reed and SITAR models on the 0-2 years males data of the three cohorts (with the SITAR model fitted using Model 2 specification on the log weight and age scales; we do not show the results for JB because it failed on the 0-2 years NINFEA data). Note that, even if the

Figure 2: Population predicted weight curves from the best specification of the three models fitted on the 0-2 years GXXI data by gender

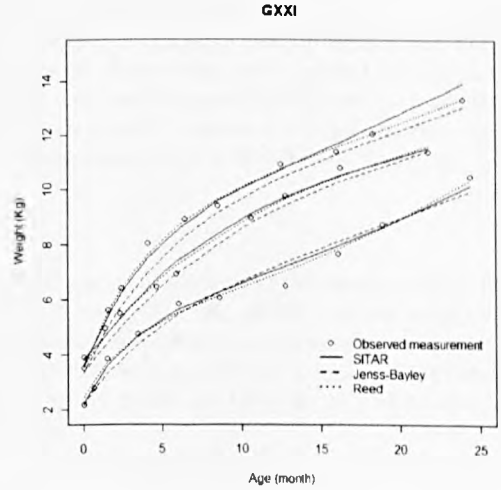


Figure 3: Predicted weight growth curves of a stratified random selection of GXXI females, superimposed on the observed data

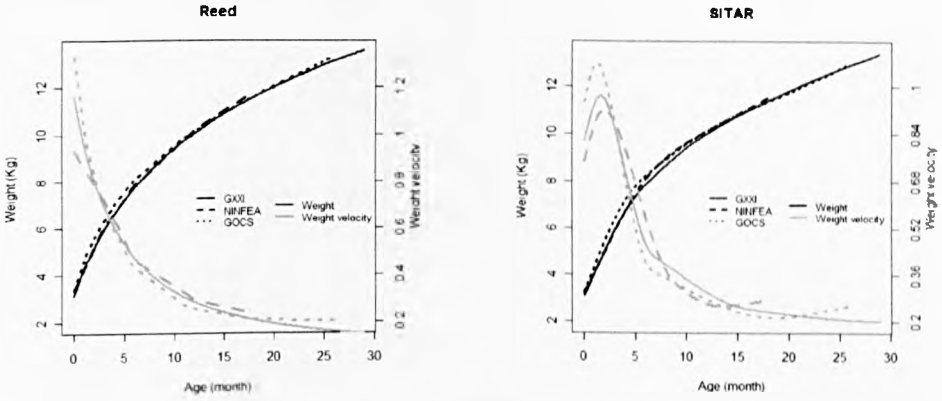


Figure 4: Population predicted weight and velocity growth curves from the Reed and the SITAR Model 2 log-weight and age models fitted on 0-2 years males from the three cohorts

4-parameter Reed model in general allows for an inflection point, our transformation of the time scale, which we used to include measures taken at birth, did not lead to an estimation of a maximum for the growth velocities. According to both models, the typical trajectory of GOCS males (i.e. the trajectory of a child with zero random effects) is predicted to be slightly heavier than those from the other cohorts especially during the first year of life, possibly reflecting again the inclusion criteria used in this study. This is also seen in the velocity curves obtained from the SITAR model, with that for GOCS boys having their peak at a slightly earlier age compared to the rest (see also the predicted APWV reported in Table 3), and reaching a much higher size at that peak. The predicted APWV for NINFEA are larger compared to those of the other cohorts, but may be biased upwards because its growth data are collected only at birth, 3 and 6 months. To validate this we refitted the SITAR model on GXXI males after excluding all observations in the period 0.1-2.9 months to resemble the data in NINFEA: the new predictions of APWV were 2.17 months in males (instead of 1.65) and 2.13 months in females (instead of 1.90), becoming similar to those obtained in NINFEA (1.98 in males and 2.04 in females, see Table 3).

Random effects

Table 5 reports the SDs and correlations of the random effects from the three models when fitted on GOCS males (SITAR: Model 2 on log weight and age scales). The results obtained from the other cohorts and for females are similar and are not reported for simplicity. The table shows that the correlations of the JB and Reed random effects are extremely high (especially for the Reed model) while those of the SITAR model are substantially lower.

	Jenss-Bayley				Reed				SITAR		
	SD	Correlations			SD	Correlations			SD	Correlations	
0-2 years											
a (a') (α)	0.36	a	b	c	11.18	a'	b'	c'	0.16	α	β
b (b') (β)	0.11	0.91			4.1	0.91			0.60	0.76	
c (c') (γ)	1.81	0.99	0.90		16.57	-0.96	-0.98		0.30	-0.77	-0.70
d (d')	0.07	-0.73	-0.74	-0.66	14.92	-0.99	-0.95	0.99			
0-4 years											
a (a') (α)	0.26	a	b	c	6.65	a'	b'	c'	0.15	α	β
b (b') (β)	0.07	0.82			1.52	0.85			0.64	0.74	
c (c') (γ)	1.35	0.98	0.70		7.54	-0.95	-0.95		0.26	-0.73	-0.66
d (d')	0.05	-0.49	-0.52	-0.40	7.86	-0.99	-0.90	0.98			

Table 5: Standard deviations (SD) and correlations of the random effects estimated by the three models fitted on GOCS males by age range

4.4 Model interpretation

Random effects

To aid the interpretation of the parameters of the three models, we re-fitted them on the combined males and females data, separately by cohort, and then compared the gender-specific distributions of their predicted random effects (BLUPs, Pinheiro and Bates (2000)). In Table 6 we report only the descriptive results obtained on the 0-2 years data separately by cohort. In each cohort the mean of the JB predicted random effects c_{1i} and a_{1i}, corresponding to the intercept of the asymptote and the intercept for the exponential term respectively, were higher in males than females. However b_{1i} and d_{1i} were not different. The Reed model random effects that most differed in terms of gender were a'_{1i}, which was on average higher among males, and c'_{1i} and d'_{1i}, the two deceleration terms, on average higher among females.

Results from the SITAR model varied across cohorts: while on average the predicted growth velocity (γ_{1i}) were higher among males in each cohort, size (α_{1i}) was higher in males in NINFEA and GOCS but not in GXXI, and tempo (β_{1i}) was higher in females only in GXXI (and to a lesser extent in NINFEA), in agreement with the corresponding differences in predicted APWV (see Table 2).

The SITAR model parameters

Both Beath (Beath (2007)) and Cole (Cole et al. (2010)) stress that a biological interpretation could be attached to the coefficients derived from the SITAR model. To clarify this we examined the model predicted trajectories and random effects for a selection of children and compared them to their respective population predicted curve. Here we report the findings for a GOCS male who had a very high birth weight (4.55 kg) and who was still considerably overweight at 24 months. His predicted random coefficients are $\hat{\alpha}_{1j} = 0.44$ (the maximum of the distribution of $\hat{\alpha}_{1i}$), $\hat{\beta}_{1j} = 0.88$ ($\simeq$95th percentile of the distribution of $\hat{\beta}_{1i}$) and $\hat{\gamma}_{1j} = -0.51$ ($\simeq$5th percentile of the $\hat{\gamma}_{1i}$ distribution). While we expected the values of the first two predicted random

		a (a') (α)		b (b') (β)		c (c') (γ)		d (d')	
		M	*F*	*M*	*F*	*M*	*F*	*M*	*F*
		Mean	*Mean*	*Mean*	*Mean*	*Mean*	*Mean*	*Mean*	*Mean*
GXXI	**Jenss-Bayley**	0.05	-0.05	0.00	0.00	0.30	-0.27	0.00	0.00
	Reed	0.95	-1.19	-0.05	-0.03	-0.25	0.57	-0.85	1.16
	SITAR	0.00	0.00	-0.05	0.06	0.06	-0.07		
GOCS	**Jenss-Bayley**	0.05	-0.05	0.00	0.00	0.24	-0.23	0.00	0.00
	Reed	1.17	-1.15	0.08	-0.07	-0.66	0.65	-1.20	1.19
	SITAR	0.01	-0.02	-0.04	-0.04	0.04	-0.02		
NINFEA	**Jenss-Bayley**								
	Reed	1.24	-1.38	0.23	-0.25	-0.98	1.06	-1.39	1.53
	SITAR	0.02	-0.03	-0.04	0.02	0.04	-0.04		

M: Males; F: Females

Table 6: Mean of the predicted random effects by gender, model and cohort (0-2 years data)

effects to be high, the negative value for $\hat{\gamma}_{1j}$ was somewhat surprising, due to the large size of this child, observed at all ages. However Figure 5 clarifies how this happens as it shows that the child-specific predicted curve (dark dotted line) is first shifted downward when his departure from the population average size, due to his larger weight, $\hat{\alpha}_{1j}$, is removed (solid thin line), and then shifted to the left when his departure from the population tempo, due to his late APWV (estimated to be at around 3 months), $\hat{\beta}_{1j}$, is removed (solid medium thick line). Because of these realignments the curve is less steep than the predicted population curve, explaining the negative value of $\hat{\gamma}_{1j}$. When the contribution of $\hat{\gamma}_{1j}$ is also removed the individual predicted curve and the population predicted curve, as expected, overlap (solid thick line).

This example illustrates how to attribute biological meanings to random effects predicted from a SITAR model. They also highlight their close inter-relationships and the care needed for their interpretation: the value of each parameter is strongly influenced by the values of the other two. This was also evidenced by the correlations shown in Table 5. Estimated correlations between α_{i1} and β_{i1} are positive, suggesting that children with lower weight are also those with an earlier tempo of growth (and earlier APWV). Estimated correlations between α_{i1} and γ_{i1} and between β_{i1} and γ_{i1} are instead of similar strength but negative, suggesting that smaller children, or children with an earlier, i.e. negative, *tempo*, experience greater growth velocities.

These observations lead to the conclusion that examination of the SITAR random effects parameters, even if biologically interpretable, requires consideration of their conditional distributions. In this light we examined the marginal associations between the gender of the child and each of *size*, *tempo* and *velocity* and also the same associations but conditional on the other random effects (Table 7). In GXXI and in NINFEA the conditional effects differ from the marginal ones, with males having larger *size* when difference in the other two dimensions are accounted for. The latter means that comparing the growth curve of a female with that of a male with similar velocity and tempo, the male curve will lie above the female one or, in other

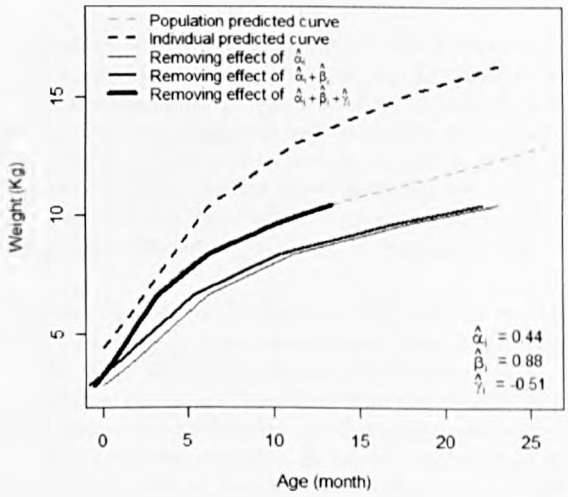


Figure 5: Predicted individual growth curve of a selected GOCS male and new curves created after progressively removing the three random effects from the SITAR model (fitted on the GOCS males aged 0-2 year on the log weight and age scale)

words, that the mean of the size coefficient, stratified by - for example - quartiles of the velocity and tempo distributions is expected to be greater in males than in females. In NINFEA the conditional effects of gender on *tempo* and *velocity* are also stronger than the corresponding marginal effects. In contrast in GOCS there is no evidence of a difference in *tempo* by gender, with the differences in *size* and *velocity* only marginally increased by adjustment. Given that GOCS children are selected to be more similar in terms of gestational age and birth weight than those from the other cohorts, this indicates that the degree of variation in gestational age and birth weight both contribute to the random effects correlations.

	Effect of gender (Male vs Female)					
	GXXI		GOCS		NINFEA	
Dependent variable	Marginal *Coefficient (SE)*	Adjusted[1] *Coefficient (SE)*	Marginal *Coefficient (SE)*	Adjusted[1] *Coefficient (SE)*	Marginal *Coefficient (SE)*	Adjusted[1] *Coefficient (SE)*
Size (α_{15})	-0.001 (0.012)	0.049 (0.008)	0.030 (0.008)	0.039 (0.005)	0.015 (0.008)	0.068 (0.006)
Tempo (β_{15})	-0.112 (0.048)	-0.095 (0.041)	-0.0001 (0.031)	-0.003 (0.019)	-0.056 (0.042)	-0.129 (0.038)
Velocity (γ_{15})	0.132 (0.021)	0.129 (0.015)	0.063 (0.015)	0.078 (0.009)	0.071 (0.016)	0.116 (0.009)

SE = Standard Error
[1] *Adjusted for the other SITAR random effects*

Table 7: Marginal and mutually adjusted effects of gender on estimated *size*, *tempo* and *velocity*, by cohort (fitted on the 0-2 years data)

The influence of gestational age

Several results, as observed above, identified differences between estimates obtained when analysing children in the GOCS cohort as opposed to the other cohorts. The

most likely explanation is that participation in GOCS was restricted to the term babies. Indeed it is well known that gestational age has a very strong influence not just on size at birth but also on the growth trajectory after birth, Sullivan et al. (2008). To examine whether such heterogeneity had an impact on the estimated parameters and therefore on the individual growth curves predicted by the SITAR model, we compared the original results obtained when analysing the GXXI data with results obtained when either, (a) pre-term babies were excluded, or (b) the age scale was re-defined as time since conception (i.e. age + gestational age, both measured in weeks).

Excluding pre-term babies (61 children) did not lead to any changes in the predicted population curves, while some changes were found in the estimated measures of variation: the SD of $\hat{\beta}_{1i}$ and its correlation with the other two random effects were smaller, becoming more similar to the values found in GOCS. In contrast the average APWVs were practically unchanged: 1.66 in males and 1.95 months in females (instead of 1.65 and 1.90, respectively). When we re-examined the conditional associations between the gender of the child and each of *size*, *tempo* and *velocity*, we observed that the effect on $\hat{\beta}$ decreased from -0.095 (SE=0.041) to -0.066 (SE=0.033), while the effects on the others two parameters were substantially unchanged.

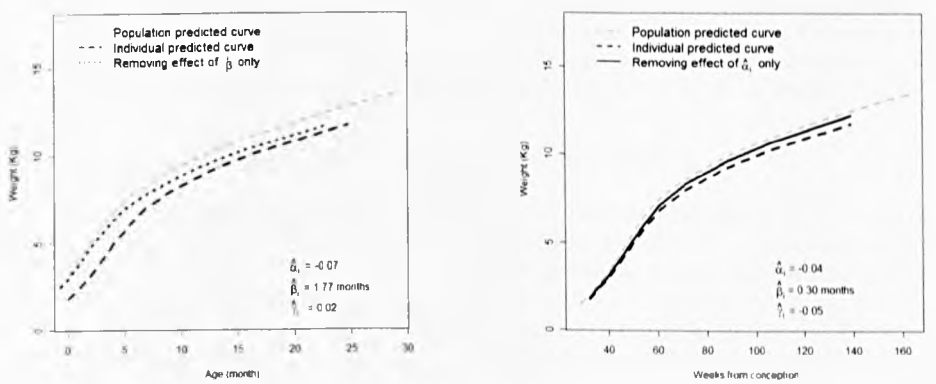


Figure 6: Growth curve for a selected child (GXXI cohort) predicted when fitted on the age time scale (left) or on the weeks since conception time scale (right)

When analyses were re-run on the new age scale of time since conception (in weeks), part of the information held in the *tempo* parameter, β_{1i}, appeared to have been removed by the new time scale. Figure 6 illustrates this. The left hand side panel shows the predicted growth curve for a male from the GXXI cohort whose gestational age was 32 weeks and whose weight at birth was 1.86Kg. On the right hand side panel the growth curve of the same child is shown when predicted by the model fitted on time since conception. The respective predicted population curves are also shown. The estimated child-specific parameters from the two models are $\hat{\alpha}_{1i} = -0.07$, $\hat{\beta}_{1i} = 1.77$, $\hat{\gamma}_{1i} = 0.02$, when fitted on the original age scale and $\hat{\alpha}_{1i} = -0.04$, $\hat{\beta}_{1i} = 1.29$ (in weeks, which corresponds to about 0.3 months), $\hat{\gamma}_{1i} = -0.05$, for the time since

conception scale (centered at about 39 weeks). Hence *tempo* is reduced to about a sixth when gestational age is accounted for. Indeed when the effect of $\hat{\beta}_{1i}$ is removed from the predicted individual curve on the left panel, the shape of the derived curve (dotted line) resembles the predicted individual curve on the right panel (dashed line). Removing only the effect of $\hat{\alpha}_{1i}$ from the latter curve (weeks from conception scale) was indeed sufficient to reach a near superimposition with the predicted population curve (solid line in the right panel of Figure 6).

5 Discussion

In this exploration of models for infant weight data we have used three recently established birth cohorts to compare three models where individual variations in growth are accounted for by specific random effects parameters. Our main finding is that the choice of which model to adopt varies with the aim of the study and, less crucially, on the richness of the available data. If interest focuses on describing individual early weight growth trajectories and/or typical profiles then all the three models considered in this paper are suitable. Among them the 4-parameter Reed model performed best in terms of standard fit criteria and, unlike the other two models, is linear in its parameters and is therefore easier to estimate. However, if the focus is on extracting salient features of the growth trajectories in order to include them in further analyses where growth is either the exposure or the outcome, then the SITAR model may be the preferred option because of the biological interpretability of its parameters. However several other factors should be considered before a choice is made.

The range and frequency of observations over time

We have shown that including measures taken at birth requires the transformation of the time scale when fitting the Reed model because it contains a logarithmic and an inverse function of age. We have adopted one particular transformation that gives a value of 1 to time of birth (and a value of 0 to nine months before birth, i.e. roughly at 'conception'). An alternative transformation had been suggested in the literature that adds a value of 1 to time only to the terms that involve the logarithmic and the inverse function of time. This hybrid solution was, in our view, unsatisfactory and indeed performed slightly worse in terms of goodness of fit than the one we have adopted. A drawback of the latter however is that, when applied to our data, the model's derivative (i.e. the weight velocity curve) did not present a maximum and therefore did not allow the identification of an APWV, if such peak was actually present. This was not an issue with the SITAR model even when a transformation was required to meet the distributional assumptions, because the peak in the weight velocity curve was still identifiable. Hence, if APWV is of interest, and if data include measurements at birth, the SITAR model, fitted on the appropriate growth and time scale to meet the assumptions, should be preferred.

Range, frequency and regularity of the observations over time are other important aspects to consider. Spline functions such as the one used by the SITAR model are

necessarily influenced by data points that are isolated and this was the case with the analyses of the NINFEA 0-4 years data.

Interpretation of the child-specific parameters

By comparing the distribution of random effects predicted for males and females by the three models, when fitted on the combined data, we aimed to examine the discriminatory values of these parameters. Those predicted from the JB and Reed models that correspond to their intercepts ($\hat{c}_{1i}$, and $\hat{a}'_{1i}$), or correction to the intercept ($\hat{a}_{1i}$), did show differences between the genders. However on their own these parameters are unlikely to discriminate the growth trajectories of subgroups defined by weaker predictors of growth than gender (for more discussion of this see the Appendix). In contrast each of the three SITAR random effects appeared to have distinct means and distributions between the genders, especially when their inter-correlations were taken into account (note that accounting for correlations among the parametric models random effects was not possible because of extreme collinearity).

The interpretability of the SITAR model however requires additional care. In particular the interpretation of the *tempo* parameter, which represents the shift on the time axis necessary to synchronize the curves which are centred on the tempo milestone (the APWV in this setting), depends on the time scale used and the setting analysed. When changed by setting the time origin at conception, we found that gender differences with respect to *tempo* were slightly reduced. Thus this parameter represents an adjustment necessary to better proxy true biological age (hence measuring growth adjusted for maturation/developmental status). This is in line with results obtained for the cohort including only term babies with birth weight between 2500 and 4500 grams (GOCS), for which we observed that estimate of *tempo* did not vary across genders when using chronological age as time scale, and it is also in line with the fact that gender effect on tempo decreased when estimated using only term babies in GXXI compared to using the whole cohort. Note that in other settings, with different growth variables and especially with different age ranges, the *tempo* parameter represents adjustments of the time scale by other factors, such as hormonal levels in puberty, Cole et al. (2010).

A final note of caution with regards to the interpretability of the SITAR parameters derives from its extreme flexibility. Unlike the JB and Reed models, SITAR is not restricted to an age range, however the wider the range of ages included in a SITAR model, the less clear will be the interpretation of its parameters. This would affect the *tempo* and *velocity* parameters in particular because individual trajectories over longer age ranges may include several inflection points, leading to multiple changes in velocity and therefore to the parameters representing a complex average of several, and possibly very different, departures from the population time scale. This would happen for example if SITAR were used to model body mass index (BMI) (weight/height^2) from age 0 to 10 years, say. During this time children would experience both a peak velocity at around 1 year and a minimum (negative) velocity at about 5 years, Rolland-Cachera et al. (1987). Moreover, a single parameter *tempo* would represent an average of the child's departure from possibly two separate biological time scales:

the one linked to the infancy BMI peak and the one linked to the childhood BMI rebound.

Issues related to life course epidemiology

Much interest in modelling growth data comes from research in life course epidemiology, Kuh and Ben-Shlomo (2004), where summary growth parameters such as those described in this paper are used as explanatory variables for the onset of a later outcome. Thus analyses consist of two components, carried out in stages: a first stage focussed on modelling the growth data, and a second stage aimed at relating parameters from the growth model to the distal health outcome. For such analyses to succeed it is crucial that salient features of growth are estimated for all children in the study and that the statistical uncertainties of the parameters extracted in the first stage are properly accounted for, Molinari and Gasser (2004). The former may not be possible if parameters used to summarise the growth data are not available for all children (e.g. see Silverwood et al. (2009)), with the second stage analyses consequently being affected by selection bias. Adopting the SITAR modelling approach, and using the three random parameters of *size*, *tempo* and *velocity* as explanatory variables would not suffer from this. Moreover, using all the three SITAR parameters instead of a single growth indicator, such as APWV, would lead to a more comprehensive summary of the individual growth patterns.

The additional statistical uncertainties arising from using estimated growth parameters in the second stage of the analyses is rarely addressed in applications. However it can be dealt with, at least approximately, by correcting the standard errors of the second stage estimates using the bootstrap, Efron and Tibshirani (1993), or weighting the second stage according to the amount of child-specific information available in the first stage (e.g. the number of observations, see Tzoulaki et al. (2010)). The former approach would be quite cumbersome given the complexities of estimating these models. A one-step estimation of the two model components, growth and distal outcome modelling, would not require such inferential adjustments. This is where specifying the growth model using random effects polynomial curves would be advantageous as they could be estimated jointly with the distal outcome model using structural equation models, De Stavola et al. (2006). This approach however would suffer from lack of biological interpretability of the random effects parameters. Moreover it is not easily implemented when fitting more complex growth models, at least using standard software.

Another aspect of growth modelling relevant to life course epidemiology is that focussed on identifying factors that influence childhood growth. Again such analyses could be carried out in two stages, where some growth parameters are first estimated and then modelled in terms of exposures of interest (as we did in section 4.4). Alternatively models could be specified so that estimation is carried out in a single step. This is easily achieved within the framework of random effects polynomial models (e.g. see dos Santos Silva et al. (2002)). Of the models reviewed here, however, only the SITAR model has been generalised to include explanatory factors, Beath (2007); Hui et al. (2010); Johnson et al. (2011), although such specification allows only the

inclusion of explanatory variables for *size* and for the shifted and re-scaled time scale (denoted by Beath as the "growth rate"). Thus the interpretation of estimated effects of explanatory variables for the SITAR growth parameters needs particular care. Moreover, given the non-linearity of the SITAR model, the explanatory variable parameters should not be used to infer population-average effects, Fitzmaurice et al. (2009).

Finally in this field, there is much interest in comparing and summarising evidence from different studies. However, not one but several study-specific parameters for each growth dimension would need to be compared, i.e. *size*, *tempo* and *velocity*. Simultaneous analysis of all parameters could be carried out by fitting multivariate meta-regression models that account for the covariance structure of the parameters estimated within each study, van Houwelingen et al. (2002).

Conclusions

In this paper we have examined alternative modelling approaches to the estimation of the salient features of growth in weight in infancy/early childhood and discussed the difficulties, advantages and disadvantages of choosing each model. Our conclusion is that there is a range of options available to researchers but that each necessitates careful understanding of assumptions and parameterizations. Of the three models discussed, the SITAR model is certainly the most flexible and the most useful for life course enquiries, as it allows the identification of important features of the children's growth trajectories. Its application however requires careful selection and substantive understanding of the model parametrization. This is true for the other models too but possibly to a lesser extent. However, the potential for extracting the most salient features of the growth patterns will depend on the richness and quality of the data, and this is where researchers should make the greatest investment.

Funding

NINFEA cohort was partially funded by Compagnia San-Paolo Foundation, Piedmont Region and Italian Ministry of University and Research. GXXI cohort was funded by Programa Operacional de Sade – Saúde XXI, Quadro Comunitário de Apoio III and by Administração Regional de Saúde Norte. GOCS cohort was funded by Fondecyt projects 1060785, 1090252 and 1120326.

Acknowledgements

The authors wish to thank Milena Maule, University of Turin, for many relevant insights. We are also grateful to all participants of the three cohorts.

References

Asefa, M., R. Drewett, and J. Hewison (1996). An Ethiopian birth cohort study. *Paediatr Perinat Epidemiol 10*(4), 443–62.

Baird, J., D. Fisher, P. Lucas, J. Kleijnen, H. Roberts, and C. Law (2005). Being big or growing fast: systematic review of size and growth in infancy and later obesity. *BMJ 331*(7522), 929.

Barker, D. J. P. (1998). *Mothers, Babies and Health in Later Life.* Edinburgh: Churchill Livingstone.

