Adiposity and behavioural self-regulation in the early years of Quebec's universal childcare policy: a population-based, longitudinal study of children age 0 to 13 years

Tanya Jeannine Murphy

Department of Epidemiology, Biostatistics, and Occupational Health

McGill University, Montreal

December 2020

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfilment of the requirements of the degree of Doctor of Philosophy

© Tanya J. Murphy, 2020

Dedication

In loving memory of my father, Lorenzo Murphy.

1949 – 2017

Table of Contents

Abstract

Not all Canadian children are provided with the resources and support needed for healthy development. The high prevalence of childhood obesity in Canada and worldwide is one such unjust burden on youth, and socioeconomically disadvantaged children disproportionately suffer. Early childhood weight gain and behavioural self-regulation are independent predictors of adolescent health, and self-regulation may play a role in weight gain. Affordable high-quality preschool childcare programs should possibly be part of the public health approach to prevention of obesity and the promotion of child well-being. However, Quebec's universal childcare policy, *les Centres de la petite enfance* (CPE), has had mixed results. This thesis examines the effect of preschool childcare on adiposity and self-regulation, and of self-regulation on adiposity in Quebec children age 6 to 13 years.

The study population included 1657 children who participated in *l'Étude longitudinale du développement de l'enfant au Québec* (ELDEQ), a representative survey of Quebec children born in 1997-98. ELDEQ coincided with the early years of Quebec's CPE program and the participants became eligible for it at 2 years old. This context and the rich longitudinal data from ELDEQ provided an opportunity to estimate overall and sub-group causal effects of childcare on self-regulation and adiposity, and whether effects changed with age. Nevertheless, to minimize confounding, account for serial outcomes, and maintain representativeness in these observational studies, a variety of sophisticated statistical methods was needed.

In manuscript 1, serial measures of BMI z-score and obesity status from 6 to 13 years of age were regressed on detailed measures of childcare used from 2 to 5 years, and adjusted for pre-childcare variables using Bayesian multilevel (generalized) linear models. Population-averaged treatment effects were estimated for four counterfactual childcare profiles: 1) parental care, 2) center-based, 3) CPE-regulated home-based, and 4) unregulated home-based. Centre-based childcare led to higher adiposity than CPE-regulated home-based care and parental care. However, CPE-regulated home-based care was slightly favourable compared to unregulated home-based care.

In manuscript 2, I used a similar analytic strategy to estimate the effect of preschool childcare on self-regulation from age 6 to 12 years. First, I estimated plausible values of latent self-regulation

scores at 6, 7, 8, 10 and 12 years from manifest behaviours of inattentiveness, hyperactivity, and impulsivity rated by mothers and teachers. CPE-regulated centre- and home-based care had small detrimental effects on self-regulation compared parental care. In both manuscripts 1 and 2, effect measure modification by family disadvantage and sex was somewhat unclear, but less advantaged children were generally not at higher risk of adiposity or self-regulation deficits due to childcare type.

In manuscript 3, I estimated the association between serial measures of adiposity from age 7 to 13 years and the 3-year mean of self-regulation score prior to each adiposity measure, controlling for potential confounders using inverse probability of exposure weighting. Self-regulation was not associated with mean BMI z-score or probability of obesity, regardless of age and sex.

High-quality public childcare has many demonstrated benefits for children's cognitive and social development, as well as for mothers' labour force participation, but there is little evidence from past research or the new results of this thesis that public childcare programs have had benefits for adiposity or self-regulation in school-aged children, generally, or in disadvantaged children. Likewise, while the promotion of self-regulation may have many benefits, it is unlikely to be an effective obesity prevention strategy in early and middle childhood. The intensification of the services offered in public childcare may achieve positive results on adiposity and self-regulation, but broader structural and environmental improvements offer more promise.

Résumé

Les enfants canadiens ne bénéficient pas tous des ressources et du soutien nécessaires pour un développement sain. La forte prévalence de l'obésité juvénile au Canada comme ailleurs dans le monde est l'un de ces fardeaux injustes pour les jeunes, et les enfants défavorisés sur le plan socio-économique en souffrent de manière disproportionnée. La prise de poids et l'autorégulation comportementale pendant la petite enfance sont des prédicteurs indépendants de la santé des adolescents, et l'autorégulation peut avoir un rôle dans la prise de poids. Des programmes de services de garde abordables et de haute qualité devraient possiblement faire partie du projet de la santé publique pour la prévention de l'obésité et la promotion du bien-être de l'enfant. Cependant, la politique universelle de services de garde au Québec, les Centres de la petite enfance (CPE), a démontré des résultats mitigés. La présente thèse doctorale examine l'effet de la fréquentation des services de garde préscolaire sur l'adiposité et l'autorégulation, et de l'autorégulation sur l'adiposité chez les enfants québécois âgés de 6 à 13 ans.

La population étudiée comprenait 1657 enfants qui ont participé à l'Étude longitudinale du développement de l'enfant au Québec (ELDEQ), une enquête représentative des enfants québécois nés en 1997-1998. L'ELDEQ a coïncidé avec les premières années du programme CPE, et les participants y sont devenus admissibles à l'âge de 2 ans. Ce contexte et les riches données longitudinales de l'ELDEQ ont permis d'estimer les effets des services de garde sur l'ensemble et sur les sous-groupes d'enfants sur l'autorégulation et l'adiposité, et si les effets évoluaient avec l'âge. Néanmoins, pour minimiser les facteurs de confusion, pour tenir compte des résultats en série et pour maintenir la représentativité dans ces études observationnelles, une gamme de méthodes statistiques sophistiquées était nécessaire.

Dans le premier article, j'ai estimé les effets des services de garde utilisé entre l'âge de 2 et 5 ans, y compris des mesures détaillées de la fréquentation, sur des mesures en série de l'indice de masse corporelle (IMC, valeurs du z) et du statut d'obésité de l'âge de 6 à 13 ans. Les effets ont été estimés par régression linéaire (généralisée) sous un cadre hiérarchique bayésien, ajustées pour variables antécédentes. Les effets moyens marginaux pour la population ont été estimés pour quatre profils de garde hypothétiques : 1) la garde parentale, 2) en garderie, 3) en CPE à domicile, ou 4) à domicile non réglementée. Les garderies entraînaient une adiposité plus élevée

que la garde parentale et les domiciles non réglementés. Cependant, les soins en CPE à domicile étaient légèrement plus favorables par rapport aux soins à domicile non réglementés.

Pour le deuxième article, j'ai utilisé une stratégie analytique semblable pour estimer l'effet de la garde préscolaire sur l'autorégulation entre l'âge de 6 et 12 ans. D'abords, j'ai estimé des valeurs plausibles du score d'autorégulation latente à l'âge de 6, 7, 8, 10 et 12 ans à partir de comportements manifestes d'inattention, d'hyperactivité et d'impulsivité, évalués par les mères et les enseignants. Les garderies et CPE à domicile avaient de légers effets néfastes sur l'autorégulation par rapport à la garde parentale. Dans ces deux premiers articles, la modification des effets du moyen de garde en fonction du sexe de l'enfant ou du milieu familiale n'était pas tout à fait claire, mais les enfants issus de milieux désavantagés n'étaient pas, en général, plus à risque d'adiposité ou de déficits d'autorégulation en raison du moyen de garde.

Pour le troisième article, j'ai estimé l'association entre une série de mesures de l'adiposité entre l'âge de 7 et 13 ans et la moyenne sur 3 ans du score d'autorégulation précédant chaque mesure d'adiposité, en contrôlant les facteurs de confusion. L'autorégulation n'était pas associée aux valeurs de z IMC moyen ou à la probabilité d'obésité, quels que soient l'âge ou le sexe de l'enfant.

Les services de garde publics de haute qualité ont démontré de nombreux bienfaits pour le développement cognitif et social des enfants, ainsi que pour la participation des mères dans le marché du travail, mais il y a peu de preuves issues de recherches antérieures ou des nouveaux résultats de cette thèse que les services de garde règlementés ont eu des effets bénéfiques sur l'adiposité ou l'autorégulation chez les enfants d'âge scolaire, en général, ou chez les enfants issus de milieux familiaux défavorisés, en particulier. De la même manière, bien que la promotion de l'autorégulation puisse avoir de nombreux avantages, il est peu probable qu'elle présente une stratégie efficace de prévention de l'obésité chez les jeunes enfants ainsi que les enfants préadolescents. C'est possible que l'augmentation des services offerts dans les garderies publiques pourrait mener à des résultats positifs sur l'adiposité et l'autorégulation, mais des améliorations politiques et environnementales sont plus prometteuses.

Acknowledgements

This thesis was possible because of the contributions of many, many generous and brilliant people. The professors, staff, and colleagues in the Department of Epidemiology, Biostatistics, and Occupational Health (EBOH) at McGill University gave me excellent training and support during my doctoral training. Foremost, I thank my co-supervisors Dr. Seungmi Yang and Dr. Jay Kaufman who have guided me through my doctoral research with wisdom and kindness. They provided me with all the mentorship and instruction a PhD student could hope for—outstanding methodologic expertise, balance between intellectual freedom and direction, patience, and moral support—times two! In addition, Seungmi helped me learn the fundamentals of child growth and behavioural development, which was a new substantive area for me, as well as psychometric methods.

I am grateful for the rich longitudinal data from the Quebec Longitudinal Study of Child Development (QLSCD). Participants, their parents, and teachers have donated hours of their time in providing this data. Thank you to the *Centre d'accès aux données de recherche de l'Institut de la statistique du Québec* (CADRISQ) and the study investigators for access to the well-organized and documented data, particularly Dr. Gilles Paradis.

I have also been helped with the methodological and subject-matter topics of my research by my committee members Drs. Patricia Li, Russell Steele, Daniel Weinstock. Dr. Sahir Bhatnagar, Institute of Health and Social Policy members, Dr. Arijit Nandi and Thomas Charters, and other members of 3PO have given me valuable feedback while I worked through study design challenges. Dr. Erin Strumpf was posing thought-provoking questions from the beginning when she reviewed my research proposal. While deciding whether to pursue a PhD and choosing a research area, I was greatly helped by regular conversations with Drs. Nandini Dendukuri, Lawrence Joseph, and James Hanley. My classmates and fellow perinatal journal club members brightened my days and supported me through periods of painful professional development and personal grief.

Last but not least, infinite thanks to my family: my husband, Eric Shintani; my children, Audrey, Alister, Michael, and Dylan; my parents, Francine and Lorenzo Murphy; my brothers, and parents-, brother-, and sisters-in-law. Your love, humour, and support carry and inspire me.

1.1 Statement of financial support

My studies were partially funded by a 2-year doctoral bursary from les Fonds de research du Québec Santé (FRQS). Prior to that, Seungmi Yang and Daniel Weinstock, through their CIHR grant, Nandini Dendukuri (through her Grand Challenges grant), and EBOH provided me with funding.

Contribution to original knowledge

The work in this thesis makes an original contribution to the existing literature on early childhood environments and child development. The first manuscript extends the study of individual-level use of childcare in Quebec children in the era of the *Centres de la petite enfance* program (CPE) and its effects on children's adiposity in their elementary school years. Previous work had not examined differences between CPE-regulated centre- and home-based care, separately, versus unregulated and parental care, and its differential effects by age and family disadvantage. The detailed measure of childcare use addresses the vague definition of childcare frequently used in past studies and cited as a major hindrance to progress by several review articles. Likewise, manuscript 2 is the first study to separately estimate the effects of the CPE-regulated centre- and home-based alternatives, compared to unregulated care, on behavioural self-regulation. The manuscript 3 examines the longitudinal effect of self-regulation on adiposity. To my knowledge only two studies have estimated the longitudinal relationship while adjusting for potential confounders. I estimated the relationship at multiple time points between 7 and 13 years of age, using flexible models for age and self-regulation.

This body of work is the realization of my own ideas and analyses.

Contribution by authors

Manuscript 1

Tanya Murphy (ORCID: 0000-0002-9643-8730): Conceptualization, Methodology, Formal analysis, Writing- Original draft preparation, Writing- Reviewing and Editing. **Jay Kaufman:** Conceptualization, Writing- Reviewing and Editing, Supervision. **Patricia Li** Writing-Reviewing and Editing. **Russell Steele** Methodology. **Seungmi Yang:** Conceptualization, Writing- Reviewing and Editing, Supervision.

Manuscript 2

Tanya Murphy: Conceptualization, Methodology, Formal analysis, Writing- Original draft preparation, Writing- Reviewing and Editing. **Jay Kaufman:** Conceptualization, Writing-Reviewing and Editing, Supervision. **Patricia Li** Guidance on literature search. **Russell Steele** Guidance on methodology. **Seungmi Yang:** Conceptualization, Writing- Reviewing and Editing, Supervision.

Manuscript 3

Tanya Murphy: Conceptualization, Methodology, Formal analysis, Writing- Original draft preparation, Writing- Reviewing and Editing. **Jay Kaufman:** Conceptualization, Writing-Reviewing and Editing, Supervision. **Patricia Li** Guidance on literature search. **Russell Steele** Guidance on methodology. **Seungmi Yang:** Conceptualization, Writing- Reviewing and Editing, Supervision.

List of Tables

List of Figures

List of Abbreviations

BMI	Body mass index
BMIz	Body mass index z-score, age- and sex-standardized according to the WHO 2007 standard curves.
CPE	Centre de la petite enfance
CI	Confidence interval estimated from classical or bootstrap methods
CrI	Credible interval estimated from Bayesian methods
ELDEQ	Étude longitudinale du développement de l'enfant au Québec
IOTF	International obesity task force
IQR	Interquartile range
ISQ	Institut de la statistique du Québec
NLSCY	National Longitudinal Survey of Children and Youth (Statistics Canada)
PSR	Poor self-regulation score
RD	Risk difference
SD	Standard deviation
SECCYD	Study of Early Child Care and Youth Development (National Institute of Child Health and Human Development, USA)
SR	Self-regulation

1 Introduction and background

Childhood obesity is a major health burden in Canada and worldwide, and excess weight gain often starts in early childhood. The prevalence of childhood obesity follows a social gradient that reflects inequities because obesity is preventable (Appleyard et al., 2005; Asada, 2005; Boone-Heinonen et al., 2016; Jester et al., 2005; Kakinami et al., 2014). Canada's wealth and long-standing welfare programs have not ensured health equity for children (Adamson & UNICEF, 2013). This inequity is seen at school entry (Carpiano et al., 2009; Laurin et al., 2012). Investment in early childhood education and care has long been believed to be important for promoting early learning and school readiness. Public childcare is also thought to be an important lever for equitable health promotion (Dornan & Woodhead, 2015; Mazza et al., 2017; McLaren & McIntyre, 2014).

In early childhood, a child's weight gain is influenced by the food and opportunities for physical activity that are provided by caregivers. As a child ages, his or her choices play a larger role in maintaining a healthy weight. Behavioural self-regulation, one's ability to moderate emotions and direct actions in the service of goals, is increasingly studied as an explanation for health behaviours and obesity (McClelland et al., 2018; A. L. Miller et al., 2018; Smithers et al., 2018). As a possible antecedent to healthy behaviours and obesity, and because children develop some capacity for reflection in the preschool years, more deliberate support for self-regulation development—beyond traditional child socialization and education—should perhaps start early. Early childhood educators are trained to expose children to a variety of day-to-day opportunities to make choices, interact with others, and to guide them on how to achieve desired outcomes. Hence, the preschool environment may influence adiposity and self-regulation in kindergarten and beyond.

The geographic and historical context of my thesis research is Quebec 1997 to 2010, a time of great change for Quebec families because of the implementation of the monumental provincial childcare policy, *les Centres de la petite enfance*, CPE. Historically, targeted intensive interventions in the USA have suggested childcare programs could improve child development and reduce disparities between more and less advantaged children at school entry, but the benefits of those early interventions have not been consistently replicated in large scale public childcare programs. The promotion of better self-regulation, as a possible antecedent to healthy behaviours and body weight is, in part, a response to unsuccessful obesity prevention efforts. However, the role of behavioural self-regulation development in adiposity differences is unclear.

1.1 Research objectives

My first objective was to estimate how much Quebec children's adiposity in kindergarten differed by type of preschool childcare, if differences persisted through the elementary school years, and whether adiposity-by-childcare patterns differed between children from more and less advantaged families.

Similarly, for my second objective, I estimated how much Quebec children's behavioural self-regulation in kindergarten differed by preschool childcare type and age of initiation of centre-based childcare, whether those differences persisted through the elementary school years, and whether the effect of childcare on self-regulation differed between children from more and less advantaged families.

Finally, for my third objective, I estimated how much differences in adiposity between age 7 to 13 years were attributable to differences in the 3-year mean behavioural self-regulation prior to each adiposity measure, and whether the association differed by age and sex.

1.2 Childhood obesity

The prevalence of obesity in Canadian children doubled from 6.3% in 1979 (WHO standard) (Shields & Tremblay, 2010) to 12% by 2004, and has since remained stable (Rao et al., 2016). But the prevalence has been much higher in some populations; for example, in 1997 approximately 20% of children age 10 to 12 year in Montreal inner city neighbourhoods were obese (O'Loughlin et al., 2000). Unlike in historically low-resource settings, the prevalence of under-nourishment and stunting in Canada is very low[1] (e.g. <2% (Roberts et al., 2012)).

International and Canadian health agencies advocate early obesity prevention (Abarca-Gómez et al., 2017; Bundy et al., 2017; Morinis et al., 2012; Public Health Agency of Canada, 2011). Obesity puts children at increased risk for type 2 diabetes, cardiovascular disease, and sleep and orthopedic problems (Atay & Bereket, 2016). Early childhood obesity, and adiposity more generally, tends to track over time (Geserick et al., 2018; Loeffler-Wirth et al., 2018; Singh et al., 2008). A large proportion of adolescents who are obese experienced their greatest acceleration in weight gain in early childhood (Geserick et al., 2018).

Obesity is known to be important but continuous changes may also be important. Adiposity is inherently a quantitative trait, but the importance of increases in different ranges of adiposity is unclear. Many studies of the risk of disease in overweight and obese children group overweight and obese into one category. In a statistical analysis, this grouping increases the number of "cases," hence, statistical power. However, Pearce et al. (2016) found young children who were overweight and normal weight had equivalent risk of developmental vulnerabilities, whereas the risk was higher for obese children. From a population point of view, generally, an upward shift in the distribution of a continuous trait is expected to generate more cases of disease (Rose, 1985). Metabolic and cardiovascular problems in children were found to rise continuously with BMI (Bell et al., 2007; Lawlor et al., 2010). Therefore, both continuous and categorical measures of adiposity are of interest, and when categories are used, overweight and obesity should not be grouped without substantial justification.

Adiposity is expected to change over the course of childhood. Mean adiposity is at its lowest around 3–5 years, then rises gradually until puberty, when it accelerates before stabilizing around 18–20 years (Wang & Chen, 2012). Patterns differ somewhat by sex. Therefore, age- and sex-

[1] However, the data are poor, and Canada does not submit prevalence estimates to international data repositories (e.g. the World Bank).

specific standards are used. The measurement of adiposity is also complex. The gold standard is fat mass as a percent of total body mass, which can be accurately measured by technologies such as dual-energy radiograph absorptiometry (DEXA), but collecting percent body fat is often not feasible for large epidemiologic studies. Body mass index (BMI; weight in kg divided by height in meters squared) is used as an adequate substitute (Freedman & Sherry, 2009). The concordance between DEXA and BMI categories is high, but it is poorer for overweight than for obesity; that is, many children classified as overweight according to BMI categories do not have excess body fatness according to DEXA. Additional measures such as tri-ponderal mass or waist circumference can improve discrimination in the overweight range (Peterson et al., 2017), but they do not have widely accepted standards. The most widely accepted normative age- and sex-specific standards for BMI are the International Obesity Task Force (IOTF) cut-offs (Cole et al., 2000; Cole et al., 2007) and WHO z-scores (Wang & Chen, 2012; WHO, 2019; WHO Multicentre Growth Reference Study Group, 2006). Appendix B.2 shows the relation between height and weight, or BMI, and WHO and IOTF cut-offs. The US Center for Disease Control (CDC) percentiles, a descriptive standard, are also commonly used. Height and weight should be measured by trained study staff because BMI calculated from parent-reported height and weight tends to be underestimated and the accuracy varies by SES or mother's education, and the child's age and sex (Dubois & Girard, 2007; Ontario Agency for Health Protection and Promotion et al., 2015).

1.3 Effects of preschool childcare

1.3.1 Quebec's *Centres de la petite enfance*

State-sponsored universal childcare in North America is recent and rare (McCain et al., 2007). In the early 2000s, Quebec was the only province or state in North America with a network of state-sponsored universal home- and centre-based childcare providers, les *Centres de la Petite Enfance* (CPE). Initiated in 1997, successively younger cohorts were eligible, regardless of parents' income or employment status. Parents were charged very low fees of $5 per day (until 2004) with additional incentives for parents on social assistance to enroll their children. Before the CPE program, eligible parents could apply for means-tested subsidies on an individual and annual basis (Friendly & Childcare Resource and Research Unit, 1997); therefore, the CPE program brought changes in the ease of access to low cost childcare for poor families but greatly changed childcare expenses for the middle class. The CPE program was part of a larger family policy aimed at anti-poverty, especially for lone parent households, and increased labour force participation by mothers (Jenson, 2002). One of the main goals of the policy was to eliminate the socioeconomic gradient in children's school readiness by providing high-quality care and an educational curriculum (Forest et al., 2007). However, unlike the early experimental or reformed Head Start programs discussed below, CPEs have not formally included coordination of health and social services or parenting classes.

A CPE is a childcare centre and the home-based providers under its supervision, governed by a parent-majority board. The centers are kindergarten-like environments with children grouped into classes by age. A single centre can care for up to 80 children, but the adult-to-child ratios are relatively high (1:8 at 1.5-3 years and 1:10 at 4-5 years versus 1:18 in 5-year-old kindergarten). In CPE-regulated home childcare, one caregiver may have a maximum 6 children of different ages, including his or her own, with maximum 2 under the age of 18 months. If a second caregiver is in the home, the limit is 9 and 4, respectively. See Friendly (2002) for more details. In addition to relatively low child-adult ratios, the program was educational (play-based learning)—the centres provided resources to home-based providers—and a primary goal was school-readiness. Therefore, there were requirements (and support) for higher teacher credentials and higher wages. The objectives were child-centered, evidence-based, and aimed for high-quality.

There have been very few studies assessing the quality of the CPE-regulated care, but they suggest that CPE-regulated providers achieved moderately better quality, overall and on some specific features that may be expected to affect children's weight gain. Regarding the early years of the CPE program, two studies reported multiple dimensions of quality: *Grandir en qualité* assessed CPE centre- and home-based providers and private day care centres in Spring 2003 including a representative sample of private daycares and CPE-regulated childcare providers that achieved a 88% participation rate (Institut de la statistique du Québec). Japel et al. (2005) also rated quality of Quebec childcare providers, including unregulated home-based providers; however, the participating providers were recruited from consenting families enrolled in ELDEQ, possibly resulting in a selected sample. The ISQ study, using a custom-designed rating system, found that although physical characteristics of the centres were generally good, equipment for outdoor play to foster gross motor and other development was low. Also, on average, educators' participation in children's active play was poor. In contrast, snacks and meals were rated very good and lunch was taken in a relaxed atmosphere. Activities and discipline meant to foster emotional and social development were, on average, fair, often unsatisfactory; although, home-based providers did establish good relationships with children and parents. Scores were similar across geographic regions.

Japel et al. (2005) rated quality using the widely used Early Childhood Environment Rating Scale (Harms et al., 1998) for centres and the Family Day Care Rating Scale (Harms, 1993) for CPE-regulated and unregulated home-based care providers. These instruments have fewer items specific to nutrition and physical play compared to the ISQ rating scales. Japel et al. (2005) found, unlike the ISQ study, the quality of snacks and meals was intermediate in regulated centres and homes, but higher than non-CPE providers. Spaces and furnishings for gross motor play were intermediate but higher than in for-profit centres and unregulated home-based settings. Centre-based CPEs achieved a high rate of very good ratings for features aimed to promote social development. Global quality ratings were equivalent for the centre-based CPEs attended by children from low and high SES families. However, the ratings for the for-profit centres and home-based providers attended by children from lower SES families were lower than those for the children from high SES families, especially among unregulated home-based providers.

The CPE program has not provided enough places for all children whose families wanted to use the service, especially in the early years (Lefebvre et al., 2011), but thousands of children have participated. In 2001, approximately 133,000 0–4 year old childcare spaces were regulated and subsidized by the Quebec government (Friendly et al., 2002), which rose to ~285,000 spaces by 2016 (Friendly et al., 2018), representing space for about 29% and 44% percent of 0–4-year-old Quebec children, respectively. Childcare in Quebec has also been available through unsubsidized day care centres and informal home-based providers that are subject to some regulations such as maximum child-adult ratios and minimal health and safety standards (Friendly et al., 2002).

The CPE centre- and home-based arrangements have been the favoured option for most parents seeking preschool childcare, but whether the program has had a net and equitable benefit for the health of Quebec's children is still debated (Baker et al., 2015; Fortin, 2018; Lefebvre et al., 2011). The CPE program (or policy) has been extensively evaluated in econometric studies, but these studies collapse all childcare arrangements (e.g. parental care, centre- and home-based; regulated and unregulated) into the broad categories of Quebec versus the rest of Canada, pre- and post-CPE reform—in other words, comparing the CPE policy to the *status quo*. They do not estimate effects attributable to specific childcare features.

1.3.2 Historically important childcare studies

Historically, the main empirical basis for the claim that state-sponsored preschool would have long-term benefits has come from studies of targeted intensive interventions in the USA. The increase in "human capital" attributed to the Carolina Abecedarian Project and the HighScope/Perry Preschools Project has been extensively analyzed by James Heckman and colleagues (e.g. Campbell et al., 2014, Conti et al., 2015). The Chicago Longitudinal Study has evaluated the effects of the Child-Parent Centers preschool component (e.g. Reynolds et al. 2011). These studies generally show that intensive preschool interventions can have lasting positive impacts on health and health behaviours, and they suggest high quality preschool can compensate for some social and material deprivation. However, those three seminal experiments enrolled only children from very disadvantaged families; therefore, they offer little direct evidence that such programs would benefit children from more advantaged families, as well. In either case, the transportability of the evidence to widespread public childcare programs is questionable because the interventions included components such as the coordination of social services and pediatric care, and parenting classes that are not provided by most childcare arrangements, even regulated programs (Zoritch et al., 2009). Head Start is a targeted (means-tested) public program that was developed in the in the 1960s in the USA and is now widespread in all states. Head Start enrolment was randomized for the Head Start Impact Study (HSIS) (U.S. Department of Health and Human Services, Administration for Children and Families 2010), greatly strengthening the causal evidence from the ensuing studies. However, the relevance of the evidence from Head Start studies to my research questions is also limited by the targeted enrolment and because the comparator for Head Start has often been a mix of all non-Head Start childcare.

Targeted childcare programs aim to protect children deemed at higher risk for poor development. Factors that flag children as “at-risk” are generally family or maternal characteristics such as low income, teen or single parenthood, parental depression and activity-limiting diseases, and little social support (e.g., due to recent immigration). Targeted (or means-tested) social programs aim to reduce health and social disparities by enrolling high-risk groups while limiting program costs, but many social scientists believe that universal programs are more just and effective (Barnett et al., 2004; McLaren & McIntyre, 2014). This had been a core principle of welfare policy in much of Europe and Quebec circa 1960 to 2000 (Michel & Mahon, 2002).

In many European countries, 3- and 4-year-old preschool is nearly universal. Ironically, the high rate of uptake and long history of European programs limits research opportunities because an appropriate and large enough comparison group is harder to find; therefore, studies of 0- to 3-year-old infants, who vary more in their childcare exposure, predominate the evidence from Europe. Nevertheless, several relevant European studies are discussed later in the chapter.

Many large general surveys of child development have included data collection on childcare use. The US National Institute of Child Health and Human Development Study of Early Child Care and Youth Development (NICHD SECCYD) was a large survey of child development with a focus on childcare; it enrolled a non-representative sample of children born in 1991 (Peth-Pierce, 1998). In Canada, the National Survey of Children and Youth (NLSCY) has been a major source of childcare and child development data; it is a representative complex biennial survey (1994-2008) including multiple birth cohorts. The NLSCY instruments were adapted for the Quebec survey, *l’Étude longitudinale du développement de l’enfant au Québec* (ELDEQ), a representative sample of children born in 1997-98 who became eligible for CPE spaces at 2 years old. ELDEQ is the data source for my thesis (see Chapter 2).

For more on the history of childcare research see Melhuish et al. (2015).

1.3.3 Strengths and weaknesses of childcare research designs

Selection into childcare

Many family and environmental factors affect child development; therefore, the association between childcare and outcomes, independent of other causes, must be isolated to infer causation

(McCartney et al., 2010). But administrative and economic barriers to childcare are ubiquitous. Thus, if the characteristics of the children and their families in the different levels of the childcare exposures being compared do not sufficiently overlap, the independent effect of childcare features cannot be estimated—the exchangeability, positivity, or consistency assumptions of causal inference are violated (Hernán & Robins, 2020; discussed in more detail in section 2.2.2). Random assignment to well-defined types of childcare is the gold standard, but it has rarely been done for realistic public childcare programs. In observational studies, studying a targeted program such as Head Start increases the probability of overlap in the characteristics of study participants in the different childcare levels. However, as discussed above, the evidence is then limited to children from low-income families. Universal subsidized childcare programs offer an opportunity to study children from more and less advantaged families in the same types of childcare. But preschool programs that guarantee space for all children tend to have such high uptake that children who do not attend tend to be quite unique. Universal programs ramping up access, especially with geographically varied roll-out will be more evenly distributed across family types while leaving a large comparison group, temporarily. In addition, when the administrative changes are arguably independent of local changes in child health or parenting trends, the data provide an opportunity to gather evidence of the effects of childcare policies under stronger causal identification conditions than observational studies without this context. Unfortunately for impact evaluation, Quebec's CPE program was not rolled-out and monitored in this way. Hopefully new Canadian programs will be.

Net effects of programs versus childcare features

Experimental studies and econometric studies that exploit administrative differences usually compare a homogeneous program or policy to a mix of diverse childcare arrangements. Therefore, the evidence for the causal effect of a program may be strong but may not provide evidence for more generic features of childcare such as centre versus home care or full-time versus part-time care that may inform new program design. The variety in features of childcare across studies—such as timing (age of attendance), intensity (hours per week of attendance), or type (centre versus home childcare)—has been the biggest barrier to meta-analysis of results and generalized conclusions (Black et al., 2017; Swyden et al., 2017).

1.3.4 Preschool childcare and adiposity

There have been several systematic or scoping reviews of the literature on the association between childcare before the age of 6 years and adiposity indices (e.g. age- and sex-standardized BMI z-scores, obesity status). Recent reviews include:

- Black et al. (2017) and Costa et al. (2017) for observational studies;
- Swyden et al. (2017) for both observational and intervention studies;
- Volger et al. (2018) and Wolfenden et al. (2020) for intervention studies.

But no meta-analysis of results has been reported because of the heterogeneity in exposure and outcome definitions across studies. The scope of Black et al. (2017) corresponded most closely with my first research question; their last search was January 2017. I re-examined the most relevant of the reviewed studies in more depth and new studies published between January 2017 and May 2020 that included childcare exposure from 18 months to the start of kindergarten and adiposity indices in elementary school (i.e. up to grade six or approximately age 12 years) and controlled for pre-childcare confounders.

Non-intervention studies have found center-based childcare to be associated with slightly higher mean BMI z-score or increased risk of overweight and obesity more often than not. In Denmark, children from high SES families who attended 3-6 year-old “kindergarten,” controlling for intensity of attendance, had higher BMI z-score at 7 years; whereas children from low and mid SES families or who had had other types of preschool childcare (including regulated home care) did not differ (Benjamin Neelon et al., 2018). Among the few observational studies with prospective follow-up and a rigorous approach to minimize confounding, were three Canadian studies. Dubois et al. (2006) found Quebec children (ELDEQ participants) had a higher prevalence of overeating at 4.5 years if they had attended centre-based childcare. McLaren et al. (2012) found that Canadian children who had attended centre care at age 2 or 3 years, compared to parental care, had a higher BMI percentile at age 6–7 years; however, height and weight were parent-reported in this study. Geoffroy et al. (2013) found a positive association between centre-based childcare and the prevalence of overweight/obesity in Quebec children (ELDEQ participants) age 5–10 years, and the effects of childcare did not diminish significantly over that time. However, those three studies did not distinguish between regulated and unregulated centre-or home-based childcare providers, nor did they test effect measure modification (EMM) by any indicators of family disadvantage.

In a quasi-experimental study of the Quebec policy, Bruce (2019) found that 12–14 year old Quebec children who had access to the program did not have a higher mean self-reported BMI or risk of obesity than Quebec children who were too old to have had access to the program and children in the rest of Canada. However, Bruce (2019), like other econometric studies of the CPE program, examined childcare as exposure to the policy, not actual use, and did not differentiate CPE-regulated centre- and home-based care.

High-quality childcare is believed to be especially beneficial for children from disadvantaged families (Campbell et al., 2014; Laurin et al., 2016). However, few studies have directly compared the effects of childcare on adiposity by familial socioeconomic circumstances. Bruce (2019) stratified estimation of the effect of access to the Quebec policy by family income and did not find differences by family income; however, income was measured at the time of the outcome, not pre-exposure. No studies from other countries shed light on my secondary research questions.

In summary, Canadian and international studies present a vague and inconsistent picture of the association between childcare and adiposity indices, but there is little evidence that regulated childcare has been beneficial. Poorly defined exposures, cross-sectional designs, little control of confounding and selection bias reduced the number of relevant studies to a handful. Although there have been two strong studies of the Quebec experience that partially answer my research questions—showing that CPE-regulated childcare has not *reduced* childhood adiposity—there is more to be gleaned from the Quebec experience.

1.3.5 Preschool childcare and behavioural self-regulation

School readiness has been a major impetus for childcare studies. Although the focus has been on cognitive skills, school readiness also includes some aspects of self-regulation. Although, very broadly speaking, “self-regulation” could include physiological responses, here I focus on the psychological construct and its manifestation in behaviour. Behavioural self-regulation (SR) is also known by other terms (e.g. effortful control) and is related to other constructs such as executive function (broader) or self-control (narrower) (Rademacher & Koglin, 2018; Smithers et al., 2018).

SR is mainly measured by one of three methods: classic lab-based assessment such as the marshmallow test (Mischel et al., 1989), observation of the child by a trained rater, and parent and teacher ratings on behaviours believed to be caused by the underlying construct, usually hyperactivity, inattention, and impulse control (McClelland et al., 2018; McCoy, 2019). The latter is partly captured in related ratings such as "externalizing behaviours," but externalizing behaviours often also includes disobedience and proactive physical aggression.

Many preschool- or kindergarten-based interventions have had a beneficial influence on SR in young children (Pandey et al., 2018; Sezgin & Demiriz, 2019) or shown long-term reductions in related outcomes such as less contact with the criminal justice system and unemployment (Kautz et al., 2014). However, the results are less consistently beneficial in non-experimental childcare studies. Many studies showed disadvantaged children benefited from regulated infant and preschool childcare on cognitive, language, and social skills. But non-parental care—particularly group care (centre-based or larger peer groups)—has been associated with more externalizing behaviours more often than a reduction in externalizing behaviours or improved SR (Melhuish et al., 2015).

Age of childcare exposure is an important distinguishing feature in these studies because of the rapid change in emotional and psychological development from infancy to the later preschool ages (3–5 years). Effects of childcare on executive function are more directly salient at older ages (e.g. at least 2 years and older). Other developmental constructs, especially attachment, have been outcomes of interest in studies of non-maternal care in infancy. I am only interested in studies that include childcare exposure past the age of 2 years and SR (or related constructs) measured between the ages of 5 and 14 years.

Comparing Quebec to the rest of Canada, pre- and post-CPE reform, several Canadian studies found that Quebec children, post-CPE reform, had slightly more frequent externalizing behaviours (hyperactivity, inattention, and physical aggression), but not consistently for all ages and cohorts (Baker et al., 2019; Haeck et al., 2018; Kottelenberg & Lehrer, 2013). Haeck et al. (2018) found the association was attenuated in exposed cohorts after the policy had been in place for about 10 years. Baker et al. (2019) found the effect was concentrated in children who had a high frequency of behaviour problems at 2 years old. Kottelenberg et al. (2014) found the negative impact was isolated to children who had access to the CPE program at a young age. Haeck et al. (2018) also found that exposed children of less educated mothers had equal or less frequent externalizing behaviours compared to unexposed children; that is, the small detrimental effects were mainly in children of highly educated mothers.

Within Quebec, Laurin et al. (2016) found, compared to low-income children who did not use childcare, children from higher-income families or who attended regulated childcare had lower odds of being vulnerable in two or more of the five developmental domains studied (physical health and well-being, social competence, emotional maturity, cognition and language, and general knowledge); they did not show results for individual domains. The conflicting results may be partly explained by whether or not the outcome included only externalizing problems or was grouped with internalizing problems and social skills. Social skills have been quite consistently improved by regulated centre-based childcare (Melhuish et al., 2015). Yang et al. (unpublished) estimated the effects of individual-level exposure in Quebec children using measures of externalizing behaviours equivalent to Haeck et al. (2018). They found children who mainly used CPE care between 2 and 5 years old had more frequent hyperactivity/impulsivity, as rated by teachers in kindergarten, than children who mainly had unregulated or parental care. However, with mothers' ratings of behaviour, the differences were negligible. Centre- and home-based CPE were not separately evaluated.

Comparing Head Start to other preschool and informal care in boys, Carneiro et al. (2014) exploited eligibility criteria and between-state differences in the availability of Head Start (1985-2000) in a regression discontinuity study, and showed that Head Start led to fewer behaviour problems at 12–13 years old. However, externalizing behaviours were grouped with social and internalizing problems. Surprisingly, after the expansion of services in Head Start to more closely match the comprehensive services included in Perry preschools, Head Start was less beneficial to overall behaviour than the control treatment—a mix of other preschool, informal, and parental care (E. B. Miller et al., 2016).

In contrast, the following studies had more detailed childcare measures. McCartney et al. (2010) estimated the association between time in childcare and externalizing behaviors in the US SECCYD sample using a detailed measure of childcare and causal inference principles (with extensive discussion of the limits of their observational data). Independent of type and quality, hours in non-parental care was detrimental, and the relation did not differ by poverty level. Higher quality and smaller peer groups partially counteracted the effect of more hours in care. Center-based care appeared protective; however, the overlap of center-based care (versus home-based) and larger peer groups was unclear. European studies of centre-based childcare exposure between 0–5 years have shown mixed results. In France, there was no difference in hyperactivity-inattention between centre-based and informal care from 0–3 years old, overall, but when centre care was used for more than a year, it was protective (Gomajee et al., 2018). In studies that adjusted type of care for intensity as mean hours per week in care, centre care was not associated with externalizing behaviours in Denmark (Datta Gupta & Simonsen, 2010) and Norway (Solheim et al., 2013), but had a small negative association in England (Stein et al., 2013) and Australia (Gialamas et al., 2015). In France, Denmark, and Norway—where home-based caregivers are licensed—home-based care was not different from informal or parental care (Gomajee et al., 2018; Solheim et al., 2013), except in boys of less educated mothers in Denmark, who had more behavioural problems with home care than parental (Datta Gupta & Simonsen, 2010).

In summary, regulated childcare rarely had a beneficial impact on SR or externalizing behaviours. The Canadian quasi-experimental econometric studies (Baker et al., 2019; Haeck et al., 2018; Kottelenberg & Lehrer, 2013) provide strong evidence that Quebec's CPE policy had a small negative impact on externalizing behaviours that are related to SR, but the ecological effect estimates and broad childcare categories do not provide estimates of the differences between childcare types independent of intensity. In addition, few studies that modelled childcare features reported results separately for childcare over the age of 1 or 2 years, effect measure modification by family disadvantage, or outcomes measured at kindergarten entry and older ages.

1.4 The association between behavioural self-regulation and adiposity

While conventional wisdom suggests SR (or at least self-control) plays a role in maintaining a healthy weight, the empirical evidence for the causal effect of SR on adiposity is sparse and conflicting, especially in young children. The lack of success in childhood obesity prevention and favourable association between SR and adiposity (albeit, cross-sectional or unadjusted) reported in observational studies (Francis & Susman, 2009; Fuemmeler et al., 2011) inspired interventions improving SR to prevent childhood obesity. Although childcare-based interventions that promoted SR as a lever for obesity-prevention (or as an endpoint) have indeed successfully improved SR, they did not lower obesity rates (Lumeng et al., 2017; Verbeken et al., 2018). However, the Lumeng et al. (2017) intervention successfully improved nutrition and physical activity.

Nevertheless, recent longitudinal studies that controlled for potential confounders did find a longitudinal association between higher adiposity indices and prior low self-control (Datar & Chung, 2018) or self-regulation (Howard & Williams, 2018), or symptoms of ADHD (Bowling et al., 2018). Howard and Williams (2018) found a small increase in the odds of overweight and obesity at 14 years old with poorer SR at 4-6 years in Australian children. Datar and Chung (2018) found the association was more consistent at older ages (BMI z-score and obesity incidence in grade 8 versus grade 5), but Anderson and Whitaker (2018) detected a U-shaped relation between SR at 2 years old and obesity prevalence at 5 years old in girls. In addition, Piche et al. (2012) found that higher impulsivity in kindergarten predicted less of an increase in BMI by grade four, controlling for several potential confounders. Hence, SR may be longitudinally related to adiposity, independent of many common causes, but the association may vary with age and sex.

1.5 Summary of the relevant literature

I found little published evidence for my research questions. Early intensive interventions provided the proof-of-concept that early childhood preschool-based interventions can have short- and long-term benefits for very disadvantaged children, but few universal programs offer the same comprehensive services. The claim that high-quality childcare especially benefits children from low socioeconomic backgrounds does not seem to be supported by empirical studies of children from high- and low-SES families attending comparable, realistic childcare programs. Despite the advantages of experimental studies for causal inference, the transportability of the evidence for the effect of childcare-based interventions on adiposity or SR is questionable because interventions have typically been much more intensive than large-scale, public childcare programs.

Canadian quasi-experimental econometric studies (Haeck et al. 2018; Kottelenberg and Lehrer 2013) provide strong evidence that Quebec's CPE policy, in its early years, had a small negative impact on externalizing behaviours related to SR—and to a more limited extent, that the CPE policy had no net effect on adiposity (Bruce, 2019)—but the ecological effect estimates do not provide estimates of the differences between childcare type (centre- versus family-based versus no childcare use). As will be discussed in section 2.2.2, obtaining causal evidence for the effects of childcare from observational studies is challenging. To be relevant for Canadian policy, childcare exposure contrasts should include quality-controlled center- and home-based programs spanning ages at least 1.5 years of age (given Canada's parental leave program) to the age of full-time kindergarten eligibility, and should not group disparate types of childcare or examine only infant care or 4-year old programs. Compared to experimental programs, Head Start programs are more comparable in scope to CPE-like programs except that they mainly start at 4 years (some at 3 years) and are means-tested. The Quebec CPE program provided a good opportunity to study my first two research questions. In the literature related to my third research question, the causal effect, or even longitudinal association, between SR and adiposity is unclear. A few high-quality observational studies and systematic reviews, while providing only partial answers to my research questions, raise important methodological issues that we addressed in manuscript 3.

2 Overview of the data source and analytic methods

2.1 L'étude longitudinale du développement de l'enfant au Québec

All of the analyses in this thesis were conducted using data from *l'Étude longitudinale du développement de l'enfant au Québec* (ELDEQ; the Quebec Longitudinal Study of Child Development). ELDEQ is a birth cohort of singleton children born between October 1997 and July 1998 at 24 to 42 weeks gestational age (Jetté, 2002). It is administered by *l'Institut de la statistique du Québec* (ISQ). The sampling frame was the Quebec birth registry for that period, excluding children living on First Nations reserves and in the health regions of Nord-du-Québec, Nunavik and Terres-Cries-de-la-Baie-James (Figure 2.1). The ISQ estimated that the sampling strategy represented approximately 94% of Quebec children born at that time. The participation rate was 83% and 2120 children were enrolled.

Annual study visits were conducted from age 5 months to 8.5 years, and then approximately biennially. The children, now young adults of 22 years, are still participating in the study, but I had access to data collected up until age 16 years. See Figure 2.2 for the visit schedule of the original study and definition of baseline, exposure, and outcome visits for my thesis studies. By design, these participants would all become eligible for kindergarten at the same time, September 2003, so it is a school cohort as much as a birth cohort.

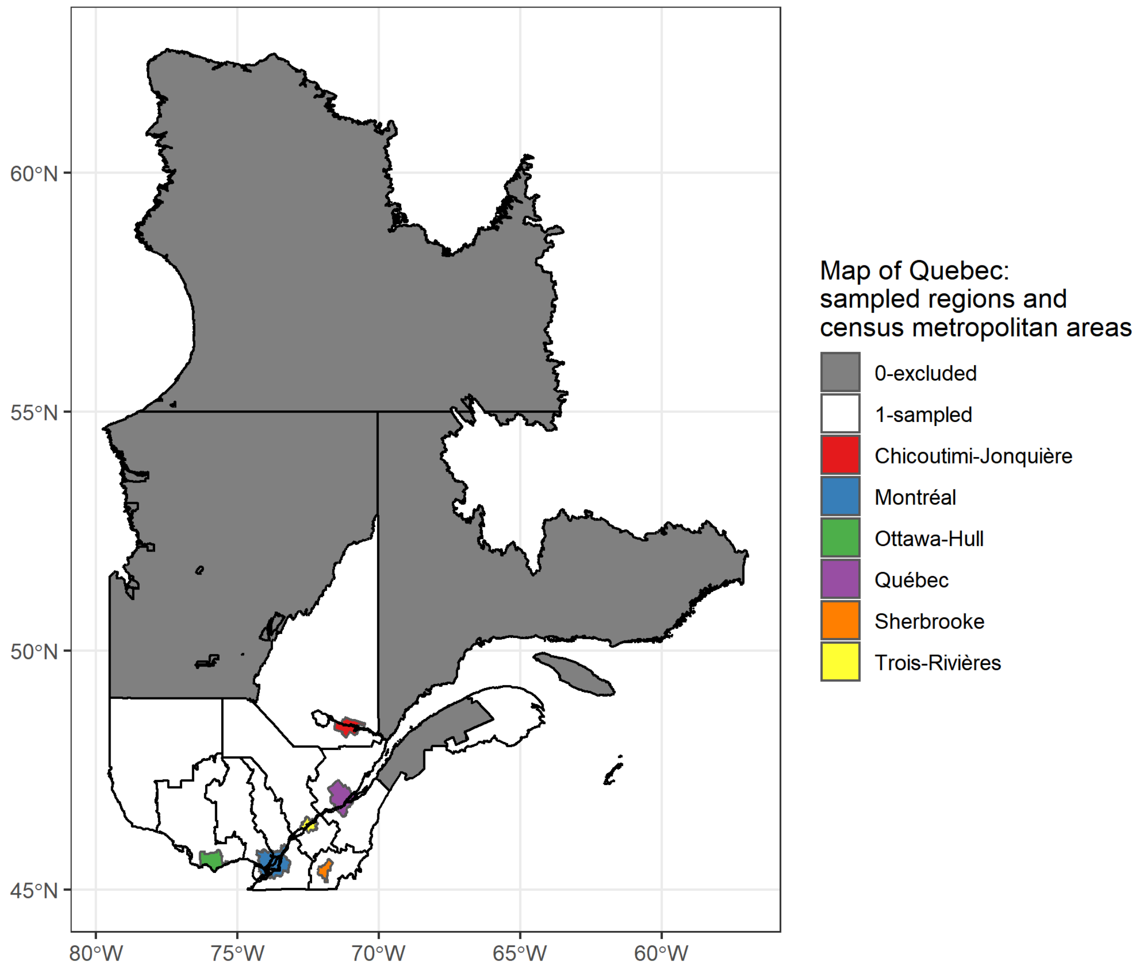


Figure 2.1. Map of the sampled health regions for l'Étude longitudinale du développement de l'enfant au Québec

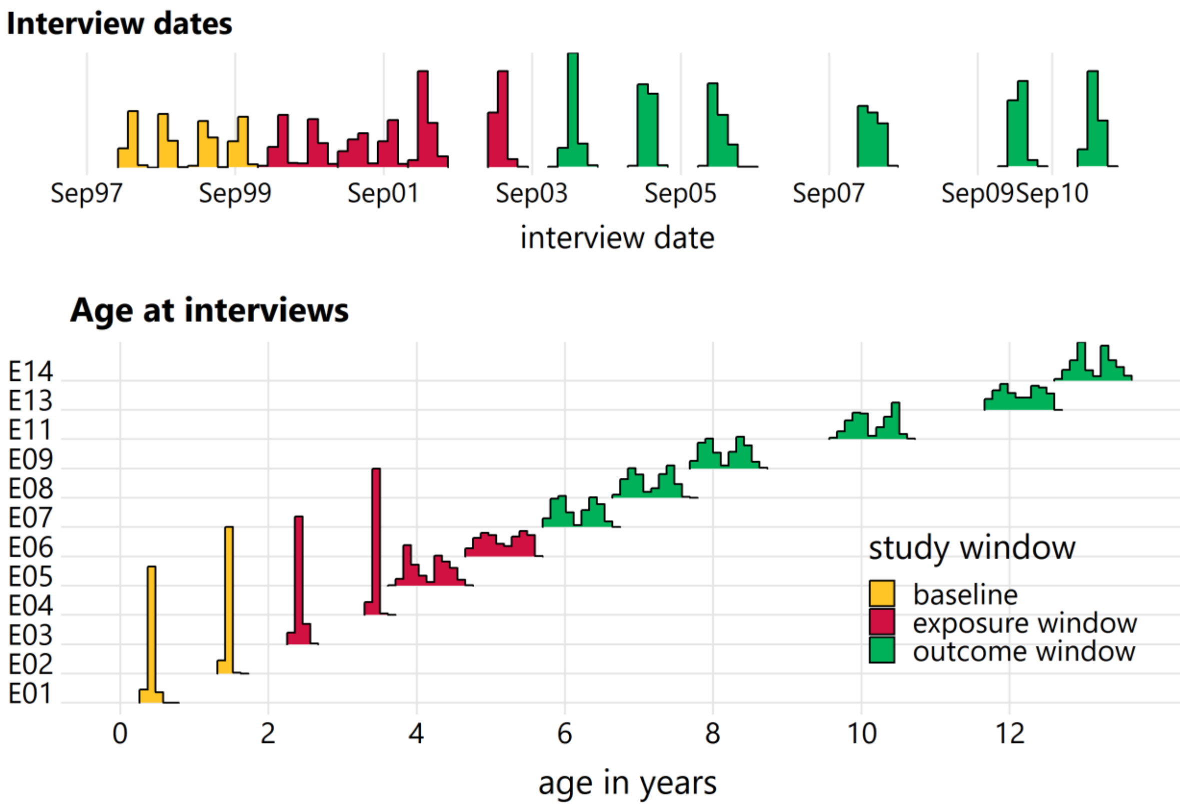


Figure 2.2. Distribution of interview dates and age at interviews

2.1.1 Ethics approval and data use agreement

My thesis studies were granted ethics approval by the institutional review board (IRB) of the Faculty of Medicine (McGill University). A copy of the certificate is included in Appendix A.

Initially, data access was granted by *le Centre d'accès aux données de recherche de l'Institut* (CADRISQ). Later I was granted use outside of CADRISQ through the McGill University co-principal investigator for ELDEQ, Dr. Gilles Paradis. The same data use agreement applies to both modes of access. In brief, to protect participant identity, data have been stored on a password protected hard drive (no web- or cloud-based storage) and will be until the planned manuscripts have been published, at which time I will destroy my copy of the data. In reporting results, cross-tabulated cells and percentiles must have a minimum of five participants or be grouped with adjacent cells. Relatedly, minima and maxima of continuous variables may not be reported, and individual values may not be plotted in figures. I cannot grant access to the data, but in the spirit of open and reproducible science practices, I will share my analysis code with reviewers upon request and post it on a public GitHub repository once manuscripts are accepted for publication. Data access can be requested through the ISQ (https://www.stat.gouv.qc.ca/research/#/demarche/etape-par-etape).

2.1.2 Exclusion criteria for my thesis studies

For substantive and practical reasons, I used the same exclusion criteria for all three studies. The opportunity to use any type of childcare was critical to the study design of objectives 1 and 2, and, later, attending a typical elementary school (public or private) was important for context in all studies. Therefore, children who had severe developmental disabilities (n = 18) or who did not start elementary school by 8 years for other reasons such as being homeschooled (n = 7) were excluded. I also excluded children not living with a mother at visits 1 or 2, age 5 and 17 months, respectively (n = 7). The Person Most Knowledgeable (PMK) about the child was the mother at the first two visits for over 95% of participants and many potential confounders related to the PMK. Single fathers, foster parents, or other guardians serving as PMK were too rare to model as separate strata. The eligible population was 2091 at baseline, but 434 children would be lost to follow-up and will be discussed below. Figure 2.3 shows the participant flow chart.

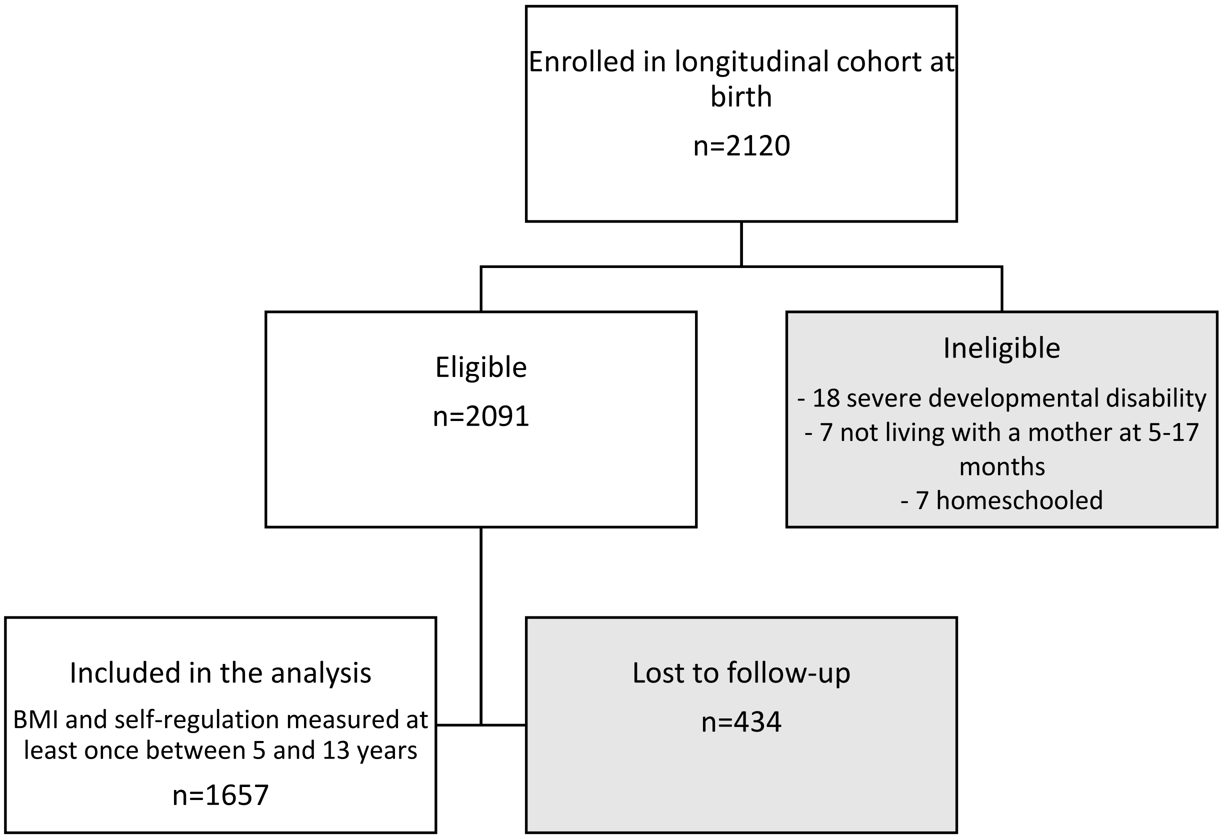


Figure 2.3 Participant flow chart

2.1.3 Measures and variables

Computer-assisted interviewer- and self-administered questionnaires were used to collect data about the child, household, and mother mainly from the PMK. Fathers living with the mother and child (biological or step) and teachers (starting in kindergarten, visit 7) provided additional data about themselves or ratings of the child's behaviour. Perinatal data were collected from obstetrical medical charts when the mother consented (98.2%). Anthropometric measures were directly measured by interviewers starting at age 6 (visit 7) until age 13 (visit 14), when most participants were in kindergarten to secondary 1, respectively. The ISQ provides the data collection instruments, questionnaire modifications between study waves, and other technical documentation on the survey website (www.jesuisjeserai.stat.gouv.qc.ca).

Several of the main study variables—main childcare type, BMI z-score, poor self-regulation, family disadvantage, and time in elementary school—were derived by me (as opposed to variables derived by the ISQ). I describe them briefly in section 2.2.1 in relation to the missing data imputation procedure, which is only described briefly in the manuscripts. More generally, analysis variables are described in the manuscripts, and Table B.1 shows a master list of variables used in my thesis studies. Appendix B.3 shows how latent poor self-regulation scores were derived.

2.2 Methodology

To estimate the answers to my research questions, I analysed the data as cohort studies with repeated measures of the outcome. Specific analytic procedures are described in each manuscript or in appendices, but here I summarize the approach I took to data preparation and modelling.

2.2.1 Missing data: survey weights adjustment and multiple imputation

It was practical to handle losses to follow up and other missing data with a single procedure for the three studies, creating a master set of adjusted survey weights and imputed data sets, since multiple imputation was computationally intensive and the three studies had the same eligibility criteria.

Adjusted survey weights. The ISQ provided the complex survey design and post-stratification weights, and the strata and cluster codes, needed to construct representative estimates. As described above, the survey was designed to be representative of Quebec children (with some caveats). Representative data are an asset in this research for at least three reasons: 1) Having the population distribution of important predictors of the study outcomes allowed me to produce marginal estimates of what the source population would have experienced under different counterfactual scenarios. 2) Intended readers can judge the transportability of this evidence to their target populations. 3) While not necessary for internal validity, the well-defined sampling frame reduced the risk of selection bias. For example, great efforts were made to enroll and retain families from groups who tend to participate less in research such as single mothers, immigrants, and parents with less education—factors correlated with childcare, adiposity, and behaviour.

Nevertheless, there were visit- and item-wise missing data. If I had only used complete cases, the sample size would have been greatly reduced and the estimates would possibly not have been representative. To extract as much information as possible from the data and maintain representativeness, I used a combination of survey weight adjustment and multiple imputation. Computation time for multiple imputation is on the order of hours to days and is related to the amount of missing data. Therefore, participants who were lost to follow-up early, were missing all data on childcare from age 2 to 5 years (visits 3 to 6), or missing either outcome across all outcome visits (visits 7 to 13), were dropped before the imputation procedure because they had little data to contribute to imputation equations (n = 434). In contrast, the completeness of visit 1 and 2 data was excellent, potentially providing good predictors of future nonresponse.

Firstly, I renormalized the provided weights after my exclusions. For the 2091 eligible participants, I estimated each child's response propensity, the probability of having follow-up data, with logistic regression (Chen et al., 2015). The independent variables are flagged in Table B.1c. I then calculated an adjusted survey weight for each participant by multiplying the original survey weight and the inverse of the retention probability. Finally, the adjusted weights were renormalized to have a mean of 1 and a sum of 1657.

Multiple imputation. Missing data for the 1657 retained participants, including outcomes, were multiply imputed. I created 50 sets, as is recommended for Bayesian analyses (X. Zhou & Reiter, 2010), using the 'impute' function in IVEware v.3 (Raghunathan et al., 2016). This imputation method assumes data are missing at random conditional on the variables included in the model (MAR); therefore, it could perpetuate or fail to correct selection or information bias caused by missing information on unobserved characteristics (Raghunathan, 2015).

Missing values on derived variables were imputed or derived after imputation of the root variables depending on a few considerations: The number of variables in the data set increases computation time and can cause the software to "crash", and the IVEware can only model certain types of dependencies between variables (although, it is more sophisticated than some popular packages).

Main childcare type: At visits1 through 6, the PMK was asked whether the child regularly attended any childcare, and, if "yes", the hours per week in each type (day care centre, in a relative’s home, in a non-relative’s home, a relative in the child’s own home, a non-relative in the child’s own home). From visit 3 onwards, center-based and others’-home arrangements were further classified as subsidized (“$5/day”) or not. From these original variables, I derived the annual main type variables as described in manuscript 1. Missing values of the derived variables—rather than the original variables—were imputed because there was no mechanism to ensure consistency across the original variables (e.g. that hours for individual types of childcare would sum to the mother’s report of total weekly hours in care).

Family disadvantage: A measure of family disadvantage, a cumulative risk score, was used as a stratifying variable in manuscripts 1 and 2. It was derived from original study variables for SES, mother’s age, mother’s depression symptoms, family structure and functioning based on the Family Risk Index used in Côté et al. (2008) as described in manuscript 1. Missing values on the original individual variables were imputed.

Adiposity indices: BMI z-scores and BMI categories were derived according to the WHO and IOTF standards, as described in manuscript 1. Missing values for BMI z-scores and height-for-age-and-sex z-scores were imputed. Missing absolute values of height, weight, and BMI were back-calculated after imputation.

Poor self-regulation: I derived the poor self-regulation (PSR) latent variable, used as the outcome in manuscript 2 and the exposure in manuscript 3, using mothers’ and teachers’ ratings of child behaviour. Fourteen Likert-scale items for manifestations of inattention, hyperactivity, impulsivity, reactive aggression and tantrums were completed by mothers at 4, 5, 6 and 8 years (with a few exceptions; see Chapter 4 and Appendix B.3), and by teachers at 6, 7, 8, 10 and 12 years. The frequency of each behaviour was rated as “never”, “sometimes” or “often,” and “don’t know” was set to missing. The Bayesian hierarchical ordinal logistic model (also known as latent regression) that I used to estimate the latent PSR score for each child-visit allows for unbalanced data. Therefore, I used all completed items to estimate the latent PSR score and multiply imputed missing latent PSR scores. Imputing individual items would have greatly increased the number of variables in the imputation procedure (123 versus 9).

Ten plausible values (Gorter et al., 2015) of the latent score were sampled from the posterior predictive distribution for each child-visit (if any mother or teacher ratings were collected for that child-visit), and the ten sets of plausible values were carried into five imputations each.

Behavior ratings from fathers were not used in the latent PSR model because they were entirely missing for children living only with a mother; hence, it was not possible to use the fathers' ratings without scores being confounded by family structure. Fathers' ratings—as sum scores for inattention, hyperactivity-impulsivity, and physical aggression (derived by the ISQ)—were used as auxiliary variables in the imputation procedure, which allows for models restricted to a subgroup.

Finally, in manuscripts 1 and 2, time is measured as time since starting kindergarten. When grade was missing and could not be deduced from adjacent visits, it was imputed; although, most children followed the usual school schedule.

The imputed (and observed) outcomes were used in the modelling steps in each study. Lang and Little (2016, 2018) explain that analyses with multiply imputed dependent variables tend to achieve the same reduction in bias as models with imputed independent variables only, and sometimes increases true information extracted from the data. Precision is not artificially inflated when a large number of imputed data sets are used for estimation. This method also resulted in balanced data sets when converted to long format (i.e. the same number of observations per child), which simplifies longitudinal data modelling when outcomes may not be missing completely at random (Aloisio et al., 2014; Harrell, 2015).

Representative results. For weighted descriptive statistics, I prepared multiply imputed *weighted* data sets using the "uncomplexed" method (Dong et al., 2014; H. Zhou et al., 2016). Before multiple imputation, the original number of participants was expanded to a large number proportional to the survey weights, accounting for survey design, using the IVEware 'bbdesign' function (Raghunathan et al., 2016). The source population was about 70,000 children (Statistics Canada, 2018), but I generated rosters of 10,000 children to reduce computation time and because variance only increased slightly between trials with 10,000 and 70,000. Population estimates were then derived from numerical summaries of a large number of bootstrap samples of the size of the number of unique participants (n = 1657). I generated 2500 bootstrap samples: 50 for each of the 50 imputations.

I could have used the multiply imputed weighted data sets for the regression models, but I instead used the adjusted survey weights as a covariate (T. Zhou et al., 2019) in manuscripts 1 and 2, and as GEE 'prior weights' in manuscript 3, to obtain representative estimates while reducing computation time.

2.2.2 Causal inference from observational data

"…to aspire to know something about counterfactual worlds is…audacious."
(Kaufman, 2019, p. 8)

A potentially useful policy-oriented answer to each of my research questions would be causal effect estimates for each exposure contrast; for example, "Use of full-time centre-based childcare *caused* the mean BMI z-score in the target population to be *x* SD higher/lower than it would have been had those children only been in parental care." Therefore, causal inference was my goal despite the limits of what I might be able to estimate with these observational data. Although more inferentially conservative language such as "association" is still preferred by many experts, I agree with others that description versus causation versus prediction goals should be made explicit because they imply different analytic strategies (Hernán, 2018; Kaufman, 2019). I designed the analysis according to causal inference goals—encouraged by the knowledge that the ELDEQ data contained a rich set of variables collected prospectively—while addressing potential limitations.

The fundamental problem of causal inference is that we cannot observe an individual's response to every level of an exposure because we cannot observe an individual under each level of an exposure simultaneously (Holland, 1986). Instead, we attempt to estimate the expected causal effect in the population by creating or mimicking (statistically) exchangeable groups of individuals who stand in for each other at different levels of an exposure. In order for causal inference to be valid, three key assumptions must be true. The study participants must, on average, be 1) exchangeable and 2) have a non-zero probability of receiving all levels of the exposure under study (the positivity assumption). 3) The exposure is well-defined, which includes the assumption that one participant's exposure does not influence another participant's exposure (the stable-unit-treatment-value assumption, SUTVA) (Hernán & Robins, 2020). "Exchangeable" implies that there is no confounding—that common causes of the exposure and outcome are held constant across levels of the exposure by design (e.g. random exposure assignment) or by conditioning on other observed variables, statistically. Regression adjustment and propensity score-based methods can be effective methods to control for confounding by observed variables (Brookhart et al., 2006; Hernán & Robins, 2020; Shpitser et al., 2012). Propensity score-based methods can additionally be used to test the positivity assumption. SUTVA must be reasoned from contextual knowledge.

There is debate about the importance of representativeness in causal inference research. Estimating causal effects for the target population certainly adds complexity. Westreich et al. (2019) proposed *target validity* as a term invoking it. I do not consider this further, but it motivated my use of survey weights in the statistical models.

The exposures in each study objective were not randomly assigned; therefore, I selected variables and statistical methods that maximized the likelihood that the effect estimates approximate the true causal effects (an untestable goal). To control confounding, I used regression adjustment in manuscripts 1 and 2, and propensity-score based inverse probability of treatment weights (IPTW) in manuscript 3. IPTW reweight the data such that the potential confounders that are included in the propensity score model and the exposure are statistically independent. Also, the mean outcome, conditional on exposure level, in the weighted data equals the standardized marginal mean in the unweighted study population. (Hernán & Robins, 2020, p. 150)

In each manuscript, the exposures change over time. In the first two manuscripts, the exposure was childcare from age 2 to 5 years, but the outcomes, BMI and behavioural self-regulation, were measured after preschool childcare exposure ended. Thus, the exposure was time-invariant with respect to outcome. In manuscript 3, the exposure, weighted cumulative poor self-regulation (cPSR; details are described in manuscript 3), was time-updated with respect to the outcome, BMI z-score or obesity status. Regression adjustment cannot eliminate time-dependent confounding without the risk of introducing collider stratification bias. IPTW avoids this bias if the propensity score is correctly specified.

Because one of the criticisms of past research is the lack of clarity in childcare exposure definitions, I aimed to derive childcare variables that distinguished between type (e.g. centre- versus home-based), timing (age attended), and intensity (hours per week attended). Preliminary data analysis showed that there was enormous variety in childcare use and that the majority of use could not be summarized into a few categories. Cluster analysis suggested the optimal number of clusters is greater than 20—that is, cluster analysis-based categories would not have been a parsimonious substitution for the original variables. Therefore, I used a multivariable childcare definition largely using the original childcare variables. Either way, propensity score-based methods would not have been parsimonious. Also, I do not know of a feasible method for assessing positivity using a large number of propensity scores. Hence, regression adjustment was preferable to a propensity score-based approach to control of confounding.

For manuscript 3, although time-dependent confounding seemed unlikely because prior BMI was unlikely to affect SR, it was not impossible; therefore, IPT weighting seemed prudent. Also, cPSR was represented by a single continuous variable (per child-visit), making it amenable to a single generalized propensity score (Fong et al., 2018). It was possible to check for positivity violations and covariate balance, and the weighted outcome model was relatively parsimonious.

The exposures in manuscripts 1 and 2, childcare, and manuscript 3, self-regulation, could, in theory, be experienced at all levels by all of the included participants; therefore, I believed the positivity assumption would be met in these studies. Likewise, I did not believe SUTVA violations would occur because the study participants were recruited from a large geographical area and were very unlikely to be in contact with one another.

2.2.3 Estimators for repeated measures of the outcome

The estimator of the effect of exposure on outcome that I used in each study is the population-averaged marginal mean or proportion for exposure contrasts of interest, or their mean differences. (Specifically, total effects, the sum of direct and indirect effects, were estimated.) This approach is often called ‘g-computation’ (Snowden et al., 2011) or ‘averaged predictive comparisons' in the Bayesian literature (Gelman & Pardoe, 2007). Marginal estimates, rather than single regression coefficient estimates, provided a more direct answer to my research questions because each exposure was multidimensional or in an interaction with time; hence, the net effect of an exposure was the sum of several regression coefficients.

G-computation works equally well for single-level (e.g. OLS, logistic regression [2]) and multilevel models (e.g. Bayesian hierarchical, frequentist mixed effects, GEE), but obtaining valid standard errors of the marginal estimators varies by method. In manuscripts 1 and 2, likelihood-based models with regression adjustment were used. Population-averaged marginal means were estimated by predicting the outcome with counterfactual values: Time was set to fixed values—the median for each study visit—and exposure variables to counterfactual values, while all other covariates were left at their observed values. Uncertainty intervals (e.g. 95% credible intervals) were generated from Markov chain Monte Carlo (MCMC) samples from the posterior predictive distribution (Gelman & Pardoe, 2007) (or bootstrap samples for frequentist sensitivity analyses). In manuscript 3, the IPTW x survey weights in GEE would generate unconfounded population-standardized coefficient estimates if the models were correctly specified—no adjustment variables would be needed in the model. However, the exposure, cPSR, was modelled as splines and interacted with age and sex terms; therefore, marginal estimates were used in manuscript 3, as well.

A note about Bayesian inference. In manuscripts 1 and 2, and for the latent SR measurement model, I used Bayesian methods. Bayesian methods have the advantage of allowing prior evidence to be incorporated into the analysis in the form of informative priors, and direct probability statements to be made about the model parameters (Greenland, 2006). However, MCMC-based Bayesian modelling strategies also have several practical advantages particularly

[2] Care must be taken when interpreting logistic regression marginal effects: the population-averaged marginal effect is not the same as the marginal effect for an 'average' participant. In linear regression they are the same Muller and MacLehose (2014).

salient for the outcome models in manuscripts 1 and 2, and the latent PSR model. Regularizing prior distributions—less informative than "informative" priors, but not "flat"—can help stabilize a complex multivariable model that contains many correlated variables (Lemoine, 2019). Complex (bespoke) multidimensional distributions such as defined in the latent PSR model are usually resolved, given enough time. Once a model has been estimated (admittedly, often a long process), it is easy and quick but valid to generate the posterior predictive distribution (PPD) of the outcomes for the observed or ad hoc values of the independent variables. From MCMC samples of a PPD, one can calculate many different quantities. PPDs from multiply imputed data sets can be combined as simply as MCMC draws from a single data set; therefore, within- and between-imputation errors are easily estimated.

However, although Bayesian methods for time-dependent confounding have been proposed (e.g. Saarela et al., 2015), they were too difficult for me. Therefore, I did not perform a Bayesian analysis in manuscript 3.

2.3 Summary

The ELDEQ data included variables from many aspects of the participants lives—health, family structure, behaviour, childcare use, parents' education and income, etc.—and many were updated annually. This rich, representative data set provided an opportunity to use a causal inference approach to the study design and statistical analysis. However, missing data, the repeated measures of the outcomes, and multidimensional exposures required several analytic choices that could have been handled differently. In this chapter I described my rational for those analytic choices.

3 Effect of preschool childcare on school-aged children’s adiposity in Quebec, Canada

3.1 Preamble

This manuscript was completed to estimate the difference in adiposity indices (BMI z-score and obesity prevalence) in kindergarten attributable to type of preschool childcare used, particularly the Quebec regulated centre- and home-based type compared to informal types. The objective was also to know whether any differences present in kindergarten persisted through the elementary school years and whether effects differed in children from more and less advantaged families. Childcare advocates were confident high-quality childcare could help reduce the risk of childhood obesity. However, results reported in the literature have been disappointing yet difficult to synthesize because of weak study design or vague exposure definitions. No previous studies had reported the three aspects of our hypothesis, but they are implicit in public childcare policy decisions.

This manuscript will be submitted to the journal *Paediatric and Perinatal Epidemiology*.

3.2 Title page and footnotes

Title: Effect of preschool childcare on school-aged children's adiposity in Quebec, Canada

Running title: Preschool childcare and adiposity in Quebec

Authors

Tanya J. Murphy[1], Jay S. Kaufman[1], Patricia Li[2], Russell Steele[3], Seungmi Yang[1]

Affiliations

[1] Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, QC, Canada

[2] McGill University Health Centre, Pediatrics, and the Research Institute of the McGill University Health Centre, CORE, Montreal, QC, Canada

[3] Department of Mathematics and Statistics, McGill University, Montreal, QC, Canada

Keywords: Adiposity; Bayesian Analysis; Child; Child, Preschool; Child Care; Child Health; Epidemiologic Effect Modifier; Linear Models; Public Policy;

Corresponding author:

Tanya Murphy
1020 Pine Avenue Ouest
Montreal, Quebec, Canada H3A 1A2
tanya.murphy@mail.mcgill.ca

3.3 Abstract

Background: Ideally, childcare providers follow nutrition and physical activity guidelines, but the impact of regulated public childcare on childhood adiposity is unclear.

Objectives: We estimated the effects of universal preschool childcare on adiposity in elementary school in Quebec, Canada, and whether the effects differed in children from more or less advantaged families.

Methods: For 1657 children with annual follow-up (1998–2010) in the Quebec Longitudinal Study of Child Development, BMI z-scores from 6 to 13 years were regressed on the childcare used from 2 to 5 years, adjusted for pre-childcare variables. Average treatment effects were estimated using Bayesian multilevel linear regression and g-computation for four childcare profiles: 1) parental care (P) or full-time (35 hours/week) of care in a 2) centre-based (CB), 3) regulated home-based (RH), or 4) unregulated home-based (UH) arrangement.

Results: Had all participants attended CB, mean BMIz in kindergarten would have been 0.38 (95% credible interval [CrI]: 0.23, 0.52), which was 0.40 SD higher than RH (95% CrI: 0.14, 0.65), 0.20 higher than UH (95% CrI: -0.04, 0.43), and 0.36 higher than P (95% CrI: 0.11, 0.60). By 12 years, mean BMIz had increased for all childcare profiles and differences diminished.

Conclusions: Although CB was associated with an earlier rise in adiposity, it had no large, enduring effect, overall, or for less advantaged children, in particular, compared to informal care.

Keywords: Adiposity; Bayesian Analysis; Child; Child, Preschool; Child Care; Child Health; Epidemiologic Effect Modifier; Linear Models; Public Policy;

3.4 Introduction

The Canadian prevalence of childhood obesity doubled in the last 35 years (Rao et al., 2016; Shields & Tremblay, 2010) and disadvantaged children were at higher risk (Boone-Heinonen et al., 2016; Kakinami et al., 2014). The prevention of obesity in early childhood is a public health priority (Abarca-Gómez et al., 2017; Public Health Agency of Canada, 2011), in part, because early childhood obesity, and adiposity more generally, track over time (Geserick et al., 2018; Simmonds et al., 2016; Singh et al., 2008). Preschool childcare may play a role in weight gain, and experts have speculated that high-quality public childcare could equitably promote healthy body weight (Jones-Taylor, 2015; McLaren & McIntyre, 2014; Story et al., 2006).

Intensive childcare-based interventions designed to improve child development had small effects on BMI (Campbell et al., 2014; Larson et al., 2011; Volger et al., 2018). The widespread Head Start program for low income children in the US has also shown small beneficial effects in subgroups (Carneiro & Ginja, 2014; Lumeng et al., 2017). However, general population studies have found centre-based childcare to be associated with slightly higher BMI or risk of overweight and obesity (Benjamin Neelon et al., 2018; Geoffroy et al., 2013), or not to be associated (Isong et al., 2016). Comparisons across those few longitudinal studies are hampered by the heterogeneity in childcare definitions and study designs (Black et al., 2017; Costa et al., 2017; Swyden et al., 2017).

High-quality childcare is believed to be especially beneficial for children from disadvantaged families (Campbell et al., 2014; Laurin et al., 2016), but few studies have directly compared the effects of childcare on adiposity by familial socioeconomic circumstances. Quebec's universal centre- and home-based childcare program, les Centres de la petite enfance (CPE), provided full-time care at a very low cost ($5 per day until 2004) regardless of parents' income or employment status (Forest et al., 2007) (although, the supply of spaces did not meet the demand (Lefebvre et al., 2011)). The CPE program offers an opportunity to directly estimate whether childcare effects differed in children from more and less advantaged families.

Using a detailed definition of childcare that accounted for type, timing, and intensity of exposure, we estimated how much children's adiposity in kindergarten differed by type of preschool childcare in a birth cohort of Quebec children. We also examined whether those differences persisted through the elementary school years, and whether adiposity-by-childcare patterns differed between children from more and less advantaged families.

3.5 Methods

3.5.1 Cohort selection

We analysed data from children participating in the Quebec Longitudinal Study of Child Development (*l'Étude longitudinal du development de enfant au Québec*; ELDEQ), a representative birth cohort of singleton children born at 24 to 42 weeks of gestation between October 1997 and July 1998 (Jetté, 2002). The sampling frame was the birth registry of the Canadian province of Quebec, excluding children living in some remote regions and Indigenous reserves (Figure 2.1), and those with incomplete birth records. The design represented approximately 94% of target population. The participation rate was 83%, and 2120 children were enrolled in the longitudinal cohort. This study was approved by the McGill University Faculty of Medicine institutional review board. As per the ELDEQ data use agreement, descriptive statistics must not include cross-tabulations or percentiles with fewer than five participants and figures must not plot individual values.

We used data collected at study visits from age 5 months to 13 years. The baseline variables were from the 5- and 17-month visits, the childcare exposure variables were from the 2.5-, 3.5-, 4- and 5-year visits, and BMI outcomes were from visits at 6, 7, 8, 10, 12 and 13 years when most participants were in kindergarten, grade 1, 2, 4, 6, and secondary 1, respectively. See Figure 2.2 for the visit schedule. We excluded children who were not living with a mother at the 5- or 17-month visits (n < 10); did not regularly attend elementary school (e.g., home-schooled n < 10, or had a severe developmental disability n=18); or had no BMI measures from kindergarten to grade 6 (n=434). The final analytic sample included 1657 participants; see Figure 2.3.

The Person Most Knowledgeable (PMK) about the child (the biological mother for >99%), responded to annual interviewer- and self-administered questionnaires. Separate consent was obtained to retrieve perinatal medical records for the birth of the target child (98.2%). Questionnaires, modifications between study waves, and other technical documentation can be found on the survey website (Institut de la statistique du Québec, 2016).

3.5.2 Exposures

At each preschool visit (2.5, 3.5, 4, and 5 years), the PMK was asked whether the child regularly attended any childcare, and, if "yes", the hours per week. Centre- and home-based arrangements were further classified as regulated under the subsidized ("$5/day") provincial program or not. For a home-based arrangement, the subsidy was synonymous with being CPE-regulated. Most subsidized centre-based arrangements were CPE-regulated (Lefebvre et al., 2011), but the precise proportion is not known in this sample. Because of this lack of specificity and because only a small proportion of centre-based arrangements were unsubsidized (~10%), we grouped all centre-based arrangements together.

For each preschool year, we summarized childcare as a quasi-continuous variable for hours per week in childcare and a nominal variable for the main childcare type. We defined the main type as the arrangement in which the child spent 70% or more of the total hours, which resulted in a nominal variable with seven categories: 1) none/parental (P), 2) centre-based (CB), 3) regulated home-based (RH), 4) unregulated home-based (UH). When no single arrangement accounted for 70% or more of the total hours per week, the main type was labelled as a mix that included 5) centre-based (mCB), 6) regulated home-based but no centre-based (mRH), or 7) only unregulated home-based (mUH). In the last preschool year, prekindergarten was a possible eighth category. To control for childcare stability, the total number of different arrangements used across the four preschool years, including changes within main type, was categorized as 0–2, 3–4, 5–6, or more than 6 arrangements. In summary, the childcare exposure was modelled using nine variables: four categorical annual 'main type', four quasi-continuous annual 'hrs/week', and one categorical 'total arrangements.' For descriptive summaries and a sensitivity analysis, we derived a single categorical variable summarizing the type of childcare used most over the preschool years (P, CB, RH, or UH) and a single 'mean hrs/wk', similar to Geoffroy et al. (2013). The participants were eligible for the CPE program at 2 years old; therefore, childcare at 5 and 17 months were used as baseline adjustment variables.

Other available pre-childcare variables included: perinatal conditions; infant health, temperament, and behaviour; parents' self-reported height and weight (converted to BMI), health and smoking, employment, demographics, and parenting style; municipality type, social support, and housing conditions (see Table B.1 for details).

Family disadvantage: Baseline covariates were used to classify a child as more or less advantaged using a Family Risk Index (Côté et al., 2008). It is a sum of the points assigned to the following categorized variables:

- Two biological parents (0) vs. single-parent or step/blended family (1);
- Mother's age at child's birth (<25 years = 1; 25+ years = 0);

- SES index (z-score categorized by quartiles where lowest = 3 and highest = 0)—an index derived from family income, parents' education and occupational prestige (Willms & Shields, 1996);
- Family functioning scale (categorized by quartiles where lowest = 0 and highest = 3)—sum score of twelve mother-reported 4-point Likert items (Appendix B.4);
- Mother's depression risk score (categorized by quartiles where lowest = 0 and highest = 3)—10-item version of the Center for Epidemiological Study Depression scale (Carpenter et al., 1998).

The mean of the two baseline visit scores was used for stratification (as an interaction with main childcare type variables), but the original variables were also included as covariates.

3.5.3 Outcome

Adiposity was represented by age- and sex-specific BMI z-score (BMIz). Height and weight were the mean of three measures measured by trained interviewers at 6, 8, 10, 12, and 13 years. At 7 years, interviewers collected a single measure of height and weight. BMI was converted to z-scores according to the World Health Organization's 2007 standard curves (WHO, 2019). Obesity was defined as a z-score of >2 SD. Figure B.1 shows the relation between, height, weight, and WHO cut-offs.

3.5.4 Statistical analysis

Missing data

Prior to estimation, survey weights were adjusted for losses to follow-up (Chen et al., 2015) and item-wise missing variables were multiply imputed 50 times using sequential regression and IVEware software (Raghunathan et al., 2016) (see details in Appendix C). The proportion of missing data on each variable is shown with the descriptive statistics of the study population (Table 3.1). Weighted percentages (or weighted mean) for the target population were estimated accounting for losses to follow-up and the complex survey design (Dong et al., 2014; H. Zhou et al., 2016).

Regression analysis

Step 1: We estimated the relation between preschool childcare and BMIz from kindergarten (median age = 6.1 years) to secondary 1 (median age = 13.1 years) using multilevel linear regression to account for the correlation in repeated measures of BMIz (level 1) within children (level 2). Although the exposure, preschool childcare, occurred over years (age 2–5 years), it was complete by the time children started kindergarten; therefore, the exposure was time-invariant with respect to the outcome measures. Time in elementary school, measured in years, was the only time varying covariate. Childcare 'main type' dummies were interacted with time in elementary school to allow the time slope to differ by 'main type'. The models included child-specific random intercepts and were adjusted for pre-exposure variables (flagged in Table B.1c). We also included attrition-adjusted survey weights as cubic basis splines (Zheng & Little, 2003).

We estimated Bayesian models with regularizing priors (Lemoine, 2019), four chains, and 3000 iterations in the R (R Foundation for Statistical Computing, Vienna, Austria) package 'brms' (Bürkner, 2017, 2018), version 2.10, an interface to Stan (Stan Development Team, 2018). Prior distributions: Student t(df = 5, mean = 0, SD = 1) for the adjustment variable coefficients; Student t(df = 3, mean = 0, SD = 2) for the annual childcare type and their interaction term coefficients (i.e., less informative); Normal(mean = 0, SD = 2) for the population-level intercept; Cauchy(1, 2) for the standard deviation of the random intercepts; Cauchy(0, 2) for the residual errors (however, the software restricts standard deviation to positive values).

Step 2: To estimate the net effects of childcare type, we used the method known as averaged predictive comparisons (Gelman & Pardoe, 2006) or g-computation (Keil et al., 2018; Snowden et al., 2011) to obtain population-averaged marginal means for idealized counterfactual childcare profiles. From the models estimated in Step 1, we generated 100 predicted BMIz for each child at 0.5, 1.5, 2.5, 4.5, and 6.5 years in elementary school for each imputed set (100 x 50 = 5000 samples) setting childcare variables to counterfactual values, while leaving other covariates at their observed values. We then calculated the differences between BMIz predicted under different childcare profiles. We worked with the full predictive distributions (fixed-effects, random effects, and individual residuals) to also estimate BMIz percentiles corresponding to undernourished (< -2 SD), normal (-2 to 1 SD), overnourished (>1 to 2 SD), and obese (> 2 SD) (WHO, 2019).

The counterfactual childcare profiles were, in each of the four preschool years: CB as main type, 35 hrs/wk; RH as main type, 35 hrs/wk; UH as main type, 35 hrs/wk; P as main type, 5 hrs/wk (because hrs/wk had been truncated at <10 hrs in some years). The population-level estimates were calculated as the grand mean of each set of 5000 means or differences for each counterfactual childcare profile and elementary school time point. We used the 2.5th and 97.5th percentiles of the posterior predicted samples as the measure of uncertainty due to estimation and multiple imputation (i.e. 95% credible intervals [CrI]) (X. Zhou & Reiter, 2010). For effect measure modification estimates of the childcare effects by family disadvantage, we calculated mean subgroup differences.

In addition to our main analysis, we performed two *post hoc* comparisons: the above childcare differences by sex, and CB started between 2.5 and 3.5 years (with parental or RH at 2.5 years) versus CB started between 2-2.5 years because Head Start and many European evaluated programs enrolling children after 3 years old.

Sensitivity analyses

To examine the sensitivity of our results to Bayesian modelling assumptions, notably regularizing priors, we repeated the analyses with frequentist methods using the R package 'lme4' v.1.1-21 (Bates et al., 2015), and obtained samples of individual-level predictions (fixed- and random-effects and individual residuals) and nonparametric bootstrapped 95% CIs.

We conducted five sensitivity analyses with frequentist models, which had shorter computation times than Bayesian models. We restricted 1) the study population to children living in urban areas because childcare centres affiliated with higher-education institutions and community services (e.g. YWCA) were early adopters of the CPE program, and 2) follow-up to grade 4 to avoid possible modelling complications associated with puberty. We explored alternative childcare definitions as 3) separating centre-based main type into subsidized and unsubsidized (an imperfect measure of CPE regulation); 4) the single summary variable of childcare over 2–5 years (adjusted for mean total hours). 5) We used general estimating equations (GEE) with a probit link to estimate risk of obesity.

3.6 Results

Our study included 1657 children of the original eligible cohort (1657 / 2091 = 79.2%) and 7859 BMIz observations. Table 3.1 shows characteristics of the participants and their parents. Figure 3.1 shows the flow of children between the childcare types over the two baseline and four preschool visits; the proportions and 95% confidence intervals, and other summaries of the childcare variables used in the regression models, are listed in Table D.1. About half of the children had been in some full-time childcare by the 17-month visit (48.8%), but only 10% had used any centre-based arrangements. Children from less advantaged families were less likely to have been in childcare before 2 years of age, but use was similar from age 2 to 5 years (Table D.2).

On average, observed BMIz increased from 0.19 (SD=1.15) in kindergarten to 0.58 (SD=1.19) in grade 6 (Figure 3.2). Figure D.1 and Table D.3 show the prevalence of obesity at each study visit. Modelled BMIz for the counterfactual childcare exposures are shown in Figure 3.3. CB for 35 hours per week from 2–5 years predicted a mean BMIz of 0.40 (95% CrI: 0.22, 0.59) in kindergarten, which was 0.40 SD higher (95% CrI: 0.14, 0.65) than RH, 0.20 SD higher (95% CrI: -0.04, 0.43) than UH, and 0.36 SD higher (95% CrI: 0.11, 0.60) than P. Adjustment for baseline covariates did not substantially affect predictions for CB or RH but lowered the mean BMIz for UH and P (compare Figure 3.3 panels A and B). The differences between the childcare types diminished over time (Table D.4). In contrast to regulated centre-based care, regulated home-based care appeared protective against high mean BMIz.

Table 3.1. Baseline characteristics of the participants of the Quebec Longitudinal Study of Child Development. Children eligible for this study (n = 1657) and inference to the source population (N ~ 70,000), Quebec singleton children born in 1997-98 excluding children from Indigenous reserves and the remote North. Data presented here were collected at visits 1 (5 months) and 2* (17 months).