Beath, K. J. (2007). Infant growth modelling using a shape invariant model with random effects. *Stat Med 26*(12), 2547–64.

Berkey, C. S. (1982). Comparison of two longitudinal growth models for preschool children. *Biometrics 38*(1), 221–34.

Berkey, C. S. and R. B. Reed (1987). A model for describing normal and abnormal growth in early childhood. *Hum Biol 59*(6), 973–87.

Box, G. and D. Cox (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological) 26*, 211–252.

Brown, R. C., T. Dwyer, C. Kasten, D. Krotoski, Z. Li, M. S. Linet, J. Olsen, P. Scheidt, D. M. Winn, and C. International Childhood Cancer Cohort (2007). Cohort profile: the International Childhood Cancer Cohort Consortium (I4C). *Int J Epidemiol 36*(4), 724–30.

CHICOS (2010). CHICOS: Developing a Child Cohort Research Strategy for Europe (http://www.chicosproject.eu/).

Cole, T. J., M. D. Donaldson, and Y. Ben-Shlomo (2010). SITAR–a useful instrument for growth curve analysis. *Int J Epidemiol 39*(6), 1558–66.

Count, E. W. (1943). Growth patterns of the human physique: an approach to kinetic anthropometry. *Hum Biol 15*, 1–32.

De Stavola, B. L., D. Nitsch, I. dos Santos Silva, V. McCormack, R. Hardy, V. Mann, T. J. Cole, S. Morton, and D. A. Leon (2006). Statistical issues in life course epidemiology. *Am J Epidemiol 163*(1), 84–96.

Deming, J. and A. H. Washburn (1963). Application of the jenss curve to the observed pattern of growth during the first eight years of life in forty boys and forty girls. *Hum Biol 35*, 484–506.

dos Santos Silva, I., B. L. De Stavola, V. Mann, D. Kuh, R. Hardy, and M. E. Wadsworth (2002). Prenatal factors, childhood growth trajectories and age at menarche. *Int J Epidemiol 31*(2), 405–12–.

Efron, B. and R. Tibshirani (1993). *An introduction to the Bootstrap.* Chapman & Hall.

Fitzmaurice, G., M. Davidian, G. Verbeke, and G. Molenberghs (Eds.) (2009). *Longitudinal data analysis*. Chapman and Hall/CRC.

Goldstein, H. (2011). *Multilevel statistical models, Fourth Edition*. London: Arnold.

Hui, L. L., G. M. Leung, B. J. Cowling, T. H. Lam, and C. M. Schooling (2010). Determinants of infant growth: Evidence from Hong Kong's "Children of 1997" birth cohort. *Ann Epidemiol 20*(11), 827–35.

Huxley, R. R., A. W. Shiell, and C. M. Law (2000). The role of size at birth and postnatal catch-up growth in determining systolic blood pressure: a systematic review of the literature. *J Hypertens 18*(7), 815–31.

Jenss, R. M. and N. Bayley (1937). A mathematical method for studying the growth of a child. *Human Biology 9*(4), 556–563.

Johnson, L., C. H. Llewellyn, C. H. van Jaarsveld, T. J. Cole, and J. Wardle (2011). Genetic and environmental influences on infant growth: prospective analysis of the Gemini twin birth cohort. *PLoS One 6*(5), e19918.

Kuh, D. and Y. Ben-Shlomo (2004). *A Life Course Approach to Chronic Disease Epidemiology*. Oxford University Press, USA.

Lindstrom, M. L. and D. M. Bates (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics 46*(3), 673–87.

Manwani, A. H. and K. N. Agarwal (1973). The growth pattern of indian infants during the first year of life. *Hum Biol 45*(3), 341–9.

Molinari, L. and T. Gasser (2004). The human growth curve: distance, velocity and acceleration. In R. Hauspie, N. Cameron, and L. Molinari (Eds.), *Methods in human growth research*. Cambridge University Press.

Pinheiro, C. and D. M. Bates (2000). *Mixed-Effects Models in S and S-PLUS*. Springer.

Richiardi, L., I. Baussano, L. Vizzini, J. Douwes, N. Pearce, and F. Merletti (2007). Feasibility of recruiting a birth cohort through the Internet: the experience of the NINFEA cohort. *Eur J Epidemiol 22*(12), 831–7.

Richter, L. M., C. G. Victora, P. C. Hallal, L. S. Adair, S. K. Bhargava, C. H. Fall, N. Lee, R. Martorell, S. A. Norris, H. S. Sachdev, A. D. Stein, and C. G. the (2011). Cohort Profile: The Consortium of Health-Orientated Research in Transitioning Societies. *Int J Epidemiol [Epub ahead of print]*.

Rolland-Cachera, M. F., M. Deheeger, M. Guilloud-Bataille, P. Avons, E. Patois, and M. Sempe (1987). Tracking the development of adiposity from one month of age to adulthood. *Ann Hum Biol 14*(3), 219–29.

Rubin, D. (1976). Inference and missing data. *Biometrika 63*, 581–592.

Ruppert, D., M. P. Wand, and R. J. Carroll (2003). *Semiparametric regression*. Cambridge University Press.

Silverwood, R. J., B. L. De Stavola, T. J. Cole, and D. A. Leon (2009). BMI peak in infancy as a predictor for later BMI in the Uppsala Family Study. *Int J Obes (Lond) 33*(8), 929–37.

Simondon, K., F. Simondon, F. Delpeuch, and A. Cornu (1992). Comparative study of five growth models applied to weight data from congolese infants between birth and 13 months of age. *American Journal of Human Biology 4*, 327–335.

Sullivan, M., M. McGrath, K. Hawes, and B. Lester (2008). Growth trajectories of preterm infants: birth to 12 years. *Journal of Pediatric Health Care 22*(2), 83–93.

Tzoulaki, I., U. Sovio, D. Pillas, A. L. Hartikainen, A. Pouta, J. Laitinen, T. H. Tammelin, M. R. Jarvelin, and P. Elliott (2010). Relation of immediate postnatal growth with obesity and related metabolic risk factors in adulthood: the northern Finland birth cohort 1966 study. *Am J Epidemiol 171*(9), 989–98.

van Dommelen, P., S. van Buuren, G. R. Zandwijken, and P. H. Verkerk (2005). Individual growth curve models for assessing evidence-based referral criteria in growth monitoring. *Stat Med 24*(23), 3663–74.

van Houwelingen, H. C., L. R. Arends, and T. Stijnen (2002). Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat Med 21*(4), 589–624.

Verbeke, G. and G. Malenberghs (2000). *Linear mixed models for longitudinal data*. Springer-Verlag.

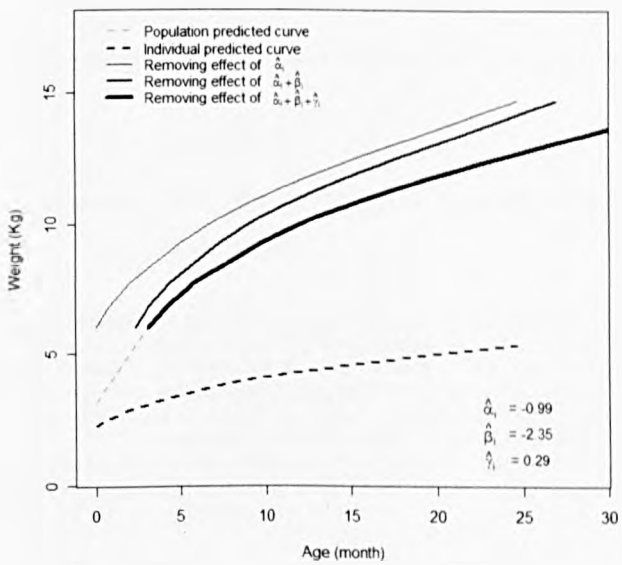
Yang, S., K. Tilling, R. Martin, N. Davies, Y. Ben-Shlomo, and M. S. Kramer (2011). Pre-natal and post-natal growth trajectories and childhood cognitive ability and mental health. *Int J Epidemiol 40*(5), 1215–26.

Appendix: extreme growth trajectories

Another interesting comparison of the three growth models concerns how well they perform in predicting unusual trajectories. In the main text (Section 4.4) we have already discussed the estimated SITAR random effects for the child depicted in Figure 5. This child has a birth weight of 4.55 kg and weighed 17.4 kg at 2 years. The SITAR predicted random coefficients for this child were the maximum of the distribution of *size* (due to his large weight), the ≃95th percentile of the distribution of *tempo* (due to his delayed pick weight velocity, occurring at about 3 months of age), and the ≃5th percentile of the distribution of *velocity* (due to his relative slow growth rate). The predicted random effects for the same child derived from the JB model corresponded to the ≃5th percentile of the distribution of a_{1i}, the ≃25th percentile of the distribution of b_{1i}, the ≃95th percentile of the distribution of c_{1i}, and the ≃5th percentile of the distribution of d_{1i}, with c_{1i} and d_{1i} - the intercept and the slope of the asymptote respectively (see Section 3.1 of the main text) - capturing the large weight of the child and his slow growth rate. The corresponding figures for the Reed model were approximately the median of the distribution of a'_{1i}, the ≃85th percentile of the distribution of b'_{1i}, the ≃85th percentile of the distribution of c'_{1i}, and about the median of the distribution of d'_{1i}. Reed parameters were close to the centre of their distribution and therefore not very informative.

The same comparisons was also drawn for a boy weighting 5.3 kg at 2 years and 2.3 kg at birth who clearly experienced severe growth problems after birth (Appendix Figure 1). His predicted SITAR random effects were, for *size* and *tempo*, the minimum values of their predicted distributions (capturing his low weight and his delayed developmental status), while for *velocity* it was above the 75th percentile of its distribution. The predicted random effects for this child derived from the JB model were, for a_{1i} and b_{1i}, about the 75th percentile of their distributions, while for c_{1i} and d_{1i} they corresponded to about the 85th and 95th of their distributions respectively, suggesting that none of the JB parameters captures the salient features of the weight trajectory of this child. Finally, the predicted Reed random coefficients were again not informative corresponding, for a'_{1i}, to the ≃25th percentile of its distribution, while for b'_{1i}, c'_{1i} and d'_{1i} to approximately the median of their distributions.

These examples (plus other that we do not report here) show that the SITAR parameters consistently identify outlying growth patterns via extreme predicted random effects. In contrast those derived by the JB and Reed models for the same children were less informative.



Appendix Figure 1 Predicted individual growth curve of a selected GXXI male and new curves created after progressively removing the three random effects from the SITAR model (fitted on the GXXI males aged 0-2 year on the log weight and age scale)

Example codes for fitting the three growth models

Pizzi C. et al. "On modelling early life weight trajectories".

RELEVANT PROGRAMS:

Three programs, produced with R version 2.11.1, are included as supplements to this paper. These include example codes for fitting the three growth models compared in the paper.
Note that in each program, wt indicates weight in kg, month indicates age since birth (in months), while data is the name of the data frame that holds these variables.

1. jb.R

This includes the example code for fitting the Jenss-Bayley random effects model.

2. reed.R

This includes the example code for fitting the Reed random effects model.

3. sitar.R

This includes the example code for fitting the SITAR shape invariant random effects model. In this example the SITAR model is fitted on the log-kg and month scales, including fixed effects for the size and velocity parameters, and with a spline function with 4 knots.
NOTE: Please note that the following code, based on that published by Beath (Beath K. J. 2007, Stat Med 26(12),2547-64), is a simplified version of the actual function used in the paper. The latter is available on request from Professor Tim Cole, who is currently developing a dedicated R library (tim.cole@ucl.ac.uk).

NOTE ABOUT THE DATA: None of the three datasets analyzed for this paper can be made directly accessible to the Journal's readership because of confidentiality restrictions. Data from each cohort study are available on direct request to the study coordinators.
Interested readers should use these auxiliary programs to understand how the analyses in the paper were conducted.

Costanza Pizzi
Cancer Epidemiology Unit,
Department of Medical Sciences
University of Turin

Italy
E-mail:

Jenss-Bayley model

```
# This is an example of the R program used to fit the Jenss-Bayley (JB)
# random effects model examined in the JRSS-A Paper "On modelling early
# life weight trajectories".

# Author: Costanza Pizzi (costanza.pizzi@lshtm.ac.uk)

# Date: February 14, 2013.



#Read the data
  data <- read.table("data", header=TRUE, sep="\t",na.strings=c(""))


#Load the nlme library
  library(nlme)


#The starting values for the JB random effects model(a0, b0, c0, d0)
#are obtained by fitting models without random effects using the nls
#function
  jb.nls <- nls(wt ~ c+d*month-exp(a+b*month),
            data= data,
            start= c( a=1,b=-0.2,c=8,d=0.1),
            trace= T
            )


#JB random effects model
  jb      <- nlme(wt ~ c+d*month-exp(a+b*month),
            data= data,
            fixed= a+b+c+d ~ 1,
            random= a+b+c+d ~ 1 | id,
            start= c(a=a0,b=b0,c=c0,d=d0),
            correlation= NULL,
            weights= NULL,
            na.action= na.fail,
            )
```

Reed model

```
# This is an example of the R program used to fit the Reed random effects
# model examined in the JRSS-A Paper "On modelling early life weight
# trajectories".

# Author: Costanza Pizzi (costanza.pizzi@lshtm.ac.uk)

# Date: February 14, 2013.



#Read the data
data <- read.table("data", header=TRUE, sep="\t",na.strings=c(""))


#Load the nlme library
  library(nlme)


#Age since birth (month) was rescaled so that the new age (month.r) is 0
#at conception (assumed occurring 9 months before delivery) and is 1 at
#birth.
  data<- transform(data, month.r=(month+9)/9)


#The starting values for the Reed random effects model (a0, b0, c0, d0)
#are obtained by fitting models without random effects using the nls
#function
  reed.nls <- nls(wt ~ a+b*month.r+(c*log(month.r))+(d/(month.r)),
              data= data,
              start= c(a=10,b=1,c=-4,d=-10),
              trace= T
              )


#Reed random effects model
  reed     <- nlme(wt ~ a+b*month.r+(c*log(month.r))+(d/(month.r)),
              data= data,
              fixed= a+b+c+d ~ 1,
              random= a+b+c+d ~ 1 | id,
              start= c(a=a0,b=b0,c=c0,d=d0),
              correlation= NULL,
              weights= NULL,
              na.action= na.fail,
              )
```

SITAR model

```
# This is an example of the R program used to fit the SITAR model
# examined in the JRSS-A Paper "On modelling early life weight
# trajectories".

# Author: Costanza Pizzi (costanza.pizzi@lshtm.ac.uk)

# Date: February 14, 2013.



#Read the data
  data <- read.table("data", header=TRUE, sep="\t",na.strings=c(""))


#Load the libraries
  library(splines)
  library(nlme)


#The starting values for the spline function (s01,s02,s03,s04) and for
#the size parameter (a0) are obtained by fitting a linear regression of
#the log-weight variable on the age variable, the latter transformed via
#a natural cubic spline. The starting value for the velocity parameter
#(c) is set at 0. The internal knots of the spline (kn) are placed at the
#quartiles of the age distributions (as in this example the spline
#function has 4 knots), while the boundary knots (bou) are placed at
#points defined as the minimum (and maximum)
#of the age distribution minus (plus) 4% of the total range.
  kn <- quantile(data$month, (1:3)/4)
  bou <- range(data$month) + 0.04 * c(-1,1) * diff(range(data$month))
  lm(log(data$wt) ~ ns(data$month, knots=kn, Boundary.knots=bou))


#Create the function for a spline with 4 knots
  model <- function(time, s1,s2,s3,s4, sa,sb,sc) {
         splcoef <- as.matrix(cbind(s1,s2,s3,s4))
         as.vector(sa+t(matrix(rep(1,4), ncol=4) %*%
         t(splcoef*as.matrix(ns((time-sb)/exp(-sc), knots=kn,
         Boundary.knots=bou)))))
         }


#Apply the function "model" to the data on log-kg and month scales,
#including fixed effects for the size and velocity parameters and random
#effects for all the 3 parameters
  sit    <- nlme(log(wt)~ model(month, s1,s2,s3,s4,a,b,c),
            data=data,
            fixed=s1+s2+s3+s4+a+c ~ 1,
            random=a+b+c ~1 |id,
            start=c(s01,s02,s03,s04,a0,0),
            na.action=na.fail,
            correlation = NULL,
            weights = NULL,
            control=nlmeControl(returnObject=TRUE)
            )
```

5.2.1 Addendum to Research Paper III

Another interesting comparison of the three growth models concerns how well they perform in predicting the weight values across different age intervals. In order to address this issue the observed and expected mean weight values (in kg), where the latter are the individual-level predictions, as well as the mean and SD of the difference between the observed and expected values were compared across a-priori defined age groups. Results are shown in Table 5.1 for the models fitted on the 0-2 years male data and in Table 5.2 for the models fitted on the 0-4 years male data (thus only for the GOCS and NINFEA studies). The results obtained for females are similar and are not reported for simplicity.
SITAR appears to perform better than the two biological models both at birth and within the first three months of life, when the non-linearity of the weight trajectories is expected to be larger. In contrast SITAR performs slightly worse than the JB and Reed models beyond 12 months of life, especially when fitted on the NINFEA data - probably due to the small and isolated number of measurement times available for this cohort - and on the 0-4 years data. In particular the SDs increase for the oldest age intervals, especially when modelling the data up to 4 years of age.
As already discussed in Research Paper III, the JB and Reed models perform much better when fitted on the 0-4 years data compared with the 0-2 years data. In particular the SD of of the difference between the observed and expected weight is considerably lower in Table 5.2 than in Table 5.1 across all age intervals. Both approaches perform clearly better when fitted on the GXXI study compared to those derived from the other two cohorts (Table 5.1), probably reflecting the fact that GXXI is the richest cohort in terms of weight measurements gathered.

		Jenss-Bayley		Reed		*SITAR*	
	Observed	*Expected*†	*Observed-Expected*‡	*Expected*†	*Observed-Expected*‡	*Expected*†	*Observed-Expected*‡
				0-2 years			
GXXI							
Birth	3.23	3.1	0.13 (0.24)	3.12	0.11 (0.21)	3.19	0.04 (0.11)
0.1- 3 months	4.64	4.71	-0.06 (0.29)	4.74	-0.09 (0.27)	4.64	0.01 (0.20)
3.1 - 6 months	6.94	6.90	0.04 (0.28)	6.89	0.05 (0.25)	6.94	-0.01 (0.27)
6.1 - 12 months	8.77	8.77	-0.004 (0.32)	8.74	0.02 (0.31)	8.79	-0.03 (0.34)
12.1+ months	11.41	11.41	0.003 (0.29)	11.42	-0.01 (0.26)	11.39	0.03 (0.39)
GOCS							
Birth	3.44	3.28	0.16 (0.28)	3.36	0.09 (0.27)	3.43	0.01 (0.15)
0.1- 3 months	4.99	5.02	-0.02 (0.67)	5.11	-0.11 (0.71)	4.98	0.02 (0.18)
3.1 - 6 months	7.45	7.29	0.16 (0.80)	7.37	0.07 (0.86)	7.49	-0.04 (0.34)
6.1 - 12 months	9.23	9.13	0.10 (1.36)	9.2	0.02 (1.36)	9.14	0.09 (0.39)
12.1+ months	11.84	11.81	0.03 (1.65)	11.85	-0.01 (1.65)	11.85	-0.01 (0.43)
NINFEA							
Birth	3.34	–	–	3.33	0.01 (0.23)	3.32	0.02 (0.09)
0.1- 3 months	5.98	–	–	6.04	-0.07 (0.68)	5.93	0.05 (0.25)
3.1 - 6 months	7.95	–	–	7.86	0.09 (0.78)	8.04	-0.09 (0.36)
6.1 - 12 months	10.15	–	–	10.19	-0.04 (0.99)	10.17	-0.03 (0.36)
12.1+ months	11.82	–	–	11.8	0.01 (1.15)	11.72	0.09 (0.46)

† *Average of the child-specific predictions*

‡ *Mean (SD)*

Table 5.1: Observed and expected mean weight values (kg) from the best specification of the three models fitted on the 0-2 years male data by cohort and age range

		Jenss-Bayley		Reed		*SITAR*	
	Observed	Expected†	*Observed-Expected*‡	*Expected*†	*Observed-Expected*‡	*Expected*†	*Observed-Expected*‡
				0-4 years			
GOCS							
Birth	3.44	3.35	0.10 (0.33)	3.37	0.07 (0.28)	3.41	0.04 (0.15)
0.1- 3 months	4.99	5.08	-0.09 (0.32)	5.09	-0.10 (0.33)	4.99	0.01 (0.25)
3.1 - 6 months	7.45	7.38	0.06 (0.33)	7.36	0.09 (0.30)	7.46	-0.01 (0.37)
6.1 - 12 months	9.23	9.22	0.001 (0.33)	9.2	0.02 (0.32)	9.24	-0.02 (0.40)
12.1 - 24 months	11.61	11.63	-0.03 (0.38)	11.63	-0.03 (0.36)	11.6	0.01 (0.45)
24.1 + months	16.37	16.37	0.004 (0.41)	16.36	0.01 (0.34)	16.32	0.05 (0.79)
NINFEA							
Birth	3.34	3.27	0.07 (0.37)	3.33	0.01 (0.40)	3.31	0.03 (0.09)
0.1- 3 months	5.98	6.01	-0.03 (0.41)	6.05	-0.07 (0.36)	5.97	0.01 (0.26)
3.1 - 6 months	7.95	7.87	0.08 (0.40)	7.86	0.09 (0.35)	7.98	-0.03 (0.39)
6.1 - 12 months	10.15	10.16	-0.02 (0.47)	10.18	-0.04 (0.39)	10.18	-0.03 (0.45)
12.1 - 24 months	11.82	11.68	0.13 (0.43)	11.81	0.01 (0.29)	11.73	0.09 (0.51)
24.1+ months	16.88	16.89	-0.01 (0.13)	16.88	0.001 (0.12)	16.84	0.04 (0.99)

† *Average of the child-specific predictions*

‡ *Mean (SD)*

Table 5.2: Observed and expected mean weight values (kg) from the best specification of the three models fitted on the 0-4 years male data by cohort and age range

5.3 Research Paper IV

Research paper cover sheet

For a 'research paper' already published

1.1. Where was the work published?
1.2. When was the work published?
1.2.1. If the work was published prior to registration for your research degree, give a brief rationale for its inclusion
1.3. Was the work subject to academic peer review?
1.4. Have you retained the copyright for the work?
If yes, attach evidence of retention
If no, or if the work is being included in its published format, attach evidence of permission from copyright holder (publisher or other author) to include work

2. For a 'research paper' prepared for publication but not yet published

2.1. Where is the work intended to be published?
PLOS ONE

2.2. List the paper's authors in the intended authorship order
Costanza Pizzi, Tim J Cole, Lorenzo Richiardi, Isabel dos Santos Silva, Camila Corvalan, Bianca L De Stavola

2.3. Stage of publication – Not yet submitted/Submitted/Undergoing revision from peer reviewers' comments/In press
UNDER REVIEW

3. For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)

I led the study design. Bianca De Stavola, Isabel dos Santos Silva and Lorenzo Richiardi provided advice on study design. Tim Cole provided the program code in R used to fit the SITAR model. I gave substantial contribution in checking and improving the code. I performed the statistical analyses and wrote the first draft of the article. All co-authors helped interpret the findings and revised the manuscript, providing useful comments.

Candidate's signature

Supervisor or senior author's signature to confirm role as stated in (3) (Bianca De Stavola)

Original Article

Prenatal influences on size, velocity and tempo of infant growth: findings from three contemporary cohorts.

Authors: Costanza Pizzi[1,2], Tim J Cole[3], Lorenzo Richiardi[1], Isabel dos Santos Silva[4], Camila Corvalan[5], Bianca De Stavola[2]

[1] Cancer Epidemiology Unit, Department of Medical Sciences, University of Turin and CPO-Piemonte, Italy
[2] Centre for Statistical Methodology, London School of Hygiene and Tropical Medicine, UK
[3] MRC Centre of Epidemiology and Child Health, UCL Institute of Child Health, UK
[4] Non-communicable Disease Epidemiology Department, London School of Hygiene and Tropical Medicine, UK
[5] Institute of Nutrition and Food Technology, University of Chile, Chile

Running head: Prenatal influences on weight trajectories in infancy.

Conflict of Interest and Source of Funding:
NINFEA cohort was partially funded by Compagnia San-Paolo Foundation, Piedmont Region and Italian Ministry of University and Research. GXXI cohort was funded by Programa Operacional de Sade - Saude XXI, Quadro Comunitario de Apoio III and by Administracao Regional de Saude Norte. GOCS cohort was funded by Fondecyt projects 1060785, 1090252 and 1120326.
No author has conflicts of interest to declare.

Acknowledgements:
The authors are grateful to all participants of the three cohorts.

Abstract

Background: Studying prenatal influences of early life growth is relevant to life-course epidemiology as some of its features have been linked to the onset of later diseases. **Methods**: We studied the association between prenatal maternal characteristics (height, age, parity, education, pre-pregnancy body mass index (BMI), smoking, gestational diabetes and hypertension) and offspring weight trajectories in infancy using SuperImposition by Translation And Rotation (SITAR) models, which parameterize growth in terms of three biologically interpretable parameters: *size*, *velocity* and *tempo*. We used data from three contemporary cohorts based in Portugal (GXXI, n=738), Italy (NINFEA, n=2,925), and Chile (GOCS, n=959). **Results**: Estimates were generally consistent across the cohorts for maternal height, age, parity and pre-pregnancy overweight/obesity. Some exposures only affected one growth parameter (e.g. maternal height (per cm): 0.4% increase in *size* (95% confidence interval (CI):0.3; 0.5)), others were either found to affect *size* and *velocity* (e.g. pre-pregnancy underweight vs normal weight: smaller *size* (-4.9%, 95% CI:-6.5;-3.3), greater *velocity* (5.9%, 95% CI:1.9;10.0)), or to additionally influence *tempo* (e.g. pre-pregnancy overweight/obesity vs normal weight: increased *size* (7.9%, 95% CI:4.9;10.8), delayed *tempo* (0.26 months, 95% CI:0.11;0.41), decreased *velocity* (-4.9%, 95% CI:-10.8;0.9)). **Conclusions**: By disentangling the growth parameters of size, velocity and tempo, we found that prenatal maternal characteristics, especially maternal smoking, pre-pregnancy overweight and underweight, parity and gestational hypertension, are associated with different aspects of infant weight growth. These results may offer insights into the mechanisms governing infant growth.

Birth size and early life growth trajectories have been found to be important predictors for the onset and development of a wide range of later diseases,[1-6] with early postnatal weight gains becoming the focus of research into the development of overweight and obesity later in childhood and adulthood.[7-11] As a consequence there is also growing interest in prenatal predictors of rapid weight gain in infancy[12,13] and overweight and obesity later life.[14-17]

A wide-ranging literature exists on the association of prenatal exposures - such as maternal, environmental and social factors, pregnancy conditions, parental age, health status and life-style - with birth outcomes, mainly birth size and gestational age.[18-22] More recently, the association of these prenatal exposures with early life growth trajectories has also been investigated,[13,23-25] particularly with reference to features of postnatal rapid weight gain.[12,26,27] A limitation of most of these analyses is that they focus on relatively simple aspects of growth, such as differences in size at pre-specified age intervals. In addition such comparisons can only be performed when growth data are available at fixed time points and therefore may involve only a subset, often unrepresentative, of the original cohort.[12]

In this paper we examine the association between several prenatal maternal exposures with weight trajectories of infants (0-2 years) from three recent cohorts based in countries with diverse socio-economic backgrounds (Portugal, Italy and Chile) using the shape-invariant random effects model called SuperImposition by Translation And Rotation (SITAR).[23,28] This approach allows the capture of individual trajectories, from irregularly spaced observations, through three parameters that have a direct biological interpretation - *size, velocity and tempo*. SITAR has been used before to model individual growth data,[22-30] and is extended here to include multiple explanatory variables for each of its three parameters. The focus is on infant growth because of its relevance in life-course epidemiology research, while the inclusion of data

from different cohorts is aimed at evaluating the robustness of the results, given the expected differences across the three source populations in distribution of the exposures (as well as their correlations with potential confounders).

MATERIAL AND METHODS

The cohort studies

GXXI. GXXI was established in 2005 in the Porto region of Portugal. All children born of women resident in the region and admitted to one of its five public hospitals for delivery, with a gestational age at birth greater than 24 weeks, were eligible to participate. Recruitment lasted from April 2005 to August 2006. Women were enrolled few days before their due date and, in the majority, completed baseline questionnaires between 24 and 72 hours after delivery. In total the baseline data consist of 8,311 singleton children. Children were actively followed-up through interviewer-administered questionnaires planned at 3, 6, 12/15 and 24 months of age. Due to logistic and financial constraints a restricted time window was allocated for each follow-up occasion. The present analyses are based on the information collected at baseline and at the 2-years follow-up, which is available for 738 singleton babies of likely Portuguese origin with known gestational age at birth, and consists of growth data from the child's health records, obtained prospectively by health professionals. These include anthropometric measures taken at about 1, 2, 4, 6, 9, 12, 15, 18 and 24 months of age, together with the actual dates of measurement. Up to 6 additional measurements and dates reported in the health records were also entered into the database. The median number of measurements per child is 10.

NINFEA. NINFEA is an on-going Italian web-based cohort study which started in 2005 and aims to recruit pregnant women via the Internet and follow up their children (more details in

[31,32]). Enrolment is carried out at the study website (www.progettoninfea.it) where women complete the first questionnaire (Q1) at any time during their pregnancy. Active follow up is via online questionnaires administered at around 6 (Q2), 18 (Q3), 48 (Q4) months and 7 years (Q5) of age of the child. For these analyses we used the database version 12.03 (downloaded in March 2012) which includes 2,925 singleton children with available data on gestational age at birth and whose mothers were born in Italy. At Q2 women were asked to report the child's anthropometric measurements at birth, 3 and 6 months, while at Q3 they were asked to report the measures at 12 and 18 months. Revisions of these questionnaires, undertaken after approximately the first 1,500 mothers enrolled, led to inclusion of additional questions on the child's measures at the time of their completion. The present analyses involve growth data up to around age 2 years, resulting in a median of 4 (range 1-7) measurements per child.

GOCS. GOCS is an on-going Chilean cohort aiming to study the association of early growth with children's maturation, adiposity and associated metabolic complications (more details in [33]). The study was initiated in 2006 when all children aged 2.6-4 years attending public nursery schools in six counties of Santiago were invited to participate if they were singleton births with a gestational age at birth between 37 and 42 weeks, and birth weight between 2500 and 4500 grams. Among the 1,498 eligible children 1,195 (80%) accepted the invitation. The present analyses includes all 959 children of non-indigenous origin without missing growth and exact gestational age data. Weight and height measurements from birth up to 36 months of life were extracted from routinely-completed health records; from the time of recruitment onwards, children were measured yearly at their nursery by a dietician. For these analyses only growth data up to around age 2 years were used, yielding a median of 6 (range 1-8) measurements per child.

Prenatal exposures

The following background maternal exposures were studied in relation to weight trajectories over the first 2 years of life: height, age, educational level and parity at the time of birth of the child. Pre-pregnancy body mass index (BMI), smoking status during pregnancy and pregnancy complications, namely gestational diabetes and pregnancy hypertension/eclampsia, were instead considered intermediate exposures as their values are likely to be affected by the background variables above. Data on prenatal variables were derived from questionnaires administered during pregnancy in NINFEA, at birth in GXXI, and when the children were approximately 3-4 years old in GOCS. Coding and further details are given in Table 1. Because of missing values a *core dataset* for each cohort was defined as the subset of records with complete information on the following core exposure variables: maternal height, age, education, parity, pre-pregnancy BMI and smoking status during pregnancy.

Statistical methods

SITAR model. The observed weight trajectories were modelled using a recently developed shape invariant random effects model. It was introduced by Cole [28] to study height trajectories in puberty, following the model proposed by Beath to analyse weight growth in infancy.[23] Let y_{it} be the weight of child i at age t, then SITAR is specified as:

$$y_{it} = a_i + h\left(\frac{t - b_i}{e^{-c_i}}\right) + \varepsilon_{it} \qquad (1)$$

where $h(z)$ is a natural cubic spline of transformed age z, α_i, β_i and γ_i are subject-specific growth parameters, and ε_{it} is the residual error term assumed to have mean zero and constant variance. The three parameters correspond respectively to the *size*, *tempo* and *velocity* of growth specific to each child: α_i represents the shift in the weight axis, while β_i and γ_i represent the change in location and scale to be applied to the age scale, respectively, in order for all children to share the same shape (mean spline curve $h(z)$). *Size* is expressed in units of weight, *tempo* in units of age, while *velocity* is a multiplier, and therefore is scale-free and reported as a percentage. Pizzi et al[34] discuss in detail how these parameters are to be interpreted given their close correlations. In brief they can be parameterized as follows: let $\alpha_i = \alpha_0 + \alpha_{1i}$, where α_0 is a fixed parameter, representing the size of a reference child, and α_{1i} a random, normally distributed variable with mean zero and constant variance, and let similar specifications for β_i and γ_i, then estimation can be carried out by maximum likelihood as for any (non-linear) mixed effects model.[35] Irregular observations can be handled under the assumption of missing at random.[36] From a biological perspective α_{1i} will be positive for heavier children, while β_{1i} is related to the timing of maximum growth velocity and therefore will be negative for children whose growth is more advanced at earlier ages (earlier velocity peak), and γ_{1i} will be positive for children with faster growth.[28]

A covariate X with observed value x_i on subject i can be included in the model by specifying the three growth parameters as follow:

$$\begin{aligned} \alpha_i &= \alpha_0 + \delta_\alpha x_i + \alpha_{1i} \\ \beta_i &= \beta_0 + \delta_\beta x_i + \beta_{1i} \\ \gamma_i &= \gamma_0 + \delta_\gamma x_i + \gamma_{1i} \end{aligned} \qquad (2)$$

where δ_α, δ_β and δ_γ represent the contribution of the covariate to a child's *size*, *tempo* and *velocity*, respectively. Generalization of equation (2) to multiple covariates is straightforward. This is a slightly different parameterization from the one adopted by Beath.[23]

Analyses. Weight was log-transformed to aid meeting the distributional assumptions of the model. As a consequence δ_α is to be interpreted as percentage changes in size relative to the reference child.[37] Age was measured in months, hence δ_β is also expressed in months. The spline function *h(z)* was defined by placing the internal knots at quantiles of the age distribution, appropriate for each cohort because of varying richness and spread of the available weight measurements (four knots were used for analyses of GXXI and GOCS data and three for analyses of NINFEA). The complexity of the SITAR model relatively to the available data led to imposing constraints on its parameters, namely that the *tempo* of the standard child, β_0, was zero. Furthermore to be able to compare the three cohorts, δ_β, the contribution of each covariate on a child's *tempo*, was also constrained to be zero. These constraints were relaxed in analyses specific to GXXI as it had more weight growth measurements.

Models were initially fitted separately by study. We first included one explanatory variable at a time, with adjustment by gender and gestational age (we will refer to the latter results as "minimally-adjusted estimates"). We used all available data and also just the *core datasets* to allow comparisons between unadjusted and adjusted estimates for each of these variables. Fully-adjusted estimates were obtained by fitting two separate models to the *core datasets*: (i) the background explanatory variables were mutually adjusted, as well as adjusted for gestational age and gender; (ii) the intermediate explanatory variables were mutually adjusted, as well as adjusted for the background variables, gestational age and gender.

Models were also refitted on the pooled data from the three cohorts, assessing evidence of heterogeneity via significance tests of the interaction between each covariate and the study indicators (one covariate at a time, using the Wald test).

RESULTS

Descriptive results

There is considerable variation in the distribution of the prenatal exposures across the three birth cohorts (Table 1); in particular, Chilean and Portuguese mothers are on average 8 and 3 cm shorter, and 6 and 3 years younger at birth, respectively, than their Italian counterparts. Despite being on average younger, the proportion of multiparous mothers is higher among GOCS participants. Educational level strongly differs across cohorts, with only 5% of the NINFEA mothers being in the lowest educational category as opposed to 36% in GOCS and almost 50% in GXXI, and with almost 60% highly educated women in NINFEA compared to 27% and 20% in GXXI and GOCS, respectively. Because of the study design, education is a strong predictor of participation into NINFEA,[32] and this explains many of the differences observed. The prevalence of overweight/obese women is much lower in NINFEA, while prevalence of underweight is slightly higher. Approximately 20% of GXXI women smoked during pregnancy with the corresponding figure in the other two populations lower than 10%. Gestational diabetes was less frequently diagnosed in GOCS, while gestational hypertension/eclampsia was less frequently diagnosed in GXXI.

Cohort-specific analyses

Table 2 presents the estimated minimally-adjusted and fully-adjusted covariate-specific parameters (i.e. the relevant δ_α and δ_γ), by cohort, obtained from models fitted to the *core datasets*. The minimally-adjusted estimates obtained when fitting the models to each whole cohort are reported in the eTable 1: they are generally close to the minimally-adjusted estimates reported in Table 2 indicating that the *core datasets* are likely to be representative of the corresponding whole cohorts. The minimally-adjusted and fully-adjusted estimates reported in Table 2 are very similar, indicating little reciprocal confounding among these variables. Despite some between-cohort differences, the findings overall are consistent with size being positively associated with maternal height (NINFEA, fully-adjusted: δ_α=0.4%; similarly in the other cohorts), pre-pregnancy overweight/obesity (NINFEA: δ_α=2.1%; similarly in the other cohorts) and parity (GXXI: δ_α=4.5%; similarly in NINFEA), but negatively associated with smoking during pregnancy ($\delta_\alpha\approx$-3% in GXXI and NINFEA) and maternal pre-pregnancy underweight ($\delta_\alpha\approx$-4% in each cohort). Post-natal growth velocity was positively associated with maternal smoking (GXXI: δ_γ=13.2%; NINFEA: δ_γ=6.5%), and possibly maternal underweight (NINFEA: δ_γ=4.4%), but negatively associated with parity (GXXI: δ_γ=-6.1%; similarly in NINFEA). The results for education were quite heterogeneous: while in GXXI medium/highly educated women have bigger children who tend to have slower growth velocity, and in GOCS the children from less educated mothers have slower growth rate compared to those in the reference group, in NINFEA no association was found.

The model that examined pregnancy complications showed that, when fully adjusted for the other characteristics, gestational diabetes was not associated with infant weight growth (Table 3). In contrast, children from mothers with gestational hypertension were smaller and with a steeper

growth curve (GXXI: δ_α=-6.4%, δ_γ=12.8%; similarly in NINFEA), although this pattern was not present in GOCS.

Pooled analyses

Pooled analyses of the three cohorts show significant heterogeneity of effects for some covariates (smoking, gestational hypertension and gestational diabetes), with the differences arising from GOCS, unsurprisingly given the results of Tables 2-3, the retrospective collection of its prenatal data and its inclusion criteria. As there was no evidence of heterogeneity between GXXI and NINFEA, their data were pooled with results reported in Table 4. The estimated coefficients for pre-pregnancy BMI confirm that babies from underweight mothers are smaller but with a greater postnatal growth rate (i.e. *velocity*), while children from overweight/obese women have a bigger *size* without evidence of decreased postnatal growth rate. Results for maternal education show that less educated mothers have smaller children that however have the same growth velocity as children of more educated mothers (Table 4).

In order to examine whether the observed heterogeneities across the cohorts were due to differences in entry criteria, we replicated cohort-specific analyses on the subset of GXXI and NINFEA children who were born at term and with a birth weight of 2500-4500 grams, using the same entry criteria as GOCS. The results pointed to much more similar effects across the cohorts.

Explanatory variables for tempo

Finally we rerun the analyses allowing for covariate effects on *tempo*, restricting them to the GXXI cohort because of its rich number of repeated weight observations (similar analyses for the

other cohorts did not lead to estimation convergence). The results are reported in Table 5. There is no evidence of an effect of maternal height, age, pre-pregnancy underweight or smoking on *tempo*, and therefore no change in the estimated effects on *size* or *velocity*. However parity, pre-pregnancy overweight/obesity, and hypertension do influence *tempo* of growth. Infants of parous mothers have relatively earlier growth spurts by about 5 days (δ_β= -0.17 months, 95% CI -0.34; -0.01). Allowing for this association 'explains away' some of the earlier associations found between parity and *size* and *velocity* (both are substantially reduced; see Table 2 and Table 5). In contrast infants have delayed *tempo* by about 8 days (δ_β=0.26, 95% CI 0.11; 0.41) if their mother is overweight/obese. As for parity, given the correlations among the three growth parameters, including maternal overweight/obesity in the specification of *tempo* changes its association with *size* and *velocity*. In particular that for *velocity* becomes negative (δ_γ=-4.9, 95% CI -10.8; 0.9) implying that infants of overweight/obese mothers not only have a later peak, but also have slower velocity than that of a reference child. For hypertension too the association with *size* and *velocity* is reduced when an association with *tempo* is allowed. The latter is found to be positive (δ_β=0.31, 95% CI 0.08; 0.53) indicating a delay in peak velocity of almost 10 days (Table 5).

DISCUSSION

In this paper we have investigated prenatal influences on weight growth in infancy in order to contribute to the understanding of its role in the development of a wide range of later diseases. We have used data on children belonging to three contemporary cohorts based in Portugal, Italy and Chile in order to compare effects across socio-economically and geographically diverse populations and gain a more robust understanding of these associations, while accounting for potentially different confounding patterns. The individual weight trajectories were modelled

using SITAR,[28] a model that provides biologically interpretable growth parameters, extended here to include multiple explanatory variables.

Our analyses indicate that prenatal exposures affect different dimensions of the weight trajectories. In all cohorts, *size* was positively associated with maternal height, parity and pre-pregnancy overweight/obesity, and negatively with pre-pregnancy underweight. Additionally in all cohorts parity negatively affected *velocity*. In contrast, only for infants from the two European studies, maternal smoking and gestational hypertension were associated with reduced *size* and increased *velocity*, while pre-pregnancy underweight was positively associated with *velocity*. Maternal education was only a moderate predictor of *size* in the European cohorts and of *velocity* in the Chilean cohort. When *tempo* was modeled in terms of covariates in analyses restricted to GXXI, we found that part of the impact on *size* and *velocity* observed for parity, maternal overweight /obesity and hypertension was captured by their influence on the *tempo* dimension. In particular, infants of parous mothers were found to have an earlier timing of growth, while those of overweight/obese or gestational hypertensive mothers to have it delayed. We found instead no evidence of an effect of maternal height, age, or smoking on *tempo*.

While some of these results are not new - e.g. the relation between parity[13,24,26] and smoking[13,14] with infant *size* and weight *velocity*, the positive association between maternal overweight/obesity with increased *size* (which corroborates the existing evidence on an intergenerational transmission of obesity[17]) - other findings are of interest, in particular the association between gestational hypertension and reduced *size*, delayed *tempo* and increased *velocity*, and the effect of maternal underweight on *size* and *velocity*. The former is consistent with current evidence of an association of hypertension with fetal growth retardation.[38] For the latter, while the consequences of maternal obesity have been extensively investigated, less

evidence is currently available on the effect of maternal pre-pregnancy underweight, especially on postnatal growth rate in economically developed countries. What we found in the European cohorts is that maternal underweight was associated with reduced *size* and increased *velocity*, while in the Chilean cohort only an effect on *size* was observed. We also found only a weak association between gestational diabetes and *size* in GXXI and NINFEA, despite previous findings linking it with increased birth weight and adiposity later in life.[17,39] This is possibly due to the self-reported and coarse (i.e. no distinction in severity) nature of the information available in all three cohorts.

A strength of these combined results is that they are derived from modelling the joint association of multiple exposures on multiple growth parameters simultaneously. Another strength of the approach adopted in this paper is that we used all the available growth data (assuming that the frequency and timing of the observations do not depend on the values that are not observed, i.e. that data are missing at random, MAR[36]). This is in contrast to the most common approach used in the epidemiological literature to analyse growth data which consists of comparing anthropometric measures taken at two fixed time points across subgroups of children (e.g. those defined by maternal characteristics). Such comparisons can only be performed for participants with observations at both occasions, therefore involving only a subset of the original cohort which leads to unbiased results only if missingness is completely at random.[12] Specifications of mixed effects models other than SITAR have been used to study growth data that are irregularly spaced, such as linear splines models.[40] Similarly to SITAR they require MAR.[36] However, they are not as flexible in modelling non-linear growth (linear mixed models) or not as interpretable (linear splines models) as SITAR. More specifically, the advantage of SITAR is the ability to naturally deal with the non-linear shape of the weight trajectories - via the use of a cubic spline -

and to summarize the growth process via three biologically meaningful parameters, two of which - *velocity* and *tempo* - separate the growth rate into specific components when trajectories are non-linear. This has given us insights into what governs the timing of peak growth velocity in infancy when we were able to fit the expanded model with explanatory variables for *tempo*, as well as *size* and *velocity*. Moreover our study showed that SITAR can be successfully fitted to dataset with relatively sparse data, such as NINFEA, providing results consistent with those obtained with richer dataset. However, when examining the association between prenatal factors and growth, we had to impose some constraints allowing for an effect on *size* and *velocity* only, as the model also including a tempo effect failed when fitted to the NINFEA and GOCS data. This is likely to be due to lack of heterogeneity in GOCS, which only include term children, and to lack of sufficient growth observations in NINFEA. The fully specified model was instead successfully fitted to the GXXI cohort, which has the greater number of weight growth measurements.

As expected given the different source populations, we found some heterogeneities across the three cohorts, mainly in relation to the effect of maternal smoking and hypertension. Some of these variations were shown to be partly due to differences in inclusion criteria and could also derive from differences in quality and coarseness of the available data in particular in relation to pregnancy complications. An additional limitation could be that we have not investigated to what extent the associations found for the prenatal exposures could be explained by relevant early postnatal factors, such as feeding. However this is beyond the scope of this paper.

In summary, our findings are that growth trajectories in contemporary infants from economically and geographically diverse countries such as Portugal, Italy and Chile share some common features, in particular with respect to the effect of maternal height, maternal overweight/obesity

and parity. In the two European cohorts we also found interestingly separate effects of maternal underweight, smoking and hypertension on the child's *size* and *velocity*, and when growth data were rich and the effect on *tempo* could also be examined, we found that parity, maternal overweight/obesity and gestational hypertension had important effects on the timing of growth. Our analytical approach therefore succeeded in separating the relationships between prenatal maternal characteristics and infant growth into different components, and may inform new biological insights into the mechanisms governing infant growth.

1. Barker DJP. Mothers, Babies and Health in Later Life. 2nd ed. Edinburgh: Churchill Livingstone, 1998.
2. De Stavola BL, dos Santos Silva I, McCormack V, et al. Childhood growth and breast cancer. Am J Epidemiol 2004;159(7):671-82.
3. Gamborg M, Andersen PK, Baker JL, et al. Life course path analysis of birth weight, childhood growth, and adult systolic blood pressure. Am J Epidemiol 2009;169(10):1167-78.
4. Huxley RR, Shiell AW, Law CM. The role of size at birth and postnatal catch-up growth in determining systolic blood pressure: a systematic review of the literature. J Hypertens 2000;18(7):815-31.
5. Raikkonen K, Forsen T, Henriksson M, et al. Growth trajectories and intellectual abilities in young adulthood:The Helsinki Birth Cohort study. Am J Epidemiol 2009;170(4):447-55.
6. Whincup PH, Kaye SJ, Owen CG, et al. Birth weight and risk of type 2 diabetes: a systematic review. JAMA 2008;300(24):2886-97.
7. Baird J, Fisher D, Lucas P, et al. Being big or growing fast: systematic review of size and growth in infancy and later obesity. BMJ 2005;331(7522):929.
8. Monteiro PO, Victora CG. Rapid growth in infancy and childhood and obesity in later life--a systematic review. Obes Rev 2005;6(2):143-54.
9. Ong KK, 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.
10. Ong KK, Loos RJ. Rapid infancy weight gain and subsequent obesity: systematic reviews and hopeful suggestions. Acta Paediatr 2006;95(8):904-8.
11. Tzoulaki I, Sovio U, Pillas D, et al. Relation of immediate postnatal growth with obesity and related metabolic risk factors in adulthood: the northern Finland birth cohort 1966 study. Am J Epidemiol 2010;171(9):989-98.
12. Mihrshahi S, Battistutta D, Magarey A, Daniels LA. Determinants of rapid weight gain during infancy: baseline results from the NOURISH randomised controlled trial. BMC Pediatr 2011;11:99.
13. Ong KK, Preece MA, Emmett PM, et al. Size at birth and early childhood growth in relation to maternal smoking, parity and infant breast-feeding: longitudinal birth cohort study and analysis. Pediatr Res 2002;52(6):863-7.
14. Dubois L, Girard M. Early determinants of overweight at 4.5 years in a population-based longitudinal study. Int J Obes (Lond) 2006;30(4):610-7.
15. Monasta L, Batty GD, Cattaneo A, et al. Early-life determinants of overweight and obesity: a review of systematic reviews. Obes Rev 2010;11(10):695-708.
16. Ong KK. Early determinants of obesity. Endocr Dev 2010;19:53-61.
17. Fall CH. Evidence for the intra-uterine programming of adiposity in later life. Ann Hum Biol 2011;38(4):410-28.
18. Dietz PM, Callaghan WM, Cogswell ME, et al. Combined effects of prepregnancy body mass index and weight gain during pregnancy on the risk of preterm delivery. Epidemiology 2006;17(2):170-7.
19. Harville EW, Madkour AS, Xie Y. Predictors of birth weight and gestational age among adolescents. Am J Epidemiol 2012;176 Suppl 7:S150-63.
20. Hindmarsh PC, Geary MP, Rodeck CH, Kingdom JC, Cole TJ. Factors predicting ante- and postnatal growth. Pediatr Res 2008;63(1):99-102.

21. Kramer MS. Intrauterine growth and gestational duration determinants. Pediatrics 1987;80(4):502-11.
22. Martin RM, Smith GD, Frankel S, Gunnell D. Parents' growth in childhood and the birth weight of their offspring. Epidemiology 2004;15(3):308-16.
23. Beath KJ. Infant growth modelling using a shape invariant model with random effects. Stat Med 2007;26(12):2547-64.
24. Hui LL, Leung GM, Cowling BJ, Lam TH, Schooling CM. Determinants of infant growth: Evidence from Hong Kong's "Children of 1997" birth cohort. Ann Epidemiol 2010;20(11):827-35.
25. Touwslager RN, Gielen M, Derom C, et al. Determinants of infant growth in four age windows: a twin study. J Pediatr 2011;158(4):566-572 e2.
26. Bulk-Bunschoten AM, van Bodegom S, Reerink JD, de Jong PC, de Groot CJ. Weight and weight gain at 4 months (The Netherlands 1998): influences of nutritional practices, socio-economic and ethnic factors. Paediatr Perinat Epidemiol 2002;16(4):361-9.
27. Regnault N, Botton J, Forhan A, et al. Determinants of early ponderal and statural growth in full-term infants in the EDEN mother-child cohort study. Am J Clin Nutr 2010;92(3):594-602.
28. Cole TJ, Donaldson MD, Ben-Shlomo Y. SITAR--a useful instrument for growth curve analysis. Int J Epidemiol 2010;39(6):1558-66.
29. Johnson L, Llewellyn CH, van Jaarsveld CH, Cole TJ, Wardle J. Genetic and environmental influences on infant growth: prospective analysis of the Gemini twin birth cohort. PLoS One 2011;6(5):e19918.
30. Prentice A, Dibba B, Sawo Y, Cole TJ. The effect of prepubertal calcium carbonate supplementation on the age of peak height velocity in Gambian adolescents. Am J Clin Nutr 2012.
31. Pizzi C, De Stavola BL, Pearce N, et al. Selection bias and patterns of confounding in cohort studies: the case of the NINFEA web-based birth cohort. J Epidemiol Community Health 2011.
32. Richiardi L, Baussano I, Vizzini L, et al. Feasibility of recruiting a birth cohort through the Internet: the experience of the NINFEA cohort. Eur J Epidemiol 2007;22(12):831-7.
33. Corvalan C, Uauy R, Stein AD, Kain J, Martorell R. Effect of growth on cardiometabolic status at 4 y of age. Am J Clin Nutr 2009;90(3):547-55.
34. Pizzi C, Cole TJ, Corvalan C, et al. On modelling early life weight trajectories. J R Stat Soc Ser A Stat Soc. IN PRESS.
35. Skrondal A, Rabe-Hesketh S. Generalized latent variable modeling: multilevel, longitudinal, and structural equation models. New York: Chapman & Hall/CRC, 2004.
36. Rubin D. Inference and missing data. Biometrika 1976;63:581-592.
37. Cole TJ. Sympercents: symmetric percentage differences on the 100 log(e) scale simplify the presentation of log transformed data. Stat Med 2000;19(22):3109-25.
38. Vohr BR, McGarvey ST, Tucker R. Effects of maternal gestational diabetes on offspring adiposity at 4-7 years of age. Diabetes Care 1999;22(8):1284-91.
39. Duley L. The global impact of pre-eclampsia and eclampsia. Semin Perinatol 2009;33(3):130-7.
40. Howe LD, Tilling K, Galobardes B, et al. Socioeconomic disparities in trajectories of adiposity across childhood. Int J Pediatr Obes 2011;6(2-2):e144-53.