		Sample		Population	
		frequency (n)	% missing	frequency (%)	95% CI
Children		*1657*			
Girl		852	0	49.3	46.7, 52.3
Preterm birth (<37 weeks)		80	0	6.6	4.2, 8.4
Sibling rank	1st	732	0	45.6	43.0, 48.3
	2nd	666		37.4	33.9, 40.3
	3rd +	259		17.1	15.4, 19.1
Person of colour		98	<1	8.9	7.5, 10.4
Breastfed*	no	482	<1	30.1	28.6, 32.1
	less than 6 months	674		40.1	37.9, 42.3
	6-12 months	397		22.3	20.9, 24.3
	more than 12 months	100		7.2	6.1, 8.2
General health less than very good*		171	<1	10.9	9.6, 13.2
Household					
Parents employed*	two-parents, both employed	1089	<1	63.1	61.3, 65.7
	two-parents, one employed	381		23.8	21.5, 25.6
	two-parents, neither employed	35		3.4	2.7, 4.1
	single mother, employed	70		3.9	2.9, 4.7
	single mother, unemployed	74		6.0	4.7, 7.1
Insufficient income* [1]		309	1.6	23.2	21.7, 25.1
Residence: Municipality type*	metropolitan area	1074	1.2	66.8	63.9, 68.7
	pop. 10,000+	195		11.6	9.8, 13.4
	rural	368		21.7	19.6, 24.0
Language most spoken at home: French		1442	0	80.6	77.8, 83.0
Crowded housing (>2 PPB)*		51	<1	6.7	5.7, 8.1
Smokers in the home*	none	1084	<1	63.5	61.2, 65.8
	one adult	351		23.8	22.1, 25.8
	both parents	211		12.6	10.9, 13.8
Mothers					
Mean age (SD)		29.6 (5.2)	0	29.3	29.1, 29.6
Education	no high school diploma	281	0	18.0	16.2, 19.6
	high school diploma	422		26.1	24.3, 27.9
	college certificate [2]	487		29.5	27.6, 31.0

		Sample		Population	
		frequency (n)	% missing	frequency (%)	95% CI
	university degree	467		26.5	24.7, 27.8
Immigration status	Canadian-born	1515	0	85.6	83.8, 87.3
	10+ years in Canada	59		5.3	4.1, 6.3
	5-9 years in Canada	42		4.0	3.0, 5.4
	<5 years in Canada	41		5.5	3.3, 7.0
Fathers		*1569*			
Type	biological	1283	<1	81.0	79.4, 82.8
	stepfather	162		10.3	8.7, 11.8
	changes between visits 1 and 2	116		8.7	7.2, 10.0
Mean age (SD)		32.2 (5.5)	1.9	32.5	32.2, 32.9
Education	no high school diploma	303	<1	20.2	17.9, 22.5
	high school diploma	397		28.1	23.9, 30.5
	college certificate [2]	440		27.3	25.2, 29.9
	university degree	387		24.7	23.2, 26.8
Immigration status	Canadian-born	1394	<1	84.7	82.8, 86.2
	10+ years in Canada	89		8.5	7.1, 10.3
	5-9 years in Canada	30		3.6	2.3, 4.5
	<5 years in Canada	33		3.2	2.5, 3.9

Abbreviations: CI, confidence interval; PPB, persons per bedroom. Notes: [1] household income is below the low-income cut-off (adjusted for geography and household size; calculated by l'Institut de la statistique du Québec); [2] Quebec high school ends at secondary 5 (grade 11) and the provincial college system (CEGEP) typically includes a 2-year pre-university or 3-year vocational program.

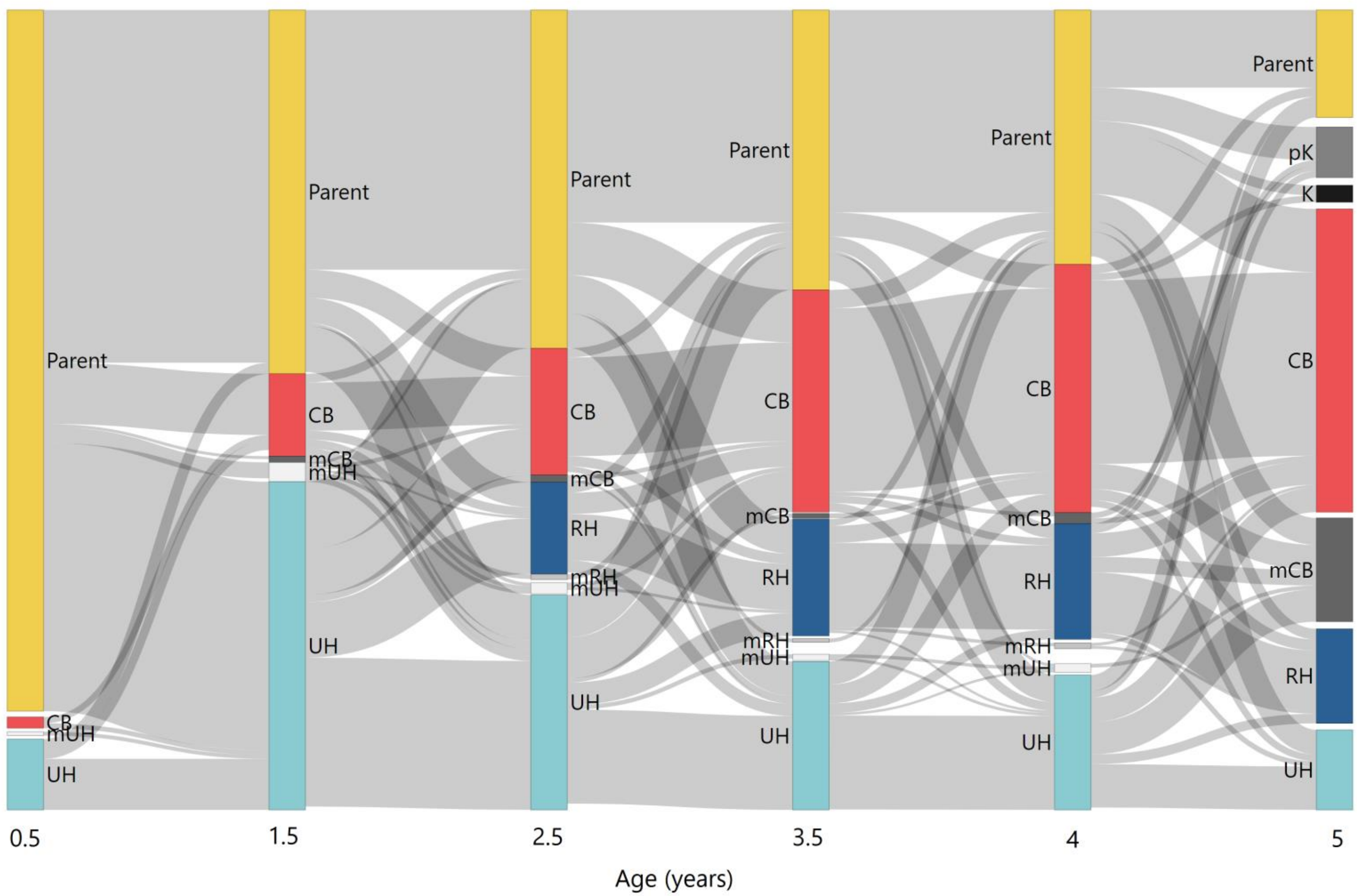


Figure 3.1. Sankey plot of childcare use in participants of the Quebec Longitudinal Study of Child Development (1997-2003; n = 1657). The proportions were derived from study sample of n = 1,657, averaged over imputations and weighted to the target population. Abbreviations: Main childcare types where a single type accounted for >70% of total hours: CB = center-based; RH = regulated home-based; UH = unregulated home-based; pK = public, part-time pre-kindergarten; K = kindergarten (exceptionally, some children started before the age of 5 yrs). When multiple arrangements were used in approximately equal proportions: mCB = a mix that included CB(s); mRH = included RH(s) (but no CB); mUH = mix of UHs (no CB or RH).

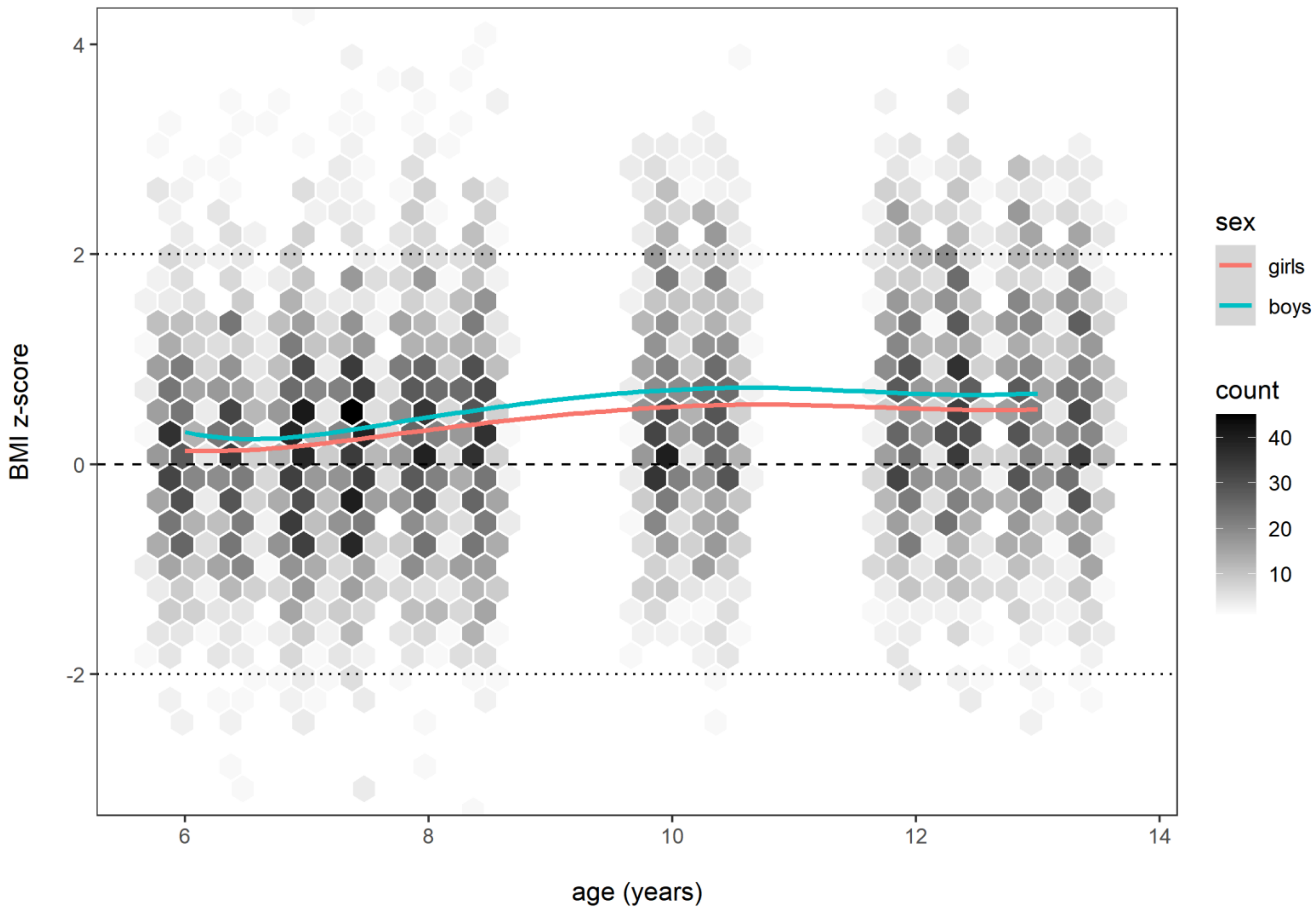


Figure 3.2. BMI z-score by age and sex in participants of the Quebec Longitudinal Study of Child Development (2003-2010; n = 1657). WHO 2007 standard. Cells show observed values, lines show means from survey-weighted data with missing values multiply imputed (m=50). Number of observed values by sex and visit: 6-yr, 542 boys and 606 girls; 7-yr, 687 boys and 769 girls; 8-yr, 673 boys and 743 girls; 10-yr, 617 boys and 687 girls; 12-yr, 635 boys and 691 girls; 13-yr, 558 boys and 651 girls.

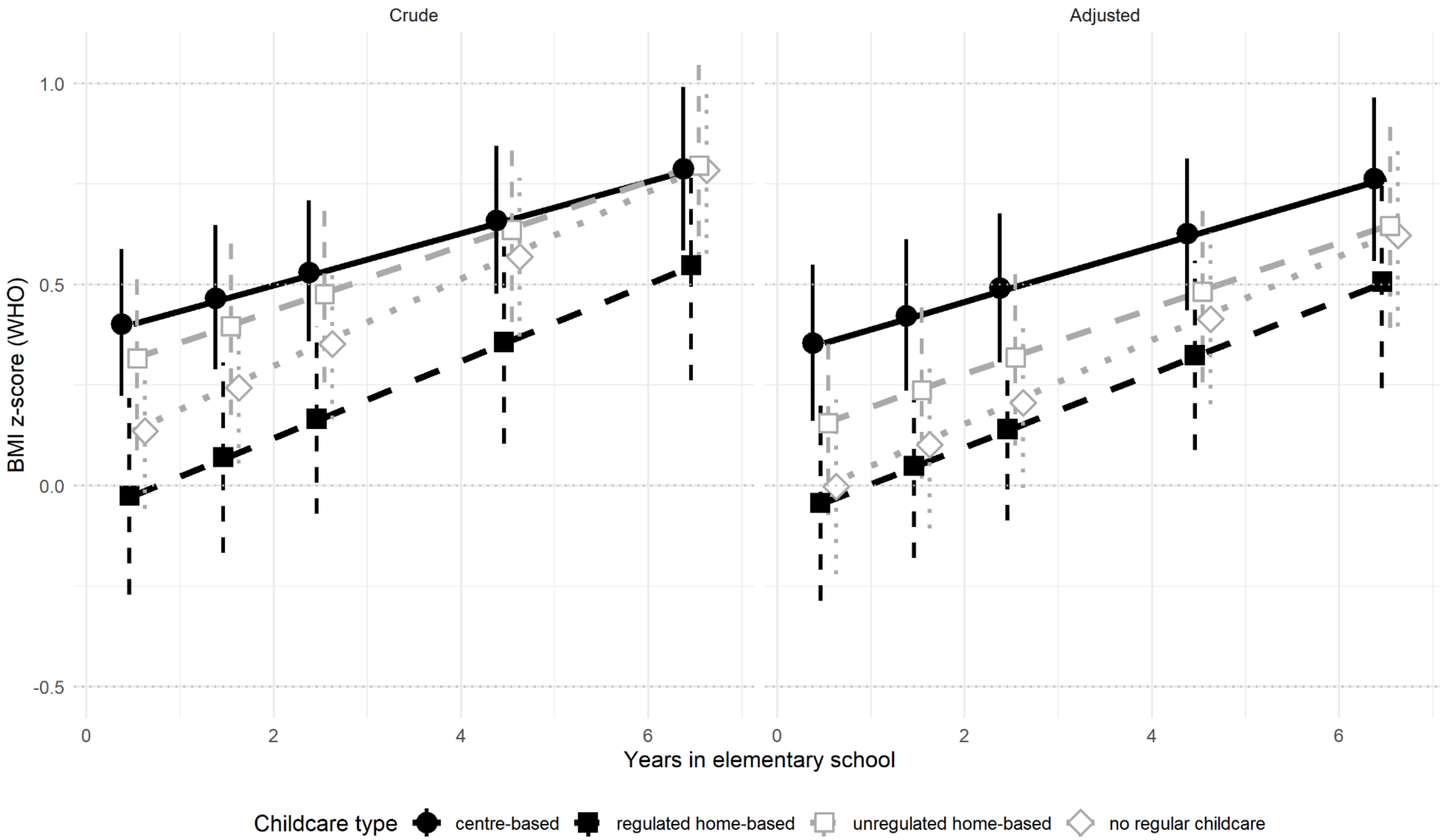


Figure 3.3. Population-averaged marginal BMI z-score by counterfactual childcare use: crude and adjusted. Estimated from participants of the Quebec Longitudinal Study of Child Development (2003-2010; n = 1657). Intensity of use for non-parental childcare was set to 35 hours per week. Mean and 95% credible interval derived from the posterior predictive distribution of Bayesian linear multilevel models. Adjusted model adjustment variables included measures of socioeconomic position; perinatal conditions; child's behaviour; childcare; parents' BMI, education, employment and general health before childcare use at 2 years, when then participants became eligible for Quebec's *Centres de la petite enfance* program (see Appendix Table B1c for the complete list).

The mean BMIz in kindergarten under the CB counterfactual did not differ by family disadvantage, but the relative differences between childcare types did (Figure 3.4). The main difference was the effect of RH. For more advantaged children, mean BMIz in kindergarten would have been 0.53 SD lower (95% CrI: 0.21, 0.84) had they used RH instead of CB, but the difference diminished over time (Table D.4). The opposite trend was predicted for less advantaged children, who would have had a mean BMIz 0.41 SD lower (95% CrI: -0.05, 0.88) in grade 6 had they used RH instead of CB. However, the differences varied by sex (see also Figure D.2) and the double-interaction estimates had wide 95% credible intervals (~1 SD).

Sensitivity analyses. The results from the frequentist linear mixed model for the main hypotheses were nearly identical to the Bayesian model results (Table D.4). When the study population was restricted to urban children, the differences in kindergarten between CB and P or RH were slightly larger. The childcare summary variable produced attenuated results. The effect of CB was smaller when children started CB between 2.5 and 3.5 years instead of between 1.5 and 2.5 years, regardless of whether they spent their 3rd year in RH or P (Figure D.3).

The predicted probabilities of obesity by observed and counterfactual childcare profiles are shown in Table D.5. The risk of obesity would have been higher had all children used CB compared to RH, but CB compared to UH and parental care varied over time. However, estimated differences varied by model type (multilevel linear regression versus GEE-probit; see Table D.6).

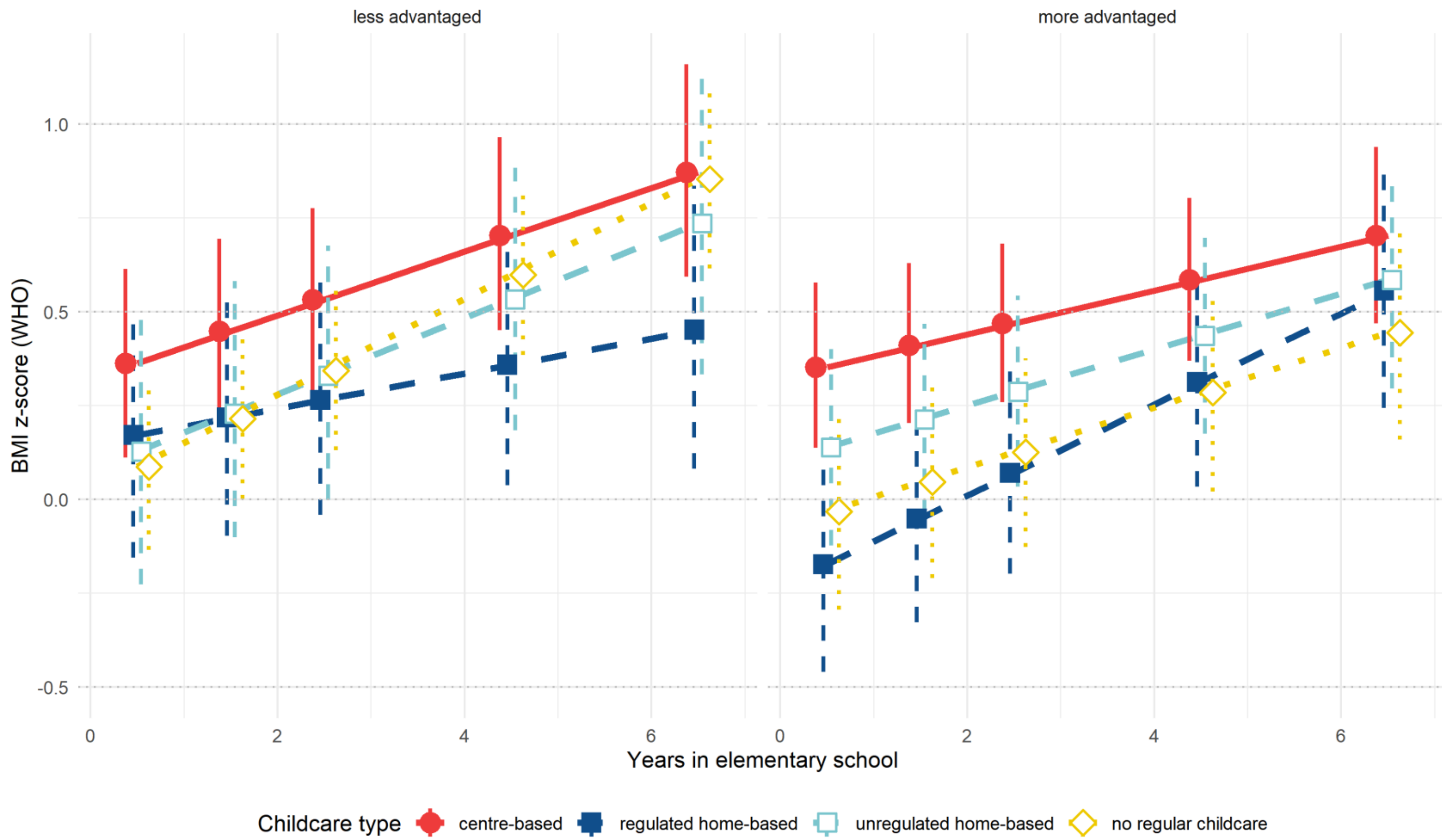


Figure 3.4. Mean BMI z-score by counterfactual childcare use, by family disadvantage. Estimated from participants of the Quebec Longitudinal Study of Child Development (2003-2010; n = 1657). Intensity of use for non-parental childcare was set to 35 hours per week. Family disadvantage measure based on the Côte et al. (2008) Family Risk Index—which includes family structure and functioning, mother's age, SES (parents' relative income, education, and occupational prestige), mother's depression symptoms. The total Family Risk Index scores for the two baseline visits ranged from 0 to 22 points; "More advantaged" ≤ 10 (median), "Less advantaged" > 10. Mean and 95% credible interval derived from the posterior predictive distribution of the Bayesian linear multilevel model. The model was adjusted for measures of socioeconomic position; perinatal conditions; child's behaviour; childcare; parents' BMI, education, employment and general health before childcare use at 2 years, when then participants became eligible for Quebec's *Centres de la petite enfance* program (see Table B.1c for complete list).

3.7 Comment

We studied the relation between preschool childcare and adiposity in elementary school in a representative birth cohort of Quebec children who became eligible for Quebec's universal CPE program at 2 years old. Using multilevel linear models and g-computation, we estimated that attending full-time centre-based care from age 2 to 5 years was associated with higher adiposity in elementary school than full-time regulated home-based care or parental care. The differences diminished over the elementary school years, overall. Our results suggest that the overall effects mask differences by family disadvantage and sex, and that a larger study sample will be needed to precisely estimate the heterogeneity. We believe our results approximate the causal effects of the main types of childcare used in Quebec in the early years of the CPE program; however, there are limits to this causal interpretation for reasons we discuss later.

Geoffroy et al. (2013) previously showed in this study population that children had a higher odds of overweight/obesity in the elementary school years, on average, when they had been in centre-based childcare from age 1 to 4 years, compared to parental care. We also found that continuous BMI z-scores were higher when children had attended centre-based care from age 2 to 5 years, using a more granular measure of childcare that isolated the effect of the CPE programs from other types of preschool childcare. However, we found that the relation varied with time, sex, and family disadvantage; higher BMI in kindergarten following centre-based childcare was mainly seen in more advantaged children, for whom childcare type mattered little by grade 6. Less advantaged children showed little difference in BMI in kindergarten by childcare type, but they may have had less of an increase in BMI by grade 6 if they had used regulated home-based childcare. We cannot explain the latter result and, to our knowledge, effects of regulated home-based care versus centre-based or informal care have not been evaluated in other studies.

Some of our results differed from those of previous studies. Carneiro and Ginja (2014) found that boys who attended the Head Start program at 3–4 years old had lower rates of obesity at 12–13 years old. We did not find that centre-based care was protective against higher BMI (or obesity) by grade 6 for any subgroups. However, boys were less susceptible to the centre-based effect than girls, and the effect was smaller when children started centre-based care around three years old instead of two years old. We did not find a positive association between adiposity and informal home-based care, compared to parental care, unlike some studies (McLaren et al., 2012; Pearce et al., 2010).

Sample size was a limitation in this study. Our estimates for the modifying effects of family disadvantage and sex were imprecise relative to meaningful population-level differences in BMI z-score (~0.25 SD or more). A smaller set of adjustment variables and data reduction techniques for the childcare exposure did not have a large impact on the precision of the predicted BMI outcomes. Reporting marginal estimates of effect, not coefficients, allowed us to model a large set of correlated childcare features and potential confounders without greatly inflating the variance of the marginal effect estimates; therefore, we believe a larger sample size for each subgroup of interest, and not necessarily a more parsimonious set of independent variables, will be required in future research. Despite the lack of precision for subgroups, our results have good generalizability because the participants were representative of the general population and Quebec's regulated childcare program is a feasible model for other jurisdictions.

Other possible limitations regard the causal interpretation of our results. Statistical adjustment can result in as-if-random assignment if all confounders are correctly modelled (Rubin, 2007; Shpitser et al., 2012). However, our results may include residual confounding by incompletely captured latent lifestyle preferences or skills that are common causes of choice of childcare (or ability to secure choice) and adiposity. Our adjustment for parents' education, income, social support, and municipality type probably controlled for many latent causes. We did not have data on parents' dietary and physical activity habits, but we do not think they were strong direct determinants of childcare choice. Adjustment for smoking in the home, which was a strong independent predictor of child's BMIz, probably served as a good proxy for health behaviours.

Our multidimensional definition of childcare exposure was not amenable to a propensity score-based approach to assess the degree of conditional covariate overlap (positivity). We relied on theoretical support for positivity: All participants in our study were eligible for all childcare types and the Quebec CPE program greatly increased the equitable distribution of regulated care (Lefebvre et al., 2011). A benefit of the separate variables for the timing (age of attendance), intensity (hrs/wk), and type of childcare was that multiple features of childcare could be independently modelled.

3.8 Conclusions

In conclusion, we found that children who attended centre-based childcare from age 2 to 5 years, in the early years of Quebec's universal CPE childcare program (2000–2003), had higher adiposity in kindergarten than children who had used CPE-regulated home-based or parental care. Although, centre-based childcare may have caused adiposity to rise at a younger age, it did not have large, enduring effects on adiposity, overall, or in less advantaged children, in particular, compared to informal care.

3.9 Manuscript references

Abarca-Gómez, L., Abdeen, Z. A., Hamid, Z. A., Abu-Rmeileh, N. M., Acosta-Cazares, B., Acuin, C., Adams, R. J., Aekplakorn, W., Afsana, K., Aguilar-Salinas, C. A., Agyemang, C., Ahmadvand, A., Ahrens, W., Ajlouni, K., Akhtaeva, N., Al-Hazzaa, H. M., Al-Othman, A. R., Al-Raddadi, R., Al Buhairan, F., . . . Ezzati, M. (2017). Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128.9 million children, adolescents, and adults. *The Lancet*, *390*(10113), 2627–2642. https://doi.org/10.1016/S0140-6736(17)32129-3

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Benjamin Neelon, S. E., Schmidt Morgen, C., Kamper-Jørgensen, M., Oken, E., Gillman, M. W., Gallis, J. A., & Sørensen, T. I. A. (2018). Childcare before age 6 and body mass index at age 7 years in a cohort of Danish children. *Pediatric Obesity*, *13*(5), 307–311. https://doi.org/10.1111/ijpo.12206

Black, L., Matvienko-Sikar, K., & Kearney, P. M. (2017). The association between childcare arrangements and risk of overweight and obesity in childhood: A systematic review. *Obesity Reviews : An Official Journal of the International Association for the Study of Obesity*, *18*(10), 1170–1190. https://doi.org/10.1111/obr.12575

Boone-Heinonen, J., Messer, L., Andrade, K., & Takemoto, E. (2016). Connecting the Dots in Childhood Obesity Disparities: A Review of Growth Patterns from Birth to Pre-Adolescence. *Current Epidemiology Reports*, *3*(1), 113–124. https://doi.org/10.1007/s40471-016-0065-9

Bürkner, P.-C. (2017). brms : An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*(1). https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, *10*(1), 395. https://doi.org/10.32614/RJ-2018-017

Campbell, F., Conti, G., Heckman, J. J., Moon, S. H., Pinto, R., Pungello, E., & Pan, Y. Y. (2014). Early Childhood Investments Substantially Boost Adult Health. *Science*, *343*(6178), 1478–1485. https://doi.org/10.1126/science.1248429

Carneiro, P., & Ginja, R. (2014). Long-Term Impacts of Compensatory Preschool on Health and Behavior: Evidence from Head Start. *American Economic Journal: Economic Policy*, *6*(4), 135–173. https://doi.org/10.1257/pol.6.4.135

Carpenter, J., Andrykowski, M., Wilson, J., Hall, L., Kay Rayens, M., Sachs, B., & Cunningham, L. (1998). Psychometrics for Two Short Forms of the Center for Epidemiologic Studies-Depression Scale. *Issues in Mental Health Nursing*, *19*(5), 481–494. https://doi.org/10.1080/016128498248917

Chen, Q., Gelman, A., Tracy, M., Norris, F. H., & Galea, S. (2015). Incorporating the sampling design in weighting adjustments for panel attrition. *Statistics in Medicine*, *34*(28), 3637–3647. https://doi.org/10.1002/sim.6618

Costa, S., Adams, J., Gonzalez-Nahm, S., & Benjamin Neelon, S. E. (2017). Childcare in Infancy and Later Obesity: A Narrative Review of Longitudinal Studies. *Current Pediatrics Reports*, *5*(3), 118–131. https://doi.org/10.1007/s40124-017-0134-7

Côté, S. M., Borge, A. I., Geoffroy, M.-C., Rutter, M., & Tremblay, R. E. (2008). Nonmaternal care in infancy and emotional/behavioral difficulties at 4 years old: Moderation by family risk characteristics. *Developmental Psychology*, *44*(1), 155–168. https://doi.org/10.1037/0012-1649.44.1.155

Dong, Q., Elliott, M. R., & Raghunathan, T. E. (2014). A nonparametric method to generate synthetic populations to adjust for complex sampling design features. *Survey Methodology*, *40*(1), 26–46.

Forest, V., Lanthier, C., Nelissen, M., & Roy, J. (2007). *Meeting Early Childhood Needs: Québec's educational program for childcare services*. Québec, Montréal. Ministère de la famille et des aînés.

Friendly, M., Beach, J., & Turiano, M. (2002). *Early childhood education and care in Canada 2001* (5th ed.). Childcare Resource and Research Unit, Centre for Urban and Community Studies, University of Toronto.

Gelman, A., & Pardoe, I. (2006). Bayesian Measures of Explained Variance and Pooling in Multilevel (Hierarchical) Models. *Technometrics*, *48*(2), 241–251. https://doi.org/10.1198/004017005000000517

Gelman, A., & Pardoe, I. (2007). Average Predictive Comparisons for Models with Nonlinearity, Interactions, and Variance Components. *Sociological Methodology*, *37*(1), 23–51. https://doi.org/10.1111/j.1467-9531.2007.00181.x

Geoffroy, M.-C., Power, C., Touchette, E., Dubois, L., Boivin, M., Séguin, J. R., Tremblay, R. E., & Côté, S. M. (2013). Childcare and overweight or obesity over 10 years of follow-up. *The Journal of Pediatrics*, *162*(4), 753-758.e1. https://doi.org/10.1016/j.jpeds.2012.09.026

Geserick, M., Vogel, M., Gausche, R., Lipek, T., Spielau, U., Keller, E., Pfäffle, R., Kiess, W., & Körner, A. (2018). Acceleration of BMI in Early Childhood and Risk of Sustained Obesity. *The New England Journal of Medicine*, *379*(14), 1303–1312. https://doi.org/10.1056/NEJMoa1803527

Institut de la statistique du Québec. (2016). *Data Collection Tools.* Gouvernement du Québec. I am, I'll be (QLSCD). https://www.jesuisjeserai.stat.gouv.qc.ca/informations_chercheurs/outils_collecte/outils_collecte_an.html

Isong, I. A., Richmond, T., Kawachi, I., & Avendaño, M. (2016). Childcare Attendance and Obesity Risk. *PEDIATRICS*, *138*(5). https://doi.org/10.1542/peds.2016-1539

Jetté, M. (2002). *Survey Description and Methodology – Part I: Logistics and Longitudinal Data Collections* (Québec Longitudinal Study of Child Development (QLSCD 1998-2002) – From Birth to 29 Months Vol. 2, No. 1). Québec. Institut de la statistique du Québec. http://www.jesuisjeserai.stat.gouv.qc.ca/publications/bebe_v2no1_an.html

Jones-Taylor, M. (2015). Early Care and Education Settings Are Vital for Childhood Obesity Prevention. *Childhood Obesity*, *11*(5), 489–490. https://doi.org/10.1089/chi.2015.0156

Kakinami, L., Séguin, L., Lambert, M., Gauvin, L., Nikiéma, B., & Paradis, G. (2014). Poverty's latent effect on adiposity during childhood: Evidence from a Québec birth cohort. *Journal of Epidemiology & Community Health*, *68*(3), 239–245. https://doi.org/10.1136/jech-2012-201881

Keil, A. P., Daza, E. J., Engel, S. M., Buckley, J. P., & Edwards, J. K. (2018). A Bayesian approach to the g-formula. *Statistical Methods in Medical Research*, *27*(10), 3183–3204. https://doi.org/10.1177/0962280217694665

Larson, N., Ward, D. S., Benjamin Neelon, S. E., & Story, M. (2011). What role can child-care settings play in obesity prevention? A review of the evidence and call for research efforts. *Journal of the American Dietetic Association*, *111*(9), 1343–1362. https://doi.org/10.1016/j.jada.2011.06.007

Laurin, I., Guay, D., Fournier, M., Bigras, N., & Solis, A. (2016). La fréquentation d'un service éducatif préscolaire : un facteur de protection pour le développement des enfants de familles à faible revenu? [Not Available]. *Canadian Journal of Public Health / Revue Canadienne de Sante'e Publique*, *106*(7 Suppl 2), eS14-20. https://doi.org/10.17269/cjph.106.4825

Lefebvre, P., Merrigan, P., & Roy-Desrosiers, F. (2011). *Quebec's Childcare Universal Low Fees Policy 10 Years After: Effects, Costs, and Benefits* (Working Paper 11-01). Montreal. CIRPEE: Centre Interuniversitaire sur le Risque, les Politiques Economiques et l »emploi.

Lemoine, N. P. (2019). Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. *Oikos*, *128*(7), 912–928. https://doi.org/10.1111/oik.05985

Lucie and André Chagnon Foundation. (2019). *Preschool programs*. https://naitreetgrandir.com/en/feature/kindergarten/preschool-programs/

Lumeng, J. C., Miller, A. L., Horodynski, M. A., Brophy-Herb, H. E., Contreras, D., Lee, H., Sturza, J., Kaciroti, N., & Peterson, K. E. (2017). Improving Self-Regulation for Obesity Prevention in Head Start: A Randomized Controlled Trial. *PEDIATRICS*, *139*(5). https://doi.org/10.1542/peds.2016-2047

McLaren, L., & McIntyre, L. (2014). Conceptualizing child care as a population health intervention: can a strong case be made for a universal approach in Canada, a liberal welfare regime? *Critical Public Health*, *24*(4), 418–428. https://doi.org/10.1080/09581596.2013.803035

McLaren, L., Zarrabi, M., Dutton, D. J., Auld, M. C., & Emery, J. C. H. (2012). Child care: Implications for overweight / obesity in Canadian children? *Chronic Diseases and Injuries in Canada*, *33*(1), 1–11.

Pearce, A., Li, L., Abbas, J., Ferguson, B., Graham, H., Law, C., & Millenium Cohort Study Child Health Group (2010). Is childcare associated with the risk of overweight and obesity in the early years? Findings from the UK Millennium Cohort Study. *International Journal of Obesity*, *34*(7), 1160–1168. https://doi.org/10.1038/ijo.2010.15

Public Health Agency of Canada. (2011). *Curbing Childhood Obesity: A Federal, Provincial and Territorial Framework for Action to Promote Healthy Weights*. Ottawa. Public Health Agency of Canada.

Raghunathan, T. E., Solenberger, P., Berglund, P. A., & Hoewyk, J. van. (2016). *IVEware* (Version 0.3) [Computer software]. The Regents of the University of Michigan. https://www.src.isr.umich.edu/software/iveware-documentation/

Rao, D. P., Kropac, E., Do, M. T., Roberts, K. C., & Jayaraman, G. C. (2016). Tendances en matière d'embonpoint et d'obésité chez les enfants au Canada [Childhood overweight and obesity trends in Canada]. *Health Promotion and Chronic Disease Prevention in Canada : Research, Policy and Practice*, *36*(9), 194–198. https://doi.org/10.24095/hpcdp.36.9.03

Rubin, D. B. (2007). The design versus the analysis of observational studies for causal effects: Parallels with the design of randomized trials. *Statistics in Medicine*, *26*(1), 20–36. https://doi.org/10.1002/sim.2739

Shields, M [Margot], & Tremblay, M. S. (2010). Canadian childhood obesity estimates based on WHO, IOTF and CDC cut-points. *International Journal of Pediatric Obesity : IJPO : An Official Journal of the International Association for the Study of Obesity*, *5*(3), 265–273. https://doi.org/10.3109/17477160903268282

Shpitser, I., VanderWeele, T. J., & Robins, J. M. (2012). On the Validity of Covariate Adjustment for Estimating Causal Effects. In *26th Conference on Uncertainty in Artificial Intelligence, UAI 2010*, Catalina Island, CA. http://arxiv.org/abs/1203.3515

Simmonds, M., Llewellyn, A., Owen, C. G., & Woolacott, N. (2016). Predicting adult obesity from childhood obesity: A systematic review and meta-analysis. *Obesity Reviews : An Official Journal of the International Association for the Study of Obesity*, *17*(2), 95–107. https://doi.org/10.1111/obr.12334

Singh, A. S., Mulder, C., Twisk, J. W. R., van Mechelen, W., & Chinapaw, M. J. M. (2008). Tracking of childhood overweight into adulthood: A systematic review of the literature. *Obesity Reviews : An Official Journal of the International Association for the Study of Obesity*, *9*(5), 474–488. https://doi.org/10.1111/j.1467-789X.2008.00475.x

Snowden, J. M., Rose, S., & Mortimer, K. M. (2011). Implementation of G-computation on a simulated data set: Demonstration of a causal inference technique. *American Journal of Epidemiology*, *173*(7), 731–738. https://doi.org/10.1093/aje/kwq472

Stan Development Team. (2018). *RStan: the R interface to Stan* (Version 2.18) [Computer software]. http://mc-stan.org/

Story, M., Kaphingst, K. M., & French, S. (2006). The Role of Child Care Settings in Obesity Prevention. *The Future of Children*, *16*(1), 143–168. https://doi.org/10.1353/foc.2006.0010

Swyden, K., Sisson, S. B., Lora, K., Castle, S., & Copeland, K. A. (2017). Association of childcare arrangement with overweight and obesity in preschool-aged children: A narrative review of literature. *International Journal of Obesity (2005)*, *41*(1), 1–12. https://doi.org/10.1038/ijo.2016.198

U.S Department of Housing and Urban Development. (2007). *Measuring Overcrowding in Housing* (017-002). Office of Policy Development and Research.

Volger, S., Rigassio Radler, D., & Rothpletz-Puglia, P. (2018). Early childhood obesity prevention efforts through a life course health development perspective: A scoping review. *PLOS ONE*, *13*(12), e0209787. https://doi.org/10.1371/journal.pone.0209787

WHO. (2019). *Growth reference 5-19 years: BMI-for-age (5-19 years)*. WHO. https://www.who.int/growthref/who2007_bmi_for_age/en/

Willms, J. D., & Shields, M [M.]. (1996). *A Measure of socioeconomic status for the National Longitudinal Study of Children.: Report prepared for Human Resource Development Canada*. Human Resource Development Canada.

Zheng, H., & Little, R. J. A. (2003). Penalized Spline Model-Based Estimation of the Finite Populations Total from Probability-Proportional-to-Size Samples. *Journal of Official Statistics*, *19*(2), 99–117.

Zhou, H., Elliott, M. R., & Raghunathan, T. E. (2016). A two-step semiparametric method to accommodate sampling weights in multiple imputation. *Biometrics*, *72*(1), 242–252. https://doi.org/10.1111/biom.12413

Zhou, X., & Reiter, J. P. (2010). A Note on Bayesian Inference After Multiple Imputation. *The American Statistician*, *64*(2), 159–163. https://doi.org/10.1198/tast.2010.09109

4 The Association between Preschool Childcare and Behavioural Self-Regulation in Quebec Children Aged 6 to 12 Years

4.1 Preamble

This manuscript was completed to estimate the difference in behavioural self-regulation in kindergarten attributable to type of preschool childcare used, particularly the Quebec regulated centre- and home-based type compared to informal types. The objective was also to know whether any differences present in kindergarten persisted through the elementary school years and whether effects differed in in children from more and less advantaged families. Several econometric studies of the Quebec childcare policy had found Quebec children in the post-policy era exhibited more externalizing problems. However, the results of some studies of other public programs reported beneficial effects. We hypothesized that inconsistencies may be explained by grouping of regulated centre- and home-based childcare types or age of initiation of centre-based care.

This manuscript is being submitted to the *Canadian Journal of Public Health*.

4.2 Title page and footnotes

Title: Association between Preschool Childcare and Behavioural Self-Regulation in Quebec Children Aged 6 to 12 Years

Short title: Preschool childcare and self-regulation in Quebec

Authors

Tanya J. Murphy[1], Jay S. Kaufman[1], Patricia Li[2], Russell Steele[3], Seungmi Yang[1]

Affiliations

[1] Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, QC, Canada

[2] McGill University Health Centre, Pediatrics, and the Research Institute of the McGill University Health Centre, CORE, Montreal, QC, Canada

[3] Department of Mathematics and Statistics, McGill University, Montreal, QC, Canada

Keywords: Behavior regulation; Bayesian Analysis; Child; Child, Preschool; Child Care; Child Health; Emotion regulation; Epidemiologic Effect Modifier; Executive function; Linear Models; Public Policy;

Corresponding author:

Tanya Murphy
1020 Pine Avenue Ouest
Montreal, Quebec, Canada H3A 1A2
tanya.murphy@mail.mcgill.ca

4.3 Abstract

Objectives: To estimate how type and timing of preschool childcare in Quebec—in the early years of the universal childcare program—affected behavioural self-regulation (SR) 6 through 12 years. We also estimated whether SR-childcare associations differed by sex and disadvantage.

Methods: Study participants were enrolled in the Quebec Longitudinal Study of Child Development, a representative birth cohort of singleton children born in 1997-98. We estimated each child's latent poor SR score (PSR) in kindergarten, grades 1, 2, 4, and 6, from 14 Likert scale items about hyperactivity, inattention, and impulsivity rated by the child's teacher or mother. We estimated the association between childcare type from age 2 to 5 years on PSR using Bayesian hierarchical linear regression, then summarized the association as population-averaged marginal effects for four counterfactual childcare profiles: center-based (CB), regulated home-based (RH), or unregulated home-based (UH) for 35 hours/week, or parental care only (P). We also compared starting centre-based care at 3 versus 2 years old.

Results: From 6241 outcome visits in 1657 children, we estimated that there would have been no difference in mean PSR in kindergarten had all children used RH or CB (0.00 SD, 95% CrI: -0.29, 0.28). But mean PSR would have been 0.22 SD higher with CB than UH (95% CrI: -0.04, 0.49), and 0.27 SD higher than P (95% CrI: 0.00, 0.53). Disadvantaged children experienced little difference by childcare type.

Conclusions: Quebec’s regulated centre- and home-based childcare did not improve behavioural self-regulation.

Keywords: Bayesian Analysis; Child Care; Child Health; Effect Modifier, Epidemiologic; Longitudinal Studies; Public Policy; Self-regulation;

4.4 Introduction

Behavioural self-regulation (SR) is increasingly proposed as a cause or mediator of health and well-being (Miller et al., 2018). Its operational definition is still evolving, but it generally refers to the extent an individual can modify their emotions in response to stimuli and direct their actions in the service of goals (Nigg, 2017; Rademacher & Koglin, 2018; Smithers et al., 2018). Many preschool- or kindergarten-based interventions have positively influenced SR in young children (Pandey et al., 2018; Sezgin & Demiriz, 2019) or shown long-term reductions in related outcomes such as less contact with the criminal justice system and unemployment (Kautz et al., 2014). In 1997, Quebec initiated a universal childcare program, *les Centres de la petite enfance* (CPE). Although the program was not an intervention specifically aimed at changing SR, one of its main goals was to promote development and reduce socioeconomic disparities in school-readiness (Forest et al., 2007). Children were eligible for the low-cost centre- and home-based childcare spaces ($5 per day until 2004) regardless of family income or employment status, but space was limited.

Comparing Quebec to the rest of Canada, pre- and post-CPE reform, Haeck et al. (2018) and Baker et al. (2019) found that hyperactivity, inattention, and physical aggression behaviours in children 5-9 years were slightly more frequent in Quebec children in the CPE era, but not consistently for all ages and cohorts. The small negative effects were mainly in children of more educated mothers. Effects diminished as children aged, and in later cohorts of children as the program matured (Haeck et al., 2018). Well-conducted quasi-experimental econometric studies provide strong causal evidence, but the ecological effect estimates—that is, that exposure to the *policy*—do not provide estimates of the differences between childcare type (centre- versus family-based versus no childcare use). Within Quebec, Yang et al. (unpublished) estimated the effects of individual-level exposure in Quebec children using measures of externalizing behaviours equivalent to Haeck et al. (2018). They found children who mainly used CPE care between 2 and 5 years old had, in kindergarten, slightly more problem behaviours (hyperactivity, aggression, and opposition as rated by teachers) than children who mainly had informal or parental care. However, with parents' ratings of behaviour, the differences were negligible. Centre- and home-based CPE were not separately evaluated. Outside of Canada, studies including universal childcare have shown mixed results; positive effects of centre-based care on a composite measure of externalizing behaviours and social skills were observed in France (Gomajee et al., 2018), but null to small negative effects on externalizing behaviours were observed in England (Stein et al., 2013), Norway (Solheim et al., 2013), and Australia (Gialamas et al., 2015). There is little evidence that large scale public childcare has improved the development of SR, but past studies generally used coarse measures of childcare exposure.

In this study we use a detailed measure of childcare to estimate the effect of childcare type on SR and whether any effects observed in kindergarten (6 years) persist through to grade 6 (12 years). We also estimated effect measure modification (EMM) by sex and family disadvantage.

4.5 Methods

4.5.1 Sample selection

The participants were enrolled in the Quebec Longitudinal Study of Child Development (*l'Étude longitudinale du développement des enfants du Québec*, ELDEQ), a representative birth cohort of singleton children born between October 1997 and July 1998 at 24 to 42 week gestational age (Jetté, 2002). The birth registry of the Canadian province of Quebec, excluding children living in the health regions of Nord-du-Québec, Nunavik and Terres-Cries-de-la-Baie-James (Figure 2.1), and other First Nations reserves served as the sampling frame; the survey design represented approximately 94% of Quebec children born at that time. The participation rate was 83% and 2120 children were enrolled in the longitudinal survey. The study was approved by the McGill University Faculty of Medicine institutional review board. The ELDEQ data use agreement stipulates that descriptive statistics must not include cross-tabulations or percentiles with fewer than five participants and figures must not plot individual values.