TABLE 1. Descriptive Statistics of the Main Variables by Cohorts

	GXXI (N=738)		NINFEA (N=2,925)		GOCS (N=959)	
	N [a]	%	N [a]	%	N [a]	%
Child characteristics						
Mean gestational age *(weeks ± SD)*	738	39.1 ± 1.6	2,925	39.6 ± 1.6	959	39.6 ± 1.3
Gender						
Female	365	49.5	1,441	49.3	487	49.2
Male	373	50.5	1,484	50.7	472	50.8
Maternal characteristics						
Mean height *(cm ± SD)*	629	161.5 ± 5.9	2,836	164.7 ± 6.1	903	156.9 ± 5.8
Mean age *(years ± SD)*	737	30.3 ± 5.1	2,925	33.5 ± 4.1	888	27.0 ± 6.9
Parity [b]						
Nulliparous	462	62.9	2,105	74.1	373	58.1
Parous	272	37.1	737	25.9	517	41.9
Educational level [c]						
Low	362	49.7	147	5.1	323	36.3
Medium	172	23.6	1,053	36.4	383	43.0
High	194	26.7	1,690	58.5	184	20.7
Pre-pregnancy BMI						
<18.5	30	4.9	235	8.3	34	5.1
18.5-24.99	376	60.7	2,060	72.8	395	59.8
25+	213	34.4	533	18.9	232	35.1
Smoke during pregnancy [d]						
No	574	79.5	2,632	91.6	809	91.0
Up to 1st trimester	53	7.3	51	1.8	80	9.0
After 1st trimester	95	13.2	190	6.6	--	--
Pregnancy complications [e]						
Gestational diabetes						
No	560	92.3	2,506	92.0	913	95.2
Yes	47	7.7	218	8.0	46	4.8
Hypertension/eclampsia						
No	576	95.2	2,498	91.8	878	91.6
Yes	29	4.8	222	8.2	81	8.4

[a] Total N might vary across variables due to missing values
[b] In GOCS child order is used as a proxy for parity
[c] GXXI: Low=≤9 years, Medium=9-12 years, High=Degree or more; NINFEA : Low=None/Primary/Secondary school, Medium=High school, High=Degree or more; GOCS : Low=None/Primary/Secondary school, Medium=High school, High=High School + technical education or more
[d] In GOCS smoking during pregnancy is categorized as No/Rarely vs Frequently
[e] Mothers suffering from these diseases before pregnancy (information available only in GXXI and NINFEA) classified as "No"

TABLE 2. Estimated Coefficients and 95% Confidence Interval for the Association Between Covariates and Size and Velocity Parameters by Cohorts

	GXXI (N = 605)								**NINFEA** (N = 2,734)								**GOCS** (N = 659)							
	Minimally-adjusted [a]				Fully-adjusted [b]				Minimally-adjusted [a]				Fully-adjusted [b]				Minimally-adjusted [a]				Fully-adjusted [b]			
	Size		**Velocity**		**Size**		**Velocity**		**Size**		**Velocity**		**Size**		**Velocity**		**Size**		**Velocity**		**Size**		**Velocity**	
	%	*95%CI*	*%*	*95%CI*	*%*	*95%CI*	*%*	*95%CI*	*%*	*95%CI*	*%*	*95%CI*	*%*	*95%CI*	*%*	*95%CI*	*%*	*95%CI*	*%*	*95%CI*	*%*	*95%CI*	*%*	*95%CI*
Background																								
Maternal height	0.4	0.2; 0.5	0.2	-0.2; 0.6	0.3	0.2; 0.5	0.2	-0.2; 0.6	0.4	0.3; 0.5	0.02	-0.2; 0.2	0.4	0.3; 0.5	0.1	-0.1; 0.2	0.4	0.2; 0.5	-0.03	-0.6; 0.1	0.4	0.2; 0.5	-0.3	-0.7; 0.02
Maternal age	0.1	-0.1; 0.3	0.2	-0.3; 0.7	-0.1	-0.3; 0.1	0.5	0.0; 1.0	0.1	-0.03; 0.2	-0.2	-0.5; 0.1	0.02	-0.1; 0.1	-0.1	-0.3; 0.2	0.1	0.01; 0.2	-0.2	-0.4; - 0.1	0.1	0.0; 0.3	-0.1	-0.4; 0.3
Maternal parity [c]																								
Nulliparous	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Parous	3.4	1.2; 5.5	-3.5	-8.4; 1.5	4.5	2.2; 6.8	-6.1	-11.6; -0.6	3.0	2.0; 4.0	-5.6	-8.2; -3.1	2.8	1.8; 3.9	-5.8	-8.4; -3.2	2.0	0.3; 3.6	-3.7	-7.4; 0.3	1.0	-1.0; 2.9	-2.1	-6.7; 2.6
Maternal education [d]																								
Low	--	--	--	--	--	--	--	--	-1.5	-3.7; 0.8	2.9	-2.7; 8.6	-0.4	-2.5; 1.7	2.8	-2.6; 8.1	0.7	-1.1; 2.5	-5.7	-10.0; -1.4	0.6	-1.2; 2.4	-5.5	-9.9; -1.1
Medium	3.3	0.7; 5.9	-0.4	-6.4; 5.6	3.3	0.8; 5.9	-1.2	-7.3; 4.8	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
High	2.3	-0.1; 4.8	-2.4	-8.1; 3.2	2.9	0.4; 5.5	-5.4	-11.4; 0.6	0.6	-1.6; 0.4	-0.6	-3.0; 1.9	-0.8	-1.7; 0.2	-0.7	-3.0; 1.6	-0.7	-2.8; 1.5	-0.9	-5.9; 4.2	-1.2	-3.3; 0.9	-0.5	-5.6; 4.6
Intermediate																								
Pre-pregnancy BMI																								
<18.5	-4.4	-9.2; 0.4	2.5	-8.7; 13.6	-4.2	-8.9; 0.5	-0.7	-11.8; 10.4	-4.4	-6.0; -2.8	5.2	1.2; 9.2	-4.1	-5.7; -2.6	4.4	0.4; 8.3	-3.6	-7.2; 0.1	-5.9	-14.6; 2.9	-4.3	-7.8; -0.1	-4.5	-13.4; 4.3
18.5-24.99	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
25+	3.4	1.2; 5.6	0.9	-4.1; 5.9	4.0	1.8; 6.1	0.7	-4.3; 5.8	2.0	0.9; 3.1	-1.6	-4.4; 1.2	2.1	0.9; 3.2	-2.1	-4.9; 0.7	1.7	-0.01; 3.4	-1.4	-5.4; 2.6	1.8	0.04; 3.5	-0.3	-4.5; 3.9
Maternal smoking [e]																								
No	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
≤1st trimester	-2.8	-6.9; 1.3	6.6	-2.7; 15.9	-1.4	-5.3; 2.5	6.6	-2.7; 15.9	0.7	-2.9; 4.2	4.1	-4.7; 12.8	0.2	-3.1; 3.5	4.6	-3.7; 12.9	1.2	-1.7; 4.1	-7.4	-14.1; -0.7	0.6	-2.2; 3.5	-5.1	-12.0; 1.8
>1st trimester	-3.8	-6.8; -0.7	12.8	5.8; 19.8	-2.9	-5.8; 0.1	13.2	6.1; 20.3	-3.2	-5.1; -1.3	8.4	3.6; 13.2	-2.6	-4.3; -0.9	6.5	2.0; 10.9								

[a] Estimates derived from model adjusted for gender and gestational age fitted on the sample of data with no missing values for the following maternal variables: height, age, parity, educational level, pre-pregnancy BMI and smoking during pregnancy.

[b] Background variables are mutually adjusted and further adjusted for gender and gestational age; intermediate variables are mutually adjusted and further adjusted for background variables, gender and gestational age

[c] In GOCS child order was used as a proxy for parity

[d] GXXI: Low=≤9 years, Medium=≤12 years, High=Degree or higher; NINFEA : Low= ≤Secondary school, Medium=High school, High=Degree or higher ; GOCS : Low= None Primary Secondary school, Medium=High school, High=High School + technical education or higher

[e] In GOCS smoking during pregnancy was categorized as No Rarely vs Frequently

TABLE 3: Estimated Coefficients and 95% Confidence Interval for the Association Between Pregnancy Complications and Size and Velocity Parameters by Cohorts

	GXXI (N = 492)								**NINFEA** (N = 2,523)								**GOCS** (N = 659)							
	Minimally-adjusted [a]				Fully-adjusted [b]				Minimally-adjusted [a]				Fully-adjusted [b]				Minimally-adjusted [a]				Fully-adjusted [b]			
	Size		**Velocity**		**Size**		**Velocity**		**Size**		**Velocity**		**Size**		**Velocity**		**Size**		**Velocity**		**Size**		**Velocity**	
	%	*95%CI*	%	*95%CI*	%	*95%CI*	%	*95%CI*	%	*95%CI*	%	*95%CI*	%	*95%CI*	%	*95%CI*	%	*95%CI*	%	*95%CI*	%	*95%CI*	%	*95%CI*
Diabetes	0.8	-3.8; 5.3	4.6	-5.9; 15.0	0.6	-3.8; 4.9	5.7	-4.5; 16.0	2.0	0.3; 3.7	-5.1	-9.3; -0.9	1.5	-0.3; 3.3	-3.1	-7.6; 1.3	0.6	-3.1; 4.3	4.1	-4.7; 12.8	-0.6	-4.2; 3.1	7.0	-1.8; 15.8
Hypertension	-4.9	-10.6; 5.5	12.3	-0.6; 25.2	-6.4	-11.7; -1.0	12.8	0.3; 25.3	-3.8	-5.5; -2.0	8.7	4.4; 13.0	-4.5	-6.2; -2.7	9.0	4.7; 13.4	2.2	-0.7; 5.1	-3.8	-10.6; 3.0	1.3	-1.4; 4.2	-4.0	-10.9; 2.9

[a] Estimates derived from model adjusted for gender and gestational age fitted on the sample of data with no missing values for the following maternal variables: height, age, parity, educational level, pre-pregnancy BMI, smoking during pregnancy, gestational diabetes and gestational hypertension.

[b] Adjusted by sex, gestational age and maternal height, age, parity, educational level, pre-pregnancy BMI and smoking during pregnancy.

TABLE 4. Fully-Adjusted Estimated Coefficients and 95% Confidence Interval for the Association Between Covariates and Size and Velocity Parameters on the Pooled GXXI & NINFEA Datasets

	GXXI+NINFEA (N = 3,339) [a]			
	Size		Velocity	
	%	*95%CI*	*%*	*95%CI*
Background [b]				
Maternal height	0.4	0.3; 0.5	0.1	-0.1; 0.2
Maternal age	0.02	-0.1; 0.1	0.02	-0.2; 0.3
Maternal parity				
Nulliparous	--	--	--	--
Parous	3.1	2.1; 4.1	-5.5	-8.0; -3.0
Maternal education [c]				
Low	-2.2	-3.7; -0.6	2.4	-1.4; 6.2
Medium	--	--	--	--
High	-0.9	-1.8; 0.1	-0.7	-3.0; 1.6
Intermediate [d]				
Pre-pregnancy BMI				
<18.5	-4.9	-6.5; -3.3	5.9	1.9; 10.0
18.5-24.99	--	--	--	--
25+	2.4	1.4; 3.5	-0.5	-3.1; 2.1
Maternal smoking				
No	--	--	--	--
≤1st trimester	-0.9	-3.4; 1.7	5.9	-0.5; 12.3
>1st trimester	-3.3	-4.8; -1.7	10.2	6.3; 14.3
	(N = 3,015) [e]			
Gestational diabetes	1.1	-0.6; 2.7	-0.8	-5.0; 3.3
Gestational hypertension	-5.2	-6.9; -3.5	10.6	6.4; 14.8

[a] Model fitted on the sample of data with no missing values for the following maternal variables: height, age, parity, educational level, pre-pregnancy BMI and smoking during pregnancy

[b] Background variables are mutually adjusted and further adjusted for gender and gestational age

[c] GXXI: Low=≤9 years, Medium=≤12 years, High=Degree or higher; NINFEA : Low= ≤Secondary school, Medium=High school, High=Degree or higher

[d] Intermediate variables are mutually adjusted and further adjusted for background variables, gender and gestational age

[e] Model fitted on the sample of data with no missing values for the maternal variables: height, age, parity, educational level, pre-pregnancy BMI, smoking during pregnancy, gestational diabetes and gestational hypertension

TABLE 5. Fully-Adjusted Estimated Coefficients and 95% Confidence Interval for the Association Between Covariates and Size, Tempo and Velocity Parameters on the GXXI Data

	GXXI (N = 605) [a]					
	Size		Tempo		Velocity	
	%	95%CI	β [b]	95%CI	α_a	95%CI
Background [c]						
Maternal height	0.5	0.2; 0.7	0.01	-0.01; 0.02	-0.05	-0.5; 0.4
Maternal age	-0.1	-0.4; 0.2	-0.01	-0.02; 0.01	0.6	0.01; 1.2
Maternal parity						
Nulliparous	--	--	--	--	--	--
Parous	1.0	-2.2; 4.3	-0.17	-0.34; -0.01	-1.1	-7.4; 5.2
Maternal education [d]						
Low	--	--	--	--	--	--
Medium	3.6	0.1; 7.2	0.003	-0.18; 0.19	-2.2	-9.2; 4.7
High	3.6	0.1; 7.1	0.08	-0.11; 0.26	-7.0	-13.9; -0.1
Intermediate [e]						
Pre-pregnancy BMI						
<18.5	-5.9	-12.3; 0.5	-0.09	-0.41; 0.23	2.6	-10.2; 15.4
18.5-24.99	--	--	--	--	--	--
25+	7.9	4.9; 10.8	0.26	0.11; 0.41	-4.9	-10.8; 0.9
Maternal smoking						
No	--	--	--	--	--	--
≤1st trimester	-0.9	-6.4; 4.5	-0.02	-0.29; 0.25	6.2	-4.6; 16.9
>1st trimester	-2.8	-6.9; 1.3	-0.08	-0.28; 0.12	14.0	5.8; 22.2
			(N = 492) [g]			
Gestational diabetes	3.1	-3.9; 10.1	0.06	-0.13; 0.24	0.7	-12.7; 14.2
Gestational hypertension	-3.5	-12.2; 5.3	0.31	0.08; 0.53	8.7	-7.9; 25.3

[a] Model fitted on the sample of data with no missing values for the following maternal variables: height, age, parity, educational level, pre-pregnancy BMI and smoking during pregnancy
[b] Model is on the log-weight and age scales, thus the effect on tempo is on the age unit (months).
[c] Background variables are mutually adjusted and further adjusted for gender and gestational age
[d] GXXI: Low=≤9 years, Medium=≤12 years, High=Degree or higher
[e] Intermediate variables are mutually adjusted and further adjusted for the background variables, gender and gestational age
[g] Model fitted on the sample of data with no missing values for the following maternal variables: height, age, parity, educational level, and pre-pregnancy BMI; smoking during pregnancy, gestational diabetes and gestational hypertension

eTABLE 1. "Minimally-Adjusted" Estimated Coefficients and 95% Confidence Interval for the Association Between Covariates and Size and Velocity Parameters by Cohorts

	GXXI (n = 738)								**NINFEA** (n = 2,925)								**GOCS** (n = 959)							
	Adjusted for gender				Adjusted for gender & gestational age				Adjusted by gender				Adjusted for gender & gestational age				Adjusted for gender				Adjusted for gender & gestational age			
	Size		**Velocity**		**Size**		**Velocity**		**Size**		**Velocity**		**Size**		**Velocity**		**Size**		**Velocity**		**Size**		**Velocity**	
	%	*95%CI*	%	*95%CI*	%	*95%CI*	%	*95%CI*	%	*95%CI*	%	*95%CI*	%	*95%CI*	%	*95%CI*	%	*95%CI*	%	*95%CI*	%	*95%CI*	%	*95%CI*
Background																								
Maternal height	0.4	0.2; 0.6	0.02	-0.4; 0.4	0.4	0.2; 0.5	0.1	-0.3; 0.5	0.4	0.4; 0.5	0.1	-0.2; 0.1	0.4	0.3; 0.5	0.02	-0.2; 0.2	0.3	0.2; 0.4	-0.1	-0.4; 0.2	0.3	0.2; 0.4	-0.1	-0.4; 0.2
Maternal age	-0.01	-0.2; 0.2	0.3	-0.1; 0.8	0.1	-0.1; 0.2	0.2	-0.2; 0.6	-0.1	-0.2; 0.1	0.1	-0.2; 0.3	0.1	-0.02; 0.2	-0.2	-0.5; 0.04	0.1	0.00; 0.2	-0.2	-0.5; 0.04	0.2	0.1; 0.3	-0.3	-0.6; - 0.1
Maternal parity [a]																								
Nulliparous	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Parous	3.5	1.2; 5.7	-2.9	-7.9; 1.9	3.2	1.2; 5.1	-2.4	-6.9; 2.0	1.2	0.02; 2.3	-1.6	-4.4; 1.1	3.0	2.0; 4.0	-5.7	-8.2; -3.2	1.6	0.2; 3.1	-3.6	-7.1; -0.1	2.3	0.9; 3.7	-4.9	-8.3; -1.6
Maternal education [b]																								
Low	--	--	--	--	--	--	--	--	0.9	-1.5; 3.3	-1.6	-7.1; 4.0	-1.3	-3.5; 0.9	2.5	-3.0; 7.9	0.4	-1.2; 2.1	-4.2	-8.1; -0.2	0.3	-1.3; 1.9	-3.7	-7.4; 0.08
Medium	2.7	0.0; 5.5	-1.9	-7.9; 4.1	3.4	1.0; 5.7	-2.7	-8.1; 2.7	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
High	0.2	-2.4; 2.8	-0.8	-6.6; 4.9	1.8	-0.5; 4.1	-3.8	-8.9; 1.4	-0.2	-1.2; 0.9	-1.4	-3.9; 1.0	-0.4	-1.4; 0.6	-1.1	-3.5; 1.3	-0.7	-2.7; 1.2	-0.9	-5.6; 3.8	-0.9	-2.7; 1.0	-0.6	-5.0; 3.9
Intermediate																								
Pre-pregnancy BMI																								
<18.5	-7.6	-13.2; -2.1	8.8	-3.5; 21.1	-4.4	-9.2; 0.5	2.5	-8.6; 13.6	-4.6	-6.4; -2.8	5.3	1.0; 9.6	-5.2	-6.9; -3.6	6.8	2.6; 10.9	-4.1	-7.9; -0.3	-4.8	-13.9; 4.4	-3.6	-7.2; 1.0	-5.9	-14.7; 2.8
18.5-24.99	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
25+	4.3	1.8; 6.8	-0.3	-5.8; 5.2	3.4	1.2; 5.6	1.1	-3.8; 6.1	2.5	1.2; 3.8	-3.0	-6.0; 0.01	2.1	0.9; 3.2	-1.9	-4.9; 0.9	1.6	-0.1; 3.4	-0.7	-5.0; 3.5	1.8	0.1; 3.5	-1.4	-5.4; 2.6
Maternal Smoking [c]																								
No	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
≤1st trimester	-4.1	-8.3; 0.1	5.9	-3.3; 15.1	-3.1	-6.7; 0.5	4.1	-4.1; 12.3	0.6	-3.2; 4.4	4.1	-4.8; 13.0	0.8	-2.7; 4.3	3.4	-5.2; 12.0	-1.7	-4.2; 0.9	-0.5	-6.5; 5.6	-0.9	-3.3; 1.5	-3.1	-8.9; 2.7
>1st trimester	-3.3	-6.5; -0.04	11.6	4.4; 18.7	-4.1	-6.8; -1.3	12.8	6.4; 19.1	-2.1	-4.1; -0.1	6.1	1.3; 10.9	-2.9	-4.8; -1.1	8.2	3.6; 12.9								
Gestational diabetes																								
No	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Yes	1.4	-2.9; 5.8	-0.8	-10.5; 8.9	0.9	-3.0; 4.7	0.5	-8.2; 9.3	0.4	-1.5; 2.3	-1.9	-6.2; 2.5	1.3	-0.4; 3.1	-3.7	-8.0; 0.5	2.4	-0.9; 5.7	-1.9	-9.9; 5.9	3.1	-0.1; 6.2	-3.7	-11.3; 3.4
Gestational hypertension																								
No	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Yes	-8.3	-13.7; -2.9	19.4	7.6; 31.3	-3.0	-7.8; 1.9	9.7	-1.2; 20.6	-5.4	-7.3; -3.6	12.4	8.1; 16.7	-3.1	-4.7; -1.5	8.2	4.3; 12.1	0.7	-1.8; 3.2	0.6	-5.3; 6.5	1.5	-0.8; 3.9	-1.2	-6.9; 4.4

[a] In GOCS child order was used as a proxy for parity

[b] GXXI: Low=≤9 years, Medium=≤12 years, High=Degree or higher; NINFEA : Low= ≤Secondary school, Medium=High school, High=Degree or higher ; GOCS : Low= None Primary Secondary school, Medium=High school, High=High School + technical education or higher

[c] In GOCS smoking during pregnancy was categorized as No Rarely vs. Frequently

5.3.1 Addendum to Research Paper IV

In order to display more clearly the effect sizes associated with the exposure variables in Tables 2 and 5 of Research Paper IV, graphs showing the predicted trajectories for children with different combinations of the covariates are presented below. In particular one graph is drawn for each exposure of interest, showing the predicted weight curves for different values of that exposure, keeping all other covariates at their reference values, and using the estimates derived from the models of Table 5 (therefore only results obtained for the GXXI study are shown). These are the models which allow for the association between the exposures and the three parameters, size, tempo and velocity. In order to assess whether the inclusion of an effect on the tempo dimension (Table 5 of Research Paper IV) results in different predicted curves compared with those derived when an effect on the size and velocity parameters only is allowed for (Table 2 of Research Paper IV), the same graphs are drawn using the estimates derived from the models of Table 2 for those exposures for which a significant association with tempo was found. These covariates are: parity, pre-pregnancy BMI and gestational hypertension.

The following reference values/categories were used to draw the graphs: mean maternal height (161 cm), mean maternal age (30 years), low maternal education, no maternal smoking during pregnancy, no gestational diabetes, nulliparity, pre-pregnancy BMI between 18.5 and 24.99 and no gestational hypertension.

Figures 5.1 to 5.7 show the predicted weight trajectories associated with the effect sizes displayed in Table 5 of Research Paper IV for those variables for which no association with the tempo parameter was observed. Namely Figure 5.1 shows the predicted curves for different levels of maternal height (161 cm vs 166 cm), with the predicted curve for a maternal height of 166 cm lying slightly above the other one. This is in line with the result reported in Research Paper IV, where maternal height was found to affect only the size dimension (δ_α=0.5% in Table 5). No evidence of an effect of maternal age on the weight growth dimensions was found and this is confirmed by the predicted trajectories of Figure 5.2, which overlap.

Figures 5.3 and 5.4 display the effects associated with maternal education. In Research Paper IV it was observed that both medium and highly educated women have bigger children (δ_α=3.6% for both categories in Table 5) who tend to have slower growth velocity, with the latter result being

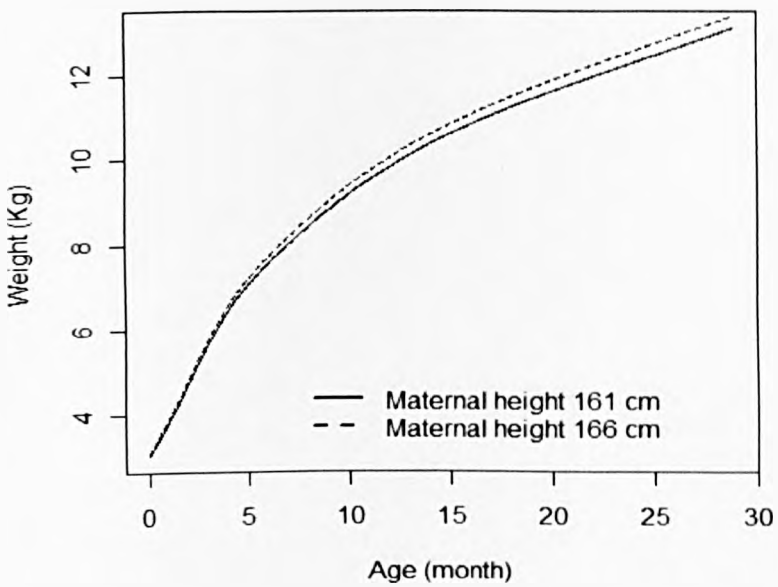


Figure 5.1: Predicted weight curves for different values of maternal height from the model allowing for the association between the covariates and **size, tempo and velocity** parameters fitted on the GXXI data.

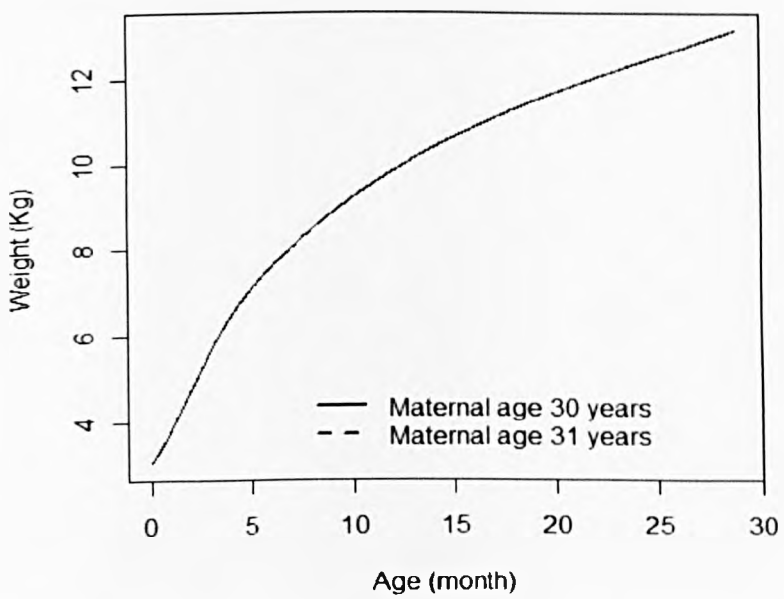


Figure 5.2: Predicted weight curves for different values of maternal age from the model allowing for the association between the covariates and **size, tempo and velocity** parameters fitted on the GXXI data.

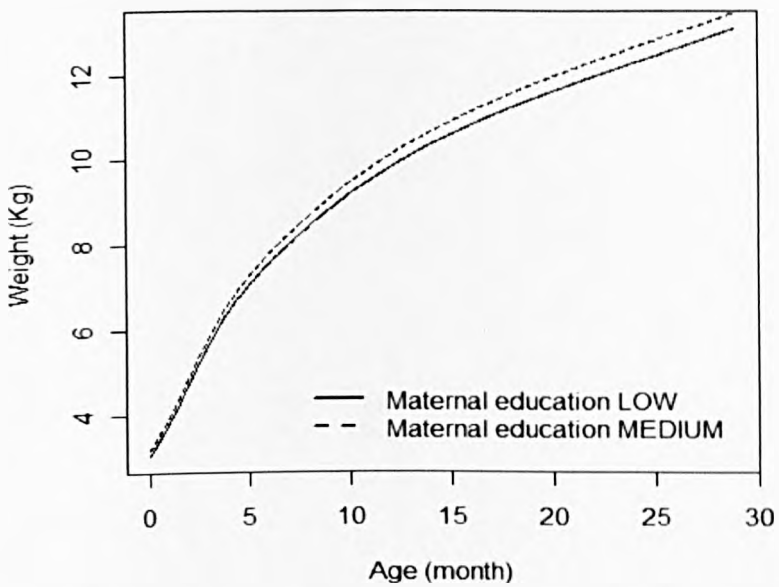


Figure 5.3: Predicted weight curves for different categories of maternal education from the model allowing for the association between the covariates and **size, tempo and velocity** parameters fitted on the GXXI data.

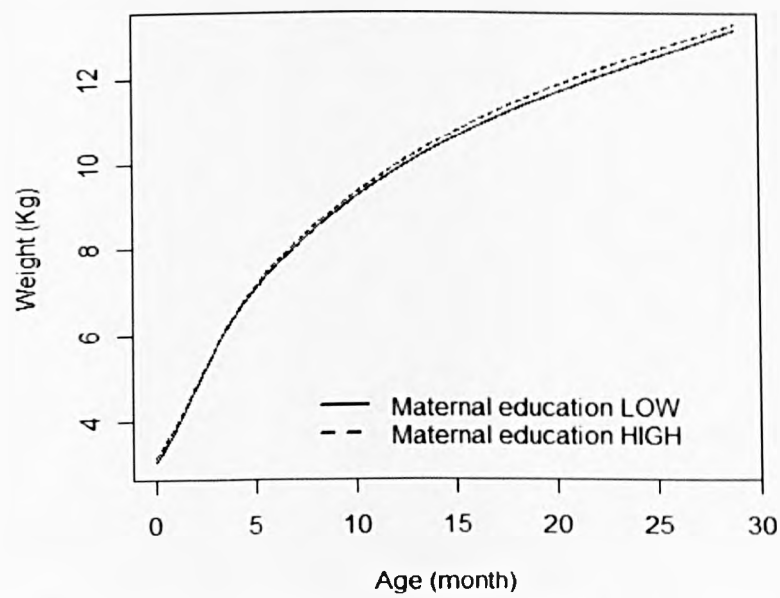


Figure 5.4: Predicted weight curves for different categories of maternal education from the model allowing for the association between the covariates and **size, tempo and velocity** parameters fitted on the GXXI data.

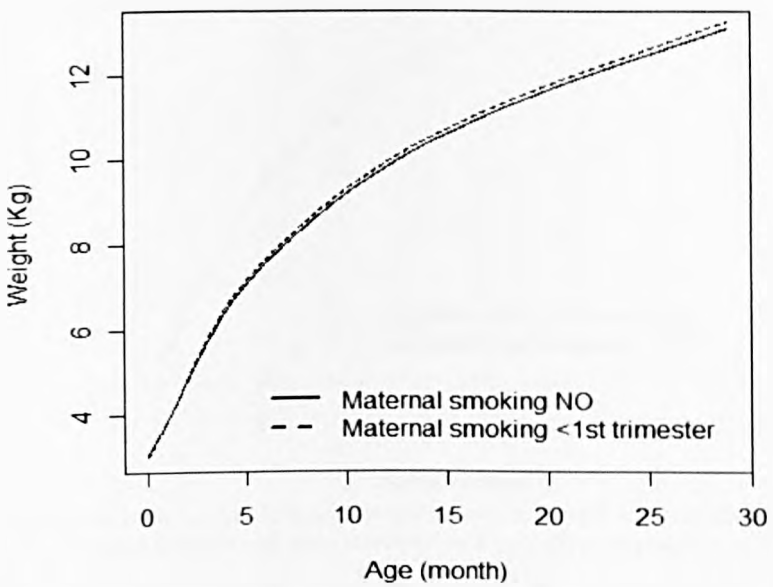


Figure 5.5: Predicted weight curves for different categories of maternal smoking from the model allowing for the association between the covariates and **size, tempo and velocity** parameters fitted on the GXXI data.

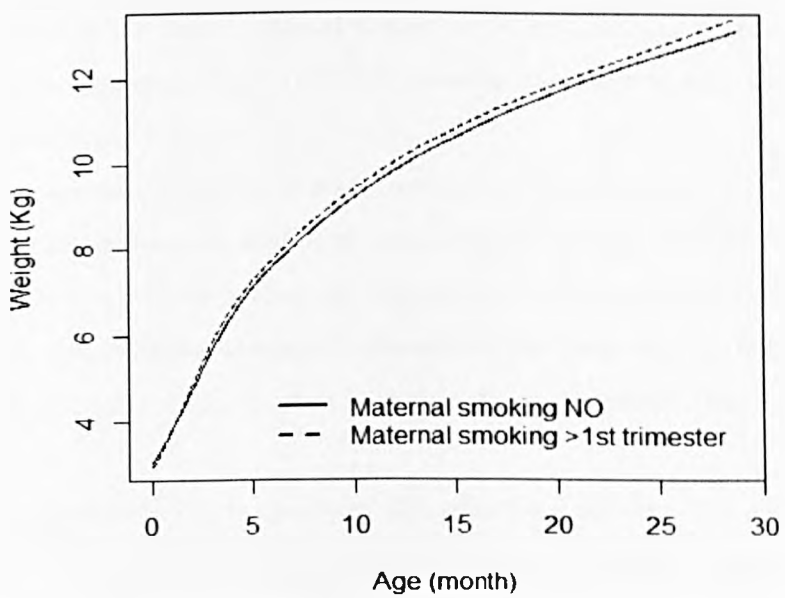


Figure 5.6: Predicted weight curves for different categories of maternal smoking from the model allowing for the association between the covariates and **size, tempo and velocity** parameters fitted on the GXXI data.

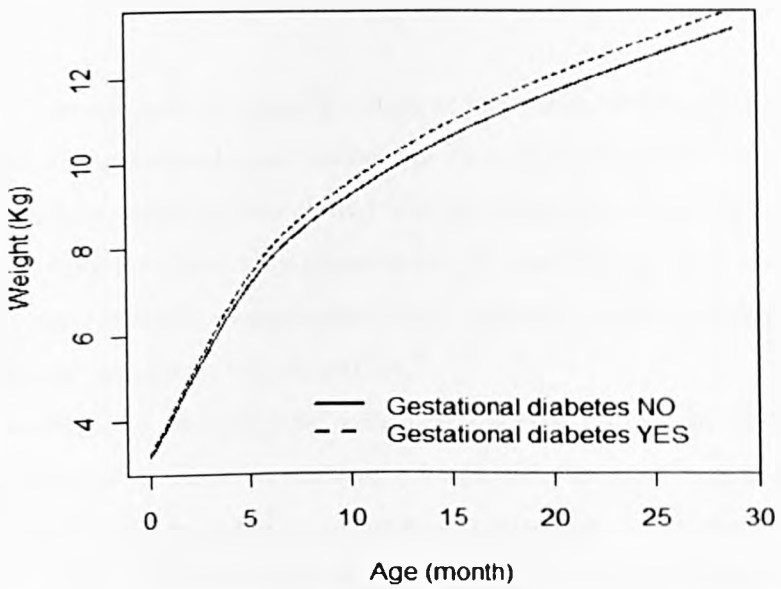


Figure 5.7: Predicted weight curves for different categories of gestational diabetes from the model allowing for the association between the covariates and **size, tempo and velocity** parameters fitted on the GXXI data.

greater among the highly educated women (δ_γ=-7.0% for highly educated vs δ_γ=-2.2% for medium educated, see Table 5). The result concerning the slower growth rate of the children with an highly educated mother becomes evident when comparing Figure 5.4 and Figure 5.3: the difference between the predicted curves for the highly educated women vs the low educated women is clearly reduced after 6 months of life compared with the difference between the trajectories for the medium educated women vs the low educated women.

Maternal smoking was associated with reduced size and increased velocity, with these effects being much stronger for those women who kept on smoking after the first trimester of the pregnancy. Again, these findings are in line with the trajectories reported in Figures 5.5 and 5.6, which show how the predicted curve for children exposed to maternal smoking during pregnancy lies below that of those not exposed for the first months of life, but then lies above after 4/5 months of life due to their increased velocity.

Finally Figure 5.7 displays the effects associated with gestational diabetes. The predicted weight curve for children exposed to gestational diabetes lies above the other one, likely reflecting the association that was observed, albeit being not significant, between gestational diabetes and size (δ_α=3.1% in Table 5).

Figures 5.8 to 5.15 concern those exposures for which an association with tempo was observed: parity, pre-pregnancy BMI and gestational hypertension. For each of these exposures two graphs are drawn: one showing the predicted weight curves derived from the models of Table 2 of Research Paper IV – which are those allowing for a covariates effect on the size and velocity parameters only – and one showing the predicted trajectories corresponding to the estimates of Table 5, where a covariates effect on all the three SITAR parameters was allowed for.

In Research Paper IV parity was associated with increased size and reduced velocity (Table 2), but when tempo was modeled in terms of covariates it was found that most part of the impact on size and velocity was captured by its influence on the tempo dimension. In particular, infants of parous mothers were found to have an earlier timing of growth (Table 5 of Research Paper IV). This is reflected in Figure 5.8 (corresponding to estimates reported in Table 2) and in Figure 5.9 (corresponding to estimates reported in Table 5): while in the first graph the curve related to the children with a parous mother lies slightly above the other one (nulliparous mother) soon after birth, with the two curves overlaping and diverging after about 5 months of life due to their decreased velocity (Figure 5.8), in the second graph the predicted curve for the children with a parous mother is slightly shifted to the left of the curve for the nulliparous mother over the full age range analyzed (Figure 5.9), reflecting the negative tempo effect.

Mother with a pre-pregnancy BMI below 18.5 were found to have children with a reduced size and no effect on the velocity dimension (δ_α=-4.2% in Table 2), and similar results when tempo was included in the model (δ_α=-5.9% in Table 5). As a consequence, the predicted curves of Figures 5.10 and 5.11 are almost identical. In contrast, Figures 5.12 and 5.13 slightly diverge. This is because infants of overweight/obese mothers were found to have a significantly delayed tempo compared with infants of mothers with a pre-pregnancy BMI between 18.5 and 25. In particular while in Table 2 of Research Paper IV maternal overweight/obesity was observed to affect only the size dimension (δ_α=4.0%), estimates of Table 5 showed an association with all the three dimensions (δ_α=7.9%, δ_β=0.26 months and δ_γ=-4.9%). The two models lead to a slightly different prediction especially after the first year of life.

Finally Figures 5.14 and 5.15 displays the effects associated with gestational hypertension. As observed

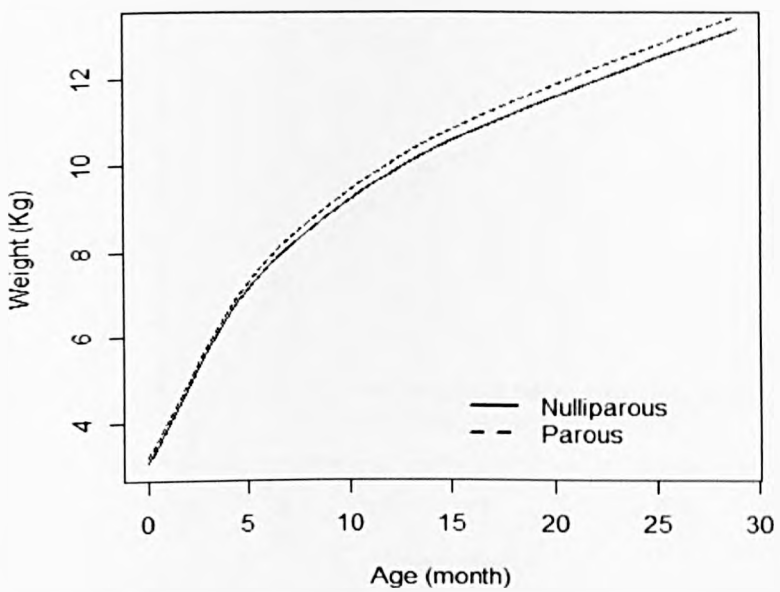


Figure 5.8: Predicted weight curves for different categories of maternal parity from the model allowing for the association between the covariates and **size and velocity** parameters fitted on the GXXI data.

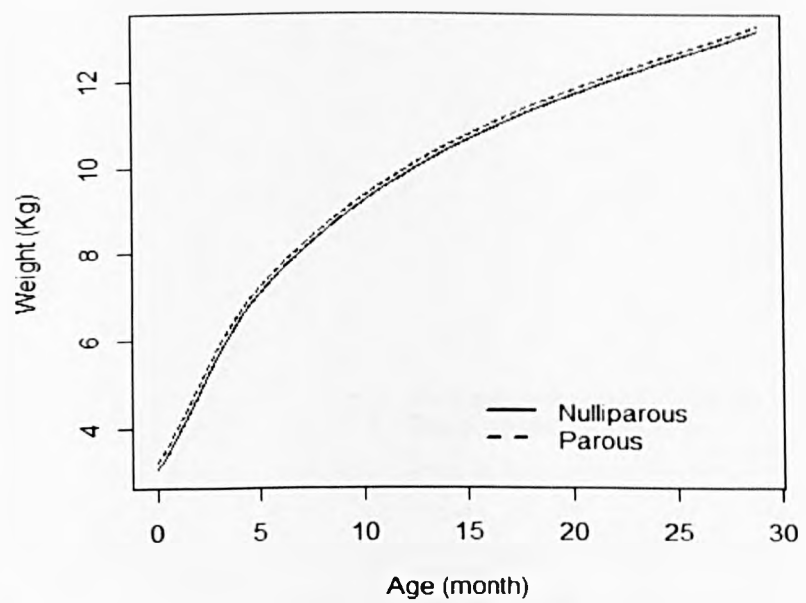


Figure 5.9: Predicted weight curves for different categories of maternal parity from the model allowing for the association between the covariates and **size, tempo and velocity** parameters fitted on the GXXI data.

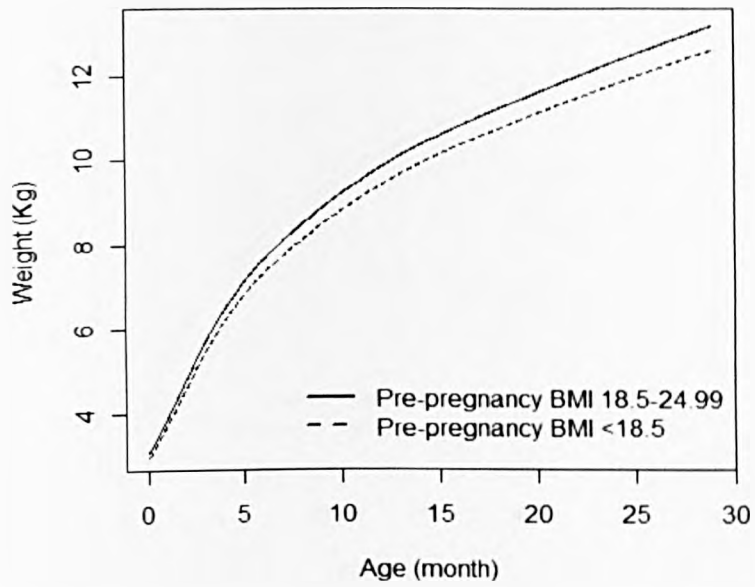


Figure 5.10: Predicted weight curves for different categories of maternal pre-pregnancy BMI from the model allowing for the association between the covariates and **size and velocity** parameters fitted on the GXXI data.

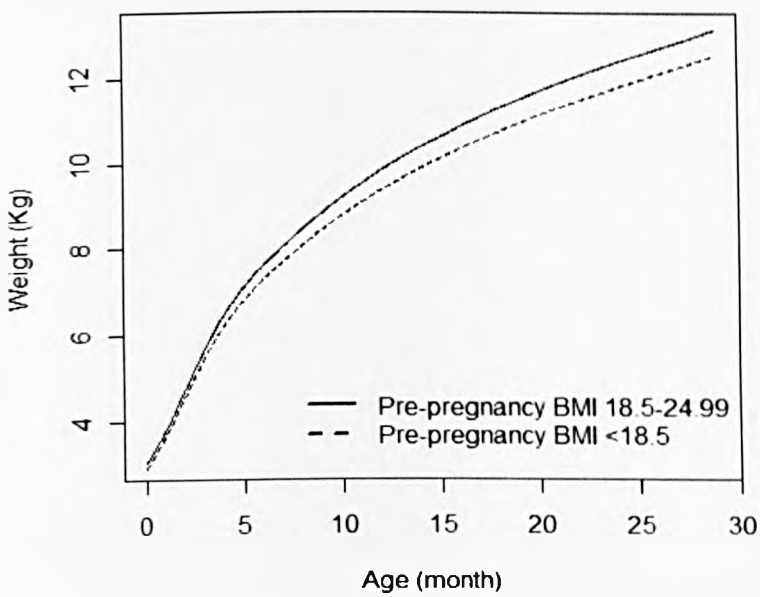


Figure 5.11: Predicted weight curves for different categories of maternal pre-pregnancy BMI from the model allowing for the association between the covariates and **size, tempo and velocity** parameters fitted on the GXXI data.

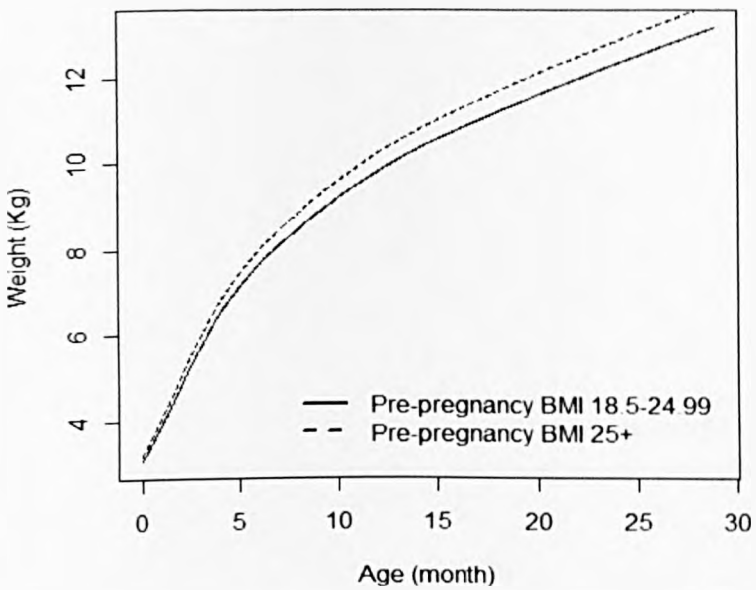


Figure 5.12: Predicted weight curves for different categories of maternal pre-pregnancy BMI from the model allowing for the association between the covariates and **size and velocity** parameters fitted on the GXXI data.

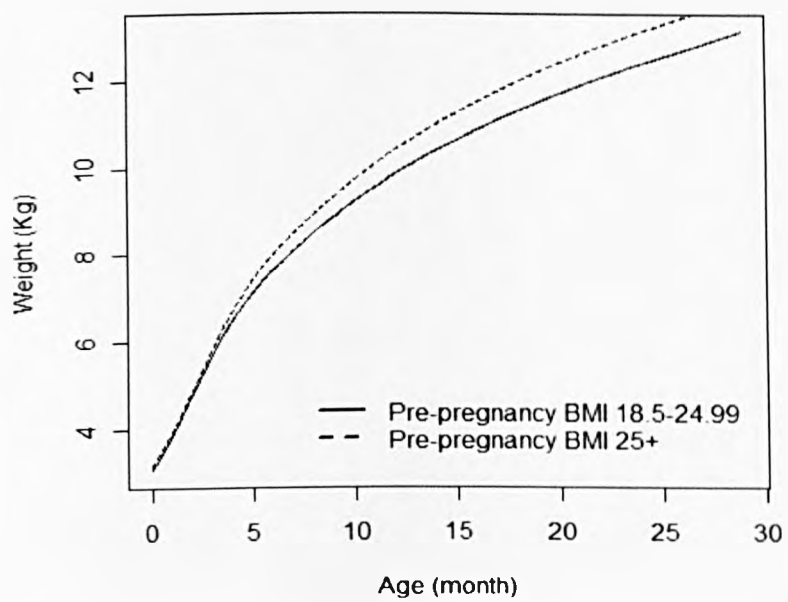


Figure 5.13: Predicted weight curves for different categories of maternal pre-pregnancy BMI from the model allowing for the association between the covariates and **size, tempo and velocity** parameters fitted on the GXXI data.

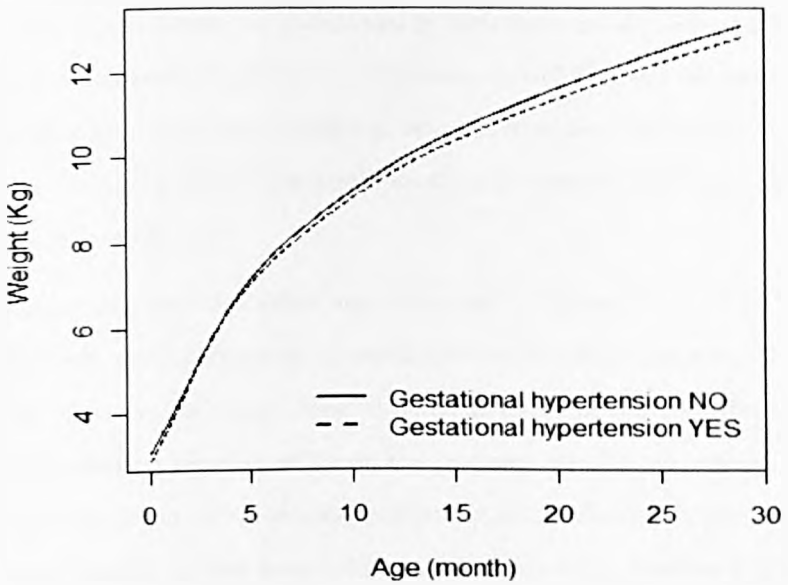


Figure 5.14: Predicted weight curves for different categories of gestational hypertension from the model allowing for the association between the covariates and **size and velocity** parameters fitted on the GXXI data.