The target population for this study was Quebec children who would have been eligible for CPE programs and other common types of childcare (unspecialized, unregulated childcare, or care by a relative or parental), which is all children free of severe impairments or who do not require specialized preschool education (e.g. deaf children). We excluded 18 children identified as having a severe developmental disability or autism before the age of 12 years. Because few children lived with a single father or other guardians, we excluded children who were not living with a mother at the 5- or 17-month visit (n < 10). Also implicit in the research question is that children start elementary school; therefore, we excluded children who did not regularly attend elementary school (e.g., home-schooled; n < 10). Finally, children with no behavior ratings between pre-kindergarten and grade 6 were dropped (n=434). We adjusted the original survey weights for the loss of these participants (see details in Appendix C). The final analytic sample included 1657 participants. Figure 2.3 shows the participant flow diagram.

4.5.2 Measures

The Person Most Knowledgeable (PMK) about the child (the biological mother for >99%), responded to annual interviewer- and self-administered questionnaires. Perinatal medical records for the birth of the target child were retrieved with the mothers' consent (98.2%). Data were collected annually until the 8-year visit, then, generally, every two years. For this study, we drew the baseline variables from the 5- and 17-month visits (the first and second visits); the childcare exposure variables from the 2.5, 3.5, 4, and 5-year visits; and the behaviour ratings from 6, 7, 8, 10, and 12-year visits when most participants were in kindergarten, grades 1, 2, 4, and 6, respectively (Figure 2.2).

Outcome. Poor self-regulation (PSR) was represented by manifest symptoms of inattention, hyperactivity, reactive physical aggression, and tantrums that were collected from teachers at each outcome study visit (6, 7, 8, 10, and 12 years), and from the PMK at 5, 6, and 8 years. However, PMK were not asked the physical aggression questions at 8 years. Fathers (biological or stepfathers) who were living with the child and mother also completed the behaviour questionnaires at 5, 6, and 10 years. The fathers' ratings were not used in latent PSR measure because they were missing for children of single mothers (n = 254 at 5 years to 380 at 10 years), but they were included in the data set for multiple imputation (see *Statistical analysis*).

Items were adapted from the Social Behavior Questionnaire (Tremblay et al., 1991) or composed for the original survey. We selected the items that corresponded closely to the inhibitory and emotional control sub-scales of the BRIEF (Roth et al., 2014), and other studies' self-regulation measures (Bailey & Jones, 2019; Lin et al., 2019). See Appendix B.3 for details.

All items were prefaced with "In the past 12 months, how often would you say that [child's name]…"

- Inattention (3 questions):
  - Was easily distracted, had trouble sticking to any activity?
  - Was unable to concentrate, could not pay attention for long?
  - Was inattentive?
- Hyperactivity (6 questions):
  - Could not sit still, was restless or hyperactive?
  - Couldn't stop fidgeting?
  - Was impulsive, acted without thinking?
  - Had difficulty waiting for [his/her] turn in games?
  - Couldn't settle down to do anything for more than a few moments?
  - Was unable to wait when someone promised [him/her] something?
- Reactive physical aggression (4 questions):
  - When somebody accidentally hurt [him/her] (such as by bumping into [him/her]), [he/she] reacted with anger and fighting?
  - Reacted in an aggressive manner when contradicted?
  - Reacted in an aggressive manner when teased?
  - Reacted in an aggressive manner when something was taken away from [him/her]?
- Tantrums (1 question): Had temper tantrums or hot temper?

Item response choices were a 3-point Likert scale ("never or not true", "sometimes or somewhat true", "often or very true", "don't know" [set to missing]). Because all items were worded in terms of the frequency of negative behaviors, lower scores indicate better SR.

To avoid the problems of sum scores (Gorter et al., 2015) and take advantage of the data from multiple raters (PMK and teachers) (Renk, 2005), latent PSR scores at age 6, 7, 8, 10 and 12 years were estimated from years were estimated from a multilevel Bayesian ordinal logistic hierarchical model (Bürkner, 2019). Individual questionnaire items for each visit and rater were the unit of analysis. Unlike classic Item Response Theory (IRT) software, this latent regression approach allows unbalanced data (different number of observations per child-visit); therefore, all non-missing observations were used. The model had fixed effects for continuous age, sex, age-sex interaction, rater type, and log of cross-sectional survey weight; and crossed random effects for child-visit (intercepts, nested) and item (item intercepts and rater-type slope). The variances of the item random intercepts could vary (equivalent to estimating the discrimination parameter in 2-parameter logistic IRT models; specifically, the generalized partial credit model). Ten plausible values of each child's latent poor SR score for each visit were drawn from the posterior predictive distribution (ignoring item random effects) and carried through to the multiple imputation and as analysis outcomes. Details of the estimation of the latent PSR score in described in Appendix B.3.

Exposure. At each preschool visit (2.5, 3.5, 4, and 5 years), the PMK was asked whether the child regularly attended any childcare, and, if "yes", the hours per week. Center- and home-based arrangements were further classified as regulated under the subsidized ("$5/day") provincial program or not. For a home-based arrangement, the subsidy is synonymous with being CPE-regulated. Most subsidized center-based arrangements were CPE-regulated (Lefebvre et al., 2011), but the precise proportion was not known in this sample. Because of this lack of specificity and because only a small proportion of center-based arrangements were unsubsidized (~10%), we grouped all center-based arrangements together.

For each preschool year, we summarized childcare as a quasi-continuous variable for total hours per week in childcare and a nominal variable for the main childcare type. We defined the main type as the arrangement in which the child spent 70% or more of the total hours, which resulted in a nominal variable with seven categories: 1) none/parental (P), 2) center-based (CB), 3) regulated home-based (RH), 4) unregulated home-based (UH). When no single arrangement accounted for 70% or more of the total hours per week, the main type was labeled as a mix that included 5) centre-based (mCB), 6) regulated home-based but no centre-based (mRH), or 7) only unregulated home-based (mUH). In the last preschool year, part-time public prekindergarten was a possible eighth category (pK). To control for childcare stability, the cumulative number of different arrangements used across the four preschool years, including changes within main type, was categorized as 0–2, 3–4, 5–6, or more than 6 arrangements. (Separate type and intensity variables for each year captured timing.) In summary, childcare exposure was modelled using nine variables: four categorical annual 'main types', four quasi-continuous annual 'hrs/week', and one categorical 'number of arrangements.' For descriptive summaries, we derived a single categorical variable summarizing the type of childcare used most over the preschool years (P, CB, RH, or UH). The participants were eligible for the CPE program at 2 years old; therefore, childcare at 5 and 17 months were used as baseline adjustment variables.

Covariates. Other available pre-childcare variables included: perinatal conditions; infant health, temperament, and behaviour; parents' self-reported height and weight (converted to BMI), health and smoking, employment, demographics, parenting style, and social support; municipality type and housing conditions (see Table B.1 for details).

Family disadvantage: Baseline covariates were used to classify a child as more or less advantaged using a Family Risk Index (Côté et al., 2008), a cumulative risk model composed from family structure, functioning, and SES index, and mother's age and depression symptoms. See section 3.5.2 for details.

4.5.3 Analysis

Prior to estimation, item-wise missing variables were multiply imputed 50 times using sequential regression and IVEware software v.0.3 (Raghunathan et al., 2016) (Appendix C).

Step 1: We estimated the relation between preschool childcare and poor SR from kindergarten (median age = 6.1 years) to grade 6 (median age = 12.1 years) using multilevel linear regression to account for the correlation in repeated measures of PSR (level 1) within children (level 2). The models included child-specific random intercepts. Although the exposure, preschool childcare, occurred over years (age 2–5 years), it was complete by the time children started kindergarten; therefore, the exposure was time-invariant with respect to the outcome measures. Time in elementary school, measured in years, was the only time varying covariate. Childcare 'main type' dummies were interacted with time in elementary school to allow the time slope to differ by 'main type'. We controlled for pre-exposure covariates (flagged in Table B.1c). We also included attrition-adjusted survey weights as cubic basis splines to minimize potential nonresponse bias (Zheng & Little, 2003).

We estimated Bayesian models with regularizing priors (Lemoine, 2019), four chains, and 3000 iterations (or 4000 when there were divergence warnings) in the R 4.0 (R Foundation for Statistical Computing, Vienna, Austria) package ‘brms’ (Bürkner, 2017, 2018), version 2.13, an interface to Stan (Stan Development Team, 2018). Prior distributions: Student t(df = 5, mean = 0, SD = 1) for the adjustment variable coefficients; (Student t(df = 3, mean = 0, SD = 2) for the annual childcare type and their interaction term coefficients (i.e., less informative); Normal(mean = 0, SD = 2) for the population-level intercept; Cauchy(1, 2) for the standard deviation of the random intercepts; Cauchy(0, 2) for the residual errors (the software restricts standard deviation to positive values).

Step 2: We estimated population-averaged marginal means for idealized or counterfactual childcare profiles using g-computation (Snowden et al., 2011) (also known as averaged predictive comparisons (Gelman & Pardoe, 2006)). The six counterfactual childcare profiles were: in each of the four preschool years CB as main type, 35 hrs/wk; RH as main type, 35 hrs/wk; UH as main type, 35 hrs/wk; or P as main type, 5 hrs/wk (because hrs/wk had been truncated at <10 hrs in some years); and CB started between 2.5 and 3.5 years with parental or RH at 2.5 years.

From the models estimated in Step 1, we generated 200 predicted PSR scores for each child at 0.5, 1.5, 2.5, 4.5, and 6.5 years in elementary school for each imputed set (200 x 50 = 10,000 MCMC samples) setting childcare variables to counterfactual values, while leaving other covariates at their observed values. The net effect of a childcare profile, compared to a reference profile, was the mean of the individual differences. For effect measure modification (EMM) estimates of the childcare effects by sex and family disadvantage, we calculated mean subgroup differences. We used the 2.5th and 97.5th percentiles of the posterior predicted MCMC samples as the measure of uncertainty due to estimation error and multiple imputation variance (i.e. 95% credible intervals [CrI]) (X. Zhou & Reiter, 2010).

Sensitivity analysis

We estimated the main models with frequentist mixed effects models and bootstrapped confidence intervals for the marginal differences between counterfactual childcare profiles. We then repeated the frequentist analysis with two changes in the modeling of the behaviour ratings at 17-month visit: 1) without adjustment for baseline externalizing behaviours; 2) baseline hyperactivity, the behaviour sub-scale most strongly associated with childcare and SR outcomes, as an effect measure modifier of main childcare type.

4.6 Results

Table 3.1 shows characteristics of 1657 study children and their families in weighted percentages (or weighted mean) with 95% CIs accounting for the complex survey design and multiple imputation variance. Figure 3.1 presents the proportion of study participants in each type of main childcare at each preschool study visit and shows that nonparental childcare increased with age. Table D.2 shows the mean (interquartile range) of the hours per week and a summary of childcare use by level of family disadvantage. Children from less advantaged families were more likely to have been in parental care for all of their preschool years than more advantaged children (17.0% versus 9.8%), but regulated care was used in approximately equal proportions.

Latent PSR scores were estimated from teachers' and mothers' ratings collected over 6241visits between kindergarten and grade six. 1106 (66.7%) children had 4-6 measures, 274 (16.6%) had 3 measures, 146 (8.8%) had 2 measures, and 110 (7.9%) had one measure, and 21 children only had measures multiply imputed from pre-kindergarten and/or fathers' ratings (and other covariates). On average, SR improved with age but mainly in girls (Figure 4.1). The response proportions to the individual items rated by mothers and teachers are shown in Appendix Table B.2.

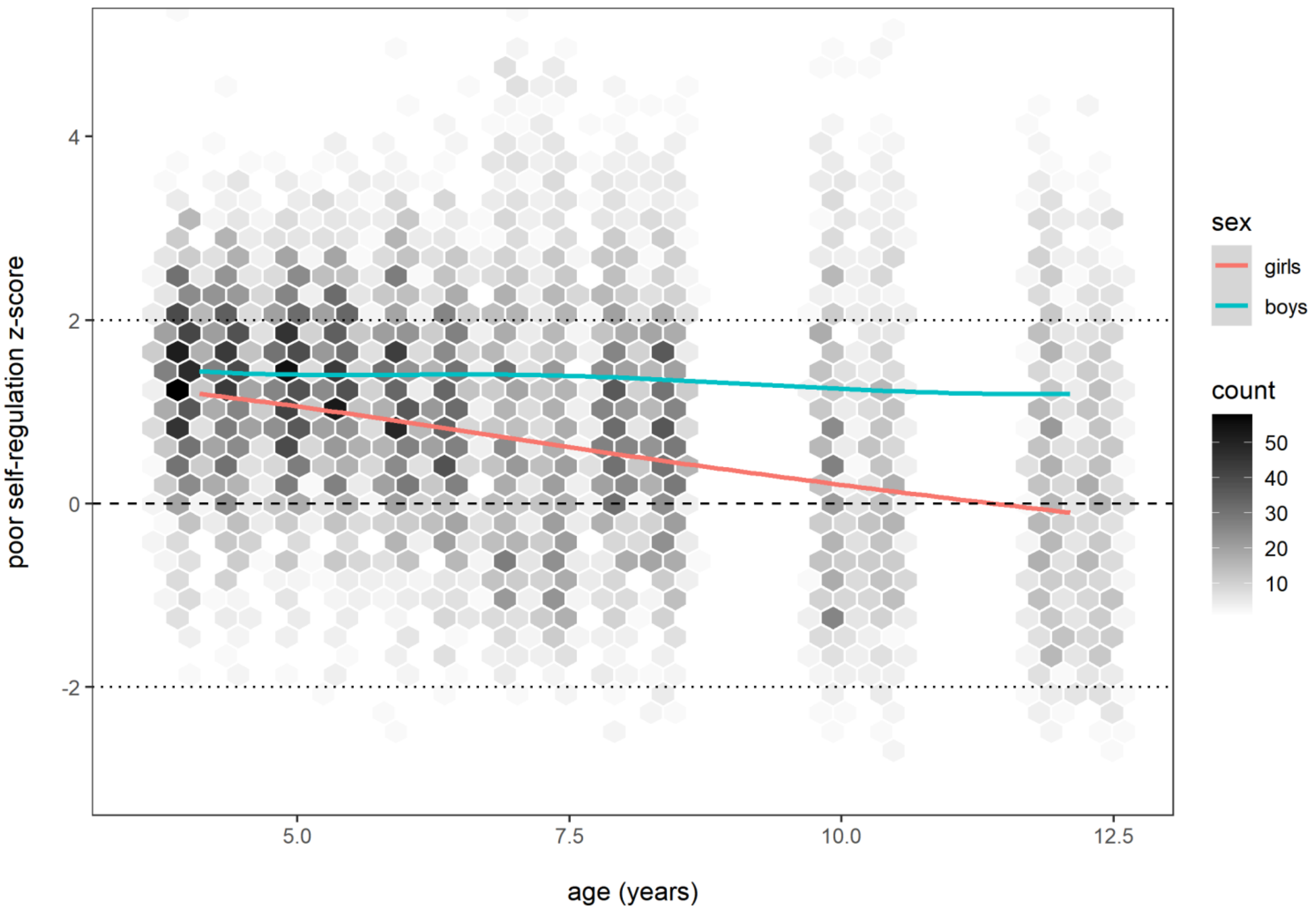


Figure 4.1. Poor self-regulation z-score by age and sex in participants of the Quebec Longitudinal Study of Child Development (2001-2009; n = 1657). Latent score estimated from Bayesian multilevel ordinal regression and 14 3-point Likert scale items rated by mothers at age 4, 5, 6, and 8 years, and teachers at age 6, 7, 8, 10, and 12 years. Lower score is better. Cells show aggregated counts at observed values, lines show smoothed means from weighted data with missing values multiply imputed (m=50). Number of latent scores based on observed values by sex and visit: 4-yr, 789 boys and 836 girls; 5-yr, 743 boys and 805 girls; 6-yr, 707 boys and 744 girls; 7-yr, 616 boys and 683 girls; 8-yr, 709 boys and 683 girls; 10-yr, 467 boys and 516 girls; 12-yr, 473 boys and 520 girls.

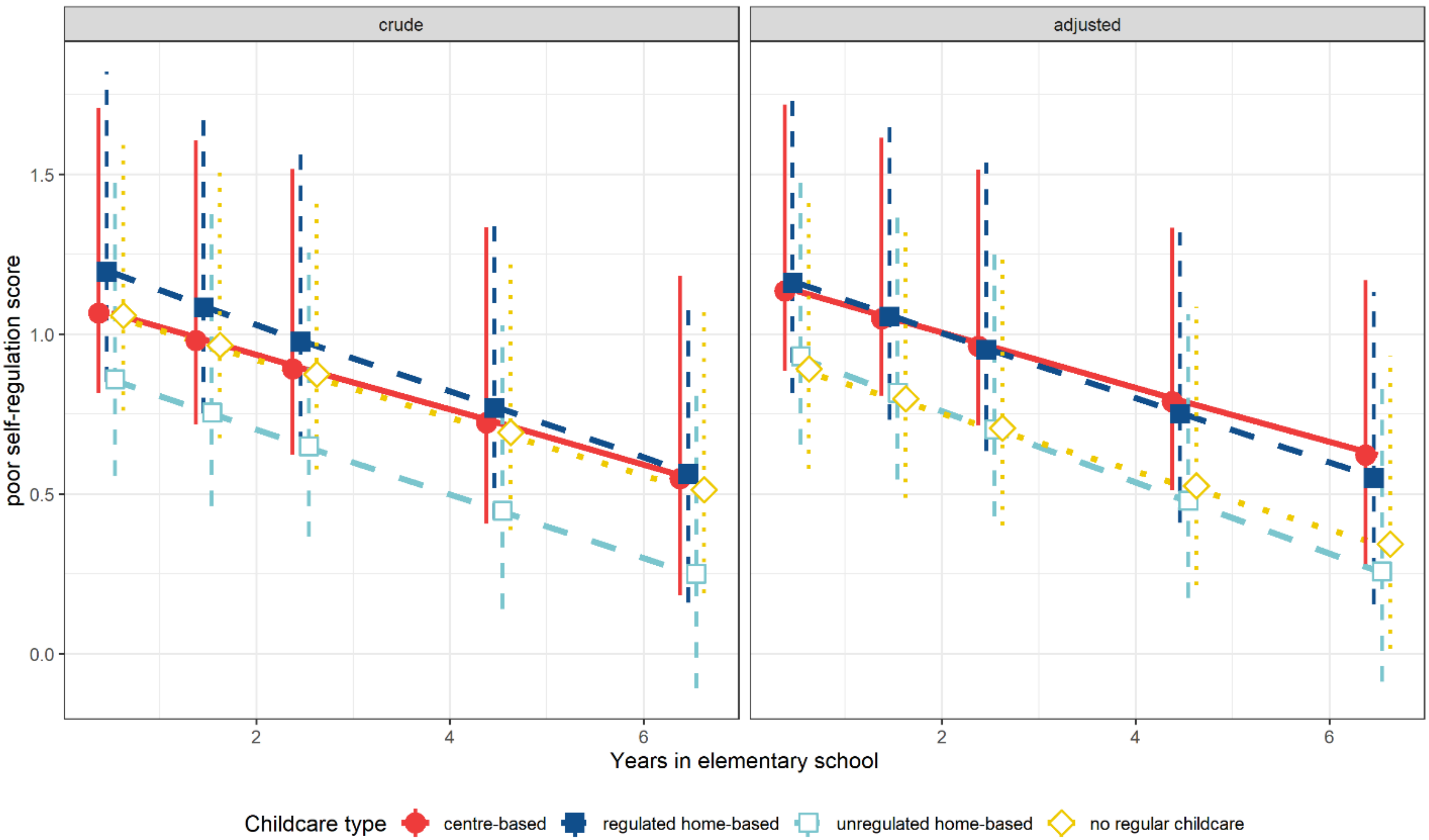


Figure 4.2. Predicted PSR for counterfactual childcare profiles from the crude and adjusted Bayesian multilevel linear models.

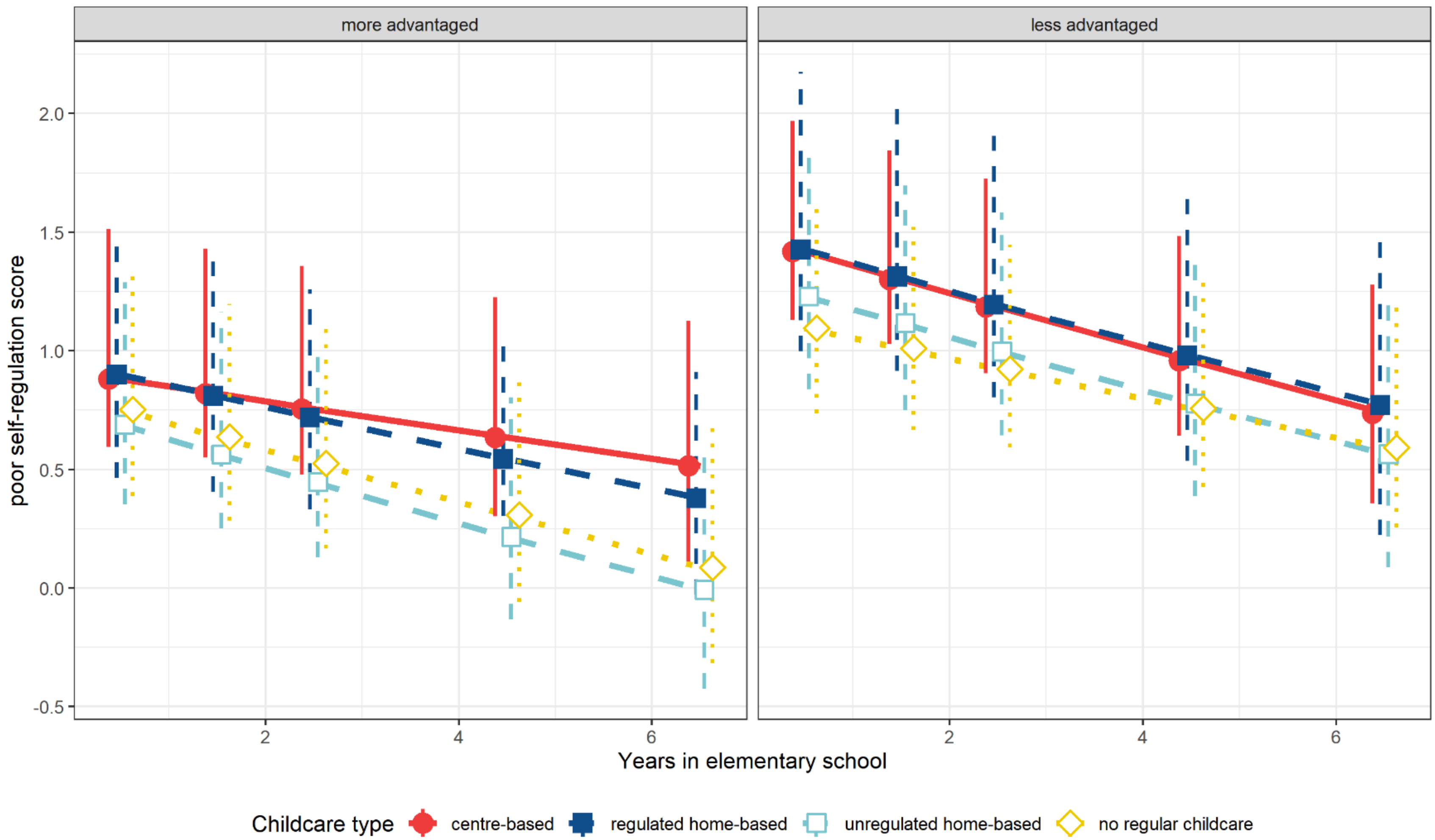


Figure 4.3. Predicted PSR for counterfactual childcare profiles, by level of family disadvantage. Adjusted Bayesian multilevel linear models.

Modeled PSR for the counterfactual childcare exposures (crude and adjusted estimates) are shown in Figure 4.2. Adjustment for baseline covariates had little effect on predictions for the population under CB, RH or UH profiles, but population-averaged parental-care predictions were, on average, lower (i.e., better SR) than unadjusted predictions. This difference suggests covariate distribution differed more between children who were mainly in parental care and those attending childcare, than between children who used different types of non-parental childcare. From the adjusted model, mean PSR score with CB for 35 hours per week from 2 to 5 years old would have been the same if all children had used RH; the difference was 0.00 SD (95% CrI: -0.29, 0.28). But mean PSR score with CB was 0.22 SD higher than UH (95% CrI: -0.04, 0.49), and 0.27 SD higher than P (95% CrI: 0.00, 0.53). The childcare-PSR relationship did not change between kindergarten and grade six. Table D.6 and Figure D.4 show the estimated differences between childcare profiles (crude, adjusted, and adjusted stratified by family disadvantage). Starting CB at 3 years instead of 2 years had no effect (0.02, 95% CrI: -0.21, 0.27, Figure D.5).

Childcare-by-PSR patterns differed somewhat by the level of a child's family's disadvantage. Although the disadvantage score was included as a continuous term (and interaction terms with childcare main type dummies), for brevity, we report marginal mean PSR scores for disadvantage scores at or below the median (more advantaged) versus higher (less advantaged). Figure 4.3 shows that, although less advantaged children had worse mean SR, childcare did not differentially affect PSR by level of disadvantage in kindergarten. However, among more advantaged children, the difference in mean PSR for regulated childcare and informal care increased (Table D.6). Similarly, boys had worse SR on average than girls, but sex hardly moderated the effect of childcare type (Figure D.6); although, CB effects were larger in boys.

The frequentist estimates for the main models were nearly identical to the Bayesian estimates (compare Tables D.6 and D.7a). Comparing frequentist main and sensitivity analysis models, had we not adjusted for baseline behaviour, overall estimates would not have changed (Table D.7b). In sex-stratified estimates, adjustment for baseline behaviour had an impact on point estimates, but changes were small relative to confidence interval widths and it would not have changed conclusions (Table D.7c). Childcare effects were not moderated by baseline behaviour (Table D.7d).

4.7 Discussion

In this study, we estimated the effect of the type of preschool childcare in Quebec children in the early years of the universal childcare program, *les centres de la petite enfance* (CPE) on the development of behavioural self-regulation. Controlling for other features of childcare use and pre-childcare variables, we found that CPE-regulated centre- and home-based use from two to five years old had equivalent effects on self-regulation (SR), and that they were associated with slightly poorer SR in kindergarten than informal care. These weak associations persisted through to grade six. There was no major effect measure modification by a child's sex or level of family disadvantage; indeed, differences were small relative to association between male sex or family disadvantage and poor SR.

Our results were consistent with recent studies of the Quebec childcare policy that found small negative on externalizing behaviours, overall. Like in Haeck et al. (2018) children from more advantaged families (or more educated mothers) exposed to regulated childcare (or having access to it) had slightly worse SR in later elementary school years compared to those having had informal care in preschool. Unsurprisingly, compared to Yang et al. (unpublished), in which teacher and ratings of externalizing behaviours were evaluated separately, our estimates were attenuated compared to estimates from the teacher ratings. Our results suggest that grouping regulated centre- and home-based care together in past studies, did not mask any heterogeneity in the effect of CPE on externalizing behaviours.

Strengths and limitations

One limitation of our study relates to our measure of behavioural self-regulation. We used the ELDEQ behaviour questions that most closely matched items in past studies, but we did not have scale validation data. Without an external standardized scoring method, the meaning of a given magnitude of difference in scores is hard to compare across studies; however, this limitation is common to these studies and will be until a standardized measure of self-regulation is accepted. Nevertheless, our latent model used the same items used in past studies of Canadian children for hyperactivity, inattention, and physical aggression sum scores, (collectively referred to as "externalizing behaviours"), with some exceptions: we only used the reactive aggression items and we added a single item about temper tantrums. We believe our psychometric model is an improvement over the sum scores for representing self-regulation. We followed recommended practice (to the extent possible with the data), integrating ratings in multiple settings (Renk, 2005) and accounting for the variable level of correlation and probability of endorsement across the items of a questionnaire (Gorter et al., 2015). By using multiple plausible values (ten for each child and visit), we propagate uncertainty from measurement error to the estimates (as well as multiple imputation and estimation variance). Finally, although the SR scores represent the distribution in the target population because the original data were drawn from a representative sample of Quebec children and we adjusted estimates for non-response.

The transportability of the results is also limited because the CPE programs (centre- and home-based) were not highly standardized. Nevertheless, the CPE childcare types imply higher quality care, on average than informal care because CPE regulation required more education in early child development and education, educational curricula. Indeed, childcare providers of the study participants had better quality ratings, on average, than unregulated providers (Japel et al., 2005). Although the measures of childcare exposure and SR outcomes are specific to this study, these limitations would not have affected the direction of any observed effects.

Other limitations relate to the causal interpretation of our results. Childcare was neither randomly assigned nor was access arbitrary because of the administrative differences between provinces such as in quasi-experimental designs. To minimize confounding, we adjusted the models for many pre-exposure variables including externalizing behaviors at 17 months, parent characteristics, and circumstances of birth. But we did not have data on some family characteristics such as parents' ADHD status, for example, which may have strongly predicted child's self-regulation and possibly affected a parent's ability to secure the childcare arrangement of choice. Overall, we believe our study largely meets the assumptions of causal inference because of 1) the temporal sequence of the adjustment variables, exposure, and outcomes; 2) the breadth of child and family characteristics used as adjustment variables; and 3) the theoretically equal opportunity for CPE care and observed similarity of the characteristics of participants using the range of childcare types—that is, enough overlap for statistical adjustment (positivity). Also, compared to summaries or ecological measures of childcare use in Quebec, childcare care in our study was well-defined.

In conclusion, regulated centre- and home-based universal childcare in Quebec did not promote the development of behavioural self-regulation better than informal or parental preschool care, overall, or in children from less advantaged families.

4.8 Manuscript references

Bailey, R., & Jones, S. M. (2019). An Integrated Model of Regulation for Applied Settings. *Clinical Child and Family Psychology Review*, *22*(1), 2–23. https://doi.org/10.1007/s10567-019-00288-y

Baker, S., Morawska, A., & Mitchell, A. (2019). Promoting Children's Healthy Habits Through Self-Regulation Via Parenting. *Clinical Child and Family Psychology Review*, *22*(1), 52–62. https://doi.org/10.1007/s10567-019-00280-6

Bürkner, P.-C. (2017). brms : An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*(1). https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (2019). Bayesian Item Response Modelling in R with brms and Stan. http://arxiv.org/pdf/1905.09501v2

Côté, S. M., Borge, A. I., Geoffroy, M.-C., Rutter, M., & Tremblay, R. E [Richard E.] (2008). Nonmaternal care in infancy and emotional/behavioral difficulties at 4 years old: Moderation by family risk characteristics. *Developmental Psychology*, *44*(1), 155–168. https://doi.org/10.1037/0012-1649.44.1.155

Forest, V., Lanthier, C., Nelissen, M., & Roy, J. (2007). *Meeting Early Childhood Needs: Québec's educational program for childcare services*. Québec, Montréal. Ministère de la famille et des aînés.

Friendly, M., Beach, J., & Turiano, M. (2002). Early childhood education and care in Canada 2001 (5th ed.). Childcare Resource and Research Unit, Centre for Urban and Community Studies, University of Toronto.

Gelman, A., & Pardoe, I. (2006). Bayesian Measures of Explained Variance and Pooling in Multilevel (Hierarchical) Models. *Technometrics*, *48*(2), 241–251. https://doi.org/10.1198/004017005000000517

Gialamas, A., Mittinty, M. N., Sawyer, M. G., Zubrick, S. R., & Lynch, J. (2015). Time spent in different types of childcare and children's development at school entry: An Australian longitudinal study. Archives of Disease in Childhood, 100(3), 226–232. https://doi.org/10.1136/archdischild-2014-306626

Gomajee, R., El-Khoury, F., Côté, S., van der Waerden, J., Pryor, L., & Melchior, M. (2018). Early childcare type predicts children's emotional and behavioural trajectories into middle childhood. Data from the EDEN mother-child cohort study. Journal of Epidemiology & Community Health, 72(11), 1033–1043. https://doi.org/10.1136/jech-2017-210393

Gorter, R., Fox, J.-P., & Twisk, J. W. R. (2015). Why item response theory should be used for longitudinal questionnaire data analysis in medical research. *BMC Medical Research Methodology*, *15*, Article 55, 1–12. https://doi.org/10.1186/s12874-015-0050-x

Haeck, C., Lebihan, L., & Merrigan, P. (2018). Universal Child Care and Long-Term Effects on Child Well-Being: Evidence from Canada. *Journal of Human Capital*, *12*(1), 38–98. https://doi.org/10.1086/696702

Japel, C., Tremblay, R. E [Richard E.], & Côté, S. M. (2005). Quality counts! Assessing the Quality of Daycare Services Based on the Quebec Longitudinal Study of Child Development. *Choices*, *11*(5), 4–42.

Jetté, M. (2002). *Survey Description and Methodology – Part I: Logistics and Longitudinal Data Collections* (Québec Longitudinal Study of Child Development (QLSCD 1998-2002) – From Birth to 29 Months Vol. 2, No. 1). Québec. Institut de la statistique du Québec. http://www.jesuisjeserai.stat.gouv.qc.ca/publications/bebe_v2no1_an.html

Kautz, T., Heckman, J. J., Diris, R., Weel, B. ter, & Borghans, L. (2014). *Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success* (NBER working paper No. 20749). OECD. https://www.nber.org/papers/w20749

Lefebvre, P., Merrigan, P [Phillip], & Roy-Desrosiers, F. (2011). *Quebec's Childcare Universal Low Fees Policy 10 Years After: Effects, Costs, and Benefits* (Working Paper 11-01). Montreal. CIRPEE: Centre Interuniversitaire sur le Risque, les Politiques Économiques et l'Emploi.

Lemoine, Nathan P. (2019): Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. In *Oikos* 128 (7), pp. 912–928. https://doi.org/10.1111/oik.05985

Lin, B., Liew, J., & Perez, M. (2019). Measurement of self-regulation in early childhood: Relations between laboratory and performance-based measures of effortful control and executive functioning. *Early Childhood Research Quarterly*, *47*, 1–8. https://doi.org/10.1016/j.ecresq.2018.10.004

Lucie and André Chagnon Foundation. (2019). Preschool programs. https://naitreetgrandir.com/en/feature/kindergarten/preschool-programs/

Miller, A. L., Gearhardt, A. N., Fredericks, E. M., Katz, B., Shapiro, L. F., Holden, K., Kaciroti, N., Gonzalez, R., Hunter, C., & Lumeng, J. C. (2018). Targeting self-regulation to promote health behaviors in children. *Behaviour Research and Therapy*, *101*, 71–81. https://doi.org/10.1016/j.brat.2017.09.008

Pandey, A., Hale, D., Das, S., Goddings, A.-L., Blakemore, S.-J., & Viner, R. M. (2018). Effectiveness of Universal Self-regulation-Based Interventions in Children and Adolescents: A Systematic Review and Meta-analysis. *JAMA Pediatrics*, *172*(6), 566–575. https://doi.org/10.1001/jamapediatrics.2018.0232

Rademacher, A., & Koglin, U. (2018). The concept of self-regulation and preschoolers' social-emotional development: a systematic review. *Early Child Development and Care*, *3*, 1–19. https://doi.org/10.1080/03004430.2018.1450251

Raghunathan, T. E., Solenberger, P., Berglund, P. A., & Hoewyk, J. van. (2016). *IVEware* (Version 0.3) [Computer software]. The Regents of the University of Michigan. https://www.src.isr.umich.edu/software/iveware-documentation/

Renk, K. (2005). Cross-Informant Ratings of the Behavior of Children and Adolescents: The "Gold Standard". *Journal of Child and Family Studies*, *14*(4), 457–468. https://doi.org/10.1007/s10826-005-7182-2

Roth, R. M., Isquith, P. K., & Gioia, G. A. (2014). Assessment of Executive Functioning Using the Behavior Rating Inventory of Executive Function (BRIEF). In S. Goldstein & J. A. Naglieri (Eds.), *Handbook of Executive Functioning* (Vol. 67, pp. 301–331). Springer New York. https://doi.org/10.1007/978-1-4614-8106-5_18

Sezgin, E., & Demiriz, S. (2019). Effect of play-based educational programme on behavioral self-regulation skills of 48-60 month-old children. *Early Child Development and Care, 189*(7), 1100–1113. https://doi.org/10.1080/03004430.2017.1369972

Smithers, L. G., Sawyer, A. C. P., Chittleborough, C. R., Davies, N., Davey Smith, G., & Lynch, J. (2018). A systematic review and meta-analysis of effects of early life non-cognitive skills on academic, psychosocial, cognitive and health outcomes. *Nature Human Behaviour*, *2*(11), 867–880. https://doi.org/10.1038/s41562-018-0461-x

Snowden, J. M., Rose, S., & Mortimer, K. M. (2011). Implementation of G-computation on a simulated data set: Demonstration of a causal inference technique. *American Journal of Epidemiology*, *173*(7), 731–738. https://doi.org/10.1093/aje/kwq472

Solheim, E., Wichstrøm, L., Belsky, J., & Berg-Nielsen, T. S. (2013). Do time in child care and peer group exposure predict poor socioemotional adjustment in Norway? Child Development, 84(5), 1701–1715. https://doi.org/10.1111/cdev.12071

Stan Development Team. (2018). *RStan: the R interface to Stan* (Version 2.18) [Computer software]. http://mc-stan.org/

Tremblay, R. E [R. E.], Loeber, R., Gagnon, C., Charlebois, P., Larivée, S., & LeBlanc, M. (1991). Disruptive boys with stable and unstable high fighting behavior patterns during junior elementary school. *Journal of Abnormal Child Psychology*, *19*(3), 285–300.

Yang, S., Charters, T., Weinstock, D., & Nandi, A. (submitted) *Impacts of Universal Childcare on Behavioral Development: Evidence from a Birth Cohort in Quebec, Canada*.

Zheng, H., & Little, R. J. A. (2003). Penalized Spline Model-Based Estimation of the Finite Populations Total from Probability-Proportional-to-Size Samples. *Journal of Official Statistics*, *19*(2), 99–117.

Zhou, X., & Reiter, J. P. (2010). A Note on Bayesian Inference After Multiple Imputation. *The American Statistician*, *64*(2), 159–163. https://doi.org/10.1198/tast.2010.09109

5 The association between poor behavioral self-regulation and adiposity in Quebec children age 7 to 13 years

5.1 Preamble

The lack of success of childhood obesity interventions prompted the design of preschool-based interventions with added behavioural self-regulation (or self-control) education components. That better self-regulation abilities would help a child achieve healthy weight goals is intuitively appealing, and some observational studies supported the relation. However, the enhanced interventions were not successful. We hypothesized that the reported associations between self-regulation and adiposity may have been confounded or only emerge in later childhood. Therefore, this manuscript estimates the longitudinal relation between serial measures of self-regulation and adiposity indices, controlling for a rich set of potential confounders.

This manuscript is being submitted to the journal *Epidemiology*.

5.2 Title page and footnotes

Title: The association between poor behavioral self-regulation and adiposity in Quebec children age 7 to 13 years

Short title: Preschool childcare and self-regulation in Quebec

Authors

Tanya J. Murphy[1], Jay S. Kaufman[1], Patricia Li[2], Russell Steele[3], Seungmi Yang[1]

Affiliations

[1] Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, QC, Canada

[2] McGill University Health Centre, Pediatrics, and the Research Institute of the McGill University Health Centre, CORE, Montreal, QC, Canada

[3] Department of Mathematics and Statistics, McGill University, Montreal, QC, Canada

Keywords: Behavior regulation; Bayesian Analysis; Child; Child, Preschool; Child Care; Child Health; Emotion regulation; Epidemiologic Effect Modifier; Executive function; Linear Models; Public Policy;

Corresponding author:

Tanya Murphy
1020 Pine Avenue Ouest
Montreal, Quebec, Canada H3A 1A2
tanya.murphy@mail.mcgill.ca

5.3 Abstract

Promotion of behavioral self-regulation (SR) in young children has been proposed for the prevention childhood obesity, but evidence for its effectiveness is limited. We aimed to estimate the effect of SR deficits on adiposity in healthy Quebec children, independent of confounders. Our study participants were children enrolled in a representative birth cohort, *l'Étude longitudinal du développement de l'enfant au Québec* (1998–2010). Repeated measures of BMI z-scores (BMIz) at age 7, 8, 10, 12, and 13 years were regressed on a flexible model for the recency-weighted cumulative mean of poor SR score (cPSR) in the 3 years before each BMIz measure, by age and sex. Weighted generalized estimating equations (GEE) were used to account for multiple measures per child. To control for confounding, models were weighted by the inverse of the generalized propensity score for PSR. The generalized propensity score was estimated with linear regression, optimized for covariate balance, and included pre-PSR variables such as perinatal conditions, child health, and family characteristics, which were multiply imputed when missing. From 7,859 outcomes in 1,657 children, we found that 1-SD increase in the population mean of cPSR was not associated with a difference in the marginal mean of BMIz or probability of obesity, regardless of age and sex. The largest difference in mean BMIz was observed in 10-year-old girls when a cPSR of 1 SD predicted a mean BMIz of 0.42 SD (95% CI: 0.25, 0.59) whereas a cPSR of 2 SD predicted a mean BMIz of 0.46 SD (95% CI: 0.25, 0.67). Differences in obesity prevalence associated with a 1-SD difference in cPSR were less than ±1.5% for each age and sex. While the promotion of SR has many potential benefits for children's well-being, it is unlikely to be effective for the prevention of obesity in early and middle childhood.

Keywords: Childhood; Self-regulation; Adiposity; inverse probability weighting.

5.4 Introduction

Obesity is a large burden on the health system and individuals that is increasingly starting in childhood, but childhood obesity prevention interventions have not been successful, generally. Since obesity prevention involves a behavioural component, supporting the development of self-regulation in children has been explored as complementary component of interventions. Behavioural self-regulation is one's ability to regulate emotions and direct behavior in the service of goals (McClelland et al., 2018); although, the definition is still being refined (Nigg, 2017; Smithers et al., 2018).

However, preschool-based obesity prevention studies that included a self-regulation development program did not achieve reductions in adiposity indices (Knowlden et al., 2015; Lumeng et al., 2017). Early observational studies motivating the hypothesis were cross-sectional or did not control for confounding (Blair et al., 2019; Francis & Susman, 2009; Fuemmeler et al., 2011; Smithers et al., 2018). However, recent longitudinal studies have reported an independent effect of SR on adiposity in childhood. Piche et al. (2012) found higher impulsivity in kindergarten predicted less of an increase in BMI by grade four. In contrast, Howard and Williams (2018) and Datar and Chung (2018) found self-regulation (or self-control) was predictive of obesity. It is possible that better SR benefits adiposity only at older ages.

In this study, we aimed to estimate the effect of SR deficits on adiposity in healthy Quebec children, and whether effects differed by age and sex.

5.5 Methods

5.5.1 Study population

The participants were enrolled in the Quebec Longitudinal Study of Child Development (*l'Étude longitudinale du développement des enfants du Québec*, ELDEQ), a representative birth cohort of singleton children born between October 1997 and July 1998 at 24 to 42 week gestational age (Jetté, 2002). The birth registry of the Canadian province of Quebec (excluding children living in some remote regions and Indigenous territories or reserves) served as the sampling frame; the survey design represented approximately 96% of Quebec children born at that time. The participation rate was 83.1% and 2,120 children were enrolled in the longitudinal survey. Data were collected annually until the 8-year visit, then, at 10, 12, and 13 years. See Chapter 2 for additional information about the original survey.

The target population for this study was Quebec children free of severe developmental disabilities or autism who attended elementary school. We excluded children who were not living with a mother at the 5- or 17-month visit because few children in the study sample lived with a single father or other guardians (n=7). We also dropped children with no BMI or behavior ratings between age 5 and 13 years from the analysis because they had little data to inform the multiple imputation or outcome models (n=434), but we adjusted the survey weights for response propensity to maintain the representativeness of the sample (Appendix C). The final analytic sample included 1,657 participants.

5.5.2 Measures

The Person Most Knowledgeable (PMK) about the child (the biological mother for >99%), responded to annual interviewer- and self-administered questionnaires. Perinatal medical records for the birth of the target child were retrieved with the mothers' consent (98.2%).

Exposure. Poor behavioral self-regulation (PSR) was represented by manifest symptoms of inattention, hyperactivity, reactive physical aggression, and tantrums that were collected from the teacher at 6, 7, 8, 10, and 12 years, and from the PMK at 4, 5, 6 and 8 years. However, PMK were not asked the physical aggression questions at year 8. Fathers (biological or stepfathers) who were living with the child and mother also completed the behaviour questionnaires at 4, 5, 6 and 10 years. The fathers' ratings were not used in latent PSR model because they were entirely missing for children of single mothers, which would have confounded family structure and rater effects. Fathers' ratings were included as auxiliary variables in multiple imputation (see Appendix C).

Items were adapted from the Social Behavior Questionnaire (Tremblay et al., 1991) or composed for the original survey. We selected the items that corresponded closely to the inhibitory and emotional control sub-scales of the BRIEF-2 (Jacobson et al., 2016; Roth et al., 2014), and current models of SR (Bailey & Jones, 2019; Lin et al., 2019).