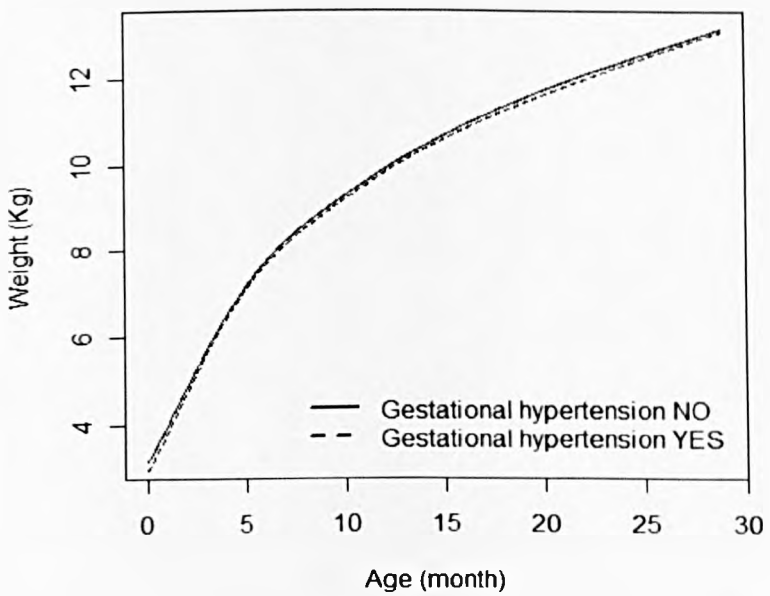


Figure 5.15: Predicted weight curves for different categories of gestational hypertension derived from the model allowing for an association between the covariates and **size, tempo and velocity** parameters fitted on the GXXI data.

for parity, most part of the impact of gestational hypertension on the size and velocity parameters observed in Table 2 of Research Paper IV (δ_α=-6.4% and δ_γ=12.8%) was captured by its influence on the tempo dimension when tempo was modeled in terms of covariates (δ_α=-3.5%, δ_β=0.31 months and δ_γ=8.7%, see Table 5). Again this is translated into slightly different predicted trajectories, especially after the first 6 months of life.

Overall these graphs show that the effect sizes reported in Tables 2 and 5 of Research Paper IV correspond to relatively small differences in population predicted trajectories. They also show that the inclusion of an effect on the tempo dimension might result in slightly different predicted curves compared with those derived when an effect on the size and velocity parameters only is allowed for. As some covariates were found to significantly affect the tempo dimension, the graphs corresponding to the more complex models are the more informative. Hence when feasible it is always advisible to include the covariates effect on all the three parameters so that the growth process can be separated into three biologically meaningful dimensions, which lead to a more comprehensive summary of the mechanisms governing the growth patterns.

Part III

Discussion

Chapter 6

Final Comments

In section 6.1 of this chapter some concluding comments on the overall findings from the two main components of this work will be presented, followed by an evaluation of the main contribution (section 6.2) and limitations of this thesis (section 6.3). In section 6.4 a discussion of possible areas of future work is then provided.

The specific results of each component of this thesis as well as their comparison with the existing literature have been reviewed both in the publications included in Chapters 4 and 5 and in the introductions to these chapters. Therefore these will not be included here.

6.1 Overall Findings of the Thesis

This Ph.D. addressed two main methodological challenges that may arise in the design and analysis of life course studies of infant growth: the potential bias that may arise due to selection at recruitment of cohort study participants and the modelling of individual growth trajectories. A brief summary of the overall findings from these two separated strands of research is presented below.

6.1.1 Selection bias

The first objective of this thesis was to examine inclusion mechanisms into cohort studies and assess their influence on the structure of the available data and consequently on the estimate of the exposure-outcome effect of interest. The main results relating to this objective were presented in Research Paper I and were based on a selection of simulated settings. The results showed that the selection process must depend upon the exposure of interest and one or more unmeasured or unknown risk factors for the outcome in order to introduce bias. Under the scenario in which the exposure of interest and an unmeasured or unknown risk factor for the outcome were independent in the general population but became associated because of the restriction of the source population, the bias in the estimated crude exposure-outcome association (expressed in terms of hazard ratio (HR)) resulted to be generally weak, with a maximum bias for the true log(HR) of ± 0.15. This corresponds to an estimated HR, when the true HR is equal to 1, of 0.86 or 1.16. This maximum bias derived from scenarios involving reasonably large associations in the selection and outcome mechanisms (e.g. exposure-selection odds ratio (OR), risk factor-selection OR and risk factor-outcome HR of 4.0 or 0.25) and a prevalence of the risk factor of about 50%. The maximum bias dropped to just ± 0.02 when these associations were decreased to values more commonly seen in epidemiological studies (e.g. ORs and HRs of 2 or 0.5).

The second objective of this Ph.D. was to investigate how selection into the web-based NINFEA birth cohort affects the confounding patterns present in the source population. This was addressed by comparing results from the NINFEA birth cohort (1,105 singleton births occurred within December 2008 in the municipality of Turin) with those from its source population, the Piedmont Birth Registry (PBR). The main findings were presented in Research Paper II and were also discussed, with reference to web-based studies, in the Book Chapter (Chapter 4), and in the Commentary (Chapter 4) where criticisms against the use of non-representative studies are rebuffed. The NINFEA participants differed substantially from the source population because low parity, high educational level, and non-smoking during pregnancy were the strongest predictors of participation. Participating mothers were also more likely to take folic acid, drink alcohol during pregnancy, and to be older at delivery. However the estimates of the association between some prenatal exposures and low birth weight and occurrence of caesarean section obtained in the sub-group of NINFEA participants did not differ considerably from

those from PBR, with relative ORs (that is, the ratios of the NINFEA-based OR over the PBR-based OR) ranging between 0.74 and 1.03. These results were analyzed in the light of the potential effect of changes in the confounding patterns due to the sample selection. The main potential confounders to be controlled for identified for the whole PBR were not the most important for analyses involving the NINFEA participants, suggesting that each population – including a selected study population – has its specific confounding pattern. Similar reasoning would apply to the role of unknown confounders. It was therefore concluded that it is not possible to predict whether the selected cohort would be more or less affected by unmeasured confounding than the equivalent population-based cohort, unless it is known whether the exposure and the potential unmeasured confounder are already associated in the source population or become associated in the selected sample. In the latter situation the selected cohort would be more affected by confounding, while in the former situation – exposure of interest and risk factor already associated in the source population – the analysis of the selected sample may be affected from increased but also from reduced unmeasured confounding, depending on the magnitude and direction of the exposure-risk factor association present in the source population and of their associations with the selection process.

Further comparisons of the data on birth weight and gestational age collected in NINFEA by self-reported questionnaires with those held in the registry showed that they were not affected by systematic measurement errors (see Chapter 3).

6.1.2 Growth modelling

The other objectives of this thesis were related to the modelling of individual growth trajectories.

In particular the third and fourth objectives of this Ph.D. were to compare the ability of alternative growth modelling approaches to fit weight trajectories in infancy identifying biologically meaningful features and to assess whether the results of fitting these models are affected by the type of available data (age range, number and timing of follow-up). These objectives were addressed comparing the random effects specifications of two models purposely developed to describe anthropometric data – the Jenss-Bayley (JB) and the four-parameters version of the Berkey-Reed model (Reed) – and a shape invariant random effects model (SITAR) recently proposed in the statistical literature, using data from

the three cohorts available for this thesis: 845 singleton NINFEA children (only growth data at fixed time points were included in this paper), 783 singleton GXXI children and 1,149 singleton GOCS children. In order to address the fourth objective, data up to 4 years of age were included for the NINFEA and GOCS studies as well as the NINFEA and GOCS data restricted to ages 0-2 years for comparison with GXXI. The main findings were presented in Research Paper III. It was concluded that the choice of which model to adopt varies with the aim of the study and, less crucially, on the richness of the available data. The Reed model performed best in terms of standard fit criteria and, being linear in its parameters, was easier to estimate. However, if the focus is on extracting salient features of the growth trajectories to be used for life course enquiries SITAR should be preferred because it allowed, unlike the other two models, the identification of important aspects of the individual growth patterns, such as the age at peak weight velocity (APWV), and because of the biological interpretability of its three parameters (size, tempo and velocity). In the data analyzed the SITAR predicted random effects were also observed to consistently identify outlying growth patterns unlike those derived by the JB and Reed models. Moreover correlations of the JB and Reed predicted child-specific random effects were extremely high while those among the SITAR predicted random effects were substantially lower. The interpretability of the SITAR parameters however required additional consideration. In particular the interpretation of the tempo parameter resulted to depend on the time scale used and the setting analysed. When the time scale was changed by setting the time origin to be at conception (instead of birth), part of the information held in the tempo parameter appeared to have been removed by this new time scale, suggesting that tempo represents an adjustment necessary to better proxy true biological age (hence measuring growth adjusted for maturation/developmental status).

The standard deviations of both the JB and Reed random effects were lower when derived from the model fitted to the 0-4 years data compared to those derived when fitted to the 0-2 years age range, and the JB model indeed failed when fitted on the NINFEA 0-2 years data, probably due to the small number of measurement times available in this cohort.

Due to its complex parametrization the SITAR models included in the analyses of this thesis required imposing some constraints on the values of the fixed effects, with fixing the tempo effect to zero leading to the best fitting models. SITAR was successfully fitted on each subsets (by cohort, gender and age

range) considered, however results were influenced by data points that were isolated, as in the case of the analyses of the NINFEA 0-4 years data.

The last objective of this thesis was to study the prenatal influences of weight trajectories in infancy and to compare them across the different cohorts. This was achieved by fitting a SITAR model, that was extended to include multiple explanatory variables for each of its three parameters, to the 0-2 years data from the three cohorts available for this thesis: 2,925 singleton NINFEA children (growth data at both fixed and varying time points were included in this paper), 738 singleton GXXI babies, and 959 singleton GOCS children. The main results related to this objective were presented in Research Paper IV. It was found that growth trajectories in contemporary infants from economically and geographically diverse countries such as Portugal, Italy and Chile share some common features, in particular with respect to the effect of maternal height, pre-pregnancy overweight/obesity and parity. In the two European cohorts opposite effects of maternal underweight, smoking and hypertension on the child's size and velocity were observed, and when growth data were rich and the effect on tempo could also be examined, parity, pre-pregnancy overweight/obesity and gestational hypertension were found to also affect the timing of growth.

An advantage of adopting the SITAR modelling approach was to separate the individual growth process into three specific components – size, velocity and tempo –, thus providing insights into the mechanisms governing infant growth. In particular the tempo dimension allowed to examine what governs the timing of peak growth velocity in infancy. However, due to its complex parametrization, it was not always possible to successfully model the association of multiple exposures with the three SITAR parameters simultaneously. In particular, the model including an effect on each of the three dimensions failed when fitted to the NINFEA dataset, due to its limited number of time points available. Convergence failures were also observed when allowing for effects on tempo in the GOCS data, probably because of lack of heterogeneity, as this study only includes term children. The fully specified model was instead successfully fitted to the GXXI cohort, which has the greater number of growth measurements. Nevertheless these results showed that the version of SITAR including fewer parameters (i.e. not allowing for the tempo effect) could be successfully fitted to dataset with relatively sparse data, such as NINFEA, providing results consistent with those obtained with richer datasets.

6.2 Main Contributions of the Thesis

6.2.1 Selection bias

This thesis contributes to the recently revived debate on potential biases that may arise from selection of study participants in cohort studies, with particular focus on web-based designs.

Specifically it provides the first simulation study which quantifies the likely extent of the bias in the exposure-outcome association (expressed in terms of crude hazard ratios) due to restrictions of the source population. A simulation study (Whitcomb *et al.*, 2009) published at the time of the submission of Research Paper I also focused on the quantification of bias due to collider-stratification, that is the bias induced by the same causal structure investigated in this thesis (see section 4.1.1). The authors analyzed the phenomenon known in the literature as "birthweight paradox". In this setting birth weight, which was treated as a continuous variable, acts as a collider and also as an intermediate variable in the relationship between a risk factor and neonatal mortality. Therefore the scenario investigated by Whitcomb et al. (Whitcomb *et al.*, 2009) differs substantially from the one analyzed in this thesis. The results obtained in Research Paper I are in agreement with those of Greenland, who studied the same setting analyzed in Research Paper I and used an analytical approach to quantify the maximum extent of the bias in the crude OR for the exposure-outcome association of interest (Greenland, 2003). Greenland investigated however only settings where the outcome prevalence was rare so that the analysis of cohort data could be performed using logistic regression. The simulation study used for Research Paper I added further insights as a wide range of outcome and selection parameters were examined, including prevalence and effects of exposure and risk factor on both the selection and outcome process, baseline incidence rate of the outcome and sample size, which allowed to underline their roles in influencing the extent of the bias.

Previous studies aiming at examining the representativeness of the study sample either compared the characteristics of participants with those of non-participants (Clarisse *et al.*, 2007; Goldberg *et al.*, 2001; Heilbrun *et al.*, 1991) or compared the estimates of exposure-outcome associations of interest obtained in the selected cohort with that obtained in the source population (Nilsen *et al.*, 2009; Nohr *et al.*, 2006). In contrast the research included in this thesis focused on the role of the confounding

patterns for the association of interest. As a consequence, it has been possible to specifically investigate the potential mechanisms through which this bias may be induced by the sample selection process. In the Commentary the mechanisms investigated have been also extended to consider the scenario, which has been less debated, in which a mediator (that is a variable that is on the pathway from the exposure to the outcome) or a cause of the mediator is associated with the probability of selection.

Moreover this Ph.D. provides insights into the use of the internet in epidemiological research, with particular attention given to web-based cohort studies. It has been shown that the main concerns raised against the use of non-representative samples, namely lack of exposure heterogeneity among the study participants and the potential for introducing bias in the estimate of interest due to the selection criteria (self-selection and restriction of the source population to internet users), are often not valid (see Commentary included in Chapter 4) and are sometime outweighed by the benefits of this approach. Indeed the use of the internet offers several advantages, including decreased costs, the possibility to reach "hidden" populations and to tailor the questionnaires to the participants' characteristics, the possibility to study rare exposures through "ad hoc" selection of the study members, and instantaneous checks to identify data inconsistencies as well as to reduce data entry errors (if transcribed). Moreover any propensity of bias arising from using non-representative samples needs to be balanced against that arising from targeting representative samples but resulting in low response rates at follow-up, as selection of the study participants is likely to create a group of more motivated persons. In this perspective in this thesis it is argued that restriction of the source population – including to internet users – does not usually hamper scientific inference, and may often enhance it.

The idea of using the internet in epidemiological research often receives skeptical reactions. Although there is the need to further investigate the impact of baseline selection in internet-based cohort studies, combining a methodological study with an application to real data (i.e. the NINFEA cohort) this thesis has generated findings that can help to overcome a-priori formed opinions on whether internet-based cohort studies are valid or not.

6.2.2 Growth modelling

This thesis also contributes to the literature on biological and statistical methods for modelling growth in early life, with particular focus on weight trajectories.

Specifically this study provides researchers with insights into the parametrization of three models, two of which belong to the family of "Biological models" described in Chapter 5. These were applied to three datasets, which differ in term of age range and number and timing of follow-up observations (regular/irregular), and therefore illustrates a variety of settings likely to be encountered in practice. In particular Research Papers III provides details on the parametrization of a novel statistical methods (SITAR), which use is rapidly increasing (Cole *et al.*, 2010; Gault *et al.*, 2011; Johnson *et al.*, 2011; Jones-Smith *et al.*, 2013; Prentice *et al.*, 2012; Warrington *et al.*, 2013). Special attention has been given to the interpretation of its predicted random effects, and among them, to the meaning of the tempo dimension, which represents the most innovative aspect of this model. Moreover the sensitivity of the three models to the varying quality of the growth data and to outlying growth patterns was compared. These findings thus contribute to guiding researchers in their application of models that properly exploit the growth data available, taking into account both the pattern and the quality of the measurements.

Research Paper IV extended this work by including multiple explanatory variables for each of the three SITAR parameters to study the influence of several maternal prenatal characteristics on the weight trajectories during infancy. Adopting the SITAR modelling approach, it has been possible to study the effect of these exposures on three different components of the growth pattern (size, tempo and velocity) instead of studying their effect on single growth indicator, such as APWV or standard growth rate. The thesis therefore provides a more comprehensive summary of the factors and mechanisms governing the infant growth process. In particular it provides the first study in which the effect of several maternal characteristics is allowed to affect the tempo dimension, as well as the size and velocity parameters. Moreover with this work it has been shown that SITAR can be used to carry out pooled analyses of studies with different number and timing of growth measurements. The latter result is particularly relevant as the increasing number of projects aimed at developing an integrated strategy for mother-child cohort research, such as the CHICOS (CHICOS, 2010) and the ENRIECO (ENRIECO, 2009)

projects, poses the question on how to combine and properly analyze these data, including those concerning growth. Finally this thesis provides new epidemiological evidence on the effect of several prenatal characteristics on infant weight trajectories, based on data from three contemporary cohorts from Portugal, Italy and Chile. Comparing their results allowed to evaluate their robustness, given the differences across the three source populations in distribution of the exposures as well as the expected differences in their confounding patterns. In particular the results on the effects of maternal parity, pre-pregnancy underweight and overweight/obesity, and of smoking and hypertension during pregnancy on the child's size, tempo and velocity are of interest.

6.3 Limitations

Some of the limitations of this thesis have been reported in the relevant Research Papers. Additional points are discussed below.

6.3.1 Selection bias

The scenario investigated in Research Paper I is limited to the DAG depicted in Figure 4.5 (Chapters 4). In this situation a binary exposure of interest and one unmeasured binary outcome risk factor, which are independent in the source population, affect the selection process and become associated in the selected sample. Examples of settings that this thesis did not examine include considering multiple risk factors; considering continuous exposure, risk factor and outcome; or considering more complex settings, such as the one in which the exposure and the risk factor are already associated in the source population (as described in Research Paper II and in the Book Chapter) or the scenario in which a mediator of the association of interest also affects the selection (as described in the Commentary).

Some of the limitations just described also apply to the analyses of Research Paper II. In particular the dichotomization of exposures (more detailed categories could have been used for parity and maternal education) and potential confounding variables (e.g. maternal age, weight gain during pregnancy) might have led to disadvantages. These include loss of statistical efficiency, residual confounding, and differential misclassification (Royston *et al.*, 2006). Moreover the dichotomization of the outcome

variable birth weight, which was thus modelled using logistic regression, could have induced some additional loss of statistical efficiency. Another limitation of this study is that some of the predictors of participations into NINFEA could be mediators of the exposure-outcome associations estimated in Research Paper II. For example maternal education could influence low birth weight (one of the outcome considered) partially via maternal smoking behaviour during pregnancy, with both maternal education and smoking during pregnancy being strong predictors of participation into the NINFEA study (Pizzi *et al.*, 2012). Under this scenario, the education-outcome association estimated adjusting for smoking (and for the other potential confounders) would be biased in both the registry and NINFEA populations. Moreover, due to the altered association between education and smoking induced by the restriction, the crude association estimated in the NINFEA sample would be biased.

6.3.2 Growth modelling

The comparison of alternative growth models to describe weight trajectories in early life involved fitting the random effects specification of two biological models, the JB and Reed models, and the SITAR model. As discussed above, the JB random effects model failed when fitted to the NINFEA 0-2 years data (Research Paper III). One of the advantages of adopting a random effects approach, as described in the introduction of Chapter 5, is that all subjects with at least one growth measure can be included in the analyses (under the assumption that observations are missing at random (Rubin, 1976)). However the decision to include in the analyses all subjects with at least one growth measure might be – at least in part – the reason for this failure. NINFEA children included in the analyses of Research Paper III had a median of 5 observations per child within the first 4 years of life, with a maximum of 6 observations per child. Additional sensitivity analyses could have thus been conducted to inform on the performance of the random effects specifications of the JB and Reed models when imposing a constraint on the minimum number of observations required to be included in the growth analyses. Also this thesis did not examine algorithms for the estimation of parameters from non-linear mixed models alternative to the one used by the `nlme` function in R, which is based on likelihood linearization (Lindstrom and Bates, 1990). Asymptotic convergence has not been proved for methods based on likelihood linearization (Samson *et al.*, 2007), and this might in part explain the convergence

failure experienced when fitting the JB model for the analyses of Research Paper III. An alternative would have been to use the Stochastic Approximation Expectation-Maximization (SAEM) algorithm, for which theoretical convergence properties have been illustrated (Samson *et al.*, 2007). Another limitation of this thesis is that, because of extreme correlations existing between the random effects parameters derived from the biological models, it was not possible to compare the marginal and the conditional associations between the gender of the child and the JB and Reed random effects parameters, as it was done for the SITAR model. These extreme correlations suggest the models were possibly overparameterised.

While the availability of three datasets with different number and timing of follow-up observations (regular/irregular) as well as different inclusion criteria (as GOCS includes only children with a gestational age at birth between 37 and 42 weeks and birth weight between 2500 and 4500 grams) was a strength in Research Paper III, part of the difference existing between these three cohorts generated some limitations in the analyses of Research Paper IV. Specifically, some of the effects estimated in GOCS differed from those obtained among the other two cohorts (particularly for gestational hypertension and smoking during pregnancy). Although sensitivity analyses were carried out replicating cohort-specific analyses on the subset of GXXI and NINFEA children who were born at term and with a birth weight of 2500-4500 grams (thus using the same entry criteria as GOCS), it was not possible to completely distinguish whether the observed heterogeneities across the cohorts were solely due to differences in entry criteria or not. Moreover the lack of heterogeneity in GOCS induced by the inclusion of term births only did not allow to estimate the effects of the covariates on the tempo dimension in this cohort. The lack of sufficient growth observations in NINFEA also led to convergence failure when attempting to estimate the effects of the exposures on the tempo parameter. Therefore it was not possible to compare this specific result (i.e. the tempo effect) across the cohorts as estimates were obtained only for the GXXI cohort. Finally differences in quality and coarseness of the available data in relation to pregnancy complications (e.g. gestational hypertension and diabetes) did not allow to examine the effect of the severity of these complications on the growth pattern. In particular it would have been interesting to distinguish between "simple" gestational hypertension, preeclampsia and eclampsia.

6.4 Areas of Future Research

This thesis identified a number of areas for further investigation. These are discussed below.

Mediators affecting the baseline selection mechanisms in cohort studies

The Commentary included in Chapter 4 has drawn attention on the role of mediators of an association of interest in the selection process of cohort studies. In that publication some relatively simple scenarios were considered, where – in the case baseline selection takes place before the mediator is manifest (see Figure 1 of the Commentary) – either a cause of the mediator or a confounder of the exposure-mediator relationship affects the selection mechanisms. Based on these DAGs, it was concluded that there is no reason to expect that non-representative cohorts have a larger exposure-mediator confounding than representative cohorts as in some of these settings the exposure-mediator association would be biased while in other the selection would decrease the exposure-mediator confounding. The scenario in which the mediator itself affects the probability of selection – when baseline selection occurs after the mediator becomes manifest (see Figure 2 of the Commentary) – was also evaluated. Although in the Commentary it was suggested that to minimize the probability of bias researchers should plan to carry out the enrolment of the participants before the mediator or its early signs could become manifest, future work on this issue is needed. In particular more complex settings, such as the one in which the selection could be affected both by the mediator and by the participant's reaction to the mediator, need to be investigated. Monte Carlo simulations could be carried out to mimic scenarios involving mediators in the selection mechanisms and assess whether the extent of the maximum bias in the association of interest is larger compared to the settings examined in Research Paper I.

Lost to follow-up in the web-based NINFEA birth cohort

In most of the publications included in Chapter 4 (Research Paper I, Book Chapter and Commentary) it has been stated that one of the main advantages of conducting cohort studies using restricted source populations is to increase the completeness of follow-up. This is because restriction to particular subgroups and baseline self-selection are likely to create a group of more motivated participants for longitudinal studies. In particular it has been argued that follow-up participation rates could be higher

in web-based cohort studies compared to "traditional" designs also because researchers have access to participants' email addresses as well as to standard contact information (e.g. mail addresses) and can keep constant contact with the study participants using social networks (see Book Chapter). Further research is therefore required on this topic. In particular determinants of completeness of follow-up in web-based cohorts could be evaluated using the data of the NINFEA birth cohort and compared with evidence gathered from traditional cohort studies (Barchielli and Balzi, 2002; Goldberg *et al.*, 2001; Howe *et al.*, 2013).

Growth modelling

Individual growth models evaluated in this Ph.D. have focused only on weight measurements in infancy. As stated in the introduction (section 1.2.2), this choice has been driven both by issues of data availability (as NINFEA length measures were affected by missingness and measurement errors, see Chapter 3) and by the fact that during infancy weight has been suggested to be more useful than length to assess poor growth, as well as to identify children experiencing a catch-up growth (Cole, 2002). However in order to understand the role of early life growth in life course epidemiology several other aspects could be examined. In particular, SITAR could be used to model linear growth and growth in BMI and to study how their potential predictors influence the size, tempo and velocity components of these growth processes.

Moreover evaluation of the performance of SITAR to model growth for longer age ranges (e.g. growth during the first 7/10 years of life) is of interest. The latter would be particularly relevant when investigating growth in BMI. As argued in the discussion of Research Paper III, interpretation of the SITAR parameters might become less straightforward when a wider range of ages is analyzed, as individual trajectories may include several inflection points, leading to multiple changes in velocity. This in turn would lead the parameters representing a complex average of several departures from the reference time scale. This is expected to affect both the SITAR velocity and, most of all, tempo parameters. When studying growth in BMI from age 0 to 10 years, both a peak velocity at around 1 year and a minimum velocity (known as adiposity rebound) at about 5 years occur (Rolland-Cachera *et al.*, 1987). Age at BMI peak has been positively associated with higher level of BMI later in life

(Silverwood *et al.*, 2009b), while age at adiposity rebound has been negatively associated with increased risk of later obesity (e.g. the earlier the adiposity rebound the higher the risk of obesity) (Ohlsson *et al.*, 2012; Rolland-Cachera, 1998; Williams and Goulding, 2012). This suggests that possibly two separate biological time scales exist: one linked to infancy BMI peak and one linked to childhood BMI rebound. Further research is therefore required to evaluate how the SITAR parameters would capture these patterns and to assess whether the tempo effect would represent an average of the child's departure from possibly two separate biological time scales. This analysis will be carried out using the GOCS data, for which growth measurements up to age 10-11 are now available. BMI growth of NINFEA children will be also modelled, as their growth measurements, up to age 5, will be retrospectively gathered from child health records and additional measurements, up to age 7, will be gathered via the 7-year follow-up questionnaire.