All items were prefaced with "In the past 12 months, how often would you say that [child's name]…"

- Inattention (3 questions):
  - Was easily distracted, had trouble sticking to any activity?
  - Was unable to concentrate, could not pay attention for long?
  - Was inattentive?
- Hyperactivity (6 questions):

  - Could not sit still, was restless or hyperactive?
  - Couldn't stop fidgeting?
  - Was impulsive, acted without thinking?
  - Had difficulty waiting for [his/her] turn in games?
  - Couldn't settle down to do anything for more than a few moments?
  - Was unable to wait when someone promised [him/her] something?
- Reactive physical aggression (4 questions):
  - When somebody accidentally hurt [him/her] (such as by bumping into [him/her]), [he/she] reacted with anger and fighting?
  - Reacted in an aggressive manner when contradicted?
  - Reacted in an aggressive manner when teased?
  - Reacted in an aggressive manner when something was taken away from [him/her]?
- Tantrums (1 question): Had temper tantrums or hot temper?

Items constituted a 3-point Likert scale with “never or not true”, “sometimes or somewhat true”, “often or very true” as response choices (“don’t know” was set to missing). Because all items were worded in terms of the frequency of negative behaviors, lower scores indicate better SR. The latent PSR score was estimated using a Bayesian hierarchical ordinal regression model, computed using R 4.0 (R Foundation for Statistical Computing, Vienna, Austria) and the package ‘brms’ v.2.12 (Bürkner, 2017) as proposed by Bürkner (2019), and is described in Appendix B.3.

Cumulative PSR (cPSR), was calculated as a recency-weighted average of the PSR scores collected in the three years prior to each BMI measure. However, three prior years of behaviour ratings were not available for every BMI collection visit, and Appendix Table D.8 shows how cPSR was calculated for each year.

Outcome. Adiposity was represented by age- and sex-specific BMI z-score (BMIz). Height and weight were measured by trained interviewers at age 8, 10, 12 and 13 years as the mean of two measures, or three if there was a discrepancy of >0.5 kg or >0.5 cm between the first two (Desrosiers et al., 2009). At 6 and 7 years, a single measure of height and weight were measured by an interviewer. To have at least one teacher rating in the 3-year PSR average, the 6-year BMI measure was not used as an outcome, but it was included in multiple imputation. BMI was calculated as weight in kilograms divided by height in meters squared and converted to z-scores according to the World Health Organization's (WHO) 2007 standard (WHO, 2019). Obesity was defined as BMIz > 2SD. Figure B.1 shows the relationship between height, weight, and the WHO BMI category cut-offs; Table B.2 also shows International Obesity Task Force (IOTF) and CDC cut-offs.

Other covariates. Time-fixed variables from the first two visits and time-updated variables from visits three to five years prior to the BMI measure visit were considered for inclusion in the propensity score (PS) model. The time-fixed variables included: perinatal conditions; parents' education, age, immigration status, BMI, and parenting style; religiosity and main language spoken at spoken, social support. Most of the same PSR items had been asked of mothers at 17 months; we did not use them in the main model because we believed it would cause an over-adjustment bias (Schisterman et al., 2009), but we present the results with 17-month behavior in the PS as a sensitivity analysis. Time-varying variables included: child's general health; parents' general health, employment status, and maternal depression; household SES index, family structure and functioning, and residence. The complete list of potential adjustment variables is shown in Appendix Table B.1.

5.5.3 Statistical analyses

Generalized propensity score for 3-yr cumulative PSR. An inverse probability of 'treatment' weight (IPTW) was estimated from a generalized propensity score (gPS) model for each child's cPSR at each outcome visit (5 per child). The covariates described above were screened as potential confounders based on background knowledge. When little was known about the relationship with exposure and outcome, the additional predictive value of variables was estimated separately for BMIz and cPSR using elastic net regression via the R package glmnet v.3.0 (Friedman et al., 2010). Variables that predicted BMIz were included in the gPS model with the exception of the following scenario: When the association with BMIz was very weak but the association with cPSR was very strong, the variable was excluded. In this scenario, the variance of the exposure coefficient tends to be inflated without much reduction in confounding bias; that is, the bias-variance trade-off is poor.

The gPS were estimated by linear regression, optimized for covariate balance and converted to IPTW (not stabilized), using the R package CBPS v.0.21 (Fong et al., 2018; Imai & Ratkovic, 2014). Covariate balance in the IPT weighted data was assessed with Pearson correlation coefficient where a correlation under 0.10 was considered balanced.

Outcome model. The relation between cPSR and BMIz was estimated with weighted generalized estimating equations (GEE) with the identity link, child as the cluster, and weighted by the IPTW multiplied by the adjusted survey weights. Cumulative PSR was modelled as natural splines with 3 degrees of freedom (df). In addition to the main cPSR score, root and interaction terms for age (as natural splines with 4 df) and sex were included as independent variables. The final estimand was the population-standardized marginal mean BMIz for each sex and target study-visit age (7, 8, 10, 12, and 13 years). 95% CIs were estimated by bootstrapping the IPTW and outcome models 100 times for each of the 50 imputed data sets (5000 samples). The procedure was repeated with obesity status as the outcome and the logit link.

Sensitivity analysis. The main model for BMIz was re-estimated without the bootstrap procedure. The R package ‘emmeans’ v.1.4.6 (Lenth, 2020) was used to obtain the point estimate and standard error of the marginal mean BMIz for each sex and target study-visit age for each imputed data set. The final estimates were calculated using “Rubin’s rules” (Rubin, 1987). This procedure had a much shorter computation time than the bootstrap procedure, allowing for comparison several alternative models: 1) instead of linear recency weights, quadratic weights that weighted the lag-1 PSR more heavily than in the main model with linear weight (see Appendix Table D.8); 2) 17-month hyperactivity was included in the gPS model.

5.6 Results

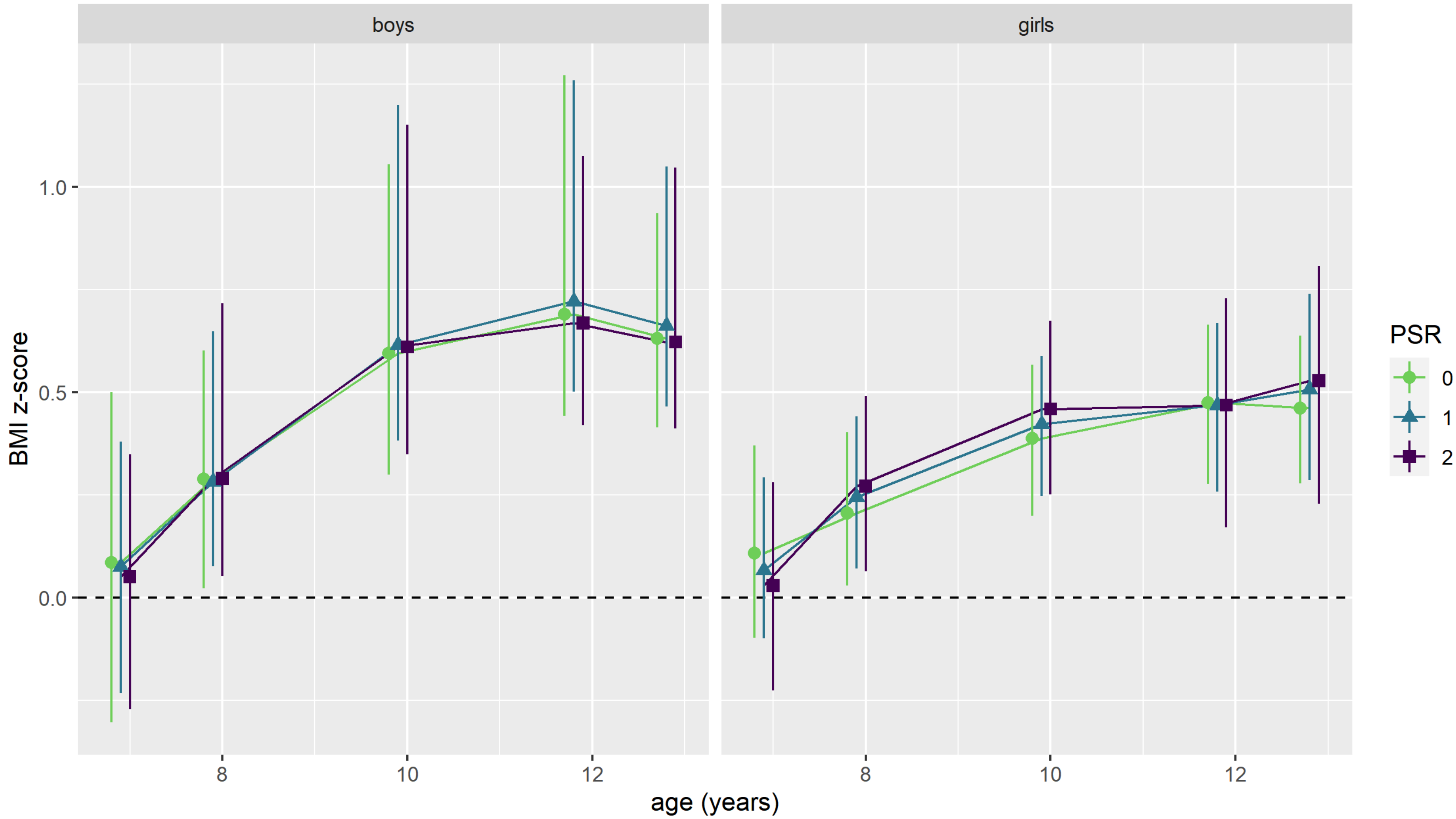
Our study included 1657 children and 7859 BMI measures. Table 3.1 shows the characteristics of the participants and their families at 0–1.5 years. The characteristics of the children in the sample closely reflected the source population as estimated by the weighted imputed data. However, children with a parent who is an immigrant or from low income families were somewhat under-represented in the study sample.

The distribution of non-missing poor self-regulation scores and the weighted, imputed trend with age are shown in Figure 4.1. Girls and boys had similar scores at age four, but the mean score for girls decreased almost linearly with age and the mean for boys decreased little. The generalized propensity score for recency-weighted 3-yr average PSR achieved satisfactory balance in the covariates in the IPT weighted sample (Appendix Figure D.7). Few of the suspected confounders were strongly correlated with SR before weighting (i.e., Pearson r > 0.1) but those that were—such as sex, parents' education, family functioning and structure—were virtually unrelated (Pearson r < 0.05) after weighting.

For the study sample and the population, Figure 3.2 shows the BMI z-scores and Figure D.1 shows the prevalence of obesity. The prevalence of obesity increased with age and was estimated to be higher in the source population than in the study sample—likely due to the association between family disadvantage, higher loss to follow-up and higher risk of overweight-obesity.

Figure 5.1 shows the mean BMI z-score from 7 to 13 years for levels of recent PSR, independent of other predictors of PSR and adiposity according to our IPTW model. The recency-weighted 3-year mean PSR was not associated with mean BMI z-score, regardless of sex or age. For example, had all girls in the population, when they were 10 years old, had a cPSR of 0 SD, mean BMIz would have been 0.39 SD (bootstrap 95% CI: 0.20, 0.57); whereas, had the same girls had a cPSR of 2 SD, mean BMIz would have been 0.46 SD (95% CI: 0.25, 0.67). A similar pattern was estimated for the probability of obesity (Figure D.8); cPSR accounted only for differences of less than ±1.5%, which was small relative to the 95% CI widths.

Compared to linear recency-weights, quadratic recency weights did not change the results (Figures D.9). Controlling for mother-rated baseline (17-month) hyperactivity did not change the results (Figure D.10).



Marginal mean (2.5th and 97.5th percentiles) from IPT-weighted GEE, BMIz = bs(SR, df=3) * girl * bs(age, df=4) (centered on 9 years),
30 imputations x 200 bootstrap samples.
SR z-score seleted from within the range of the data.

Figure 5.1. Marginal mean BMI z-score for mean values of recency-weighted 3-year cumulative mean poor self-regulation score (PSR), by sex.

5.7 Discussion

In this study we estimated how much recent level of self-regulation explained adiposity in children age 7 to 13 years. Independent of child characteristics and family conditions that pre-dated the self-regulation and adiposity measures, we found that poor self-regulation explained almost none of the variation in BMI z-score or obesity status, regardless of age, sex, or how the 3-year cumulative self-regulation mean score was weighted. This brings into question how much the improvement of self-regulation would be a useful upstream target in obesity prevention.

However, our results differ somewhat from those of some recent well-designed studies. Howard and Williams (2018) found a 1.4-fold increase in the odds of overweight and obesity at 14 years old with poorer SR at 4–6 years in Australian children. It is possible that effects of poor SR may take time to accumulate. Our measure of SR was the recency-weighted 3-year mean prior to each BMI measure, with the measure one year (or two years) prior having the most weight; therefore, we were estimating relatively proximal effects compared to Howard and Williams (2018). Datar and Chung (2018) also had a delay of several years between their measure of self-control and the BMI measures. They found no association between BMI change and self-control but found a decrease in the *incidence* of obesity with better self-control.

The preschool-based interventions that targeted SR as an intermediary to obesity prevention had short follow-up (Knowlden et al., 2015; Lumeng et al., 2017). Therefore, it is possible that if improvements in SR take several years to have an impact on adiposity, they would not have been observed in the 1-year follow-up of the preschool-based interventions. They did report an improvement in SR. We did not examine the effect of within-child changes in SR, but Howard and Williams (2018) did; they found no association with the change in SR between 4 and 6 years.

5.7.1 Study strengths

We exploited the rich longitudinal data to minimize confounding in estimating the effect of recent SR on adiposity indices. We estimated a covariate-balancing generalized propensity score and derived IPT weights, to avoid collider stratification bias. That is, IPTW was more prudent than regression adjustment because the exposure and several of the potential confounders were time-varying with respect to the outcome. We conserved the representativeness of the original survey by adjusting survey weights for attrition and multiply imputed missing data in the remaining participants.

5.7.2 Limitations of data

A limitation in our study was that we did not use a validated measure of poor self-regulation. Our manifest items closely mirrored the measure used by Moffit (and Howard), and to a slightly less extent, components of the BRIEF, whose psychometric properties have been extensively studied. Therefore, we believe the items are good indicators of self-regulation. However, the effect of combining mother and teacher ratings in a latent regression model on accuracy is unknown. In principle, ratings from more raters captures a more complete measure of children's SR because mothers and teacher observe the child in different contexts. A weakness in some aspects of SR that are present at home and at school are more likely to be SR weaknesses rather than a reaction to something about the school or home setting. By using a latent regression model, items or raters that are less correlated with the overall pattern within a child, carry less weight; therefore, our measure should be reflecting a signal within the data better than a sum score.

In conclusion, self-regulation has little effect on adiposity in children 7 to 13 years old when common causes are controlled. Leveraging interventions that improve self-regulation to prevent childhood obesity is not likely to be effective.

5.8 Manuscript references

Bailey, R., & Jones, S. M. (2019). An Integrated Model of Regulation for Applied Settings. *Clinical Child and Family Psychology Review*, *22*(1), 2–23. https://doi.org/10.1007/s10567-019-00288-y

Blair, C., Kuzawa, C. W., & Willoughby, M. T. (2019). The development of executive function in early childhood is inversely related to change in body mass index: Evidence for an energetic tradeoff? *Developmental Science*, e12860. https://doi.org/10.1111/desc.12860

Bürkner, P.-C. (2017). brms : An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*(1). https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (2019). Bayesian Item Response Modelling in R with brms and Stan. http://arxiv.org/pdf/1905.09501v2

Datar, A., & Chung, P. J. (2018). Childhood Self-Control and Adolescent Obesity: Evidence from Longitudinal Data on a National Cohort. *Childhood Obesity (Print)*, *14*(4), 238–247. https://doi.org/10.1089/chi.2017.0217

Desrosiers, H., Dumitru, V., & Dubois, L. (2009). Le surplus de poids chez les enfants de 4 à 7 ans : des cibles pour l'action. *Étude Longitudinale Du Développement Des Enfants Du Québec (ÉLDEQ 1998-2010)*, *4*(3), 1–19.

Fong, C., Hazlett, C., & Imai, K. (2018). Covariate balancing propensity score for a continuous treatment: Application to the efficacy of political advertisements. *The Annals of Applied Statistics*, *12*(1), 156–177. https://doi.org/10.1214/17-AOAS1101

Francis, L. A., & Susman, E. J. (2009). Self-regulation and rapid weight gain in children from age 3 to 12 years. *Archives of Pediatrics & Adolescent Medicine*, *163*(4), 297–302. https://doi.org/10.1001/archpediatrics.2008.579

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, *33*(1), 1–22. http://www.jstatsoft.org/v33/i01/

Fuemmeler, B. F., Østbye, T., Yang, C., McClernon, F. J., & Kollins, S. H. (2011). Association between attention-deficit/hyperactivity disorder symptoms and obesity and hypertension in early adulthood: A population-based study. *International Journal of Obesity (2005)*, *35*(6), 852–862. https://doi.org/10.1038/ijo.2010.214

Howard, S. J., & Williams, K. E. (2018). Early Self-Regulation, Early Self-Regulatory Change, and Their Longitudinal Relations to Adolescents' Academic, Health, and Mental Well-Being Outcomes. *Journal of Developmental and Behavioral Pediatrics : JDBP*, *39*(6), 489–496. https://doi.org/10.1097/DBP.0000000000000578

Imai, K., & Ratkovic, M. (2014). Covariate balancing propensity score. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *76*(1), 243–263. https://doi.org/10.1111/rssb.12027

Jacobson, L. A., Pritchard, A. E., Koriakin, T. A., Jones, K. E., & Mahone, E. M. (2016). Initial Examination of the BRIEF2 in Clinically Referred Children With and Without ADHD Symptoms. *Journal of Attention Disorders.* Advance online publication. https://doi.org/10.1177/1087054716663632

Jetté, M. (2002). *Survey Description and Methodology – Part I: Logistics and Longitudinal Data Collections* (Québec Longitudinal Study of Child Development (QLSCD 1998-2002) – From Birth to 29 Months Vol. 2, No. 1). Québec. Institut de la statistique du Québec. http://www.jesuisjeserai.stat.gouv.qc.ca/publications/bebe_v2no1_an.html

Knowlden, A. P., Sharma, M., Cottrell, R. R., Wilson, B. R. A., & Johnson, M. L. (2015). Impact evaluation of Enabling Mothers to Prevent Pediatric Obesity through Web-Based Education and Reciprocal Determinism (EMPOWER) Randomized Control Trial. *Health Education & Behavior : The Official Publication of the Society for Public Health Education*, *42*(2), 171–184. https://doi.org/10.1177/1090198114547816

Lenth, R. (2020). *emmeans* (Version 1.4.6) [Computer software]. R Foundation for Statistical Computing. https://CRAN.R-project.org/package=emmeans

Lin, B., Liew, J., & Perez, M. (2019). Measurement of self-regulation in early childhood: Relations between laboratory and performance-based measures of effortful control and executive functioning. *Early Childhood Research Quarterly*, *47*, 1–8. https://doi.org/10.1016/j.ecresq.2018.10.004

Lumeng, J. C., Miller, A. L., Horodynski, M. A., Brophy-Herb, H. E., Contreras, D., Lee, H., Sturza, J., Kaciroti, N., & Peterson, K. E. (2017). Improving Self-Regulation for Obesity Prevention in Head Start: A Randomized Controlled Trial. *PEDIATRICS*, *139*(5). https://doi.org/10.1542/peds.2016-2047

McClelland, M., Geldhof, J., Morrison, F., Gestsdóttir, S., Cameron, C., Bowers, E., Duckworth, A., Little, T., & Grammer, J. (2018). Self-Regulation. In N. Halfon, C. B. Forrest, R. M. Lerner, & E. M. Faustman (Eds.), *Handbook of Life Course Health Development* (pp. 275–298). Springer International Publishing. https://doi.org/10.1007/978-3-319-47143-3_12

Nigg, J. T. (2017). Annual Research Review: On the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk-taking, and inhibition for developmental psychopathology. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *58*(4), 361–383. https://doi.org/10.1111/jcpp.12675

Piche, G., Fitzpatrick, C., & Pagani, L. S. (2012). Kindergarten Self-Regulation As a Predictor of Body Mass Index and Sports Participation in Fourth Grade Students. *Mind, Brain, and Education*, *6*(1), 19–26.

Roth, R. M., Isquith, P. K., & Gioia, G. A. (2014). Assessment of Executive Functioning Using the Behavior Rating Inventory of Executive Function (BRIEF). In S. Goldstein & J. A. Naglieri (Eds.), *Handbook of Executive Functioning* (Vol. 67, pp. 301–331). Springer New York. https://doi.org/10.1007/978-1-4614-8106-5_18

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. *Wiley series in probability and mathematical statistics. Applied probability and statistics, 0271-6232*. Wiley.

Smithers, L. G., Sawyer, A. C. P., Chittleborough, C. R., Davies, N., Davey Smith, G., & Lynch, J. (2018). A systematic review and meta-analysis of effects of early life non-cognitive skills on academic, psychosocial, cognitive and health outcomes. *Nature Human Behaviour*, *2*(11), 867–880. https://doi.org/10.1038/s41562-018-0461-x

Tremblay, R. E., Loeber, R., Gagnon, C., Charlebois, P., Larivée, S., & LeBlanc, M. (1991). Disruptive boys with stable and unstable high fighting behavior patterns during junior elementary school. *Journal of Abnormal Child Psychology*, *19*(3), 285–300.

WHO. (2019). *Growth reference 5-19 years: BMI-for-age (5-19 years).* WHO. https://www.who.int/growthref/who2007_bmi_for_age/en/

6 Conclusions

6.1 Summary of findings

The prevalence of childhood obesity has risen alarmingly over the past 35 years (Di Cesare et al., 2019). Obesity and many of the health risks faced by Canadian children have some behavioural causes. Self-regulation (SR) abilities are believed to influence health behaviours (A. L. Miller et al., 2018). Early childhood care and education has many benefits for child development and women’s labour force participation. However, the evidence for effect of public childcare programs on adiposity and SR has been inconsistent, and the analytical methods employed have not been optimal to estimate the causal relations between dimensions of childcare arrangements and child outcomes. Likewise, the evidence for the causal effect of SR on adiposity was limited. Quebec's unique subsidized childcare policy (CPE) provided an excellent opportunity to assess whether universal preschool childcare had a beneficial impact on adiposity and SR, in particular for children from less advantaged families. However, taken together my thesis manuscripts and prior research show no compelling evidence that regulated childcare or a child’s SR abilities are major pathways to obesity prevention.

In my first manuscript, I studied how much the main types of preschool childcare in Quebec affected adiposity, measured as BMI z-scores and obesity, in kindergarten. I also estimated whether effects persisted through the elementary school years or differed between more and less advantaged children (lower SES, young maternal age, maternal depression symptoms, and biological father not living with the child and mother, and/or poor family functioning). I modelled the type, timing, and intensity of childcare exposure separately. Most prior studies had used coarse or vague exposure definitions, and the results could not be meta-analysed or generalized. I found that centre-based care, mostly regulated through the CPE program, did not lower mean BMI z-score or the risk of obesity in kindergarten compared to parental or unregulated childcare. The adiposity-raising effect of centre-based care was mainly driven by the effect in more advantaged children and it waned by grade 6. Among less advantaged children, childcare type made little difference. Adiposity was lowest when children had attended regulated home-based childcare. This is a novel result and the differential effect over time by family disadvantage needs further examination. This hypothesis—that there is a mediator between preschool childcare type and later adiposity—could be studied with the ELDEQ data.

In my second manuscript, I used a similar design to estimate how much childcare affected behavioural-emotional self-regulation (SR) in kindergarten. First, I estimated children's latent SR using a Bayesian generalized latent linear multilevel model (GLLMM), which is an alternative to classic item response theory that offers a lot of flexibility. I was able to use all of the behaviour data collected from teachers and mothers because the method did not require that all children have the same number of visits and ratings. I found CPE centre- and home-based care were equally associated with worse SR compared to informal care (unregulated home-based or parental care). This result was consistent with past research, but past studies had not evaluated CPE-regulated centre- and home-based separately. However, overall, SR was only slightly worse with CPE childcare compared to informal care (parental or unregulated home-based care). Children from less advantaged families had worse SR but they were neither especially protected nor harmed by regulated childcare. Whereas, more advantaged children had slightly worse outcomes later in elementary school had they had centre-based preschool childcare. That there were no differences in kindergarten suggests something in the intervening time.

In my third manuscript, I estimated how much SR deficits independently affected adiposity. Unlike the design of manuscript 1 and 2, the exposure and potential confounders were time-varying with respect the outcome; therefore, to avoid introducing a selection bias, I used inverse probability of 'treatment' weighting (IPTW) for control for confounding instead of regression adjustment. IPTW were derived from a covariate-balancing generalized propensity score model (Fong et al., 2018). Over a range of ages, poor SR in the three years prior to each adiposity measure did not predict BMI z-score or obesity status.

Features of the ELDEQ data that strengthened my research were: measured anthropometrics, frequent study visits (annual or biennial), representative sampling with serious efforts to recruit and retain families from sub-populations who tend to have low participation, and data from fathers. Data features that would strengthen future research are: continuation of annual visits after age 8 years, a much larger sample for precise sub-group effect estimates and to estimate cluster effects such as school and geographic areas, validation sub-studies for behaviour constructs if appropriate validated instruments are not available, and more data collected directly from birth fathers who not live with the child's mother. Population registries and supplementing study data with electronic medical records could extend study resources as is demonstrated by many European studies.

Across the manuscripts, I used a variety of methods to extract as much information as possible from the clustered (longitudinal) data, specifically multilevel generalized linear models in manuscripts 1 and 2, and general estimating equations in manuscript 3. Although measurement of SR needs to be clarified and validated against validated instruments or real-world suspected consequences of SR deficits, the use of a Bayesian multilevel generalized latent linear model to estimate plausible values of SR addressed some weaknesses of past research and was quite novel for an epidemiologic study.

I also made rigorous efforts to control for confounding and conserve representativeness of the results, which sometimes required a less conventional approach to bias-variance trade-offs. Potential confounders were neither dropped from the model when they were not statistically significant predictors of the outcome nor because of multicollinearity. The use of marginal estimates largely eliminates the potential problems of less parsimony and multicollinearity. That is, whereas the variance of the regression coefficient for the exposure tends to be inflated in the presence of multicollinearity, variance of the marginal estimate is not. (However, the conditions for this observation could be tested more formally.) Nevertheless, my variance estimates were probably conservative. My procedures were designed to capture multiple sources of uncertainty—from multiple imputation and self-regulation measurement error, as well as the usual estimation error. I believe the simulation-based approach to survey weighting, missing data, and outcome modelling was a good approach. Despite the long computation times for the chosen missing data imputation procedure, once completed, many studies could be conducted from the prepared data sets. Similarly, once Bayesian models were estimated, many numerical summaries, including valid credible intervals, could be calculated from the posterior predictive distribution for observed or set values of the independent variables.

Although the CPE program and other public childcare programs, generally, have not had a positive impact on adiposity and non-cognitive behaviour outcomes, the comprehensive early experiments such as Highscope/Perry schools did (Campbell et al., 2014). Perhaps a partnership between CPEs and primary care and social services would approximate the holistic services offered in the early experiments. For obesity prevention, however, Gortmaker et al. (2015) concluded preschool based interventions were not cost-effective compared to sugar taxes and school-lunch nutritional regulations. In addition, interventions that mainly aim to change individual behaviour without changing environmental influences are ethically dubious (Medvedyuk et al., 2018; Voigt et al., 2014; Wickins-Drazilova & Williams, 2011).

Supporting children's self-regulation development in preschool to foster long-term improved self-regulation, as well as to improve adiposity outcomes, was a reasonable hypothesis (Epstein & Anzman-Frasca, 2017). Unfortunately, 1) manuscript 2 and Canadian research suggests CPE childcare, at least in the early years of the program, did not improve self-regulation, and 2) preschool-based obesity prevention programs integrating self-regulation development interventions did not improve adiposity outcomes (Knowlden et al., 2015; Lumeng et al., 2017). Clearly, a better understanding of the causes of adiposity and self-regulation in Canadian children is important, and research should focus on, or at least include, other policy and environmental differences. However, there is little recent and representative longitudinal research data collected on Canadian children. Although there are regular repeated cross-sectional surveys such as the Quebec Survey of Child Development in Kindergarten, ELDEQ is one of the few sources of detailed data on children followed longitudinally that also includes measured anthropometric data. Canada-wide sources of adiposity data have been cross-sectional (Canadian Health Measures Survey) or parent-reported (National Longitudinal Survey of Children and Youth). More importantly, the CPE program has matured, and new longitudinal data sources are needed.

Although regulated childcare in Quebec was found to be generally beneficial (e.g. Laurin et al. 2016), it had some small negative impacts in its early years. As new public childcare programs are initiated, impact evaluation methods should be employed to periodically monitor the effects. Methods such as the lottery assignment scheme used in Head Start Impact Study (U.S. Department of Health and Human Services, Administration for Children and Families, 2010) seem feasible and would strengthen causal inference about specific features of childcare programs that cannot be disentangled in ecological estimates obtained from quasi-experimental studies such as pre/post-reform cross-provincial difference-in-differences.

In conclusion, my doctoral research contributed to the evidence about the effects of regulated childcare on child development, specifically that Quebec's CPE did not promote healthy weight gain and SR. Also, considering past studies and my results, the empirical case for childhood obesity prevention through self-regulation promotion is not compelling.

Appendix A. Ethics approval

McGill

Faculty of Medicine
3655 Promenade Sir William Osler #633
Montreal, QC H3G 1Y6

Faculté de médecine
3655, Promenade Sir William Osler #633
Montréal, QC H3G 1Y6

Fax/Télécopieur: (514) 398-3870
Tél/Tel: (514) 398-3124

CERTIFICATION OF ETHICAL ACCEPTABILITY FOR RESEARCH INVOLVING HUMAN SUBJECTS

The Faculty of Medicine Institutional Review Board (IRB) is a registered University IRB working under the published guidelines of the Tri-Council Policy Statement, in compliance with the Plan d'action ministériel en éthique de la recherche et en intégrité scientifique (MSSS, 1998), and the Food and Drugs Act (17 June 2001); and acts in accordance with the U.S. Code of Federal Regulations that govern research on human subjects. The IRB working procedures are consistent with internationally accepted principles of Good Clinical Practices.

At a full Board meeting on 11 November 2019, the Faculty of Medicine Institutional Review Board, consisting of:

Frances Aboud, PhD	Patricia Dobkin, PhD
Sally Mann, M.S.	Kathleen Montpetit, M.Sc.
Roberta Palmour, PhD	Lucille Panet-Raymond, BA
Alexandra Pasca, LL.M.	Maida Sewitch, PhD
Blossom Shaffer, MBA	Margaret Swaine, BA

Sylvia Villeneuve, PhD

Examined the research project **A10-B65-19A (19-10-028)** titled: *The influence of preschool childcare arrangements on adiposity and behavioural self-regulation in childhood: a population-based, longitudinal study in the era of Quebec's universal childcare policy*

As proposed by: Dr. Seungmi Yang (Applicant) to ____________ (Granting Agency, if any)

And consider the experimental procedures to be acceptable on ethical grounds for research involving human subjects.

11 November 2019	Roberta Palmour	
Date	Chair, IRB	Dean/Associate Dean

Institutional Review Board Assurance Number: FWA 00004545

Appendix B. Original variables and derived measures

B.1. Tables B1a-c. Data variables: master list for thesis manuscripts

Parts
a: Topic and source   b. parametrization   c. use in models

All the variables listed in the table were used in multiple imputation. Plus, the following auxiliary variables were reduced to principal components and included in the multiple imputation:
Notes:
[1] little or no data for separated fathers (bdadin = 0); assigned reference value or mean
[2] z-score derived from original 0-10 scaled sum score
[3] INSPQ = Institut national de santé publique du Québec, mtl = Montreal census metropolitan area, ocma = other census metropolitan areas, ge10k = municipality with ≥10,000 pop.
[4] ISQ-provided survey weight x estimated attrition weights; ISQ = Institut de la statistique du Québec
[5] natural log
[6] vacuum extraction was rare so grouped with vaginal delivery
[7] derived from amdeq06 (Alcohol during pregnancy), amdeq07 (Usual alcohol servings per episode), amdeq08d (Alcohol during all trimesters).
[8] derived from amdeq03 (Smoked during pregnancy) and amdeq05d (Smoked in all trimesters).
[9] crowded housing threshold of >2 persons per bedroom (PPB) based on U.S Department of Housing and Urban Development (2007)
[10] in manuscript propensity score model, weighted average of visit 2 and visit 10 parents' BMI. Weighted according to visit's proximity to measure.
[11] visit measure included if listed and if baseline (visit 1 or 2) or 3, **4 (priority)**, or 5 years prior to BMI outcome visit.
[12] other candidate variables for propensity score: contact with father in last 3 months; childcare includes afterschool care after visit 6; also from visit 7 on public or private school, active transport to or from school;

* Original variable (o) or derived by author(s) (v). Original variable definition may have included infrequent categories that we collapsed

Original variable naming scheme
character position: code
1: visit code ex. a = visit 1
2-3: topic code ex. hl = health
4: about who, e = target child, m = mother, j = father
5: variable type, q = survey question, d/s/t = derived by ISQ, v = derived by us
6-10: variable code (subtopic)

We did not consistently replicate this scheme when deriving new variables.

a. Topic and source

#	Variable label	Topic	About who	Source	Baseline collection visit(s)	Notes	Orig (o) or deriv (v) *
1	Father-s height (m)	anthropometry	father	f	2	[1]	o
2	Father-s weight (kg)	anthrop.	f	f	2	[1]	o
3	Father's BMI (centered on 24, by 5)	anthrop.	f	f	2	[1]	v
4	Mother-s height (m)	anthrop.	mother	m	2		o
5	Mother-s weight (kg)	anthrop.	m	m	2		o
6	Mother's BMI (centered on 24, by 5)	anthrop.	m	m	2		v
7	Birthweight for GA category	anthrop., perinatal	c[hild]	chart	1		v
8	Head circumference at birth (cm)	anthrop., perin.	c	chart	1		o
9	Child-s length at birth (cm)	anthrop., perin.	c	chart	1		o
10	Birthweight (kg, centered on 3.4kg)	anthrop., perin.	c	chart	1		o
11	Hyperactivity	behavior / temperament	c	m, f, t	2	[1] [2]	o
12	Inattention	behave. / temp.	c	m, f, t	2	[2]	o
13	Emotional troubles	behave. / temp.	c	m, f	2	[1] [2]	o
14	Anxiety	behave. / temp.	c	m, f	2	[1] [2]	o
15	Overall physical aggression	behave. / temp.	c	m, f, t	2	[1] [2]	o
16	Prosocial behaviour	behave. / temp.	c	m	2	[1] [2]	o
17	Opposition	behave. / temp.	c	m	2	[2]	o
18	Shyness	behave. / temp.	c	m, f	2	[1] [2]	o
19	Perception of child-s qualities	behave. / temp.	c	m, f	1	[1] [2]	o
20	Perception of difficult temperament	behave. / temp.	c	m, f	1, 2	[1] [2]	o
21	Perception of unpredictable temperament	behave. / temp.	c	m, f	2	[1] [2]	o
22	Sleeps through the night	behave. / temp.	c	m	1, 2		o
23	Childcare use	childcare	c	m	1, 1		v
24	Birth order	demogs	c	m	1		o
25	New sibling (by 17 mos)	demogs	c	m	2		v
26	Child's sex	demogs	c	m	1		o
27	Month of birth	demogs	c	m	1		v
28	Age (centered on 30 yrs, by 5 yrs)	demogs	m, f	m	1	[1]	o
29	A father in mother's home	demogs	f	m	1, 2		v

#	Variable label	Topic	About who	Source	Baseline collection visit(s)	Notes	Orig (o) or deriv (v) *
30	The father in mother's home is a stepfather	demogs	f	m	1, 2		v
31	Relation between mother and father	demogs	m-f	m	1, 2		o
32	Immigrant status	demogs	m, f	m	1	[1]	o
33	Years since immigration	demogs	m, f	m	1	[1]	o
34	Race or color: white	demogs	c, m, f	m	1		o
35	Religion	demogs	h	m	1		o
36	Freq. attend religious services	demogs	h	m	1		o
37	Languages in which father can converse	demogs	f	m	1	[1]	o
38	Language most spoken at home: French	demogs	h	m	1		o
39	Residence (4 cat.) – def. INSPQ	demogs	h	m	1, 2	[3]	o
40	Plans to have another baby	demogs	m	m	2		o
41	Survey weight (adjusted; centered on 0)	design				[4] [5]	v
42	Main activity currently	econ	m, f	m	1, 2	[1]	o
43	Worked in past 12 months	econ	m, f	m	1, 2		o
44	Main source of household income	econ	h	m	1, 2		o
45	Total household income	econ	h	m	1, 2		o
46	SES index (z-score)	econ	h	m	1, 2		o
47	Income sufficiency	econ	h	m	1, 2		o
48	Education, highest degree	educ	m, f	m	1		o
49	Child-s general health	health	c	m	1, 2		o
50	Last 12 months, freq. good health	health	c	m	1		o
51	Chronic health problems (child)	health	c				o
52	Times saw a health professional	health	c	m	1, 2	[5]	v
53	General health	health	m, f	m, f	2	[1]	o
54	Chronic health problems	health	m, f	m	1	[1]	o
55	Either parent has diabetes	health	h	m, f	2		v
56	Depression risk score	health	m	m	1, 2	[2]	o
57	Verbalisation (interviewer-rated)	parenting	m-c	i[nterviewer]	1, 2	[2]	o
58	Coercition (interviewer-rated)	parenting	m-c	i	2	[2]	o
59	Stimulation (interviewer-rated)	parenting	m-c	i	1, 2	[2]	o

#	Variable label	Topic	About who	Source	Baseline collection visit(s)	Notes	Orig (o) or deriv (v) *
60	Feeling of efficacy	parenting	m, f	m, f	1, 2	[2]	o
61	Perception of impact as a parent	parenting	m, f	m, f	1, 2	[2]	o
62	Coercive parenting	parenting	m, f	m, f	1, 2	[2]	o
63	Overprotective parenting	parenting	m, f	m, f	1, 2	[2]	o
64	Positive interactions	parenting	m, f	m, f	2	[2]	o
65	Type of delivery	perinatal	c	chart	1	[6]	o
66	Born premature (lt 37 weeks)	perinatal	c	chart	1		o
67	Neonatal cumulative risk index	perinatal	c	chart	1		o
68	Mother consumed alcohol during pregnancy	perinatal	c	m	1	[7]	v
69	Mother smoked during pregnancy	perinatal	c	m	1	[8]	v
70	Child-s health at birth	perinatal	c	m	2		o
71	Breastfeeding duration (mos)	perinatal	c	m	1-2		v
72	Family functioning	social-envir	h	m	1, 2	[1]	o
73	Alcohol is a source of tension	social-envir	h	m	1		o
74	Any smoking in the home	social-envir	h	m	2		o
75	Social support	social-envir	h	m	2	[1]	o
76	Homeowner	social-envir	h	m	1, 2		o
77	Subsidized housing	social-envir	h	m	1, 2		o
78	Housing needs repairs	social-envir	h	m	1, 2		o
79	Crowded housing (PPB > 2)	social-envir	h	m	1, 2	[9]	v
80	Family risk index over visits 1-2	social-envir	h	various	1-2		v
81	Family risk index over visits 1-2 (dichot)	social-envir	h	various	1-2		v

b. Parameterization

#	Variable label	Variable name	Variable class (q = bounded)	Factor levels (ref. cat. listed 1st)
1	Father-s height (m)	bhtjme	numeric	
2	Father-s weight (kg)	bwtjkg	numeric	
3	Father's BMI (centered on 24, by 5)	jbmic24_5	numeric	
4	Mother-s height (m)	bhtmme	numeric	
5	Mother-s weight (kg)	bwtmkg	numeric	
6	Mother's BMI (centered on 24, by 5)	mbmic24_5	numeric	
7	Birthweight for GA category	sga	factor	normal, SGA, LGA
8	Head circumference at birth (cm)	a0hlevhdcir_cm	numeric	
9	Child-s length at birth (cm)	a0htecm	numeric	
10	Birthweight (kg, centered on 3.4kg)	a0wtekg0	numeric	
11	Hyperactivity	bbeet0[1/3]a	numeric q	
12	Inattention	bbeet01b	numeric q	
13	Emotional troubles	bbeet0[1/3]c	numeric q	
14	Anxiety	bbeet0[1/3]d	numeric q	
15	Overall physical aggression	bbeet0[1/3]f	numeric q	
16	Prosocial behaviour	bbeet01g	numeric q	
17	Opposition	bbeet01h	numeric q	
18	Shyness	bbeet0[1/3]i	numeric q	
19	Perception of child-s qualities	apa[j/m]s06	numeric q	
20	Perception of difficult temperament	*tmet0[1/3]	numeric q	
21	Perception of unpredictable temperament	btmet0[1/3]a	numeric q	
22	Sleeps through the night	*qmmq0[7/5]	factor	yes, no
23	Childcare use	*crev1ab_cat	factor	none, P[art]T[ime], F[ull]T[ime]
24	Birth order	arged01	factor	1st, 2nd, ge3rd
25	New sibling (by 17 mos)	newsib	factor	no, yes
26	Child's sex	asexf	factor	boy, girl
27	Month of birth	smob	factor	oct (97), nov, dec, jan (98), mar, apr, may, jun, jul
28	Age (centered on 30 yrs, by 5 yrs)	[m/j]agec30_5y	numeric	
29	A father in mother's home	*dadin	factor	no, yes

#	Variable label	Variable name	Variable class (q = bounded)	Factor levels (ref. cat. listed 1st)
30	The father in mother's home is a stepfather	*stepdadin	factor	no, yes
31	Relation between mother and father	*re1v3	factor	married, notmarried
32	Immigrant status	asd[m/j]d1a	factor	CND[anadian]Born, EU[ropoean]imm, n[on]EUimm
33	Years since immigration	asd[m/j]d3a	factor	ge10, 5a9, lt5
34	Race or color: white	asd[e/m/j]q4aa	factor	yes, no
35	Religion	asdeq08	factor	none, Roman Catholic, other Christian, Jewish, Muslim, other
36	Freq. attend religious services	asdeq09	factor	never, 1-2py, 3-4py, 1+pm, 1+pw
37	Languages in which father can converse	asdjd05	factor	FRnENG, FRorENG, triling, FRorENGnOth
38	Language most spoken at home: French	asdmq6ab	factor	yes, no
39	Residence (4 cat.) – def. INSPQ	*gefd03	factor	mtl, ocma, ge10k, rural
40	Plans to have another baby	bqmmq04	factor	yes, no, refuse
41	Survey weight (adjusted; centered on 0)	swt1b0	numeric	
42	Main activity currently	*lf[m/j]d01	factor	remun[erated work], fam[ily care], famNremun, edu, oth, unemp
43	Worked in past 12 months	*lf[m/j]d1b	factor	yes, no
44	Main source of household income	*infd2a	factor	wage, self, welf, ei, oth
45	Total household income	*infd03		
46	SES index (z-score)	*infd09	numeric	
47	Income sufficiency	*infd05	factor	yes, no
48	Education, highest degree	aed[m/j]d02	factor	HS, noHS, col, uni
49	Child-s general health	*hleq01	factor	excel, vgood, good, fairpoor
50	Last 12 months, freq. good health	ahleq02	factor	always, often, [about] half [of the time], less [than half of the time]
51	Chronic health problems (child)	bhlev45i	factor	none, ge1
52	Times saw a health professional	*hlevhcvis	numeric	
53	Father-s general health	bhl[m/j]q01	factor	excel, very good, good, fair-poor
54	Chronic health problems	ahl[m/j]v1aq	factor	none, ge 1
55	Either parent has diabetes	pardiab	factor	no, yes
56	Depression risk score	*dpmt01	numeric	
57	Verbalisation (interviewer-rated)	*ifft01a	numeric	
58	Coercition (interviewer-rated)	bifft01b	numeric	
59	Stimulation (interviewer-rated)	*ifft01c	numeric	

#	Variable label	Variable name	Variable class (q = bounded)	Factor levels (ref. cat. listed 1st)
60	Feeling of efficacy	*pa[m/j]l01	numeric	
61	Perception of impact as a parent	*pa[m/j]l02	numeric	
62	Coercive parenting	*pa[m/j]l03	numeric	
63	Overprotective parenting	*pa[m/j]l05	numeric	
64	Positive interactions	bpret0[1/3]	numeric	
65	Type of delivery	admmta	factor	vag, csec-p[rimaire], csec-i[terative]
66	Born premature (lt 37 weeks)	prem	factor	no, yes
67	Neonatal cumulative risk index	admeicrn	numeric	
68	Mother consumed alcohol during preg.	prgalc	factor	none, some, a lot
69	Mother smoked during pregnancy	prgsmk	factor	none, some, yes
70	Child-s health at birth	bmdeq22	factor	excel, vgood, good, fairpoor
71	Breastfeeding duration (mos)	bfdur	factor	never, 0-6, 6-12, 12+
72	Family functioning	*fnft01	numeric	
73	Alcohol is a source of tension	afnfq01m	factor	completely disagree, disagree, (completely) agree
74	Any smoking in the home	bhlfv2a	factor	no, one [parent or other adult], both [parents]
75	Social support	bsuft01	numeric q	
76	Homeowner	*hhfq01		yes, no
77	Subsidized housing	*hhfq02a		no [or is homeowner], yes
78	Housing needs repairs	*hhfq02b	factor	normal maintenance, minor, major
79	Crowded housing (PPB > 2)	*hhfvppb_cat	factor	le2, gt2
80	Family risk index over visits 1-2	risksum	numeric q	
81	Family risk index over visits 1-2 (dichot)	riskbin	factor	lo [ge median], hi [gt median]

c. Variable used in analytic step…

#	**Variable label**	**Retention probability model** † = decide a priori, forced into model	**Manu. 1 final model**	**Manu. 2 final model**	**Manu 3. IPTW** [11][12]
1	Father-s height (m)	n	n	n	n
2	Father-s weight (kg)	n	n	n	n
3	Father’s BMI (centered on 24, by 5)	y (2 †)	y	y	y (2, [10])
4	Mother-s height (m)	n	n	n	n
5	Mother-s weight (kg)	n	n	n	n
6	Mother’s BMI (centered on 24, by 5)	y (2 †)	y	y	y (2, [10])
7	Birthweight for GA category	y †	y	y	y
8	Head circumference at birth (cm)	n	n	y	n
9	Child-s length at birth (cm)	n	n	n	n
10	Birthweight (kg, centered on 3.4kg)	y †	y	n	y
11	Hyperactivity	y (m2 †)	y (m)	y (m)	n (only from visit 2 in sensitivity analysis)
12	Inattention	n	n	y (m)	n
13	Emotional troubles	n	n	n	n
14	Anxiety	n	n	n	n
15	Overall physical aggression	n	n	y (m)	n
16	Prosocial behaviour	n	n	n	n
17	Opposition	n	n	n	n
18	Shyness	n	n	n	n
19	Perception of child-s qualities	n	n	n	n
20	Perception of difficult temperament	y (1)	y (m)	n	n
21	Perception of unpredictable temperament	n	n	n	y (m)
22	Sleeps through the night	y (2)	n	n	y (2)
23	Childcare use	y (1-5 †)	y	y	y (1, 3, 4, 5, 6, 7, 8, 10)
24	Birth order	y †	y	y	y
25	New sibling (by 17 mos)	y †	n	y	y (4, 5, 6, 7, 8, 10)
26	Child’s sex	y †	y	y	y
27	Month of birth	n	y	y	y
28	Age (centered on 30 yrs, by 5 yrs)	y (m)	y (m, f)	y (m, f)	y