Pre and postnatal factors influencing childhood obesity

Early life growth, and in particular rapid postnatal weight gain, has become the focus of research into the development of overweight and obesity later in childhood and adulthood (Baird *et al.*, 2005; Monteiro and Victora, 2005; Ong and Loos, 2006; Tzoulaki *et al.*, 2010). The motivation for this research arises from the extensive evidence that the prevalence of childhood obesity is very high worldwide (GRUPPO-OKkio, 2010; Ogden and Carroll, 2010). It has been suggested that the roots of this epidemic can be tracked back to fetal life (Fall, 2011), supporting a relevant role for maternal life-style and nutrition prior to and during pregnancy. In particular childhood obesity has been associated with maternal pre-pregnancy obesity, gestational diabetes and smoking during pregnancy (Monasta *et al.*, 2010). It is also recognized that early postnatal factors such as feeding type and rapid postnatal weight gain, as discussed above, play an important role in the development of this disease (Monasta *et al.*, 2010). However it would be relevant to better understand how these prenatal exposures interplay with postanatal factors in inducing obesity.

This thesis has drawn attention to a model, SITAR, particularly useful to provide a comprehensive summary of the individual early life weight trajectories, as it allows to specify the growth process in terms of three different biologically interpretable components: size, tempo and velocity of growth.

Future work will therefore include application of this model to the infancy weight data of different birth cohorts, including NINFEA, to identify children who experienced an extremely rapid postnatal weight gain, where the latter could be defined from the predicted child-specific velocity coefficients, conditional on size and tempo. In particular children with a velocity parameter above a pre-specified cutoff (e.g. the 90th percentile of the internally standardized distribution) could be defined as exposed to a rapid postnatal weight gain. The resulting variable (experience of a rapid postnatal weight gain) could then be treated as a mediator of the effect of selected prenatal exposures on childhood obesity, where the latter can be defined from BMI data at a given age (e.g. 4 or 7 years). The indirect effects of prenatal exposures on the prevalence of obesity at a certain age mediated via such a variable could then be defined – and estimated – within the framework of causal mediation analysis (Vansteelandt, 2012). These analyses will thus provides a better understanding of which prenatal factors act mainly via postnatal growth pattern and which factors act through other pathways.

6.5 Conclusion

This Ph.D. aimed at modelling individual infant weight trajectories using data from three contemporary cohort studies, with special consideration given to issues of selection bias because members of one of the studies are volunteers who participate only via a web-based system. The interest in infant weight trajectories arose from the broad and consistent evidence in the literature that fetal and early life growth are important predictors for the onset and development of a wide range of later diseases.

This thesis found that using a restricted source population to design cohort study will, under a range of sensible scenarios, produce only relatively weak bias in estimates of the exposure-outcome associations. It also showed that restriction may either increase or decrease the amount of confounding in the data and thus the resulting bias in the estimates of interest. With regards to the modelling of individual growth trajectories this study found that the choice of which model to adopt varies with the aim of the study and, less crucially, with the richness of the available data. Of the models examined SITAR was found to be the most flexible and the most useful for life course enquiries because of the biological interpretability of its parameters.

The thesis thus provides a set of results that contributes to the interpretation of findings from se-

lected cohort studies, including internet-based cohorts, and to the implementation of advanced growth modelling approaches for life course research.

Bibliography

(2012) Children's Study Row. *Nature*, **483**.

Agostoni, C., Grandi, F., Gianni, M. L., Silano, M., Torcoletti, M., Giovannini, M. and Riva, E. (1999) Growth patterns of breast fed and formula fed infants in the first 12 months of life: an italian study. *Arch Dis Child*, **81**(5), 395–9.

Baird, J., Fisher, D., Lucas, P., Kleijnen, J., Roberts, H. and Law, C. (2005) Being big or growing fast: systematic review of size and growth in infancy and later obesity. *BMJ*, **331**(7522), 929.

Barchielli, A. and Balzi, D. (2002) Nine-year follow-up of a survey on smoking habits in Florence (Italy): higher mortality among non-responders. *International journal of epidemiology*, **31**(5), 1038–1042.

Barker, D. J. P. (1998) *Mothers, Babies and Health in Later Life*. Edinburgh: Churchill Livingstone.

Batista, R. F., Silva, A. A., Barbieri, M. A., Simoes, V. M. and Bettiol, H. (2012) Factors associated with height catch-up and catch-down growth among schoolchildren. *PLoS One*, **7**(3), e32903.

Baulon, E., Fraser, W. D., Piedboeuf, B., Buekens, P. and Xiong, X. (2005) Pregnancy-induced hypertension and infant growth at 28 and 42 days postpartum. *BMC Pregnancy Childbirth*, **5**, 10.

Beath, K. J. (2007) Infant growth modelling using a shape invariant model with random effects. *Stat Med*, **26**(12), 2547–64.

Ben-Shlomo, Y. and Kuh, D. (2002) A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. *Int J Epidemiol*, **31**(2), 285–93.

Ben-Shlomo, Y., McCarthy, A., Hughes, R., Tilling, K., Davies, D. and Smith, G. D. (2008) Immediate postnatal growth is associated with blood pressure in young adulthood the barry caerphilly growth study. *Hypertension*, **52**(4), 638–644.

Benn, R. T. (1971) Some mathematical properties of weight-for-height indices used as measures of adiposity. *Br J Prev Soc Med*, **25**(1), 42–50.

Berkey, C. S. (1982) Comparison of two longitudinal growth models for preschool children. *Biometrics*, **38**(1), 221–34.

Berkey, C. S. and Reed, R. B. (1987) A model for describing normal and abnormal growth in early childhood. *Hum Biol*, **59**(6), 973–87.

Betoko, A., Charles, M. A., Hankard, R., Forhan, A., Bonet, M., Regnault, N., Botton, J., Saurel-Cubizolles, M. J., de Lauzon-Guillain, B. and the, E. M.-C. C. S. G. (2012) Determinants of infant formula use and relation with growth in the first 4 months. *Matern Child Nutr*, **[Epub ahead of print]**.

Betoko, A., Charles, M. A., Hankard, R., Forhan, A., Bonet, M., Saurel-Cubizolles, M. J., Heude, B. and de Lauzon-Guillain, B. (2013) Infant feeding patterns over the first year of life: influence of family characteristics. *Eur J Clin Nutr*, **[Epub ahead of print]**.

Binns, C. (1998) Infant-feeding and growth. In *The Cambridge Encyclopedia of Human Growth and Development* (Eds S. Ulijaszek, F. Johnston and M. A. Preece). Cambridge: Cambridge University Press.

Blair, P. S., Drewett, R. F., Emmett, P. M., Ness, A. and Emond, A. M. (2004) Family, socioeconomic and prenatal factors associated with failure to thrive in the Avon Longitudinal Study of Parents and Children (ALSPAC). *Int J Epidemiol*, **33**(4), 839–47.

Bocca-Tjeertes, I. F., Kerstjens, J. M., Reijneveld, S. A., de Winter, A. F. and Bos, A. F. (2011) Growth and predictors of growth restraint in moderately preterm children aged 0 to 4 years. *Pediatrics*, **128**(5), e1187–94.

Bogin, B. (1998) Patterns of human growth. In *The Cambridge Encyclopedia of Human Growth and Development* (Eds S. Ulijaszek, F. Johnston and M. A. Preece). Cambridge: Cambridge University Press.

Borja, J. B. and Adair, L. S. (2003) Assessing the net effect of young maternal age on birthweight. *American journal of human biology*, **15**(6), 733–740.

Botton, J., Heude, B., Maccario, J., Ducimetiere, P. and Charles, M. A. (2008) Postnatal weight and height growth velocities at different ages between birth and 5 y and body composition in adolescent boys and girls. *Am J Clin Nutr*, **87**(6), 1760–8.

Botton, J., Heude, B., Maccario, J., Borys, J. M., Lommez, A., Ducimetiere, P., Charles, M. A. and group, F. s. (2010) Parental body size and early weight and height growth velocities in their offspring. *Early Hum Dev*, **86**(7), 445–50.

Bouthoorn, S. H., van Lenthe, F. J., Hokken-Koelega, A. C., Moll, H. A., Tiemeier, H., Hofman, A., Mackenbach, J. P., Jaddoe, V. W. and Raat, H. (2012) Head circumference of infants born to mothers with different educational levels; the generation r study. *PloS one*, **7**(6), e39798.

Britton, G. R., James, G. D., Collier, R., Sprague, L. M. and Brinthaupt, J. (2013) The effects of smoking cessation and a programme intervention on birth and other perinatal outcomes among rural pregnant smokers. *Annals of human biology*, **[Epub ahead of print]**.

Brooke, O. G., Anderson, H. R., Bland, J. M., Peacock, J. L. and Stewart, C. M. (1989) Effects on birth weight of smoking, alcohol, caffeine, socioeconomic factors, and psychosocial stress. *BMJ*, **298**(6676), 795–801.

Bulk-Bunschoten, A. M., van Bodegom, S., Reerink, J. D., de Jong, P. C. and de Groot, C. J. (2002) Weight and weight gain at 4 months (The Netherlands 1998): influences of nutritional practices, socio-economic and ethnic factors. *Paediatr Perinat Epidemiol*, **16**(4), 361–9.

Cameron, N. (1998) Measurement and assessment. In *The Cambridge Encyclopedia of Human Growth and Development* (Eds S. Ulijaszek, F. Johnston and M. A. Preece). Cambridge: Cambridge University Press.

Campbell, M. K., Cartier, S., Xie, B., Kouniakis, G., Huang, W. and Han, V. (2012) Determinants of small for gestational age birth at term. *Paediatr Perinat Epidemiol*, **26**(6), 525–33.

Carter, R. C., Jacobson, S. W., Molteno, C. D. and Jacobson, J. L. (2007) Fetal alcohol exposure, iron-deficiency anemia, and infant growth. *Pediatrics*, **120**(3), 559–67.

Carter, R. C., Jacobson, J. L., Sokol, R. J., Avison, M. J. and Jacobson, S. W. (2013) Fetal alcohol-related growth restriction from birth through young adulthood and moderating effects of maternal prepregnancy weight. *Alcohol Clin Exp Res*, **37**, 452–62.

Catalano, P. M., McIntyre, H. D., Cruickshank, J. K., McCance, D. R., Dyer, A. R., Metzger, B. E., Lowe, L. P., Trimble, E. R., Coustan, D. R., Hadden, D. R. *et al.* (2012) The hyperglycemia and adverse pregnancy outcome study associations of gdm and obesity with pregnancy outcomes. *Diabetes care*, **35**(4), 780–786.

Chen, A., Pennell, M. L., Klebanoff, M. A., Rogan, W. J. and Longnecker, M. P. (2006) Maternal smoking during pregnancy in relation to child overweight: follow-up to age 8 years. *Int J Epidemiol*, **35**(1), 121–30.

CHICOS (2010) CHICOS: Developing a Child Cohort Research Strategy for Europe.

Clarisse, B., Nikasinovic, L., Poinsard, R., Just, J. and Momas, I. (2007) The paris prospective birth cohort study: which design and who participates? *European journal of epidemiology*, **22**(3), 203–210.

Cole, T. J. (1986) Weight/heightp compared to weight/height2 for assessing adiposity in childhood: influence of age and bone age on p during puberty. *Ann Hum Biol*, **13**(5), 433–51.

Cole, T. J. (2000) Secular trends in growth. *Proc Nutr Soc*, **59**(2), 317–24.

Cole, T. J. (2002) Assessment of growth. *Best Pract Res Clin Endocrinol Metab*, **16**(3), 383–98.

Cole, T. J. and Green, P. J. (1992) Smoothing reference centile curves: the LMS method and penalized likelihood. *Stat Med*, **11**(10), 1305–19.

Cole, T. J., Donaldson, M. D. and Ben-Shlomo, Y. (2010) SITAR–a useful instrument for growth curve analysis. *Int J Epidemiol*, **39**(6), 1558–66.

Corvalan, C., Uauy, R., Stein, A. D., Kain, J. and Martorell, R. (2009) Effect of growth on cardiometabolic status at 4 y of age. *The American journal of clinical nutrition*, **90**(3), 547–555.

Count, E. W. (1943) Growth patterns of the human physique: an approach to kinetic anthropometry. *Hum Biol*, **15**, 1–32.

Crump, C., Sundquist, K., Winkleby, M. A. and Sundquist, J. (2013) Early-term birth (37–38 weeks) and mortality in young adulthood. *Epidemiology*, **24**(2), 270–276.

DAloisio, A. A., DeRoo, L. A., Baird, D. D., Weinberg, C. R. and Sandler, D. P. (2013) Prenatal and infant exposures and age at menarche. *Epidemiology*, **24**(2), 277–284.

Daniel, R. M., Kenward, M. G., Cousens, S. N. and De Stavola, B. L. (2012) Using causal diagrams to guide analysis in missing data problems. *Statistical Methods in Medical Research*, **21**(3), 243–256.

De Stavola, B. L., dos Santos Silva, I., McCormack, V., Hardy, R. J., Kuh, D. J. and Wadsworth, M. E. (2004) Childhood growth and breast cancer. *Am J Epidemiol*, **159**(7), 671–82.

De Stavola, B. L., Leon, D. A. and Koupil, I. (2011) Intergenerational correlations in size at birth and the contribution of environmental factors the uppsala birth cohort multigenerational study, sweden, 1915–2002. *American journal of epidemiology*, **174**(1), 52–62.

Deardorff, J., Berry-Millett, R., Rehkopf, D., Luecke, E., Lahiff, M. and Abrams, B. (2012) Maternal pre-pregnancy bmi, gestational weight gain, and age at menarche in daughters. *Maternal and child health journal*, **[Epub ahead of print]**.

Deierlein, A. L., Siega-Riz, A. M., Adair, L. S. and Herring, A. H. (2011) Effects of pre-pregnancy body mass index and gestational weight gain on infant anthropometric outcomes. *The Journal of pediatrics*, **158**(2), 221–226.

Dewey, K. G. (2001) Nutrition, growth, and complementary feeding of the breastfed infant. *Pediatr Clin North Am*, **48**(1), 87–104.

Dietz, P. M., Callaghan, W. M., Cogswell, M. E., Morrow, B., Ferre, C. and Schieve, L. A. (2006) Combined effects of prepregnancy body mass index and weight gain during pregnancy on the risk of preterm delivery. *Epidemiology*, **17**(2), 170–7.

Dietz, W. H. (1994) Critical periods in childhood for the development of obesity. *Am J Clin Nutr*, **59**(5), 955–9.

Doll, R., Peto, R., Boreham, J. and Sutherland, I. (2004) Mortality in relation to smoking: 50 years' observations on male British doctors. *BMJ*, **328**(7455), 1519.

Dubois, L. and Girard, M. (2006) Early determinants of overweight at 4.5 years in a population-based longitudinal study. *Int J Obes (Lond)*, **30**(4), 610–7.

ENRIECO (2009) Enrieco Project: Environmental Health Risks in European Birth Cohorts.

Erickson, A. C. and Arbour, L. T. (2012) Heavy smoking during pregnancy as a marker for other risk factors of adverse birth outcomes: a population-based study in british columbia, canada. *BMC Public Health*, **12**(1), 102.

Eriksson, J. G. (2011) Early growth and coronary heart disease and type 2 diabetes: findings from the helsinki birth cohort study (hbcs). *Am J Clin Nutr*, **94**(6 Suppl), 1799S–1802S.

Fall, C. (2011) Evidence for the intra-uterine programming of adiposity in later life. *Annals of human biology*, **38**(4), 410–428.

Fawzi, W. W., Herrera, M. G., Nestel, P., el Amin, A. and Mohamed, K. A. (1998) A longitudinal study of prolonged breastfeeding in relation to child undernutrition. *Int J Epidemiol*, **27**(2), 255–60.

Feldman, H. S., Jones, K. L., Lindsay, S., Slymen, D., Klonoff-Cohen, H., Kao, K., Rao, S. and Chambers, C. (2012) Prenatal alcohol exposure patterns and alcohol-related birth defects and growth deficiencies: a prospective study. *Alcohol Clin Exp Res*, **36**(4), 670–6.

Gamborg, M., Andersen, P. K., Baker, J. L., Budtz-Jorgensen, E., Jorgensen, T., Jensen, G. and Sorensen, T. I. (2009) Life course path analysis of birth weight, childhood growth, and adult systolic blood pressure. *Am J Epidemiol*, **169**(10), 1167–78.

Gartside, P. S., Dine, M. S. and Glueck, C. J. (1984) Relative velocity of accretion of weight and height using the Benn index in the first nine years of life. *Pediatr Res*, **18**(7), 627–30.

Gault, E. J., Perry, R. J., Cole, T. J., Casey, S., Paterson, W. F., Hindmarsh, P. C., Betts, P., Dunger, D. B. and Donaldson, M. D. (2011) Effect of oxandrolone and timing of pubertal induction on final height in turners syndrome: randomised, double blind, placebo controlled trial. *BMJ: British Medical Journal*, **342**.

Gluckman, P. D. and Hanson, M. A. (2004) Maternal constraint of fetal growth and its consequences. *Semin Fetal Neonatal Med*, **9**(5), 419–25.

Gluckman, P. D., Hanson, M. A. and Mitchell, M. D. (2010) Developmental origins of health and disease: reducing the burden of chronic disease in the next generation. *Genome Med*, **2**(2), 14.

Glymour, M. (2006) Using causal diagrams to understand common problems in social epidemiology. In *Methods in social epidemiology* (Eds M. Oakes and J. Kaufman). San Francisco: Jossey-Bass.

Godfrey, K. M., Gluckman, P. D. and Hanson, M. A. (2010) Developmental origins of metabolic disease: life course and intergenerational perspectives. *Trends in Endocrinology & Metabolism*, **21**(4), 199–205.

Goldberg, G. (1998) Alcohol consumption and pregnancy outcome. In *The Cambridge Encyclopedia of Human Growth and Development* (Eds S. Ulijaszek, F. Johnston and M. A. Preece). Cambridge: Cambridge University Press.

Goldberg, M., Chastang, J. F., Leclerc, A., Zins, M., Bonenfant, S., Bugel, I., Kaniewski, N., Schmaus, A., Niedhammer, I., Piciotti, M. *et al.* (2001) Socioeconomic, demographic, occupational, and health factors associated with participation in a long-term epidemiologic survey: a prospective study of the French GAZEL cohort and its target population. *American journal of epidemiology*, **154**(4), 373–384.

Goldenberg, R. L. and Cliver, S. P. (1997) Small for gestational age and intrauterine growth restriction: definitions and standards. *Clin Obstet Gynecol*, **40**(4), 704–14.

Goldstein, H. (2010) *Multilevel statistical models, Fourth Edition.* London: Wiley.

Goldstein, H., Healy, M. J. and Rasbash, J. (1994) Multilevel time series models with applications to repeated measures data. *Stat Med*, **13**(16), 1643–55.

Green, P. J. and Silverman, B. W. (1994) *Nonparametric Regression and Generalized Linear Models: A roughness penalty approach.* Chapman and Hall/CRC.

Greenland, S. (2003) Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology*, **14**(3), 300–6.

Greenland, S., Pearl, J. and Robins, J. M. (1999) Causal diagrams for epidemiologic research. *Epidemiology*, **10**(1), 37–48.

Griffiths, L. J., Dezateux, C. and Cole, T. J. (2007) Differential parental weight and height contributions to offspring birthweight and weight gain in infancy. *Int J Epidemiol*, **36**(1), 104–7.

GRUPPO-OKkio (2010) OKkio alla SALUTE: sintesi dei risultati 2010.

van der Gugten, A. C., Koopman, M., Evelein, A. M., Verheij, T. J., Uiterwaal, C. S. and van der Ent, C. K. (2012) Rapid early weight gain is associated with wheeze and reduced lung function in childhood. *Eur Respir J*, **39**(2), 403–10.

Gurrin, L. C., Scurrah, K. J. and Hazelton, M. L. (2005) Tutorial in biostatistics: spline smoothing with linear mixed models. *Statistics in medicine*, **24**(21), 3361–3381.

Haschke, F. and van't Hof, M. A. (2000) Euro-Growth references for breast-fed boys and girls: influence of breast-feeding and solids on growth until 36 months of age. Euro-Growth Study Group. *J Pediatr Gastroenterol Nutr*, **31 Suppl 1**, S60–71.

Hauspie, R. C. Cameron, N. and Molinari, L. (eds) (2004) *Methods in Human Growth Research.* Cambridge University Press.

Hauspie, R. (1998) Curve-fitting. In *The Cambridge Encyclopedia of Human Growth and Development* (Eds S. Ulijaszek, F. Johnston and M. A. Preece). Cambridge: Cambridge University Press.

Hediger, M. L., Overpeck, M. D., Ruan, W. J. and Troendle, J. F. (2000) Early infant feeding and growth status of US-born infants and children aged 4-71 mo: analyses from the third National Health and Nutrition Examination Survey, 1988-1994. *Am J Clin Nutr*, **72**(1), 159–67.

Heilbrun, L. K., Ross, P. D., Wasnich, R. D., Yano, K. and Vogel, J. M. (1991) Characteristics of respondents and nonrespondents in a prospective study of osteoporosis. *Journal of clinical epidemiology*, **44**(3), 233–239.

Hermanussen, M., Staub, K., Assmann, C. and van Buuren, S. (2012) Dilemmas in choosing and using growth charts. *Pediatric Endocrinology Reviews (PER)*, **9**(3), 650.

Hernan, M. and Robins, J. M. (2012) Causal inference.

Hernan, M. A., Hernandez-Diaz, S. and Robins, J. M. (2004) A structural approach to selection bias. *Epidemiology*, **15**(5), 615–25.

Hindmarsh, P. C., Geary, M. P., Rodeck, C. H., Kingdom, J. C. and Cole, T. J. (2008) Factors predicting ante- and postnatal growth. *Pediatr Res*, **63**(1), 99–102.

de Hoog, M. L., van Eijsden, M., Stronks, K., Gemke, R. J. and Vrijkotte, T. G. (2011) The role of infant feeding practices in the explanation for ethnic differences in infant growth: the amsterdam born children and their development study. *Br J Nutr*, **106**(10), 1592–601.

van den Hooven, E. H., Pierik, F. H., de Kluizenaar, Y., Willemsen, S. P., Hofman, A., van Ratingen, S. W., Zandveld, P. Y., Mackenbach, J. P., Steegers, E. A., Miedema, H. M. *et al.* (2012) Air pollution exposure during pregnancy, ultrasound measures of fetal growth, and adverse birth outcomes: a prospective cohort study. *Environmental health perspectives*, **120**(1), 150.

Hoover, R. N., Hyer, M., Pfeiffer, R. M., Adam, E., Bond, B., Cheville, A. L., Colton, T., Hartge, P., Hatch, E. E., Herbst, A. L. *et al.* (2011) Adverse health outcomes in women exposed in utero to diethylstilbestrol. *New England Journal of Medicine*, **365**(14), 1304–1314.

Howe, L. D., Tilling, K., Galobardes, B., Smith, G. D., Ness, A. R. and Lawlor, D. A. (2011) Socioeconomic disparities in trajectories of adiposity across childhood. *International Journal of Pediatric Obesity*, **6**(2Part2), e144–e153.

Howe, L. D., Tilling, K., Galobardes, B., Smith, G. D., Gunnell, D. and Lawlor, D. A. (2012) Socioeconomic differences in childhood growth trajectories: at what age do height inequalities emerge? *Journal of epidemiology and community health*, **66**(2), 143–148.

Howe, L. D., Tilling, K., Galobardes, B. and Lawlor, D. A. (2013) Loss to follow-up in cohort studies: bias in estimates of socioeconomic inequalities. *Epidemiology*, **24**(1), 1–9.

Hui, L. L., Leung, G. M., Cowling, B. J., Lam, T. H. and Schooling, C. M. (2010) Determinants of infant growth: Evidence from Hong Kong's "Children of 1997" birth cohort. *Ann Epidemiol*, **20**(11), 827–35.

Hui, L. L., Leung, G. M., Wong, M. Y., Lam, T. H. and Schooling, C. M. (2012) Small for gestational age and age at puberty: evidence from hong kong's "children of 1997" birth cohort. *Am J Epidemiol*, **176**(9), 785–93.

Huxley, R. R., Shiell, A. W. and Law, C. M. (2000) The role of size at birth and postnatal catch-up growth in determining systolic blood pressure: a systematic review of the literature. *J Hypertens*, **18**(7), 815–31.

Jacobsen, T. N., Nohr, E. A. and Frydenberg, M. (2010) Selection by socioeconomic factors into the danish national birth cohort. *European journal of epidemiology*, **25**(5), 349–355.

Jacquemyn, Y., Osmanovic, F. and Martens, G. (2006) Preeclampsia and birthweight by gestational age in singleton pregnancies in flanders, belgium: a prospective study. *Clinical and experimental obstetrico & gynecology*, **33**(2), 96–98.

Jansen, P. W., Tiemeier, H., Looman, C. W., Jaddoe, V. W., Hofman, A., Moll, H. A., Steegers, E. A., Verhulst, F. C., Mackenbach, J. P. and Raat, H. (2009) Explaining educational inequalities in birthweight: the generation r study. *Paediatr Perinat Epidemiol*, **23**(3), 216–28.

Jenss, R. M. and Bayley, N. (1937) A mathematical method for studying the growth of a child. *Human Biology*, **9**(4), 556–563.

Jeric, M., Roje, D., Medic, N., Strinic, T., Mestrovic, Z. and Vulic, M. (2013) Maternal pre-pregnancy underweight and fetal growth in relation to institute of medicine recommendations for gestational weight gain. *Early Human Development*, **89**, 277–81.

Johnson, L., Llewellyn, C. H., van Jaarsveld, C. H., Cole, T. J. and Wardle, J. (2011) Genetic and environmental influences on infant growth: prospective analysis of the Gemini twin birth cohort. *PLoS One*, **6**(5), e19918.

Jolicoeur, P., Pontier, J., Pernin, M. O. and Sempe, M. (1988) A lifetime asymptotic growth curve for human height. *Biometrics*, **44**(4), 995–1003.