#	Variable label	**Retention probability model** † = decide a priori, forced into model	**Manu. 1 final model**	**Manu. 2 final model**	**Manu 3. IPTW** [11][12]
29	A father in mother's home	y † (1, 2)	y (2)	y (1, 2)	y (2, 4, 5, 6, 7, 8, 10)
30	The father in mother's home is a stepfather	y † (1, 2)	y (2)	y (1, 2)	y (2, 4, 5, 6, 7, 8, 10)
31	Relation between mother and father	n	y (1)	n	n
32	Immigrant status	y (m †, f)	y (m, f)	y (m, f)	y (m)
33	Years since immigration	y (m †, f)	y (m, f)	y (m, f)	n
34	Race or color: white	y (c †, m, f)	y (m)	y (c)	n
35	Religion	y †	n	n	y
36	Freq. attend religious services	y †	n	n	y
37	Languages in which father can converse	y	y	n	n
38	Language most spoken at home: French	y †	y	n	n
39	Residence (4 cat.) – def. INSPQ	y (1, 2)	y (2)	y (2)	y (2, 4, 5, 6, 7, 8, 10)
40	Plans to have another baby	y	n	n	n
41	Survey weight (adjusted; centered on 0)	y	y	y	n
42	Main activity currently	y (2)	y (f1)	n	y (2, 3, 4, 5, 6, 7, 8, 10)
43	Worked in past 12 months	y (m1	y (m1, m2)	y (m1, m2)	y (3, 4, 5, 6, 7, 8, 10)
44	Main source of household income	y (1, 2)	y (2)	n	n
45	Total household income	n	n	n	n
46	SES index (z-score)	y (1, 2)	y (1, 2	y (2)	y (4, 5, 6, 7, 8, 10)
47	Income sufficiency	y (1, 2)	n	n	y (2,)
48	Education, highest degree	y (m †, f)	y (m, f)	y (m)	y (m, f)
49	Child-s general health	y † (1, 2)	y	y (2)	y (1, 3, 4, 5, 6, 7, 8, 10)
50	Last 12 months, freq. good health	n	n	n	n
51	Chronic health problems (child)	y (1)	y	n	n
52	Times saw a health professional	y (1, 2)	n	n	n
53	General health	y (m1 †, m2 †)	y (f2, m2)	y (m2)	y (m4, m5, m6, m7, m8, m10, f3, f4, f5, f6, f7, f8)
54	Chronic health problems	n	n	n	n
55	Either parent has diabetes	y †	y	n	y
56	Depression risk score	y (m1 †, m2 †)	y (1, 2)	y (1)	y (3, 4, 5, 6, 7, 8
57	Verbalisation (interviewer-rated)	y (1)	n	n	n

#	Variable label	**Retention probability model** † = decide a priori, forced into model	**Manu. 1 final model**	**Manu. 2 final model**	**Manu 3. IPTW** [11][12]
58	Coercition (interviewer-rated)	y	n	n	n
59	Stimulation (interviewer-rated)	n	n	n	n
60	Feeling of efficacy	n	n	n	n
61	Perception of impact as a parent	n	n	n	n
62	Coercive parenting	y (m2)	n	n	n
63	Overprotective parenting	y (m2)	y (m2)	y (m2)	y (m2)
64	Positive interactions	y (m)	n	n	n
65	Type of delivery	y	y	n	n
66	Born premature (lt 37 weeks)	y	y	y	n
67	Neonatal cumulative risk index	n	n	n	n
68	Mother consumed alcohol during pregnancy	y	y	y	n
69	Mother smoked during pregnancy	y	y	n	y
70	Child-s health at birth	y	y	n	y
71	Breastfeeding duration (mos)	y †	y	y	y
72	Family functioning	y (1, 2 †)	y (2)	y (2)	y (2, 4, 5, 6, 7, 8, 10)
73	Alcohol is a source of tension	n	n	n	n
74	Any smoking in the home	y †	y	y	y (4, 5, 6, 7, 8, 10)
75	Social support	y (2)	y (2)	y (1, 2)	n
76	Homeowner	y (1)	n	n	y (4, 5, 6, 7, 8, 10)
77	Subsidized housing	y (1)	n	n	n
78	Housing needs repairs	y (1)	y (1)	y (2)	y (2)
79	Crowded housing (PPB > 2)	y (2 †)	y (2)	n	y (2)
80	Family risk index over visits 1-2	n	y	y	n
81	Family risk index over visits 1-2 (dichot)	n	n	n	n

B.2 BMI standard curves and cut-offs

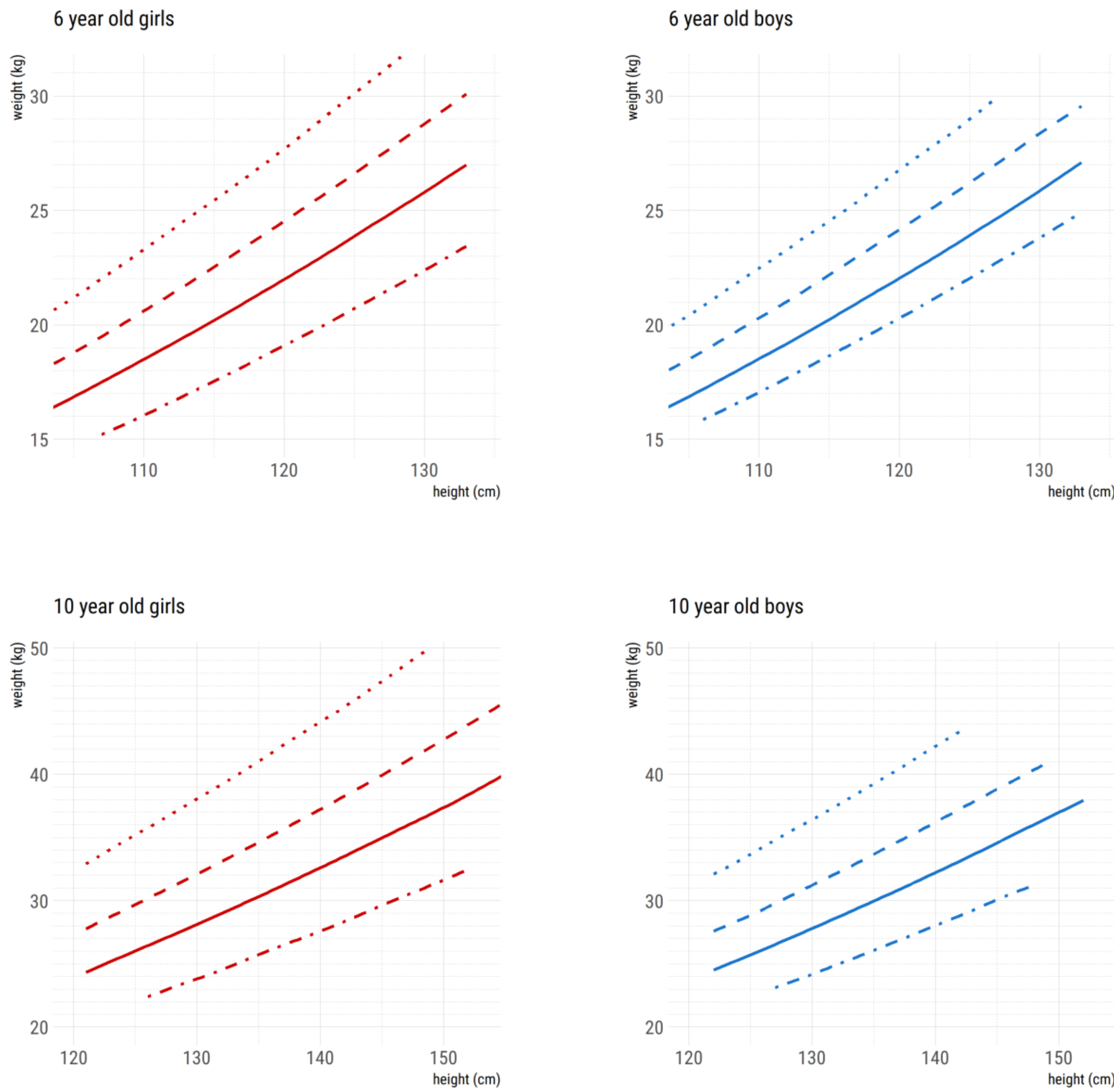


Figure B.1. Relation between weight, height, and WHO z-score cut-offs. Each panel shows the weight and height at age- and sex- standardized reference BMI z-score values (BMI in parentheses). Line, from bottom: dot-dash = -1.5 SD (lightly undernourished), solid = 0 SD (normal), dash = 1 SD (overnourished), dotted = 2 SD (obese). Healthier patterns are those in normal ranges (z-scores between -1.5 to 1). A child who stays “on his or her curve” over time would have a slope of zero in BMIz units.

Table B.2. BMI cut-offs for obesity by different health organization standards. WHO = World Health Organization (WHO, 2019), IOTF = International Obesity Task Force (Cole et al., 2000), CDC = Centers for Disease Control (US) (Vogel, 2019)

	Boys			Girls		
Age	WHO	IOTF	CDC	WHO	IOTF	CDC
6	18.4	19.8	18.4	19.1	19.7	18.8
7	18.9	20.6	19.1	19.7	20.5	19.6
8	19.6	21.6	20.0	20.4	21.6	20.6
9	20.3	22.7	21.0	21.4	22.8	21.7
10	21.3	24.0	22.0	22.4	24.1	22.9
11	22.3	25.1	23.1	23.5	25.4	24.1
12	23.4	26.0	24.1	24.8	26.7	25.2
13	24.6	26.8	25.1	26.0	27.8	26.2

B.3. Self-regulation latent score

Self-regulation was measured as a latent score for poor self-regulation (PSR). The raw collected data were mother and teacher ratings on 14 Likert scale items. The questions asked respondents to rate the frequency that the child demonstrated a behaviour; the choices were "never", "sometimes" and "often" ("don't know" was set to missing).

The items were chosen based on definitions of self-regulation in the literature and behaviour ratings used in recent studies. Piché et al. (2012) measured 'behavioural regulation' in ELDEQ participants using teacher assessments of emotional distress, physical aggression, and impulsivity (includes hyperactivity and inattention). Those emotional distress items corresponded to internalizing behaviours in other Canadian (or Quebec) studies, and the physical aggression items were the three proactive physical aggression items from the larger 9-item aggression scale. The BRIEF-2 measures self-regulation with the inhibitory and emotional control, self-monitor, and shift subscales (Jacobson et al., 2016), similar to the inhibitory, emotional control, and shift subscales of the first BRIEF (Roth et al., 2014). These do not include proactive aggression, inattention, or internalizing emotional problems, but some items for anger, moodiness, and aggressive or violent reactions to small insults. Moffitt et al. (2011) and Howard (Howard & Williams, 2018) include aggression, but not internalizing emotional difficulties. They also include (lack of) persistence and inattention. Given those behaviour rating examples and recent definitions and theories, we used the same nine 'impulsivity' items as Piché et al. (2012), but we used the item group labels, "inattention" and "hyperactivity", as also appears in the original data.

We did not the emotional distress or proactive aggression items. Instead we used the four reactive aggression items of the overall aggression scale plus an item on tantrums.

This appendix first summarizes the raw data by item, rater, and study visit. Self-regulation is probably a multidimensional construct, but it is used as a single construct in many qualitative discussions and empirical studies. Our goal was not to describe the psychometric properties of our measure of self-regulation, in detail. However, we did test the empirical support for a general factor and consistency between mother and teacher ratings. Sub-section 2 describes the latent regression model used to estimate the PSR score used in manuscripts 2 and 3.

B.3.1. Description of the original behaviour ratings

Table B.3. Number of children with a rater-visit

Visit	Mother ratings	Teacher ratings
4-year	1625 [1]	NA
5-year	1548 [1]	NA
6-year	1434	948
7-year	NA	1299
8-year	1427 [2]	1266
10-year	NA	983
12-year	NA	993

1. 'Tantrums' item not asked
2. Physical aggression items not asked

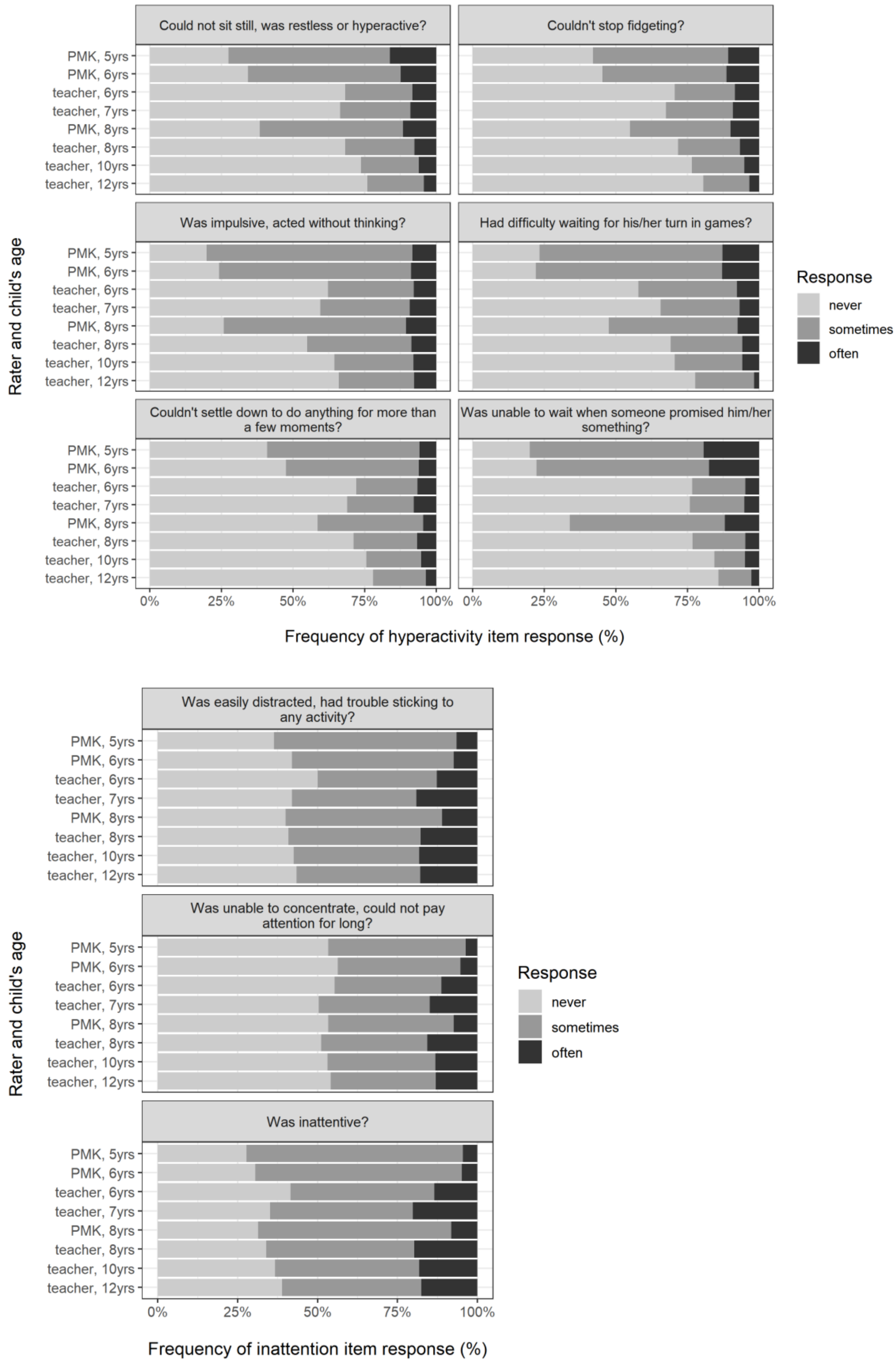


Figure B.2. Response proportions by behaviour item, visit, and rater.

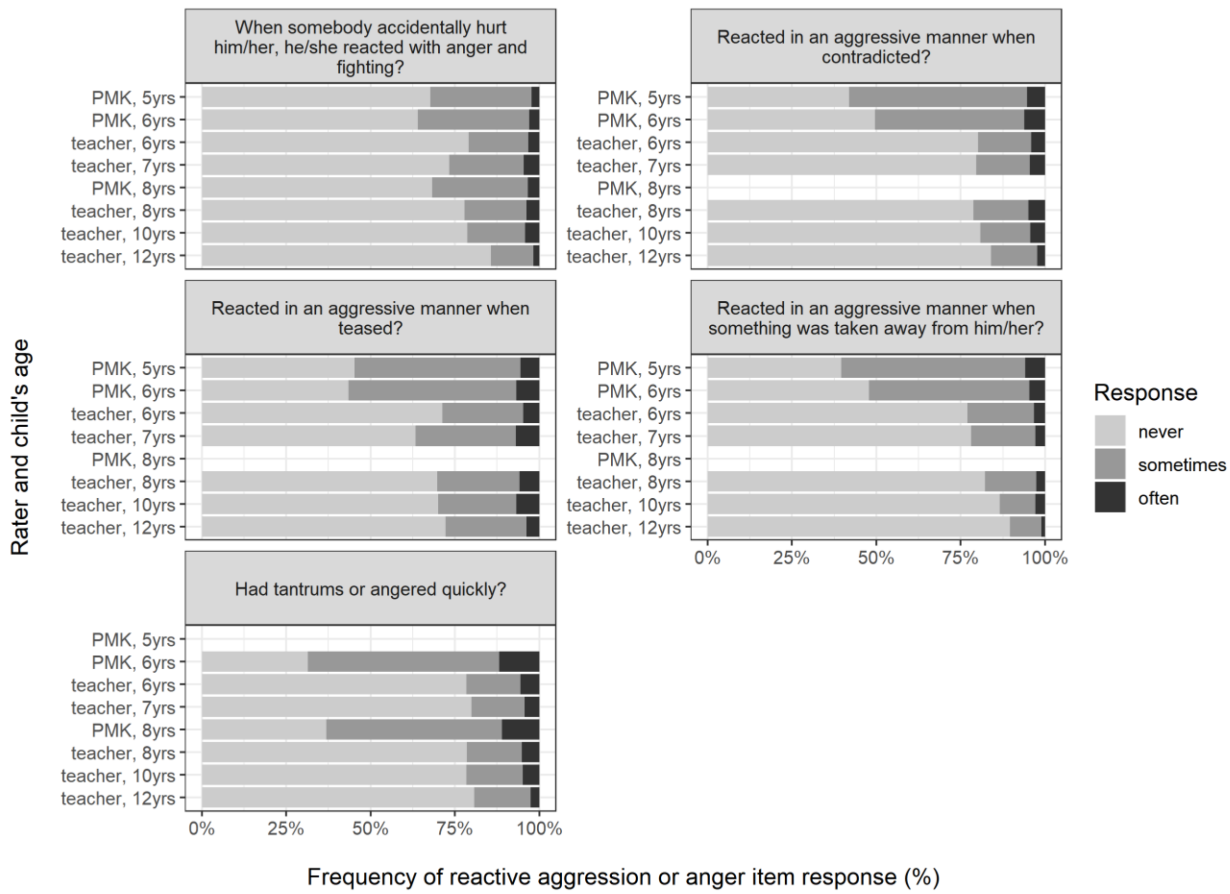


Figure B.2 (continued). Response proportions by behaviour item, visit, and rater.

Bifactor analysis suggested that the correlations between these items/subscales can be explained by a general factor. For example, Figure B.3 shows the bifactor analysis results for teacher ratings at the 8-year-old visit.

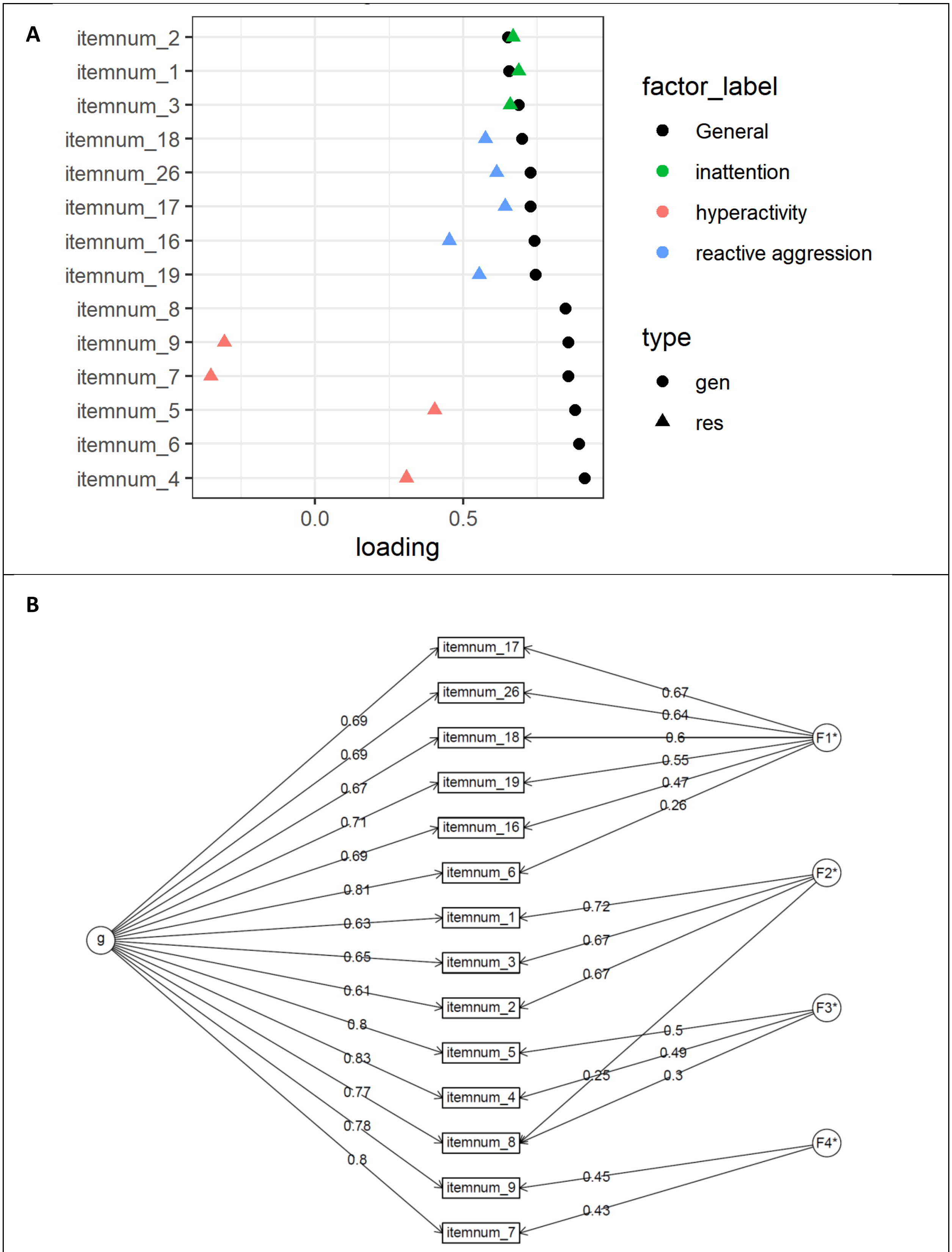


Figure B.3 Bifactor analysis for teacher ratings at the 8-year-old visit. A. Factor loadings B. Structural model.

B.3.2. Latent PSR score estimation

Latent variable models have traditionally been estimated using structural equations (SEM) or item response theory (IRT). However, generalized (latent) linear mixed models can also be used. Common advantages of GLLMM for epidemiology research include ease of use with unbalanced data (different number of time-points or items for different subjects, by design or due to missing data), and a regression framework, which is usually the ‘bread and butter’ of quantitative epidemiologic analysis. De Boek et al. (2011) provides a good introduction to the equivalence between IRT models and GLLMM with R code. Rabe-Hesketh and Skrondal describe GLLMM in more theoretical depth, including polytomous responses, and Stata code (2012, 2016). Burkner’s ‘brms’ package (Bürkner, 2017) for R (R Foundation for Statistical Computing, Vienna, Austria) provides a very flexible interface to Bayesian GLLMM in Stan (Stan Development Team, 2018), but scales the latent response slightly differently than classic IRT). The brms vignette provides an accessible introduction (Bürkner, 2019).

While sum scores are generally fine for descriptive statistics, as an outcome in a statistical model, they often have poor statistical or distributional properties. There is often a floor or ceiling, and discrete values may not be truly interval-scaled. Gorter et al (2015) also show how in repeated measures data, sum scores underestimate between-subject variability and overestimate within-subject variability. They recommend deriving the latent score (e.g. a value from a standard normal distribution) from a latent variable model. The GLLMM approach also facilitates use of data from multiple raters; however, we are not aware of any publications validating our specification of rater effects.

The ‘true’ latent poor self-regulation trait was assumed to have a standard logistic distribution. The link function and item random effect accounts for the multiple ordinal items. There are multiple rows per child-visit; therefore, ‘random’ effects for child and visit were estimated. Together, they represent a child’s ‘true’ deviation from the population mean at visit t, independent of the population-level effects for sex and age, and the child’s probability of participating (i.e., latent score is as if each child, at each visit, had a an equal probability of participating). Visits are nested within child.

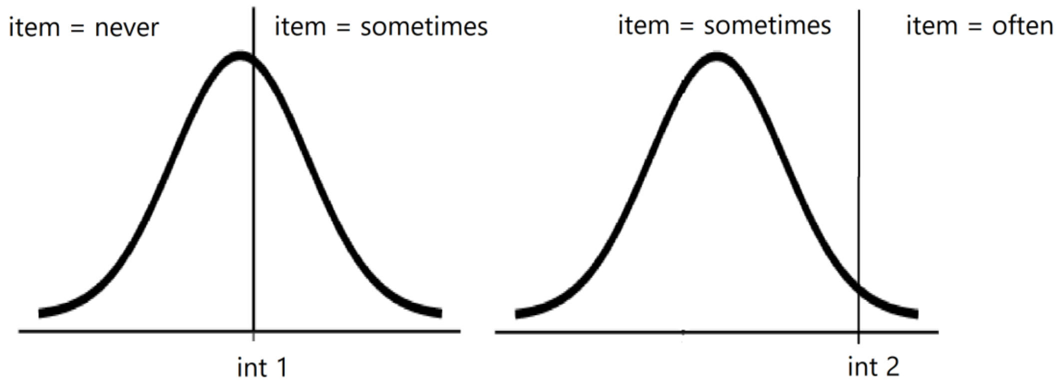


Figure B.4. Latent trait diagram for with thresholds for 3-level response (adjacent category model)

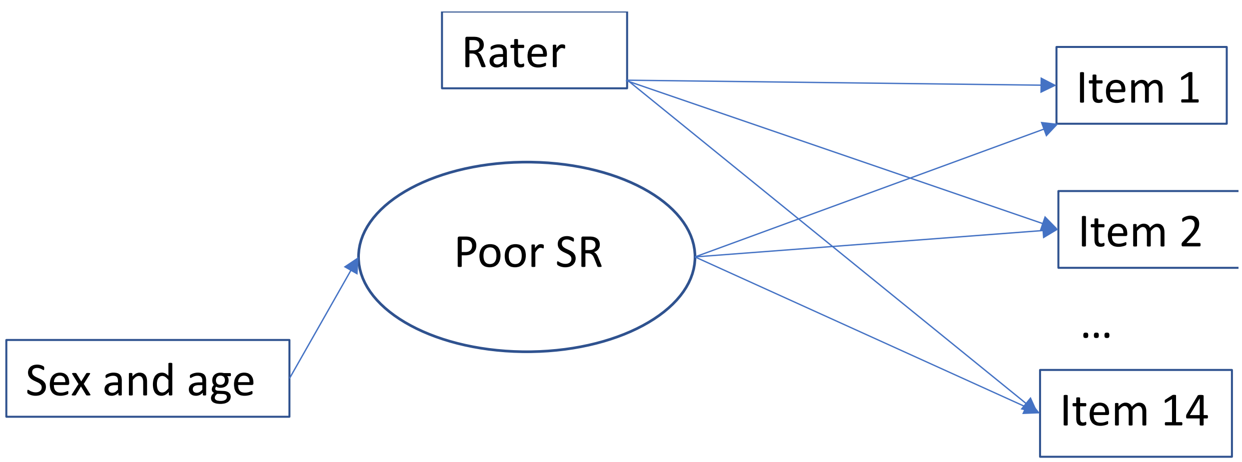


Figure B.5. Structural model for latent poor self-regulation (simplified)

Child (and nested child-visit) effects are crossed with items effects. The item random intercepts represent how much each item's intercept deviated from the mean intercept for all items. We used the adjacent category logit parametrization, which is equivalent to the generalized partial credit model.

The model estimated that, on average, teachers switched from a response of "never" to "sometimes" at 2.0 SD on the latent curve, and from "sometimes" and "often" at 4.8 SD on the latent curve. Whereas mothers switch their responses to a level higher at an average of 1 SD higher on the latent curve. However, there was considerable variation across items as seen by the 0.94 SD for the item intercept. The item-specific intercepts are inversely proportional to the 'difficulty' parameter in IRT. Bürkner calls it 'easiness' and it represents how much the latent curve would shift if measured by that item only (minus any shrinkage). Therefore, a less endorsed item such as “Had temper tantrums or hot temper” (item 26) is shifted left compared to the “true” curve.

The R code and summary output is shown below, followed by Figure B.6 showing the implied mean (bias) and variance of the items (i.e. linear combinations). Finally, the posterior distribution of population level parameters and the MCMC trace plots are shown in Figure B.6.

```
# The prior distributions:
prior.ord.2pl.mult5 <-
  prior("normal(1, 2)", class = "sd", group = "itemf") +
  prior("normal(0.2, 0.5)", class = "sd", group = "itemf",
        dpar = "disc") +
  prior("lkj(4)", class = "cor") +
  prior("constant(1)", class = "sd", group = "idme") +
  prior("normal(1, 1)", class = "sd", group = "idme:vis") +
  prior("normal(0, 2)", class = "Intercept", coef = "1") +
  prior("normal(4, 2)", class = "Intercept", coef = "2") +
  prior("normal(0, 2)", class = "b")

# the multilevel ordinal logistic model
   mpt6 <- brm(
     bf(
       scoref ~ ageyc7 * asexf + raterf +
         s1svywt0 + s2svywt0 + s3svywt0 + s4svywt0 +
         (1 + raterf |i| itemf) + (1 | idme/vis),
         disc ~ 1 + (1 |i| itemf)
     ), data = dbe3,
     family = brmsfamily("acat", "logit"),
     prior = prior.ord.2pl.mult5, cores = 4, iter = 4000,
     future = TRUE, control = list(adapt_delta = 0.9)
     )
summary(mpt6)

   Family: acat
     Links: mu = logit; disc = log
   Formula: scoref ~ ageyc7 * asexf + raterf + s1svywt0 + s2svywt0 + s3svywt0 +
   s4svywt0 + (1 + raterf | i | itemf) + (1 | idme/vis)
            disc ~ 1 + (1 | i | itemf)
      Data: dbe3 (Number of observations: 150198)
   Samples: 4 chains, each with iter = 4000; warmup = 2000; thin = 1;
            total post-warmup samples = 8000

   Group-Level Effects:
   ~idme (Number of levels: 1657)
                 Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
   sd(Intercept)     1.00      0.00     1.00     1.00 1.00     8000     8000

   ~idme:vis (Number of levels: 9382)
                 Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
   sd(Intercept)     0.97      0.02     0.92     1.02 1.00     1859     3505

   ~itemf (Number of levels: 14)
                                Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
```

```
sd(Intercept)                      0.94      0.18      0.67      1.37 1.00      2103
sd(raterfpmk)                      1.08      0.22      0.75      1.61 1.00      2194
sd(disc_Intercept)                 0.19      0.04      0.13      0.28 1.00      2646
cor(Intercept,raterfpmk)          -0.57      0.17     -0.83     -0.18 1.00      2251
cor(Intercept,disc_Intercept)      0.40      0.19     -0.02      0.72 1.00      3178
cor(raterfpmk,disc_Intercept)     -0.13      0.22     -0.53      0.31 1.00      4064
                               Tail_ESS
sd(Intercept)                      3433
sd(raterfpmk)                      3415
sd(disc_Intercept)                 3889
cor(Intercept,raterfpmk)           3505
cor(Intercept,disc_Intercept)      4496
cor(raterfpmk,disc_Intercept)      4435

Population-Level Effects:
               Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept[1]       2.00      0.33     1.37     2.64 1.00     1296     2413
Intercept[2]       4.77      0.34     4.13     5.44 1.00     1318     2504
disc_Intercept    -0.01      0.06    -0.12     0.10 1.00     1626     2879
ageyc7            -0.01      0.01    -0.03     0.00 1.00     3413     4977
asexf             -0.81      0.06    -0.92    -0.70 1.00     1805     3408
raterfpmk          1.04      0.30     0.45     1.62 1.00     1118     2308
s1svywt0           1.51      0.19     1.13     1.89 1.00     2627     3921
s2svywt0           1.23      0.15     0.94     1.53 1.00     2875     4304
s3svywt0           2.92      0.42     2.10     3.76 1.00     2756     3877
s4svywt0           0.73      0.24     0.25     1.20 1.00     4566     5630
ageyc7:asexf      -0.16      0.01    -0.18    -0.14 1.00     3102     4736
```

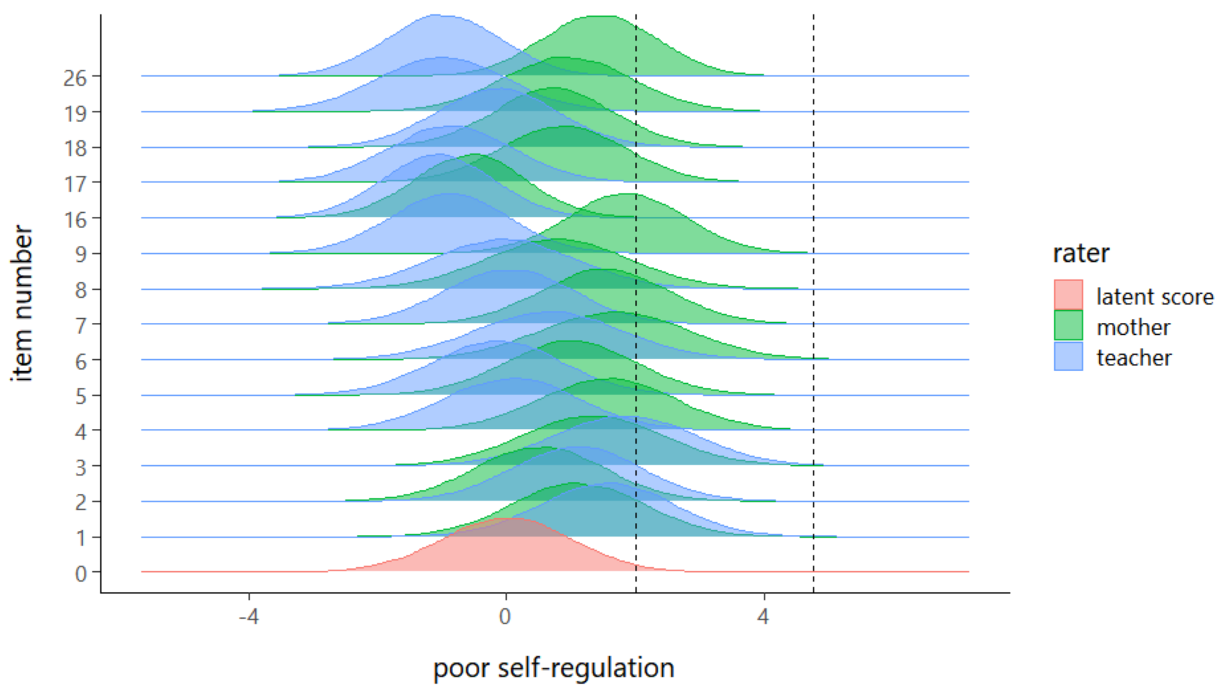


Figure B6. The item-specific marginal mean (bias) and variance compared to the latent model, by rater type

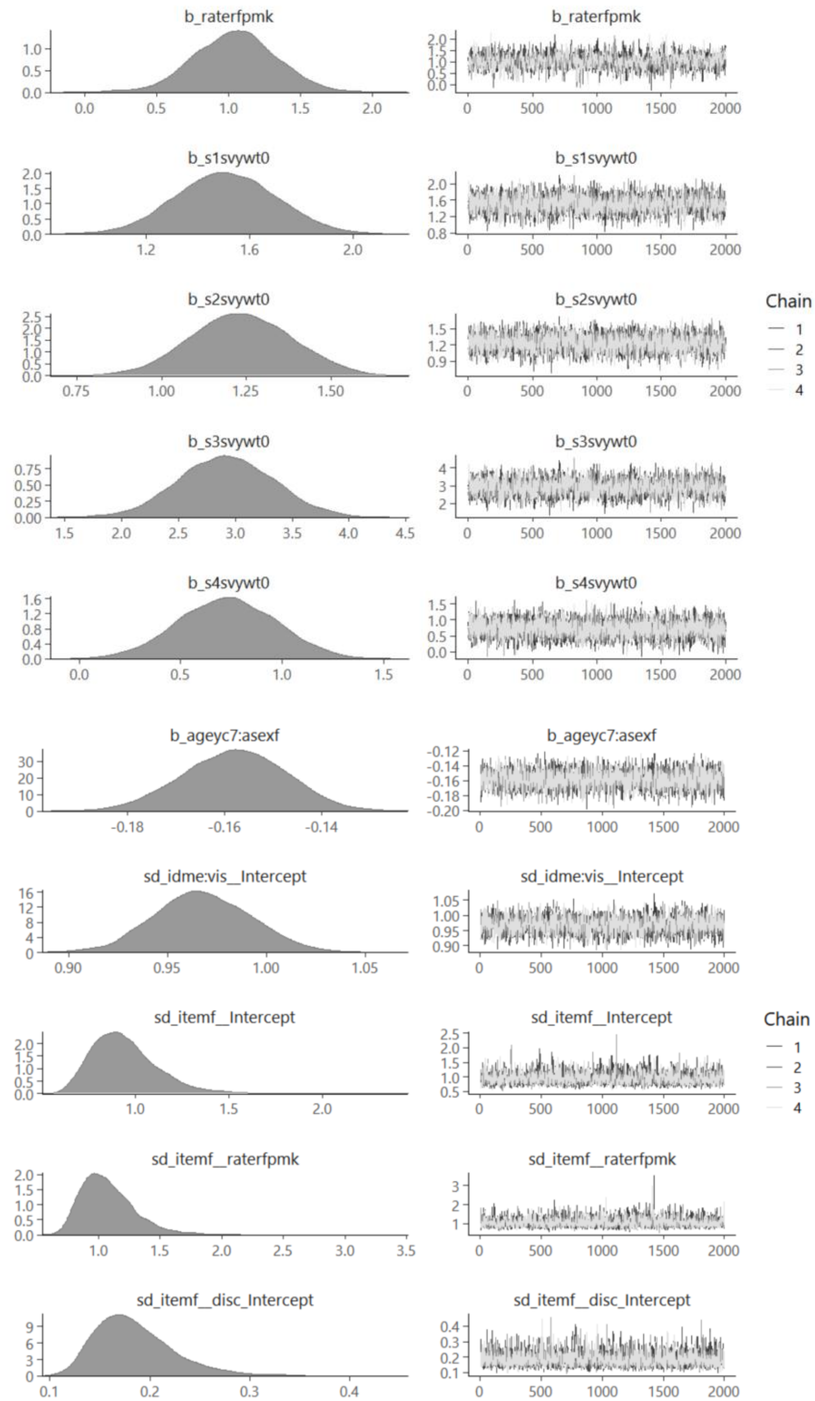


Figure B.7. Posterior distribution and trace plots for latent poor self-regulation model. Posterior distributions show that the parameters were not overly influenced by the priors (no truncation) and trace plots show that the MCMC chains mixed well.

I also checked for evidence that mothers and teachers were measuring different constructs. I compared fit for the behavioural self-regulation model with teachers and mothers, separately and together, using WAIC as a comparative stat.

I restricted the data to children who had both teacher and mother ratings in kindergarten (n = 938; same children in each model) and ran 4 models:

- Mother-ratings only   p_waic = 811 (SE = 8.7)
- Teacher-ratings only p_waic = 748 (SE = 12)
- Mother and teacher ratings but randomly sampled the teacher's set OR the mother's set so 1 set of ratings per child p_waic = 817 (SE = 11)

The fit was similar in separate and combined rater models.

B.3.3. Latent PSR scores by ADHD diagnosis and stimulant use

"In the following questions, long-term conditions refer to conditions that have lasted or are expected to last 6 months or more and have been diagnosed by a health professional (a doctor). Does [name] have any of the following long-term conditions?… Attention deficit disorder (with ot without hyperactivity)?"

"In the past 12 months, does [name] take any of the following prescribed medication on a regular basis:…Ritalin or any other medication that treat hyperactivity or inattention?"

"Does [name] still take Ritalin or any other medication that treat hyperactivity or inattention?"

At the 5-year-old visit, 109/1657 children had no data about ADHD diagnosis, 6/1542 (0.4%) PMKs reported that a health professional had diagnosed the child with ADHD. Data for 6 to 12 years are summarized in the next table.

Table B.4. Prevalence of ADHD indicators by study visit

Year	Missing (%)	ADHD, ever Dx (%)	Current use (%)	Used in past 12 months (%)
6	13.6	1.7	1.1	1.3
7	10.0	NA	3.3	3.5
8	13.9	4.8	5.0	5.4
10	21.2	7.1	7.7	8.2
12	17.4	9.6	9.2	9.9

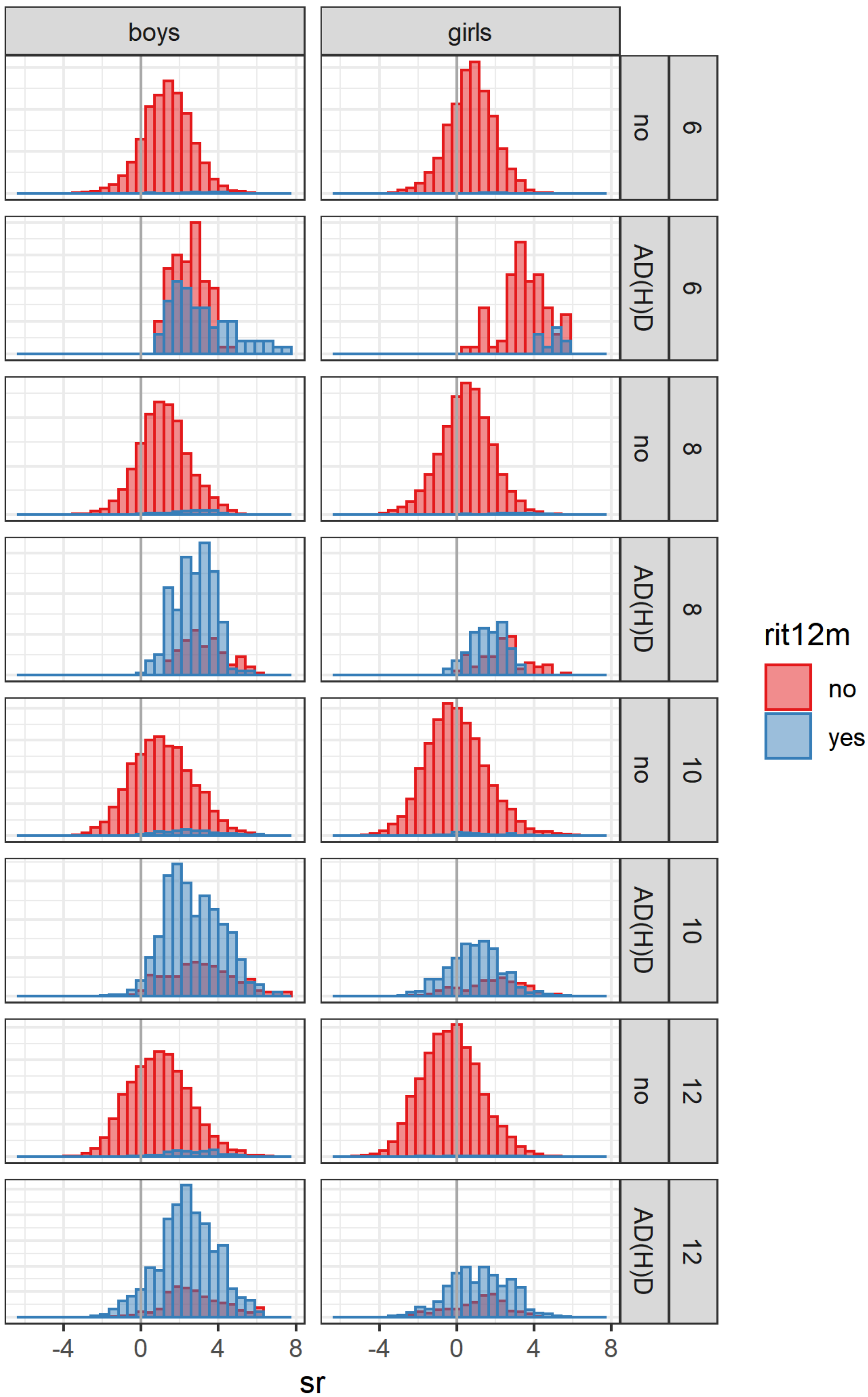


Figure B.8. Distribution of poor self-regulation scores by sex, age, AD(H)D diagnosis and stimulant use in the last 12 months.

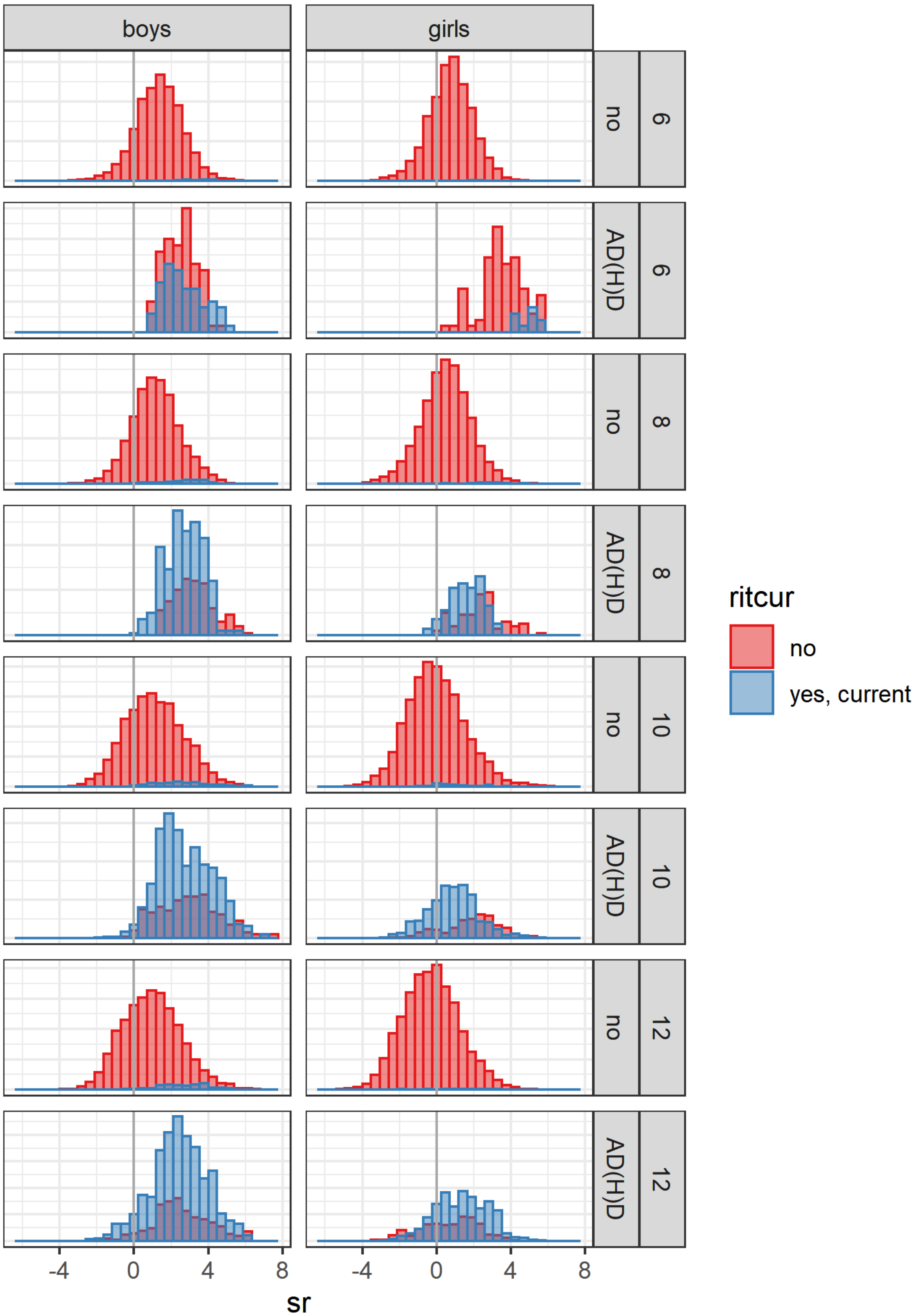


Figure B.9. Distribution of poor self-regulation scores by sex, age, AD(H)D diagnosis and current stimulant use.

B.4. Family functioning scale

Reproduced from the visit 1 QIRI ELDEQ questionnaire, pp 28-33.

https://www.jesuisjeserai.stat.gouv.qc.ca/informations_chercheurs/outils_collecte/E01_QIRI.pdf

[For consultation only]

The following statements are about families and family relationships. For each one, please indicate which response best describes your family: strongly agree, agree, disagree or strongly disagree.

1. *Planning family activities is difficult because we misunderstand each other.*
2. *In times of crisis we can turn to each other for support.*
3. *We cannot talk to each other about sadness we feel.*
4. *Individuals (in the family) are accepted for what they are.*
5. *We avoid discussing our fears or concerns.*
6. *We express feelings to each other.*
7. *There are lots of bad feelings in our family.*
8. *We feel accepted for what we are.*
9. *Making decisions is a problem for our family.*
10. *We are able to make decisions about how to solve problems.*
11. *We don't get along well together.*
12. *We confide in each other.*

Appendix C. Survey weights and multiple imputation of missing data

Adjusted regression weights

To adjust for possible effect heterogeneity that differs between the study and source populations, we adjusted estimates for survey non-response, and attrition in this study. A two-step adjustment to the survey weights that were provided by l'Institut de la statistique du Québec (ISQ) was used to standardize descriptive summaries and effect estimates to the target population. The target population was all singleton Quebec children, excluding First Nations and Inuit children (unfortunately, due to lack of data), born in 1997-98 who were free of serious illnesses that would preclude use of *Centres de la petite enfance childcare* (CPE), or parental care, only.

We estimated the retention weights as follows: The first weight was the inverse probability of eligible participants (n = 2091) having follow-up date (n = 1657) versus not (n = 434). The retention probability was estimated with logistic regression model including visit 1 variables as candidate predictor variables. Estimation and variable selection were performed simultaneously using elastic net (R package glmnet). The original survey weight was multiplied by the inverse of the retention probabilities.

Missing data imputation

Item-wise missing independent variables were multiply imputed 50 times (m = 50) using sequential regression and IVEware software, version 0.3 (Raghunathan et al., 2016). We prepared the imputed sets for multiple studies; therefore, possible predictors of childcare, and anthropometric and behavior scores in elementary school (i.e. outcome variables for multiple studies) from visits 1 to 14, were candidate variables for the regression equations. These included baseline variables listed in Web Table 1 and their updated values (when available). The sequential regression algorithm screens all listed predictors for variables that best explain the available data but under some user-specified constraints; we set the stopping MSE differential to 0.05 and a maximum of 25 predictors.

Population-weighted estimates via the "uncomplexing" method

A Bayesian bootstrap approach (BBDESIGN v.0.3 2), the adjusted survey weights, and information about the sampling design was used to expand the study population and approximate a simple random sample 3 of 10,000 of the children of the target population (N ~ 70,000). 50 expanded datasets were created and then missing data were imputed, as above. For population-weighted descriptive statistics, five bootstrap samples (n = 1657) from each imputed set—or sub-groups, if applicable—were summarized by the mean, and 2.5th and 97.5th percentiles 4,5, which obviates, but is equivalent to, "Rubin's rules" for combining imputation estimates.

Appendix D. Additional results

Table D.1. Childcare use: analysis variables (n = 1657)

	Study sample		Population	
	Summary statistic	**Missing %**	**weighted statistic**	**95% CI**
Baseline childcare (5 & 17 months, before CPE eligibility)				
5-month visit		<1		
no regular childcare	1447		88.8	86.7, 90.4
part-time	66		3.7	2.4, 4.9
full-time	127		7.7	5.9, 9.2
17-month visit		<1		
no regular childcare	715		46.9	44.1, 50.8
part-time	184		11.8	10.0, 14.1
full-time	750		40.9	38.0, 44.5
Childcare 'exposure' from 2-5 years				
Childcare at 2.5-yr visit		3		
Main type				
no regular childcare	673		42.7	40.3, 45.0
centre-based (unsubsidized)	271 (20)		16.1 (1.0)	14.1, 19.7 (0.6, 2.1)
regulated home-based	167		12.1	8.8, 14.1
unregulated home-based	452		26.4	24.1, 28.8
mix of types	45		2.6	1.9, 6.0
Hrs/wk among children with any regular childcare at 2.5 yrs (mean [IQR])	35.4 [28, 43]	1	35.5	34.7, 36.2
Childcare at 3.5-yr visit				
Main type		2.4		
no regular childcare	507		34.9	31.9, 37.4
centre-based (unsubsidized)	458 (48)		28.1 (3.0)	26.1, 31.1
regulated home-based	257		14.7	12.7, 17.3
unregulated home-based	339		19.0	16.7, 21.8
mix of types	56		3.0	1.7, 6.3
Hrs/wk among children with any regular childcare at 3.5 yrs (mean [IQR])	35.8 [30, 43]	2.9	35.8	35.0, 36.6
Childcare at 4 years				
Main type		2.6		
no regular childcare	482		31.9	29.6, 34.6
centre-based (unsubsidized)	535 (53)		31.4 (3.4)	28.7, 35.7
regulated home-based	256		14.7	12.7, 18.6
unregulated home-based	281		17.3	14.0, 20.8
mix of types	60		3.8	2.5, 7.1
Hrs/wk among children with any regular childcare at 4 yrs (mean [IQR])	36.4 [30, 42]	2.8	36.2	35.3, 37.0
Childcare at 5 years (winter/spring before kindergarten)				
Main type		8.9		

	Study sample		Population	
	Summary statistic	Missing %	weighted statistic	95% CI
no regular childcare	173		14.1	11.6, 16.4
centre-based (unsubsidized)	625 (203)		38.6 (13.6)	35.2, 41.4
regulated home-based	215		12.2	10.5, 17.8
unregulated home-based	168		9.8	7.9, 18.5
mix of types	192		13.3	9.4, 20.4
part-time, public pre-kindergarten	100		5.9	4.3, 13.4
NA, started kindergarten early	37	2.4	2.4	1.4, 4.2
Hrs/wk among children with any regular childcare at 5 yrs (median [IQR])	33.9 [25, 44]	8.4	34.0	32.5, 35.0
Summary variables for preschool childcare (from 2-5 years)				
Main type		1		
no regular childcare	131		9.4	7.1, 12.5
part-time, public preK only	80		4.3	3.3, 5.6
centre-based	677		43.0	39.5, 46.7
regulated home-based	301		17.0	15.3, 18.9
unregulated home-based	455		26.0	23.8, 28.3
Hrs/wk among children with any regular childcare (median [IQR])	28.6 [19, 39]	1	27.8	26.9, 28.7
Total arrangements		13.6		
0	101		7.2	5.0, 9.2
1-2	863		56.2	53.0, 59.7
3-4	362		24.7	22.5, 26.9
5-6	86		8.0	6.3, 10.3
7+	19		3.9	2.1, 6.0

Notes: For childcare used on a regular basis at the time of the interview, part-time = 10-23 hrs per week and full-time >23 hrs per week

Table D.2. Family disadvantage indicators at baseline (0 to 1.5 years) and childcare, stratified by family disadvantage. Family disadvantage measure based on the Côte et al. (2008) Family Risk Index—which includes family structure and functioning, mother's age, SES (parents' relative income, education, and occupational prestige), mother's depression symptoms. The total Family Risk Index scores for the two baseline visits ranged from 0 to 22 points; "More advantaged" ≤ 10 (median), "Less advantaged" > 10. [1] Most frequently reported type of childcare over the four study visits from 2 to 5 years old. [2] Public prekindergarten programs were focused on school-readiness and included <14 hrs per week. Montreal program eligibility: children from two disadvantaged neighbourhoods could be referred. Outside Montreal: *Passe-partout* (Friendly et al., 2002; Lucie and André Chagnon Foundation, 2019).

	More advantaged				Less advantaged			
	Sample (n = 945)		Population		Sample (n = 712)		Population	
	frequency or mean [IQR]	% missing	weighted % or mean	95% CI	frequency or mean [IQR]	% missing	weighted % or mean	95% CI
Single-parent or stepfamily at either visit	92	0	10.4	87.7, 91.6	277	0	37.9	34.7, 40.9
Age of mother (mean [IQR])	30.4 [27.4, 33.5]	0	30.6	30.3, 31.0	27.7 [23.1, 31.8]	0	28.1	27.6, 28.6
Socioeconomic status z-score (mean [IQR])	0.55 [0.06, 1.1]	0	55.9	50.4, 61.4	-0.64 [-1.2, -0.12]	0	-0.67	-0.75, -0.62
Family functioning problems (mean [IQR])	0.95 [0.28, 1.41]	0	0.98	0.91, 1.04	2.22 [1.31, 2.96]	<1	2.25	2.13, 2.37
Mother's depression symptoms (mean [IQR])	0.84 [0.39, 1.15]	0	0.83	0.79, 0.87	2.07 [1.20, 2.67]	0	2.08	1.98, 2.25
Childcare use from 0 to 1.5 yrs		<1				<1		
Parental care only	289		32.8	29.4, 36.4	405		57.8	54.4, 61.8
Part-time (max. 10-23 hrs/wk)	116		14.3	11.6, 17.5	70		9.5	7.5, 11.5
Any full-time (>23 hrs/wk at either visit)	540		52.5	48.7, 57.5	237		32.6	29.4, 36.4
Childcare type from 2 to 5 years [1]		<1				1.5		
Parental care, only	61		7.3	5.4, 10.3	70		11.7	8.3, 15.6
Part-time prekindergarten at 5 yrs, only [2]	31		3.0	2.6, 3.3	49		5.6	3.8, 8.3
Centre-based, mainly	394		42.6	38.9, 46.6	283		44.0	37.1, 48.3
Regulated home-based, mainly	166		17.8	14.5, 20.7	135		16.4	13.6, 19.5
Unregulated home-based, mainly	291		29.2	25.9, 32.7	164		22.5	19.1, 26.4
Mean hours per week in childcare, if any (mean [IQR])	29.8 [22.5, 39.0]	<1	29.5	28.6, 30.4	26.9 [16.6, 36.8]	1.9	25.8	24.5, 27.5
Number of arrangements		12.7				14.9		
0	47		5.4	3.8, 7.4	54		9.0	5.4, 11.8
1-2	496		56.3	52.7, 50.9	367		56.1	51.5, 60.7
3-4	224		27.6	24.7, 30.7	138		21.6	18.0, 25.0
5+	58		10.6	7.7, 13.2	47		13.7	8.6, 17.9

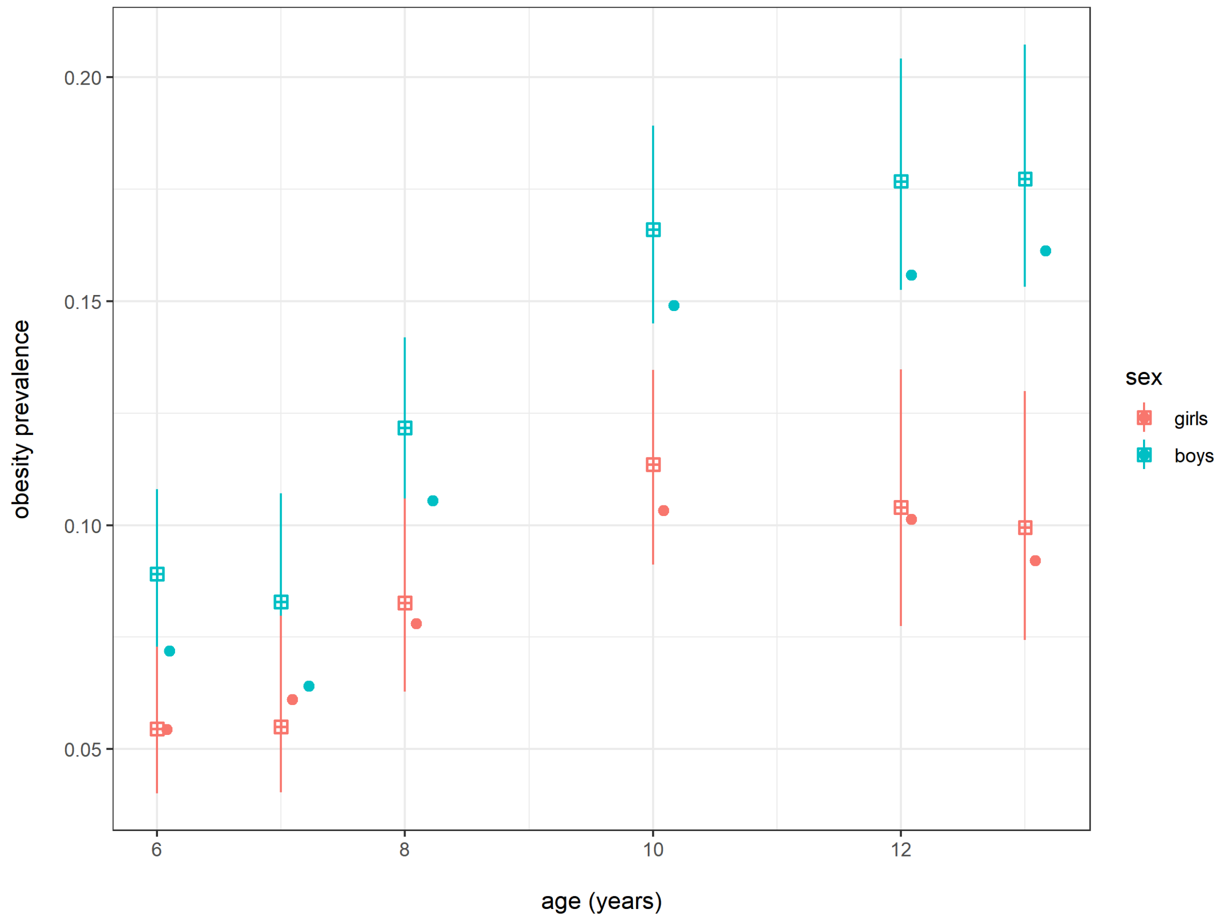


Figure D.1. Obesity prevalence by age and sex in participants of the Quebec Longitudinal Study of Child Development (2003-2010; n = 1657). WHO 2007 obesity cut-off of BMI z-score > 2. Circles show observed prevalence, squares and lines show 5th – 50th – 95th percentiles for estimated prevalence from survey-weighted data with missing values multiply imputed (m=50). Number of observed values by sex and visit: 6-yr, 542 boys and 606 girls; 7-yr, 687 boys and 769 girls; 8-yr, 673 boys and 743 girls; 10-yr, 617 boys and 687 girls; 12-yr, 635 boys and 691 girls; 13-yr, 558 boys and 651 girls.

Table D.3. Observed and weighted imputed obesity prevalence

			WHO standard			IOTF standard		
			sample			sample		
Sex	Median age at study visit	N	freq. obese	% obese	population-weighted prevalence % (95% CI)	freq. obese	% obese	population-weighted prevalence % (95% CI)
Boys	6	541	38	7.2	8.2 (6.2, 11.1)	19	3.6	5.0 (3.1, 7.1)
	7	683	44	6.4	6.3 (4.1, 8.9)	27	4.0	3.9 (2.1, 5.7)
	8	673	71	10.5	10.9 (8.0, 14.2)	27	4.0	4.4 (2.4, 6.6)
	10	617	92	14.9	15.2 (12.2, 19.3)	38	6.2	6.5 (4.2, 9.2)
	12	635	99	15.6	16.2 (12.5, 19.9)	47	7.4	8.2 (5.7, 10.8)
Girls	6	592	32	5.4	5.6 (3.3, 8.1)	26	4.4	4.5 (2.7, 6.9)
	7	764	45	5.9	5.6 (3.9, 7.6)	33	4.3	4.2 (2.5, 6.3)
	8	740	58	7.8	8.2 (6.1, 10.5)	41	5.5	5.9 (4.0, 8.1)
	10	685	71	10.4	12.2 (8.9, 15.8)	42	6.1	7.2 (4.8, 10.5)
	12	689	70	10.2	8.9 (6.8, 11.1)	47	6.8	5.6 (3.6, 7.7)

Additional results for manuscript 1 models

Table D.4. Manuscript 1: BMIz Main model and sensitivity analysis estimates for counterfactual childcare differences.
Crude, adjusted, and adjusted + EMM by family disadvantage estimates for counterfactual differences: Read as "If all children had had full-time center-based from 2-5 years instead of the reference profile childcare profile (full-time regulated home-based, full-time unregulated home-based, or no regular childcare use [parental]), the mean BMI z-score would have been…SD lower/higher"

Analysis	Reference simulated / counterfactual profile [1]	Crude mean diff (95% CrI)	Adjusted [2] mean diff (95% CrI)			EMM diff-n-diff (95% CrI)
		All	All	More advantaged	Less advantaged [3]	Less - more
Bayesian						
Main analytic	*Kindergarten (age ~6y)*					
sample	difference for RH	0.43 (0.17, 0.68)	0.40 (0.14, 0.65)	0.53 (0.21, 0.84)	0.19 (-0.20, 0.58)	-0.33 (-0.82, 0.15)
1. n=1657;	difference for UH	0.09 (-0.16, 0.33)	0.20 (-0.04, 0.43)	0.21 (-0.07, 0.49)	0.24 (-0.16, 0.65)	0.03 (-0.45, 0.50)
corresponds to	difference for parental	0.27 (0.03, 0.49)	0.36 (0.11, 0.60)	0.38 (0.07, 0.69)	0.28 (-0.05, 0.60)	-0.10 (-0.50, 0.29)
differences	*Grade 6 (age ~12y)*					
between	difference for RH	0.24 (-0.06, 0.54)	0.26 (-0.03, 0.54)	0.15 (-0.21, 0.49)	0.41 (-0.05, 0.88)	0.27 (-0.29, 0.82)
counterfactual	difference for UH	-0.01 (-0.29, 0.27)	0.12 (-0.16, 0.38)	0.12 (-0.21, 0.44)	0.15 (-0.32, 0.64)	0.03 (-0.51, 0.60)
predictions in	difference for parental	0.00 (-0.26, 0.26)	0.14 (-0.12, 0.40)	0.25 (-0.08, 0.59)	0.01 (-0.37, 0.39)	-0.24 (-0.69, 0.21)
Figures 2 &)						
1b. Boys	*Kindergarten (age ~6y)*					
(model n=805;	difference for RH	0.35 (-0.02, 0.70)	0.24 (-0.11, 0.60)	0.37 (-0.03, 0.79)	0.04 (-0.42, 0.50)	-0.34 (-0.83, 0.15)
corresponds to	difference for UH	-0.01 (-0.35, 0.34)	0.10 (-0.23, 0.43)	0.10 (-0.27, 0.48)	0.13 (-0.36, 0.59)	0.03 (-0.46, 0.50)
differences	difference for parental	0.28 (0.00, 0.55)	0.37 (0.05, 0.68)	0.40 (0.04, 0.76)	0.29 (-0.10, 0.67)	-0.10 (-0.52, 0.28)
between	*Grade 6 (age ~12y)*					
counterfactual	difference for RH	0.24 (-0.20, 0.64)	0.14 (-0.27, 0.54)	0.03 (-0.45, 0.48)	0.29 (-0.26, 0.86)	0.27 (-0.30, 0.82)
predictions in	difference for UH	-0.35 (-0.75, 0.06)	-0.22 (-0.61, 0.16)	-0.21 (-0.64, 0.22)	-0.18 (-0.73, 0.35)	0.03 (-0.52, 0.59)
Web Figure 4A)	difference for parental	-0.09 (-0.41, 0.22)	-0.06 (-0.40, 0.29)	0.07 (-0.33, 0.48)	-0.18 (-0.60, 0.26)	-0.24 (-0.70, 0.20)
1c. Girls	*Kindergarten (age ~6y)*					
(model n=852;	difference for RH	0.50 (0.15, 0.84)	0.54 (0.19, 0.88)	0.67 (0.27, 1.06)	0.33 (-0.13, 0.79)	-0.34 (-0.84, 0.16)
corresponds to	difference for UH	0.17 (-0.18, 0.50)	0.29 (-0.04, 0.62)	0.31 (-0.05, 0.68)	0.34 (-0.13, 0.81)	0.03 (-0.46, 0.50)
differences	difference for parental	0.25 (-0.03, 0.52)	0.35 (0.03, 0.67)	0.37 (0.00, 0.73)	0.26 (-0.13, 0.64)	-0.10 (-0.51, 0.30)
between	*Grade 6 (age ~12y)*					
counterfactual	difference for RH	0.24 (-0.18, 0.66)	0.36 (-0.03, 0.76)	0.26 (-0.19, 0.70)	0.52 (0.00, 1.06)	0.27 (-0.29, 0.83)
predictions in	difference for UH	0.30 (-0.10, 0.68)	0.43 (0.06, 0.80)	0.42 (0.02, 0.84)	0.45 (-0.08, 1.01)	0.03 (-0.54, 0.59)
Web Figure 4B)	difference for parental	0.09 (-0.22, 0.40)	0.32 (-0.02, 0.67)	0.43 (0.03, 0.85)	0.19 (-0.25, 0.62)	-0.24 (-0.71, 0.22)

Analysis	Reference simulated / counterfactual profile [1]	Crude mean diff (95% CrI)	Adjusted [2] mean diff (95% CrI)			EMM diff-n-diff (95% CrI)
		All	All	More advantaged	Less advantaged [3]	Less - more
Frequentist [4]						
n=1657	*Kindergarten (age ~6y)*					
	difference for RH	0.42 (0.22, 0.68)	0.42 (0.29, 0.64)	0.47 (0.29, 0.71)	0.28 (-0.01, 0.64)	-0.19 (-0.59, 0.16)
	difference for UH	0.09 (-0.15, 0.29)	0.24 (-0.02, 0.48)	0.22 (-0.06, 0.50)	0.36 (0.04, 0.77)	0.43 (-0.43, 0.56)
	difference for parental	0.23 (-0.01, 0.40)	0.38 (0.09, 0.62)	0.39 (0.09, 0.61)	0.36 (-0.06, 0.69)	-0.04 (-0.36, 0.29)
	Grade 4 (age ~10y)					
	difference for RH	0.30 (0.07, 0.59)	0.34 (0.16, 0.59)	0.35 (0.14, 0.56)	0.36 (0.01, 0.69)	0.02 (-0.36, 0.30)
	difference for UH	0.03 (-0.14, 0.26)	0.16 (-0.01, 0.37)	0.19 (-0.01, 0.40)	0.20 (-0.09, 0.50)	-0.02 (-0.34, 0.32)
	difference for parental	0.05 (-0.18, 0.21)	0.23 (-0.02, 0.52)	0.34 (0.07, 0.63)	0.10 (-0.23, 0.47)	-0.25 (-0.53, 0.04)
	Grade 6 (age ~12y)					
	difference for RH	0.25 (-0.04, 0.53)	0.29 (0.07, 0.57)	0.27 (0.03, 0.51)	0.39 (-0.10, 0.80)	0.15 (-0.30, 0.44)
	difference for UH	-0.01 (-0.20, 0.28)	0.14 (-0.03, 0.32)	0.20 (-0.02, 0.41)	0.13 (-0.17, 0.51)	-0.06 (-0.47, 0.32)
	difference for parental	- 0.03 (-0.27, 0.16)	0.17 (-0.12, 0.49)	0.33 (0.01, 0.65)	-0.05 (-0.39, 0.36)	-0.33 (-0.70, 0.03)
1. Urban only (n=1078)	*Kindergarten (age ~6y)*					
	difference for RH	0.61 (0.27, 0.94)	0.61 (0.28, 0.95)	0.64 (0.30, 0.97)	0.31 (-0.24, 0.84)	-0.33 (-0.86, 0.24)
	difference for UH	0.07 (-0.23, 0.37)	0.20 (-0.09, 0.51)	0.21 (-0.10, 0.53)	0.12 (-0.43, 0.67)	-0.09 (-0.63, 0.44)
	difference for parental	0.36 (0.04, 0.68)	0.53 (0.21, 0.85)	0.50 (0.16, 0.85)	0.66 (0.16, 1.13)	0.16 (-0.27, 0.62)
	Grade 6 (age ~12y)					
	difference for RH	0.24 (-0.17, 0.61)	0.22 (-0.15, 0.60)	0.19 (-0.20, 0.60)	0.22 (-0.42, 0.93)	0.03 (-0.64, 0.72)
	difference for UH	-0.07 (-0.40, 0.25)	0.05 (-0.30, 0.38)	0.09 (-0.28, 0.43)	-0.13 (-0.78, 0.49)	-0.22 (-0.85, 0.42)
	difference for parental	-0.06 (-0.41, 0.27)	0.15 (-0.19, 0.51)	0.23 (-0.18, 0.58)	0.02 (-0.48, 0.52)	-0.21 (-0.71, 0.30)
2. End follow-up at 10-yr visit (n=1657)	*Kindergarten (age ~6y)*					
	difference for RH	0.51 (0.24, 0.79)	0.46 (0.19, 0.71)	0.54 (0.27, 0.82)	0.20 (-0.25, 0.63)	-0.34 (-0.79, 0.07)
	difference for UH	0.08 (-0.21, 0.33)	0.18 (-0.06, 0.43)	0.21 (-0.03, 0.46)	0.24 (-0.16, 0.69)	0.03 (-0.37, 0.45)
	difference for parental	0.31 (0.07, 0.55)	0.38 (0.11, 0.66)	0.40 (0.12, 0.67)	0.31 (-0.06, 0.66)	-0.09 (-0.45, 0.28)
	Grade 4 (age ~10y)					
	difference for RH	0.30 (0.02, 0.59)	0.28 (0.01, 0.57)	0.30 (0.02, 0.59)	0.22 (-0.23, 0.65)	-0.08 (-0.54, 0.34)
	difference for UH	0.01 (-0.26, 0.28)	0.12 (-0.14, 0.38)	0.14 (-0.11, 0.40)	0.12 (-0.36, 0.58)	-0.02 (-0.46, 0.41)
	difference for parental	0.07 (-0.17, 0.34)	0.19 (-0.06, 0.46)	0.24 (-0.03, 0.53)	0.12 (-0.24, 0.49)	-0.12 (-0.48, 0.22)

Analysis	Reference simulated / counterfactual profile [1]	Crude mean diff (95% CrI)	Adjusted [2] mean diff (95% CrI)			EMM diff-n-diff (95% CrI)
		All	All	More advantaged	Less advantaged [3]	Less - more
3. Reference = subsidized center-based (unsubsidized center-based care in separate category)	*Kindergarten (age ~6y)*					
	difference for RH	0.51 (0.24, 0.78)	0.47 (0.21, 0.75)	0.55 (0.29, 0.83)	0.21 (-0.22, 0.63)	-0.35 (-0.76, 0.08)
	difference for UH	0.09 (-0.17, 0.36)	0.21 (-0.04, 0.44)	0.22 (-0.04, 0.47)	0.28 (-0.15, 0.71)	0.06 (-0.34, 0.46)
	difference for parental	0.28 (0.03, 0.54)	0.39 (0.12, 0.65)	0.39 (0.12, 0.65)	0.35 (0.00, 0.70)	-0.04 (-0.36, 0.31)
	Grade 6 (age ~12y)					
	difference for RH	0.33 (0.05, 0.63)	0.34 (0.04, 0.64)	0.30 (-0.02, 0.61)	0.47 (0.01, 0.91)	0.17 (-0.29, 0.64)
	difference for UH	0.02 (-0.29, 0.30)	0.16 (-0.11, 0.45)	0.15 (-0.11, 0.42)	0.21 (-0.25, 0.70)	0.06 (-0.39, 0.52)
	difference for parental	0.09 (-0.20, 0.35)	0.24 (-0.04, 0.52)	0.29 (0.01, 0.57)	0.16 (-0.24, 0.53)	-0.12 (-0.50, 0.26)
4. 2-5yr childcare summary variable	*Kindergarten (age ~6y)*					
	difference for RH	0.17 (0.00, 0.33)	0.12 (-0.05, 0.28)	0.12 (-0.09, 0.36)	0.12 (-0.12, 0.36)	-0.00 (-0.35, 0.32)
	difference for UH	-0.02 (-0.16, 0.13)	0.00 (-0.14, 0.14)	-0.05 (-0.23, 0.13)	0.11 (-0.11, 0.35)	0.16 (-0.13, 0.47)
	difference for parental	0.17 (-0.07, 0.41)	0.18 (-0.08, 0.43)	0.04 (-0.31, 0.39)	0.30 (-0.04, 0.65)	0.25 (-0.22, 0.72)
	Grade 6 (age ~12y)					
	difference for RH	0.10 (-0.09, 0.28)	0.09 (-0.10, 0.28)	0.03 (-0.21, 0.27)	0.17 (-0.09, 0.44)	0.14 (-0.22, 0.49)
	difference for UH	-0.00 (-0.18, 0.16)	0.02 (-0.15, 0.19)	-0.03 (-0.23, 0.15)	0.11 (-0.15, 0.38)	0.14 (-0.17, 0.47)
	difference for parental	-0.17 (-0.43, 0.08)	-0.14 (-0.43, 0.12)	-0.13 (-0.49, 0.27)	-0.18 (-0.53, 0.24)	-0.05 (-0.53, 0.45)
Boys (n=805)	*Kindergarten (age ~6y)*					
	difference for RH		0.31 (-0.06, 0.69)	0.50 (0.06, 0.93)	0.03 (-0.47, 0.52)	-0.46 (-1.03, 0.09)
	difference for UH		0.08 (-0.27, 0.42)	0.08 (-0.29, 0.49)	0.12 (-0.37, 0.62)	0.03 (-0.42, 0.55)
	difference for parental		0.41 (0.07, 0.75)	0.44 (0.05, 0.81)	0.37 (-0.04, 0.78)	-0.07 (-0.51, 0.39)
	Grade 6 (age ~12y)					
	difference for RH		0.19 (-0.21, 0.61)	0.14 (-0.35, 0.61)	0.30 (-0.24, 0.83)	0.16 (-0.43, 0.73)
	difference for UH		-0.20 (-0.60, 0.17)	-0.17 (-0.59, 0.25)	-0.17 (-0.69, 0.39)	0.01 (-0.54, 0.60)
	difference for parental		-0.01 (-0.38, 0.34)	0.14 (-0.27, 0.55)	-0.11 (-0.53, 0.34)	-0.25 (-0.72, 0.24)
Girls (n=852)	*Kindergarten (age ~6y)*					
	difference for RH		0.58 (0.20, 0.91)	0.75 (0.36, 1.14)	0.29 (-0.20, 0.78)	-0.46 (-0.99, 0.08)
	difference for UH		0.26 (-0.09, 0.58)	0.31 (-0.08, 0.69)	0.34 (-0.12, 0.83)	0.04 (-0.43, 0.58)
	difference for parental		0.36 (0.01, 0.69)	0.38 (-0.01, 0.72)	0.31 (-0.11, 0.73)	-0.07 (-0.51, 0.38)
	Grade 6 (age ~12y)					
	difference for RH		0.37 (-0.02, 0.77)	0.32 (-0.12, 0.76)	0.47 (-0.07, 1.00)	0.16 (-0.42, 0.74)
	difference for UH		0.40 (0.01, 0.79)	0.43 (-0.00, 0.90)	0.44 (-0.11, 1.02)	0.01 (-0.58, 0.60)
	difference for parental		0.32 (-0.04, 0.68)	0.44 (0.02, 0.87)	0.20 (-0.27, 0.65)	-0.25 (-0.71, 0.24)

[1] Childcare types:

RH = regulated home-based (*CPE en famille*)

UH = unregulated home-based (child's own home or others' home)

No FT = no regular non-parental childcare (<10 hrs per week)

[2] Adjusted for birth order, breastfeeding, prematurity, birthweight, type of delivery, child's health, parents' BMI and health, family structure, parents' age, SES, mother's depression symptoms, family functioning, baseline childcare, parents' employment and education, smoking, child's temperament and behaviour, parenting perceptions, parents' immigration status and languages, area of residence and housing conditions. See Table B.1 for details.

[3] Unless stated otherwise, models use Family Risk Index cut-off at ~median (>5).

[4] 95% bootstrap confidence intervals are reported for frequentist models (instead of Bayesian credible intervals)

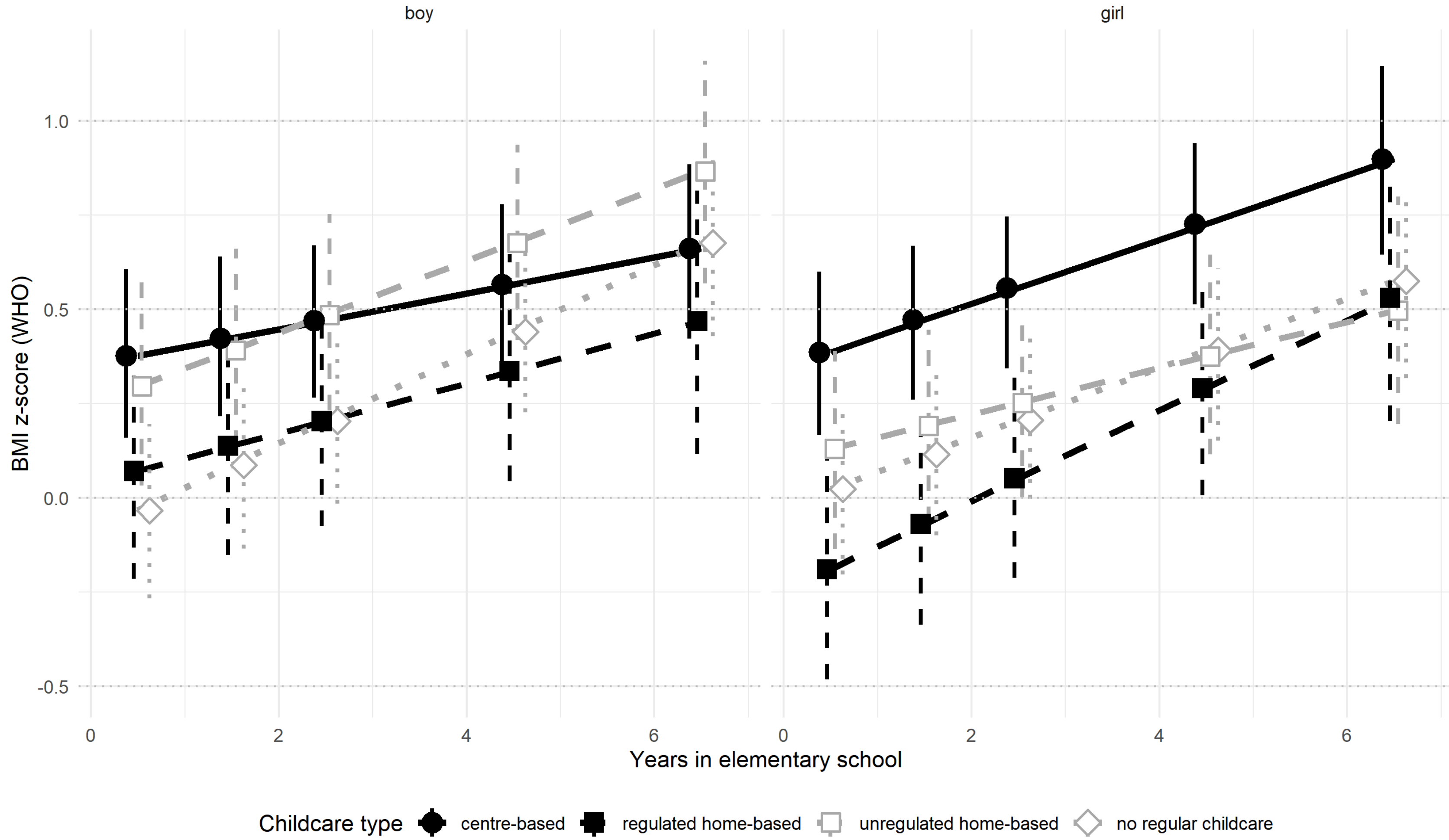


Figure D.2. Predicted BMI z-scores for counterfactual childcare profiles, by sex. Adjusted model

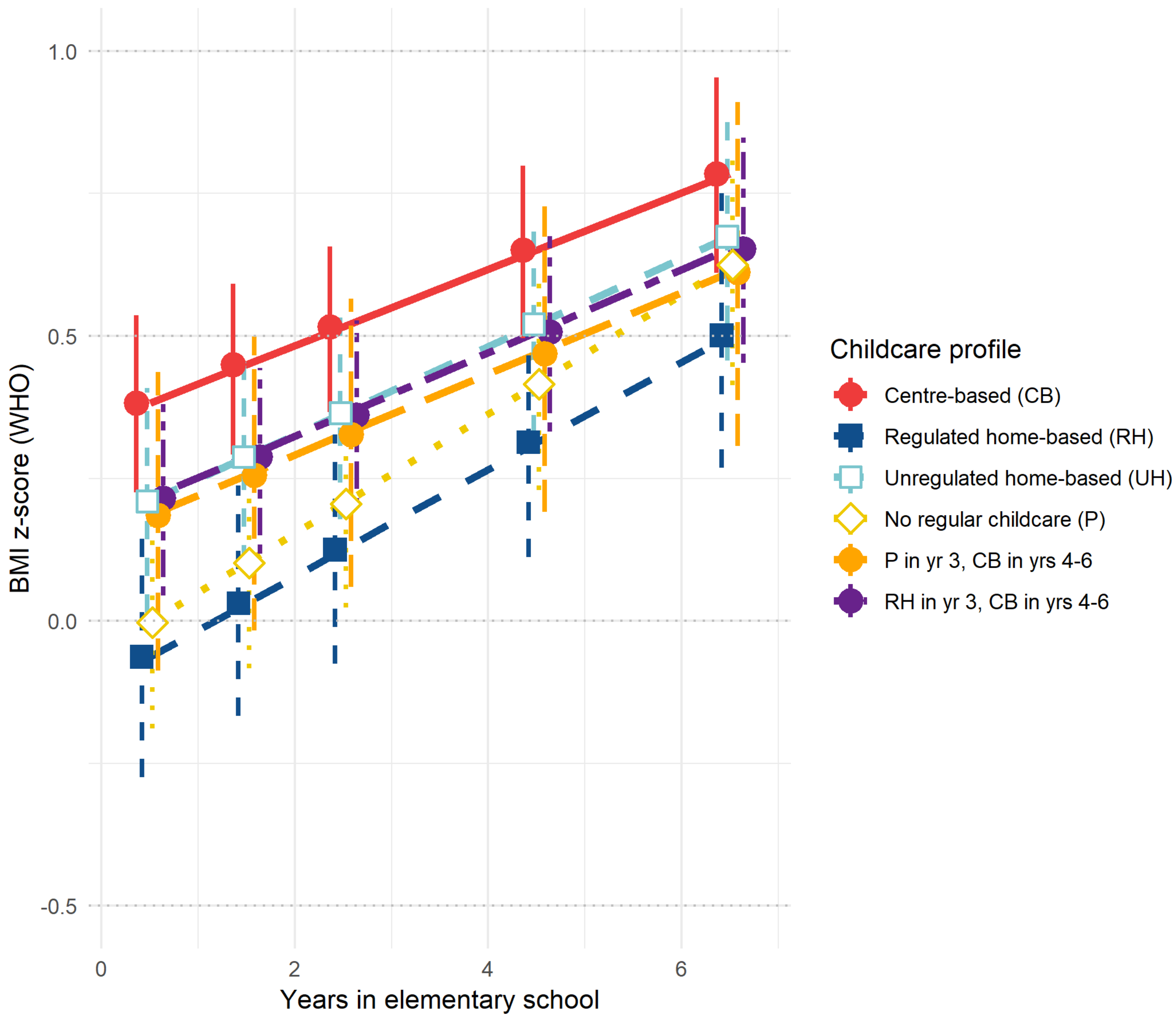


Figure D.3. Predicted BMI z-scores for additional counterfactual childcare profiles: later initiation of centre-based childcare. Adjusted model.

Table D.5. Predicted prevalence of obesity under observed and counterfactual childcare profiles. Estimated from posterior predicted distribution of the main Bayesian multilevel linear regression. Obesity (BMI z-score > 2 SD)

	Predicted under observed childcare	Predicted under counterfactual childcare			
		center-based	regulated home-based	unregulated home-based	no regular childcare
Kindergarten					
All	6.3 (5.0, 7.7)	8.3 (6.2, 10.9)	5.0 (3.4, 7.1)	6.4 (4.6, 8.9)	5.2 (3.6, 7.0)
By family risk index					
less advantaged	8.5 (6.8, 10.4)	9.8 (7.0, 13.5)	7.9 (5.1, 11.9)	7.5 (4.5, 12.0)	7.1 (5.0, 9.6)
more advantaged	4.7 (3.3, 6.3)	6.8 (4.5, 9.9)	3.1 (1.6, 5.2)	5.0 (3.1, 7.8)	3.6 (2.1, 5.8)
By sex					
boys	6.9 (5.3, 8.7)	8.6 (6.0, 12.1)	6.3 (3.8, 9.8)	7.7 (4.9, 11.3)	5.3 (3.5, 7.7)
girls	5.6 (4.2, 7.2)	7.9 (5.5, 10.9)	3.8 (2.2, 5.9)	5.3 (3.3, 8.2)	4.9 (3.3, 7.1)
Grade 6					
All	14.9 (12.5, 17.5)	17.1 (13.4, 21.2)	12.7 (9.0, 17.2)	15.2 (11.3, 19.8)	14.6 (11.1, 18.8)
By family risk index					
less advantaged	18.7 (16.0, 21.7)	20.7 (15.5, 26.9)	13.5 (8.5, 20.1)	18.1 (11.8, 26.4)	20.1 (15.7, 25.6)
more advantaged	12.6 (9.9, 15.7)	14.8 (10.7, 19.9)	12.3 (7.8, 18.6)	13.2 (8.9, 18.4)	10.7 (7.0, 15.5)
By sex					
boys	16.9 (14.1, 19.9)	15.6 (11.5, 20.6)	13.4 (8.4, 19.5)	19.6 (14.0, 26.5)	16.6 (12.2, 21.8)
girls	13.0 (10.5, 15.8)	18.3 (13.6, 24.0)	11.9 (7.6, 18.0)	11.0 (7.0, 16.2)	12.5 (8.7, 17.5)

Table D.6. Difference in population-averaged predicted probability of obesity under different childcare profiles, by WHO and IOTF standards

Results of GEE probit-binomial regression with child as cluster. IOTF standard was sensitivity analysis 5.