Jones-Smith, J., Neufeld, L., Laraia, B., Ramakrishnan, U., Garcia-Guerra, A. and Fernald, L. (2013) Early life growth trajectories and future risk for overweight. *Nutrition & diabetes*, **3**(2), e60.

de Jong, F., Monuteaux, M. C., van Elburg, R. M., Gillman, M. W. and Belfort, M. B. (2012) Systematic review and meta-analysis of preterm birth and later systolic blood pressure. *Hypertension*, **59**(2), 226–34.

Jung, T. and Wickrama, K. (2008) An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass*, **2**(1), 302–317.

Kang Sim, D. E., Cappiello, M., Castillo, M., Lozoff, B., Martinez, S., Blanco, E. and Gahagan, S. (2012) Postnatal growth patterns in a chilean cohort: The role of ses and family environment. *Int J Pediatr*, **2012**, 354060.

Karlberg, J. (1987) On the modelling of human growth. *Stat Med*, **6**(2), 185–92.

Karlberg, J. (1989) A biologically-oriented mathematical model (ICP) for human growth. *Acta Paediatr Scand Suppl*, **350**, 70–94.

Karlberg, J. (1998) The human growth curve. In *The Cambridge Encyclopedia of Human Growth and Development* (Eds S. Ulijaszek, F. Johnston and M. A. Preece). Cambridge: Cambridge University Press.

Kenny, L. C., Lavender, T., McNamee, R., ONeill, S. M., Mills, T. and Khashan, A. S. (2013) Advanced maternal age and adverse pregnancy outcome: Evidence from a large contemporary cohort. *PLoS ONE*, **8**(2).

Kim, J. and Peterson, K. E. (2008) Association of infant child care with infant feeding practices and weight gain among US infants. *Arch Pediatr Adolesc Med*, **162**(7), 627–33.

Kirchengast, S. and Hartmann, B. (2003) Impact of maternal age and maternal somatic characteristics on newborn size. *American Journal of Human Biology*, **15**(2), 220–228.

Knight, B., Shields, B. M., Turner, M., Powell, R. J., Yajnik, C. S. and Hattersley, A. T. (2005) Evidence of genetic regulation of fetal longitudinal growth. *Early Hum Dev*, **81**(10), 823–31.

Koo, Y.-J., Ryu, H.-M., Yang, J.-H., Lim, J.-H., Lee, J.-E., Kim, M.-Y. and Chung, J.-H. (2012) Pregnancy outcomes according to increasing maternal age. *Taiwanese Journal of Obstetrics and Gynecology*, **51**(1), 60–65.

Kramer, M. S. (1987) Intrauterine growth and gestational duration determinants. *Pediatrics*, **80**(4), 502–11.

Kramer, M. S. and Kakuma, R. (2012) Optimal duration of exclusive breastfeeding. *Cochrane Database Syst Rev*, **8**, CD003517.

Kramer, M. S., Guo, T., Platt, R. W., Shapiro, S., Collet, J. P., Chalmers, B., Hodnett, E., Sevkovskaya, Z., Dzikovich, I. and Vanilovich, I. (2002) Breastfeeding and infant growth: biology or bias? *Pediatrics*, **110**(2 Pt 1), 343–7.

Kramer, M. S., Guo, T., Platt, R. W., Vanilovich, I., Sevkovskaya, Z., Dzikovich, I., Michaelsen, K. F. and Dewey, K. (2004) Feeding effects on growth during infancy. *J Pediatr*, **145**(5), 600–5.

Kramer, M. S., Moodie, E. E., Dahhou, M. and Platt, R. W. (2011) Breastfeeding and infant size: evidence of reverse causality. *Am J Epidemiol*, **173**(9), 978–83.

Kramer, M. S., Moodie, E. E. and Platt, R. W. (2012) Infant feeding and growth: Can we answer the causal question? *Epidemiology*, **23**(6), 790–794.

Kuh, D. and Ben-Shlomo, Y. (2004) *A Life Course Approach to Chronic Disease Epidemiology*. Oxford University Press, USA.

Laurent, O., Wu, J., Li, L., Chung, J. and Bartell, S. (2013) Investigating the association between birth weight and complementary air pollution metrics: a cohort study. *Environmental Health*, **12**(1), 18.

Lazarus, R., Baur, L., Webb, K. and Blyth, F. (1996) Adiposity and body mass indices in children: Benn's index and other weight for height indices as measures of relative adiposity. *Int J Obes Relat Metab Disord*, **20**(5), 406–12.

Leunissen, R. W., Kerkhof, G. F., Stijnen, T. and Hokken-Koelega, A. (2009) Timing and tempo of first-year rapid growth in relation to cardiovascular and metabolic risk profile in early adulthood. *JAMA: the journal of the American Medical Association*, **301**(21), 2234–2242.

Leunissen, R. W., Kerkhof, G. F., Stijnen, T. and Hokken-Koelega, A. C. (2012) Effect of birth size and catch-up growth on adult blood pressure and carotid intima-media thickness. *Horm Res Paediatr*, **77**(6), 394–401.

Linabery, A. M., Nahhas, R. W., Johnson, W., Choh, A. C., Towne, B., Odegaard, A. O., Czerwinski, S. A. and Demerath, E. W. (2012) Stronger influence of maternal than paternal obesity on infant and early childhood body mass index: the fels longitudinal study. *Pediatr Obes*, **[Epub ahead of print]**.

Lindstrom, M. J. and Bates, D. M. (1990) Nonlinear mixed effects models for repeated measures data. *Biometrics*, pp. 673–687.

Lourenco, B. H., Villamor, E., Augusto, R. A. and Cardoso, M. A. (2012) Determinants of linear growth from infancy to school-aged years: a population-based follow-up study in urban amazonian children. *BMC Public Health*, **12**, 265.

Maisonet, M., Correa, A., Misra, D. and Jaakkola, J. J. (2004) A review of the literature on the effects of ambient air pollution on fetal growth. *Environmental Research*, **95**(1), 106–115.

Malabarey, O. T., Balayla, J., Klam, S. L., Shrim, A. and Abenhaim, H. A. (2012) Pregnancies in young adolescent mothers: a population-based study on 37 million births. *Journal of Pediatric and Adolescent Gynecology*, **25**, 98–102.

Malina, R. (1998) Post-natal growth and maturation. In *The Cambridge Encyclopedia of Human Growth and Development* (Eds S. Ulijaszek, F. Johnston and M. A. Preece). Cambridge: Cambridge University Press.

Mamun, A. A., Callaway, L. K., O'Callaghan, M. J., Williams, G. M., Najman, J. M., Alati, R., Clavarino, A. and Lawlor, D. A. (2011) Associations of maternal pre-pregnancy obesity and excess pregnancy weight gains with adverse pregnancy outcomes and length of hospital stay. *BMC pregnancy and childbirth*, **11**(1), 62.

Martin, R. M., Smith, G. D., Frankel, S. and Gunnell, D. (2004) Parents' growth in childhood and the birth weight of their offspring. *Epidemiology*, **15**(3), 308–16.

Martin Bland, J. and Altman, D. (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet*, **327**(8476), 307–310.

Matijasevich, A., Howe, L. D., Tilling, K., Santos, I. S., Barros, A. J. and Lawlor, D. A. (2012) Maternal education inequalities in height growth rates in early childhood: 2004 pelotas birth cohort study. *Paediatric and Perinatal Epidemiology*, **26**, 236–49.

Mehta, A., Hindmarsh, P. C., Stanhope, R. G., Turton, J. P., Cole, T. J., Preece, M. A. and Dattani, M. T. (2005) The role of growth hormone in determining birth size and early postnatal growth, using congenital growth hormone deficiency (GHD) as a model. *Clin Endocrinol (Oxf)*, **63**(2), 223–31.

Metzger, B. E., Lowe, L. P., Dyer, A. R., Trimble, E. R., Chaovarindr, U., Coustan, D. R., Hadden, D. R., McCance, D. R., Hod, M., McIntyre, H. D., Oats, J. J., Persson, B., Rogers, M. S. and Sacks, D. A. (2008) Hyperglycemia and adverse pregnancy outcomes. *N Engl J Med*, **358**(19), 1991–2002.

Mihrshahi, S., Battistutta, D., Magarey, A. and Daniels, L. A. (2011) Determinants of rapid weight gain during infancy: baseline results from the NOURISH randomised controlled trial. *BMC Pediatr*, **11**, 99.

Modi, N., Murgasova, D., Ruager-Martin, R., Thomas, E. L., Hyde, M. J., Gale, C., Santhakumaran, S., Doré, C. J., Alavi, A. and Bell, J. D. (2011) The influence of maternal body mass index on infant adiposity and hepatic lipid content. *Pediatric research*, **70**(3), 287–291.

Monasta, L., Batty, G. D., Cattaneo, A., Lutje, V., Ronfani, L., Van Lenthe, F. J. and Brug, J. (2010) Early-life determinants of overweight and obesity: a review of systematic reviews. *Obes Rev*, **11**(10), 695–708.

Monteiro, P. O. and Victora, C. G. (2005) Rapid growth in infancy and childhood and obesity in later life–a systematic review. *Obes Rev*, **6**(2), 143–54.

Nafstad, P., Jaakkola, J. J., Hagen, J. A., Pedersen, B. S., Qvigstad, E., Botten, G. and Kongerud, J. (1997) Weight gain during the first year of life in relation to maternal smoking and breast feeding in Norway. *J Epidemiol Community Health*, **51**(3), 261–5.

Nilsen, R. M., Vollset, S. E., Gjessing, H. K., Skjærven, R., Melve, K. K., Schreuder, P., Alsaker, E. R., Haug, K., Daltveit, A. K. and Magnus, P. (2009) Self-selection and bias in a large prospective pregnancy cohort in Norway. *Paediatric and perinatal epidemiology*, **23**(6), 597–608.

Nobile, C. G., Raffaele, G., Altomare, C. and Pavia, M. (2007) Influence of maternal and social factors as predictors of low birth weight in Italy. *BMC Public Health*, **7**, 192.

Nohr, E. A., Frydenberg, M., Henriksen, T. B. and Olsen, J. (2006) Does low participation in cohort studies induce bias? *Epidemiology*, **17**(4), 413–8.

Ogden, C. and Carroll, M. (2010) Nchs health e-stat: prevalence of obesity among children and adolescentsunited states, trends 1963–1965 through 2007–2008.

Ohlsson, C., Lorentzon, M., Norjavaara, E. and Kindblom, J. M. (2012) Age at adiposity rebound is associated with fat mass in young adult malesthe good study. *PloS one*, **7**(11), e49404.

Olsson, D., Mogren, I. and Forsberg, B. (2013) Air pollution exposure in early pregnancy and adverse pregnancy outcomes: a register-based cohort study. *BMJ open*, **3**(2).

Ong, K. K. (2007) Catch-up growth in small for gestational age babies: good or bad? *Current Opinion in Endocrinology, Diabetes and Obesity*, **14**(1), 30–34.

Ong, K. K. (2010) Early determinants of obesity. *Endocr Dev*, **19**, 53–61.

Ong, K. K. and Loos, R. J. (2006) Rapid infancy weight gain and subsequent obesity: systematic reviews and hopeful suggestions. *Acta Paediatr*, **95**(8), 904–8.

Ong, K. K., Ahmed, M. L., Emmett, P. M., Preece, M. A. and Dunger, D. B. (2000) Association between postnatal catch-up growth and obesity in childhood: prospective cohort study. *BMJ*, **320**(7240), 967–71.

Ong, K. K., Preece, M. A., Emmett, P. M., Ahmed, M. L., Dunger, D. B. and Team, A. S. (2002) Size at birth and early childhood growth in relation to maternal smoking, parity and infant breast-feeding: longitudinal birth cohort study and analysis. *Pediatr Res*, **52**(6), 863–7.

Ong, K. K., Bann, D., Wills, A. K., Ward, K., Adams, J. E., Hardy, R., Kuh, D., National Survey of, H., Development, S. and Data Collection, T. (2012) Timing of voice breaking in males associated with growth and weight gain across the life course. *J Clin Endocrinol Metab*, **97**(8), 2844–52.

Ouyang, F., Parker, M., Cerda, S., Pearson, C., Fu, L., Gillman, M. W., Zuckerman, B. and Wang, X. (2012) Placental weight mediates the effects of prenatal factors on fetal growth: the extent differs by preterm status. *Obesity (Silver Spring)*, **[Epub ahead of print]**.

Padula, A. M., Mortimer, K., Hubbard, A., Lurmann, F., Jerrett, M. and Tager, I. B. (2012) Exposure to traffic-related air pollution during pregnancy and term low birth weight: Estimation of causal associations in a semiparametric model. *American journal of epidemiology*, **176**(9), 815–824.

Pearl, J. (1995) Causal diagrams for empirical research. *Biometrika*, **82**(4), 669–688.

Pearl, J. (2000) *Causality: Models, Reasoning, and Inference*. Cambridge, UK: Cambridge University Press.

Pizzi, C., De Stavola, B., Merletti, F., Bellocco, R., dos Santos Silva, I., Pearce, N. and Richiardi, L. (2011) Sample selection and validity of exposure–disease association estimates in cohort studies. *Journal of epidemiology and community health*, **65**(5), 407–411.

Pizzi, C., De Stavola, B. L., Pearce, N., Lazzarato, F., Ghiotti, P., Merletti, F. and Richiardi, L. (2012) Selection bias and patterns of confounding in cohort studies: the case of the ninfea web-based birth cohort. *Journal of Epidemiology and Community Health*, **66**(11), 976–81.

Power, C., Li, L., Manor, O. and Smith, G. D. (2003) Combination of low birth weight and high adult body mass index: at what age is it established and what are its determinants? *Journal of epidemiology and community health*, **57**(12), 969–973.

Preece, M. A. and Baines, M. J. (1978) A new family of mathematical models describing the human growth curve. *Ann Hum Biol*, **5**(1), 1–24.

Prentice, A., Dibba, B., Sawo, Y. and Cole, T. J. (2012) The effect of prepubertal calcium carbonate supplementation on the age of peak height velocity in Gambian adolescents. *The American journal of clinical nutrition*, **96**(5), 1042–1050.

Prentice, P. and Viner, R. (2012) Pubertal timing and adult obesity and cardiometabolic risk in women and men: a systematic review and meta-analysis. *International Journal of Obesity*, **[Epub ahead of print]**.

Raikkonen, K., Forsen, T., Henriksson, M., Kajantie, E., Heinonen, K., Pesonen, A. K., Leskinen, J. T., Laaksonen, I., Osmond, C., Barker, D. J. and Eriksson, J. G. (2009) Growth trajectories and intellectual abilities in young adulthood: The helsinki birth cohort study. *Am J Epidemiol*, **170**(4), 447–55.

Regnault, N., Botton, J., Forhan, A., Hankard, R., Thiebaugeorges, O., Hillier, T. A., Kaminski, M., Heude, B. and Charles, M. A. (2010) Determinants of early ponderal and statural growth in full-term infants in the eden mother-child cohort study. *Am J Clin Nutr*, **92**(3), 594–602.

Richiardi, L., Baussano, I., Vizzini, L., Douwes, J., Pearce, N. and Merletti, F. (2007) Feasibility of recruiting a birth cohort through the Internet: the experience of the NINFEA cohort. *Eur J Epidemiol*, **22**(12), 831–7.

Risnes, K. R., Vatten, L. J., Baker, J. L., Jameson, K., Sovio, U., Kajantie, E., Osler, M., Morley, R., Jokela, M., Painter, R. C. *et al.* (2011) Birthweight and mortality in adulthood: a systematic review and meta-analysis. *International journal of epidemiology*, **40**(3), 647–661.

Roggero, P., Giannì, M., Garbarino, F. and Mosca, F. (2013) Consequences of prematurity on adult morbidities. *European journal of internal medicine*, **[Epub ahead of print]**.

Roland, M. C. P., Friis, C. M., Voldner, N., Godang, K., Bollerslev, J., Haugen, G. and Henriksen, T. (2012) Fetal growth versus birthweight: The role of placenta versus other determinants. *PloS one*, **7**(6), e39324.

Rolland-Cachera, M. F. (1998) Adiposity rebound and prediction of adult fatness. In *The Cambridge Encyclopedia of Human Growth and Development* (Eds S. Ulijaszek, F. Johnston and M. A. Preece). Cambridge University Press.

Rolland-Cachera, M. F., Deheeger, M., Guilloud-Bataille, M., Avons, P., Patois, E. and Sempe, M. (1987) Tracking the development of adiposity from one month of age to adulthood. *Ann Hum Biol*, **14**(3), 219–29.

Rothman, K., Greenland, S. and Lash, T. J. (2008) Distinguishing selection bias from confounding. In *Modern Epidemiology*, pp. 136–137. Philadelphia: Lippincott Williams & Wilkins.

Royston, P. and Altman, D. G. (1994) Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Applied Statistics*, pp. 429–467.

Royston, P., Altman, D. G. and Sauerbrei, W. (2006) Dichotomizing continuous predictors in multiple regression: a bad idea. *Statistics in medicine*, **25**(1), 127–141.

Rubin, D. (1976) Inference and missing data. *Biometrika*, **63**, 581–592.

Ruppert, D., Wand, M. P. and Carroll, R. J. (2003) *Semiparametric regression*. Cambridge University Press.

Salafia, C. M., Zhang, J., Charles, A. K., Bresnahan, M., Shrout, P., Sun, W. and Maas, E. M. (2008) Placental characteristics and birthweight. *Paediatr Perinat Epidemiol*, **22**(3), 229–39.

Salomon, L. J., Pizzi, C., Gasparrini, A., Bernard, J.-P. and Ville, Y. (2010) Prediction of the date of delivery based on first trimester ultrasound measurements: an independent method from estimated date of conception. *Journal of Maternal-Fetal and Neonatal Medicine*, **23**(1), 1–9.

Samson, A., Lavielle, M. and Mentré, F. (2007) The SAEM algorithm for group comparison tests in longitudinal data analysis based on non-linear mixed-effects model. *Statistics in medicine*, **26**(27), 4860–4875.

Schell, L. (1998) Smoking. In *The Cambridge Encyclopedia of Human Growth and Development* (Eds S. Ulijaszek, F. Johnston and M. A. Preece). Cambridge: Cambridge University Press.

Schimek, M. G. (2009) Semiparametric penalized generalized additive models for environmental research and epidemiology. *Environmetrics*, **20**(6), 699–717.

Scholl, T. (1998) Teenage pregnancy. In *The Cambridge Encyclopedia of Human Growth and Development* (Eds S. Ulijaszek, F. Johnston and M. A. Preece). Cambridge: Cambridge University Press.

Selling, K. E., Carstensen, J., Finnström, O. and Sydsjö, G. (2006) Intergenerational effects of preterm birth and reduced intrauterine growth: a population-based study of swedish mother–offspring pairs. *BJOG: An International Journal of Obstetrics & Gynaecology*, **113**(4), 430–440.

Shah, P. S. (2010) Paternal factors and low birthweight, preterm, and small for gestational age births: a systematic review. *American journal of obstetrics and gynecology*, **202**(2), 103–123.

Silva, L. M., van Rossem, L., Jansen, P. W., Hokken-Koelega, A. C., Moll, H. A., Hofman, A., Mackenbach, J. P., Jaddoe, V. W. and Raat, H. (2012) Children of low socioeconomic status show accelerated linear growth in early childhood; results from the generation r study. *PLoS One*, **7**(5), e37356.

Silverwood, R., Leon, D. and B.L., D. S. (2009a) Long-term trends in BMI: are contemporary childhood BMI growth references appropriate when looking at historical datasets? *Longitudinal and Life Course Studies*, **1**, 27–44.

Silverwood, R. J., De Stavola, B. L., Cole, T. J. and Leon, D. A. (2009b) BMI peak in infancy as a predictor for later BMI in the Uppsala Family Study. *Int J Obes (Lond)*, **33**(8), 929–37.

Spiegler, J., Stichtenoth, G., Weichert, J., Konig, I. R., Schlaud, M., A, V. D. W., Olbertz, D., Gurth, H., Schiffmann, J. H., Bohnhorst, B., Gortner, L., Herting, E., Gopel, W. and The German Neonatal Network, G. N. N. (2013) Pregnancy risk factors for very premature delivery: what role do hypertension, obesity and diabetes play? *Arch Gynecol Obstet.*

Strobino, D. M., Ensminger, M. E., Kim, Y. J. and Nanda, J. (1995) Mechanisms for maternal age differences in birth weight. *American Journal of Epidemiology*, **142**(5), 504–514.

Tanner, J. (1962) *Growth at adolescence*, volume 2. Blackwell Scientific Publications Oxford.

Tanner, J. M. (1986) Growth as a target-seeking functions:catch-up and catch-down growth. In *Human growth: a comprehensice treatise* (Ed. F. Falkner). New York: Plenum.

Tanner, J. M. (1989) *Foetus into Man.* Ware: Castlemand Publications.

Touwslager, R. N., Gielen, M., Derom, C., Mulder, A. L., Gerver, W. J., Zimmermann, L. J., Houben, A. J., Stehouwer, C. D., Vlietinck, R., Loos, R. J. and Zeegers, M. P. (2011) Determinants of infant growth in four age windows: a twin study. *J Pediatr*, **158**(4), 566–572 e2.

Traviss, G. D., West, R. M. and House, A. O. (2012) Maternal mental health and its association with infant growth at 6 months in ethnic groups: results from the born-in-bradford birth cohort study. *PLoS One*, **7**(2), e30707.

Tzoulaki, I., Sovio, U., Pillas, D., Hartikainen, A. L., Pouta, A., Laitinen, J., Tammelin, T. H., Jarvelin, M. R. and Elliott, P. (2010) Relation of immediate postnatal growth with obesity and related metabolic risk factors in adulthood: the northern Finland birth cohort 1966 study. *Am J Epidemiol*, **171**(9), 989–98.

Vansteelandt, S. (2012) *Estimation of direct and indirect effects, In Causality: Statistical Perspectives and Applications.* London: Wiley.

Victora, C. G., Morris, S. S., Barros, F. C., Horta, B. L., Weiderpass, E. and Tomasi, E. (1998) Breast-feeding and growth in Brazilian infants. *Am J Clin Nutr*, **67**(3), 452–8.

Vohr, B. R., McGarvey, S. T. and Tucker, R. (1999) Effects of maternal gestational diabetes on offspring adiposity at 4-7 years of age. *Diabetes Care*, **22**(8), 1284–1291.

Warner, M. and Ozanne, S. (2010) Mechanisms involved in the developmental programming of adulthood disease. *Biochem. J*, **427**, 333–347.

Warrington, N. M., Wu, Y. Y., Pennell, C. E., Marsh, J. A., Beilin, L. J., Palmer, L. J., Lye, S. J. and Briollais, L. (2013) Modelling BMI Trajectories in Children for Genetic Association Studies. *PloS one*, **8**(1), e53897.

Weaver, L. T. (2006) Rapid growth in infancy: balancing the interests of the child. *J Pediatr Gastroenterol Nutr*, **43**(4), 428–32.

Wehkalampi, K., Hovi, P., Dunkel, L., Strang-Karlsson, S., Jarvenpaa, A. L., Eriksson, J. G., Andersson, S. and Kajantie, E. (2011) Advanced pubertal growth spurt in subjects born preterm: the helsinki study of very low birth weight adults. *J Clin Endocrinol Metab*, **96**(2), 525–33.

Wells, J. C., Hallal, P. C., Reichert, F. F., Dumith, S. C., Menezes, A. M. and Victora, C. G. (2011) Associations of birth order with early growth and adolescent height, body composition, and blood pressure: prospective birth cohort from brazil. *American journal of epidemiology*, **174**(9), 1028–1035.

Whincup, P. H., Kaye, S. J., Owen, C. G., Huxley, R., Cook, D. G., Anazawa, S., Barrett-Connor, E., Bhargava, S. K., Birgisdottir, B. E., Carlsson, S. *et al.* (2008) Birth weight and risk of type 2 diabetes: a systematic review. *JAMA*, **300**(24), 2886–2897.

Whitcomb, B. W., Schisterman, E. F., Perkins, N. J. and Platt, R. W. (2009) Quantification of collider-stratification bias and the birthweight paradox. *Paediatric and perinatal epidemiology*, **23**(5), 394–402.

Wijlaars, L. P., Johnson, L., van Jaarsveld, C. H. and Wardle, J. (2011) Socioeconomic status and weight gain in early infancy. *International Journal of Obesity*, **35**(7), 963–970.

Williams, S. M. and Goulding, A. (2012) Early adiposity rebound is an important predictor of later obesity. *Obesity*, **17**(7), 1310–1310.

Woo, J. G., Guerrero, M. L., Ruiz-Palacios, G. M., Peng, Y. M., Herbers, P. M., Yao, W., Ortega, H., Davidson, B. S., McMahon, R. J. and Morrow, A. L. (2013) Specific infant feeding practices do not consistently explain variation in anthropometry at age 1 year in urban united states, mexico, and china cohorts. *J Nutr*, **143**(2), 166–74.

Wright, C. M., Parkinson, K. and Scott, J. (2006) Breast-feeding in a UK urban context: who breast-feeds, for how long and does it matter? *Public Health Nutr*, **9**(6), 686–91.

Xiong, X., Demianczuk, N. N., Saunders, L. D., Wang, F.-L. and Fraser, W. D. (2002) Impact of preeclampsia and gestational hypertension on birth weight by gestational age. *American journal of epidemiology*, **155**(3), 203–209.

Yang, S., Tilling, K., Martin, R., Davies, N., Ben-Shlomo, Y. and Kramer, M. S. (2011) Pre-natal and post-natal growth trajectories and childhood cognitive ability and mental health. *Int J Epidemiol*, **40**(5), 1215–26.

Zhang, J., Jiang, J., Himes, J. H., Liu, G., Huang, X., Guo, Y., Shi, J. and Shi, S. (2012) Determinants of high weight gain and high bmi status in the first three months in urban chinese infants. *Am J Hum Biol*, **24**(5), 633–9.