Centre-based care minus reference (counterfactuals)	Adjusted risk difference (bootstrapped 95% CI)	
	WHO standard (>2 SD)	IOTF standard
Kindergarten (age ~6y)		
difference for RH	8.7 (4.7, 13.3)	5.3 (1.7, 9.0)
difference for UH	5.3 (-2.2, 9.9)	0.7 (-4.6, 5.4)
difference for parental	5.3 (1.8, 12.5)	3.0 (-4.8, 7.8)
Grade 6 (age ~12y)		
difference for RH	7.2 (0.9, 16.2)	5.8 (1.5, 11.6)
difference for UH	7.5 (2.3, 14.8)	1.6 (-6.5, 10.8)
difference for parental	-2.0 (-14.6, 6.7)	-1.1 (-8.3, 10.3)

Additional results for manuscript 2 models

Table D.7. Manuscript 2: predicted PSR differences between counterfactual childcare profiles. Point estimates and 95% credible intervals (CrI) for population-averaged poor self-regulation score for centre-based care (*unless otherwise stated) for 35 hrs/week from 2-5 years minus comparator. Abbreviations: CB = centre-based (mostly CPE-regulated), RH = CPE-regulated home-based, UH = unregulated home-based. Differences correspond to differences between mean predictions plotted in Figures 4.1-2. Estimated from 6241 measures in 1657 children with missing data multiply imputed 50 times.

		Crude model		Adjusted model		Adjusted model with interaction between main childcare type and family disadvantage					
						more advantaged		less advantaged		EMM: less - more	
Grade	Comparator	E(y)	95% CrI	E(y)	95% CrI	E(y)	95% CrI	E(y)	95% CrI	E(y)	95% CrI
K	regulated home-based	-0.11	-0.42, 0.22	0.00	-0.29, 0.28	0.01	-0.35, 0.38	0	-0.44, 0.43	-0.02	-0.53, 0.52
	unregulated home-based	0.23	-0.07, 0.54	0.22	-0.04, 0.49	0.21	-0.13, 0.55	0.21	-0.19, 0.60	0.01	-0.50, 0.50
	parent	0.03	-0.26, 0.41	0.27	0.00, 0.53	0.16	-0.18, 0.49	0.34	-0.01, 0.67	0.18	-0.25, 0.58
	CB at 3 yrs, none at 2 yrs	-0.05	-0.31, 0.21	0.02	-0.21, 0.27	0.16	-0.16, 0.47	-0.12	-0.42, 0.19	-0.29	-0.66, 0.09
	RH* vs UH	0.33	-0.01, 0.70	0.22	-0.11, 0.55	0.2	-0.23, 0.59	0.21	-0.28, 0.70	0.02	-0.57, 0.61
6	regulated home-based	-0.01	-0.37, 0.37	0.07	-0.26, 0.40	0.14	-0.26, 0.57	-0.03	-0.53, 0.45	-0.18	-0.79, 0.40
	unregulated home-based	0.30	-0.06, 0.64	0.35	0.00, 0.68	0.52	0.08, 0.92	0.19	-0.28, 0.65	-0.33	-0.92, 0.30
	parent	0.03	-0.29, 0.40	0.27	-0.04, 0.58	0.42	0.02, 0.83	0.14	-0.24, 0.51	-0.29	-0.76, 0.19
	CB at 3 yrs, none at 2 yrs	0.15	-0.13, 0.42	0.20	-0.08, 0.48	0.47	0.13, 0.83	-0.01	-0.37, 0.34	-0.49	-0.92, -0.05
	RH* vs UH	0.31	-0.10, 0.71	0.28	-0.11, 0.65	0.37	-0.13, 0.84	0.22	-0.32, 0.75	-0.15	-0.80, 0.55

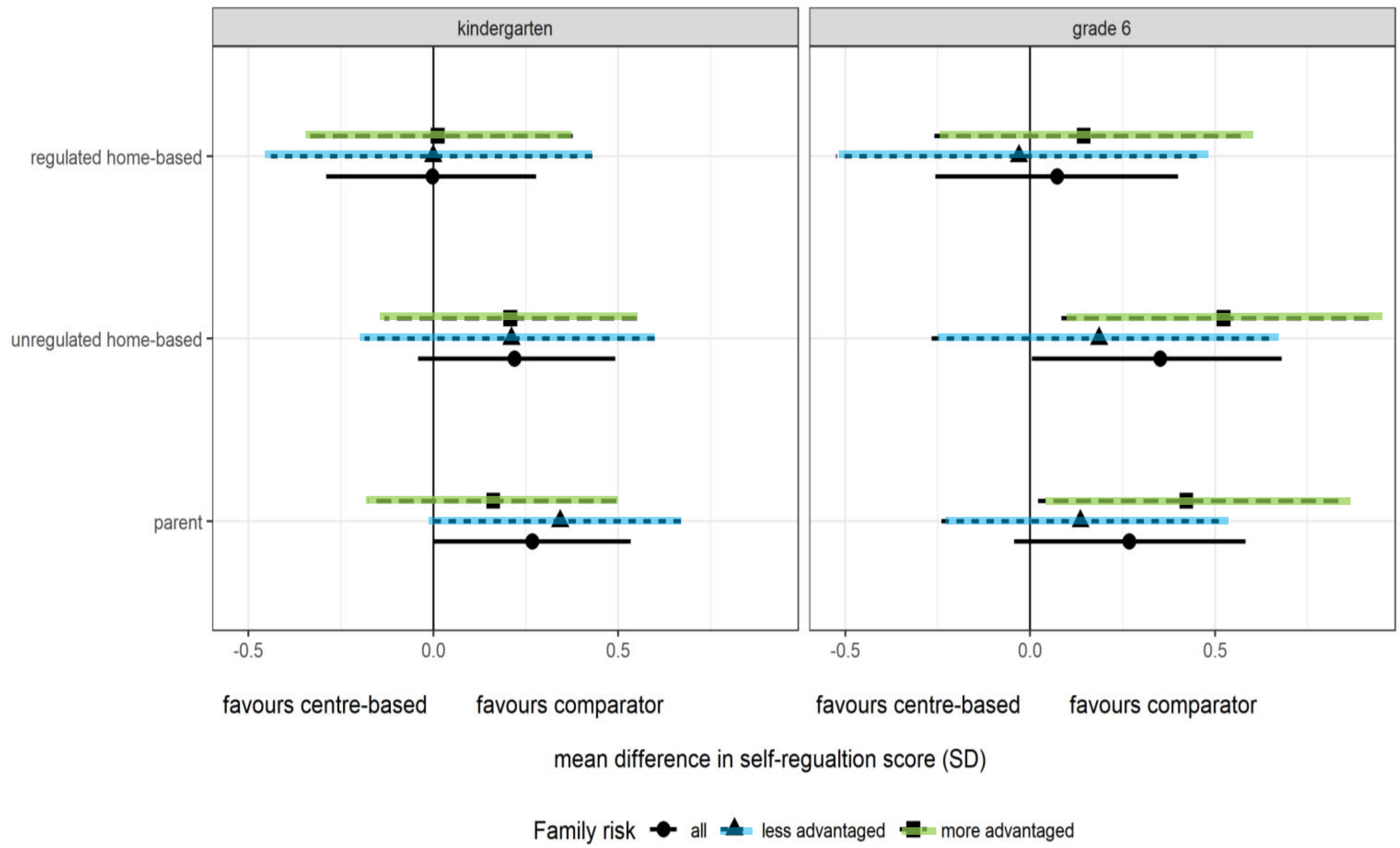


Figure D.4. Mean difference in poor self-regulation scores by counterfactual childcare type and family disadvantage.

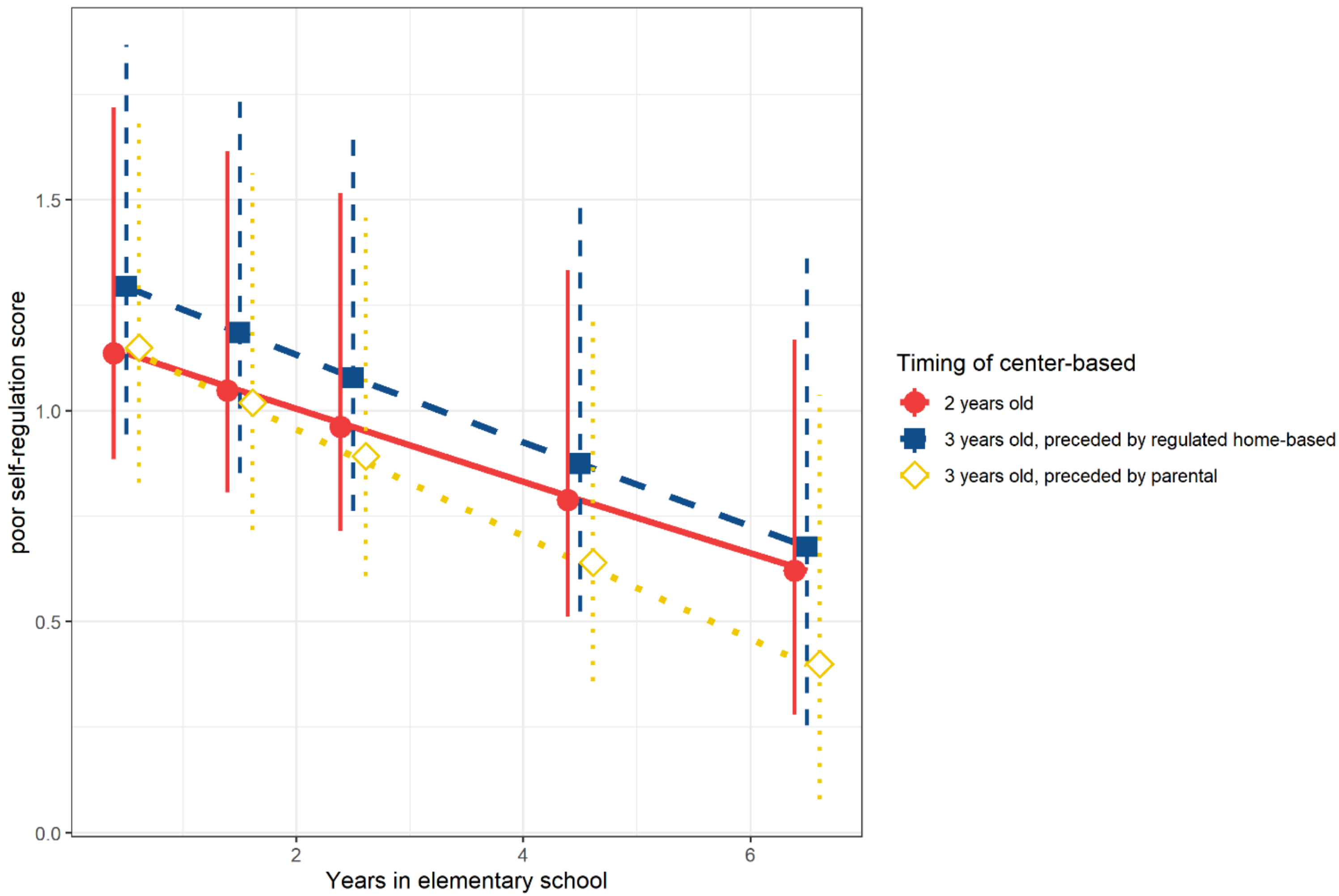


Figure D.5. Predicted poor self-regulation score by counterfactual childcare profiles varying timing of initiation of centre-based care

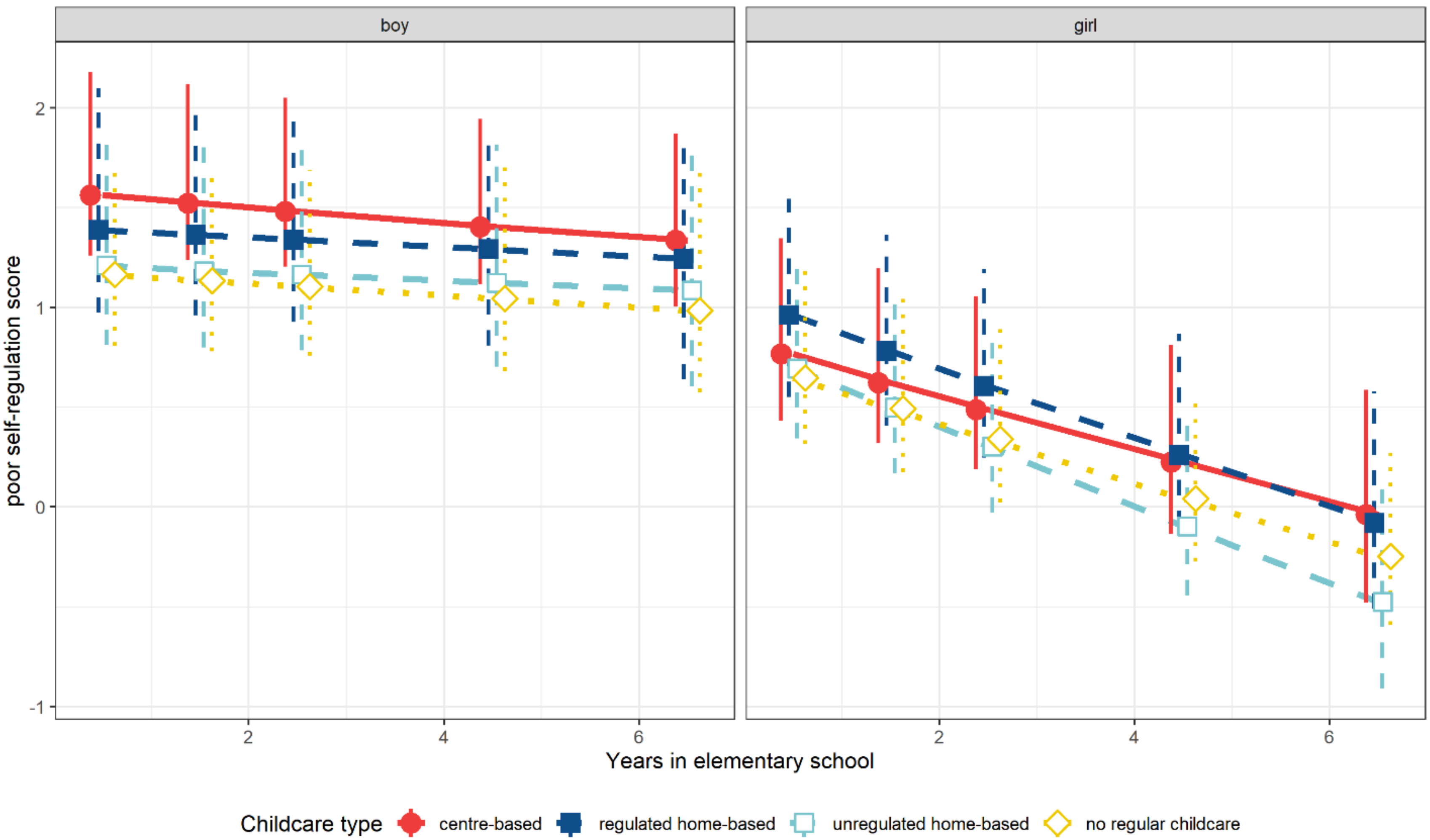


Figure D.6. Mean difference in poor self-regulation scores for counterfactual childcare type and sex.

Table D.8. Sensitivity analyses for manuscript 2 a. Frequentist main model

		Crude model		Adjusted model		Adjusted model with interaction between main childcare type and family disadvantage					
						more advantaged		less advantaged		EMM: less - more	
grade	Comparator	E(y)	95% CrI	E(y)	95% CrI	E(y)	95% CrI	E(y)	95% CrI	E(y)	95% CrI
K	regulated home-based	-0.11	-0.39, 0.17	0.02	-0.24, 0.30	0.05	-0.28, 0.40	0.03	-0.35, 0.40	-0.04	-0.47, 0.45
	unregulated home-based	0.25	-0.02, 0.53	0.23	-0.03, 0.49	0.20	-0.12, 0.51	0.24	-0.13, 0.57	0.03	-0.41, 0.50
	parent	0.02	-0.23, 0.26	0.26	-0.01, 0.52	0.17	-0.17, 0.51	0.33	0.00, 0.64	0.16	-0.21, 0.52
	CB at 3 yrs, none at 2 yrs	-0.05	-0.28, 0.26	0.01	-0.22, 0.22	0.15	-0.17, 0.44	-0.12	-0.40, 0.16	-0.27	-0.63, 0.11
	RH vs UH	0.36	0.05, 0.70	0.21	-0.09, 0.49	0.15	-0.21, 0.48	0.20	-0.21, 0.65	0.06	-0.41, 0.54
6	regulated home-based	-0.01	-0.40, 0.36	0.06	-0.28, 0.44	0.12	-0.32, 0.55	-0.03	-0.59, 0.55	-0.15	-0.79, 0.49
	unregulated home-based	0.31	-0.10, 0.69	0.32	0.05, 0.69	0.48	0.05, 0.91	0.17	-0.37, 0.68	-0.31	-0.93, 0.30
	parent	0.02	-0.33, 0.36	0.29	-0.02, 0.64	0.43	-0.01, 0.87	0.17	-0.25, 0.59	-0.25	-0.78, 0.25
	CB at 3 yrs, none at 2 yrs	0.14	-0.17, 0.44	0.18	-0.12, 0.48	0.47	0.08, 0.82	-0.03	-0.41, 0.40	-0.49	-0.93, -0.03
	RH vs UH	0.32	-0.14, 0.77	0.25	-0.19, 0.64	0.35	-0.16, 0.86	0.20	-0.46, 0.81	-0.15	-0.92, 0.54

b. Frequentist model, adjusted but no adjustment for baseline behaviour

grade	txf	Median	95% CI	qual. change
	regulated home-based	0.02	(-0.23, 0.22)	CB diff same
	unregulated home-based	0.20	(0.06, 0.47)	CB diff same
	parent	0.23	(-0.03, 0.42)	CB diff same
	CB at 3 yrs, none at 2 yrs	-0.03	(-0.23, 0.23)	CB diff same
K	RH vs UH	0.21	(-0.00, 0.46)	RH - UH diff same
	regulated home-based	0.11	(-0.21, 0.43)	CB diff same
	unregulated home-based	0.38	(0.13, 0.65)	CB diff same
	parent	0.33	(0.04, 0.62)	CB diff same
	CB at 3 yrs, none at 2 yrs	0.19	(-0.05, 0.53)	CB diff same
6	RH vs UH	0.28	(-0.20, 0.62)	RH - UH diff same

Table D.8 continued. c. Frequentist model, stratified by sex

		Main model (adjusted for baseline behaviour)				No adjustment for baseline behaviour				
		Boys (n = 805)		Girls (n = 852)		Boys (n = 805)		Girls (n = 852)		
Grade	Comparator	Median	95% CI	Median	95% CI	Median	95% CI	Median	95% CI	qual. change
K	regulated home-based	0.20	(-0.19, 0.61)	0.20	(-0.25, 0.50)	-0.13	(-0.45, 0.18)	-0.15	(-0.49, 0.22)	CB diff flipped
	unregulated home-based	0.38	(0.03, 0.75)	0.32	(0.07, 0.73)	0.13	(-0.15, 0.34)	0.09	(-0.27, 0.45)	CB diff smaller
	parent	0.39	(0.07, 0.72)	0.37	(0.08, 0.64)	0.10	(-0.25, 0.40)	0.13	(-0.21, 0.49)	CB diff smaller
	CB at 3 yrs, none at 2 yrs	0.01	(-0.31, 0.34)	-0.08	(-0.42, 0.33)	0.01	(-0.21, 0.28)	0.01	(-0.29, 0.30)	CB diff same
	RH vs UH	0.18	(-0.25, 0.59)	0.18	(-0.15, 0.48)	0.26	(-0.12, 0.54)	0.24	(-0.18, 0.65)	RH - UH diff larger
6	regulated home-based	0.14	(-0.38, 0.71)	0.12	(-0.38, 0.63)	0.17	(-0.31, 0.48)	-0.02	(-0.50, 0.48)	CB diff same
	unregulated home-based	0.24	(-0.26, 0.79)	0.21	(-0.09, 0.53)	0.55	(0.17, 0.88)	0.38	(-0.09, 0.88)	CB diff larger
	parent	0.40	(-0.07, 0.85)	0.35	(0.02, 0.73)	0.25	(-0.06, 0.63)	0.20	(-0.24, 0.62)	CB diff smaller
	CB at 3 yrs, none at 2 yrs	0.10	(-0.32, 0.54)	0.09	(-0.28, 0.40)	0.33	(-0.02, 0.69)	0.26	(-0.13, 0.64)	CB diff larger
	RH vs UH	0.08	(-0.54, 0.67)	0.17	(-0.35, 0.57)	0.35	(-0.22, 0.84)	0.40	(-0.15, 0.95)	RH - UH diff larger

d. Frequentist model, adjustment for baseline behaviour and interaction with childcare type

Grade	Comparator	Median	95% CI	Median	95% CI	Median	95% CI
K	regulated home-based	0.04	(-0.25, 0.28)	0.02	(-0.32, 0.35)	-0.03	(-0.42, 0.41)
	unregulated home-based	0.22	(0.04, 0.45)	0.31	(-0.03, 0.58)	0.10	(-0.30, 0.45)
	parent	0.24	(-0.03, 0.46)	0.20	(-0.19, 0.52)	-0.03	(-0.39, 0.25)
	CB at 3 yrs, none at 2 yrs	0.01	(-0.22, 0.24)	-0.07	(-0.36, 0.32)	-0.09	(-0.31, 0.22)
	RH vs UH	0.20	(-0.07, 0.46)	0.28	(-0.14, 0.73)	0.11	(-0.34, 0.60)
6	regulated home-based	0.13	(-0.14, 0.46)	0.12	(-0.38, 0.54)	-0.06	(-0.39, 0.35)
	unregulated home-based	0.38	(0.07, 0.73)	0.43	(0.11, 0.77)	0.03	(-0.34, 0.42)
	parent	0.35	(0.04, 0.62)	0.26	(-0.06, 0.72)	-0.05	(-0.39, 0.22)
	CB at 3 yrs, none at 2 yrs	0.21	(0.03, 0.45)	0.11	(-0.16, 0.57)	-0.11	(-0.32, 0.21)
	RH vs UH	0.25	(-0.16, 0.72)	0.34	(-0.17, 0.75)	0.06	(-0.39, 0.57)

Additional results for manuscript 3 models

Table D.9. Recency weights for 3-year mean poor self-regulation

Median age at visit	4	5	6	7	8	9	10	11	12	13
BMI collected?			[BMI]	BMI	BMI		BMI		BMI	BMI
SR collected (rater)?	SR (m)	SR (m)	SR (m, t)	SR (t)	SR (m, t)		SR (t)		SR (t)	
Linear weights for the recency-weighted cumulative 3-yr mean SR	1/9	3/9	5/9							
		1/9	3/9	5/9						
				1/9	3/9	(sr7 + 2*sr8/3) * 5/9				
							9/9			
							1/9	3/9*(sr10 + sr12)/2	5/9	
Quadratic 1 weights for the recency-weighted cumulative 3-yr mean SR	1/13	3/13	9/13							
		1/13	3/13	9/13						
				1/13	3/13	(sr7 + 2*sr8/3) * 9/13				
							13/13			
							1/13	3/13*(sr10 + sr12)/2	9/13	

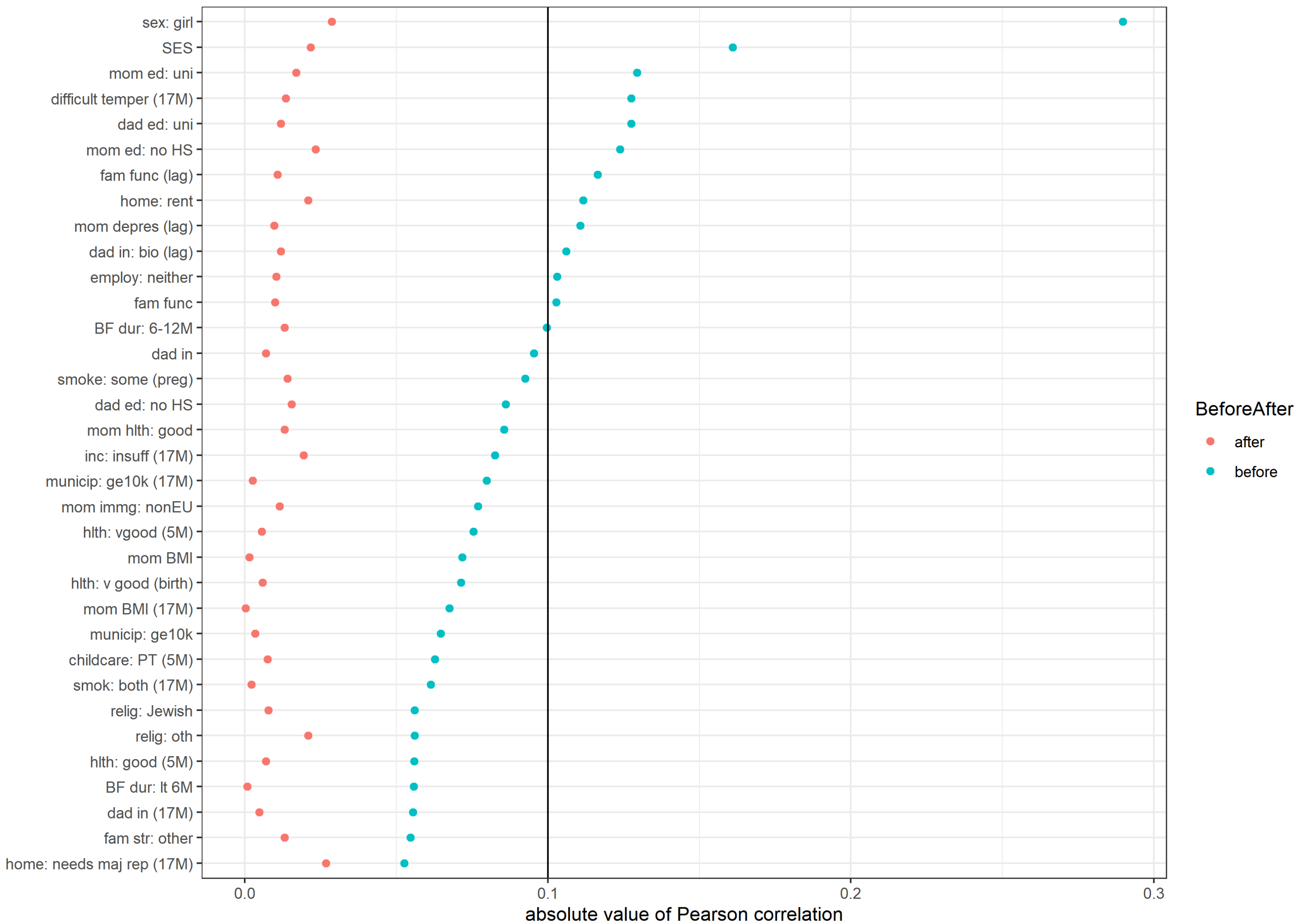
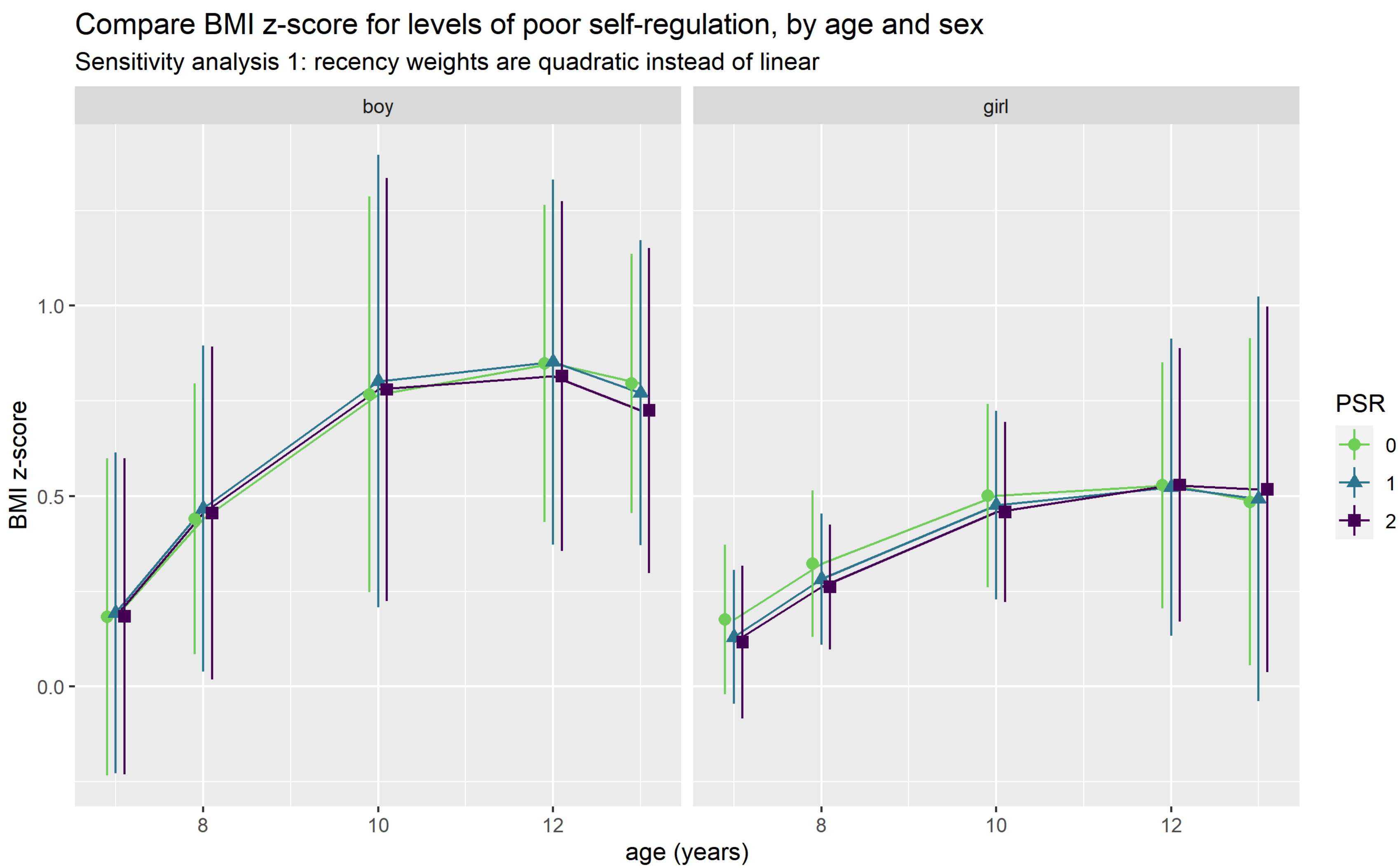


Figure D.7. Pearson correlation between exposure, recency-weighted 3-year mean poor self-regulation, and adjustment variables, before and after IPT weighting.

Figure D.8. Probability of obesity by sex, age and recent poor self-regulation score.



Marginal mean (+/- 2 * total SE) from IPT-weighted GEE, BMIz = (SR + SR^2) * girl * (age + age^2) (centered on 9 years)
50 imputations, Rubin's rules for total SE.
SR z-score seleted from within the range of the data.

Figure D.9. Mean BMI z-score by sex, age, and recent poor self-regulation score. Sensitivity analysis with quadratic recency weights (Table D.10).

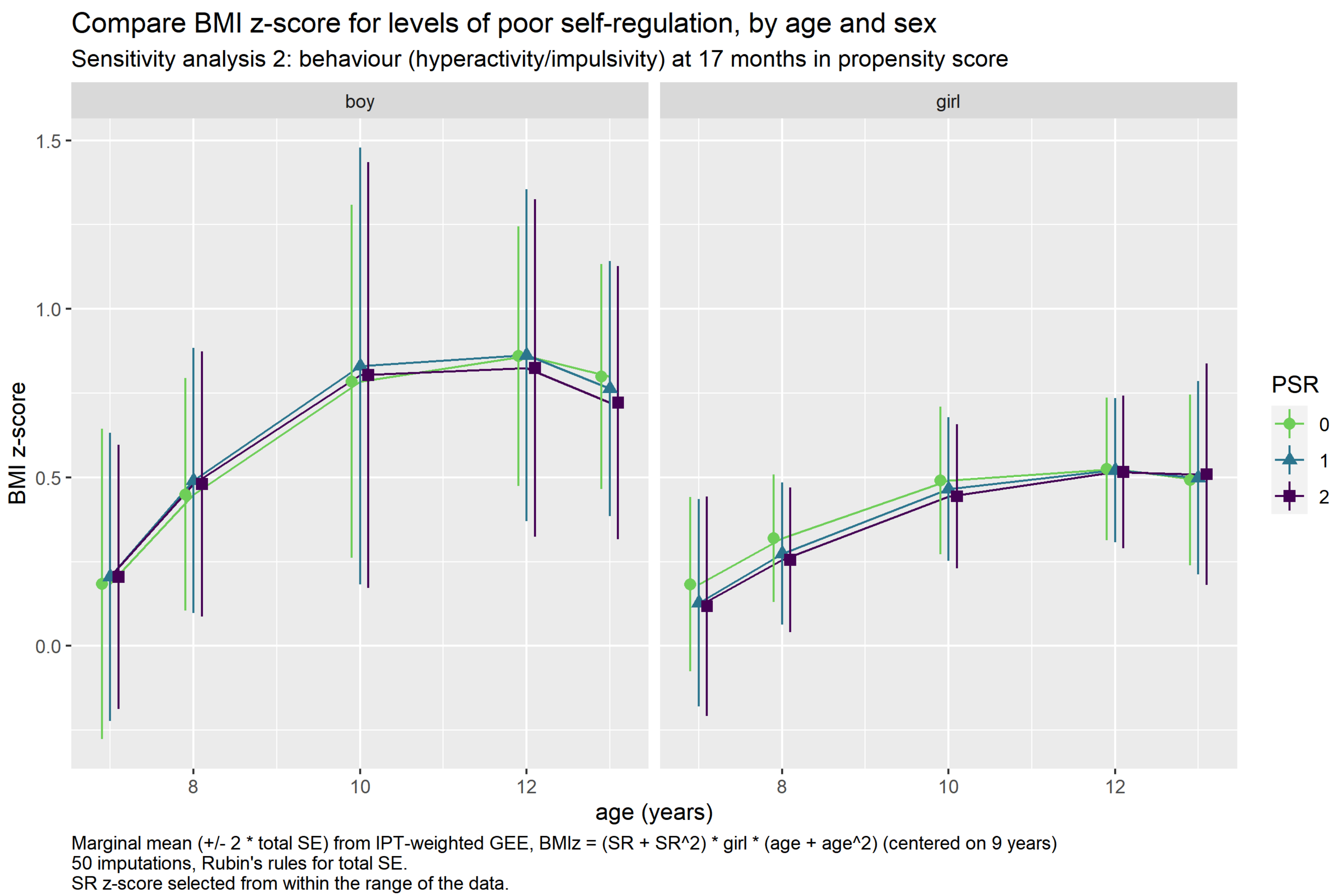


Figure D.10. Mean BMI z-score by sex, age, and recent poor self-regulation score. Sensitivity analysis: 17-month behaviour ratings included in propensity score.

7 References

Abarca-Gómez, L., Abdeen, Z. A., Hamid, Z. A., Abu-Rmeileh, N. M., Acosta-Cazares, B., Acuin, C., Adams, R. J., Aekplakorn, W., Afsana, K., Aguilar-Salinas, C. A., Agyemang, C., Ahmadvand, A., Ahrens, W., Ajlouni, K., Akhtaeva, N., Al-Hazzaa, H. M., Al-Othman, A. R., Al-Raddadi, R., Al Buhairan, F., . . . Ezzati, M. (2017). Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128.9 million children, adolescents, and adults. *The Lancet*, *390*(10113), 2627–2642. https://doi.org/10.1016/S0140-6736(17)32129-3

Adamson, P., & UNICEF. (2013). *Child well-being in rich countries: A comparative overview* (Innocenti report card). Florence.

Aloisio, K. M., Swanson, S. A., Micali, N., Field, A., & Horton, N. J. (2014). Analysis of partially observed clustered data using generalized estimating equations and multiple imputation. *The Stata Journal*, *14*(4), 863–883.

Anderson, S. E., & Whitaker, R. C. (2018). Association of Self-regulation With Obesity in Boys vs Girls in a US National Sample. *JAMA Pediatrics*, *172*(9), 842–850. https://doi.org/10.1001/jamapediatrics.2018.1413

Appleyard, K., Egeland, B., van Dulmen, M. H. M., & Sroufe, L. A. (2005). When more is not better: The role of cumulative risk in child behavior outcomes. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *46*(3), 235–245. https://doi.org/10.1111/j.1469-7610.2004.00351.x

Asada, Y. (2005). A framework for measuring health inequity. *Journal of Epidemiology & Community Health*, *59*(8), 700–705. https://doi.org/10.1136/jech.2004.031054

Atay, Z., & Bereket, A. (2016). Current status on obesity in childhood and adolescence: Prevalence, etiology, co-morbidities and management. *Obesity Medicine*, *3*, 1–9. https://doi.org/10.1016/j.obmed.2016.05.005

Bailey, R., & Jones, S. M. (2019). An Integrated Model of Regulation for Applied Settings. *Clinical Child and Family Psychology Review*, *22*(1), 2–23. https://doi.org/10.1007/s10567-019-00288-y

Baker, M., Gruber, J., & Milligan, K. (2015). *Non-Cognitive Deficits and Young Adult Outcomes: The Long-Run Impacts of a Universal Child Care Program* (NBER working paper No. 21571). https://www.nber.org/papers/w21571

Baker, M., Gruber, J., & Milligan, K. (2019). The Long-Run Impacts of a Universal Child Care Program. *American Economic Journal: Economic Policy*, *11*(3), 1–26. https://doi.org/10.1257/pol.20170603

Barnett, W. S., Brown, K., Shore, R., & (None). (2004). *The Universal vs Targeted Debate: Should the United States Have Preschool for All?* (Preschool Policy Matters, No. 6). New Jersey. The National Institute for Early Education Research.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bell, L. M., Byrne, S., Thompson, A., Ratnam, N., Blair, E., Bulsara, M., Jones, T. W., & Davis, E. A. (2007). Increasing body mass index z-score is continuously associated with complications of overweight in children, even in the healthy weight range. *The Journal of Clinical Endocrinology and Metabolism*, *92*(2), 517–522. https://doi.org/10.1210/jc.2006-1714

Benjamin Neelon, S. E., Schmidt Morgen, C., Kamper-Jørgensen, M., Oken, E., Gillman, M. W., Gallis, J. A., & Sørensen, T. I. A. (2018). Childcare before age 6 and body mass index at age 7 years in a cohort of Danish children. *Pediatric Obesity*, *13*(5), 307–311. https://doi.org/10.1111/ijpo.12206

Black, L., Matvienko-Sikar, K., & Kearney, P. M. (2017). The association between childcare arrangements and risk of overweight and obesity in childhood: A systematic review. *Obesity Reviews : An Official Journal of the International Association for the Study of Obesity*, *18*(10), 1170–1190. https://doi.org/10.1111/obr.12575

Blair, C., Kuzawa, C. W., & Willoughby, M. T. (2019). The development of executive function in early childhood is inversely related to change in body mass index: Evidence for an energetic tradeoff? *Developmental Science*, e12860. https://doi.org/10.1111/desc.12860

Boone-Heinonen, J., Messer, L., Andrade, K., & Takemoto, E. (2016). Connecting the Dots in Childhood Obesity Disparities: A Review of Growth Patterns from Birth to Pre-Adolescence. *Current Epidemiology Reports*, *3*(1), 113–124. https://doi.org/10.1007/s40471-016-0065-9

Bowling, A. B., Tiemeier, H. W., Jaddoe, V. W. V., Barker, E. D., & Jansen, P. W. (2018). Adhd symptoms and body composition changes in childhood: A longitudinal study evaluating directionality of associations. *Pediatric Obesity*, *13*(9), 567–575. https://doi.org/10.1111/ijpo.12288

Brookhart, M. A., Schneeweiss, S., Rothman, K. J., Glynn, R. J., Avorn, J., & Stürmer, T. (2006). Variable selection for propensity score models. *American Journal of Epidemiology*, *163*(12), 1149–1156. https://doi.org/10.1093/aje/kwj149

Bruce, N. (2019). *Impacts of Subsidized Childcare on Obesity in the Long-run* [BA honours]. University of Victoria, Victoria, British Columbia, Canada.

Bundy, D. A. P., de Silva, N., Horton, S., Jamison, D. T., & Patton, G. C. (2017). Child and Adolescent Health and Development. Advance online publication. https://doi.org/10.1596/978-1-4648-0423-6

Bürkner, P.-C. (2017). brms : An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*(1). https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, *10*(1), 395. https://doi.org/10.32614/RJ-2018-017

Bürkner, P.-C. (2019). Bayesian Item Response Modelling in R with brms and Stan. http://arxiv.org/pdf/1905.09501v2

Campbell, F., Conti, G., Heckman, J. J., Moon, S. H., Pinto, R., Pungello, E., & Pan, Y. Y. (2014). Early Childhood Investments Substantially Boost Adult Health. *Science*, *343*(6178), 1478–1485. https://doi.org/10.1126/science.1248429

Carneiro, P., & Ginja, R. (2014). Long-Term Impacts of Compensatory Preschool on Health and Behavior: Evidence from Head Start. *American Economic Journal: Economic Policy*, *6*(4), 135–173. https://doi.org/10.1257/pol.6.4.135

Carpenter, J., Andrykowski, M., Wilson, J., Hall, L., Kay Rayens, M., Sachs, B., & Cunningham, L. (1998). Psychometrics for Two Short Forms of the Center for Epidemiologic

Studies-Depression Scale. *Issues in Mental Health Nursing*, *19*(5), 481–494. https://doi.org/10.1080/016128498248917

Carpiano, R. M., Lloyd, J. E., & Hertzman, C. (2009). Concentrated affluence, concentrated disadvantage, and children's readiness for school: A population-based, multi-level investigation. *Social Science & Medicine*, *69*(3), 420–432. https://doi.org/10.1016/j.socscimed.2009.05.028

Chen, Q., Gelman, A., Tracy, M., Norris, F. H., & Galea, S. (2015). Incorporating the sampling design in weighting adjustments for panel attrition. *Statistics in Medicine*, *34*(28), 3637–3647. https://doi.org/10.1002/sim.6618

Cole, T. J., Bellizzi, M. C., Flegal, K. M., & Dietz, W. H. (2000). Establishing a standard definition for child overweight and obesity worldwide: International survey. *BMJ (Clinical Research Ed.)*, *320*(7244), 1240–1243. https://doi.org/10.1136/bmj.320.7244.1240

Cole, T. J., Flegal, K. M., Nicholls, D., & Jackson, A. A. (2007). Body mass index cut offs to define thinness in children and adolescents: International survey. *BMJ (Clinical Research Ed.)*, *335*(7612), 194. https://doi.org/10.1136/bmj.39238.399444.55

Conti, G., Heckman, J. J., & Pinto, R. (2015). The Effects of Two Influential Early Childhood Interventions on Health and Healthy Behaviors. Advance online publication. https://doi.org/10.3386/w21454

Costa, S., Adams, J., Gonzalez-Nahm, S., & Benjamin Neelon, S. E. (2017). Childcare in Infancy and Later Obesity: A Narrative Review of Longitudinal Studies. *Current Pediatrics Reports*, *5*(3), 118–131. https://doi.org/10.1007/s40124-017-0134-7

Côté, S. M., Borge, A. I., Geoffroy, M.-C., Rutter, M., & Tremblay, R. E. (2008). Nonmaternal care in infancy and emotional/behavioral difficulties at 4 years old: Moderation by family risk characteristics. *Developmental Psychology*, *44*(1), 155–168. https://doi.org/10.1037/0012-1649.44.1.155

Datar, A., & Chung, P. J. (2018). Childhood Self-Control and Adolescent Obesity: Evidence from Longitudinal Data on a National Cohort. *Childhood Obesity (Print)*, *14*(4), 238–247. https://doi.org/10.1089/chi.2017.0217

Datta Gupta, N., & Simonsen, M. (2010). Non-cognitive child outcomes and universal high quality child care. *Journal of Public Economics*, *94*(1-2), 30–43. https://doi.org/10.1016/j.jpubeco.2009.10.001

De Boeck, P., Bakker, M., Zwitser, R., Nivard, M., Hofman, A [Abe], Tuerlinckx, F., & Partchev, I. (2011). The Estimation of Item Response Models with the lmer Function from the lme4 Package in R. *Journal of Statistical Software*, *39*(12).

Desrosiers, H., Dumitru, V., & Dubois, L. (2009). Le surplus de poids chez les enfants de 4 à 7 ans : des cibles pour l'action. *Étude Longitudinale Du Développement Des Enfants Du Québec (ÉLDEQ 1998-2010)*, *4*(3), 1–19.

Di Cesare, M., Sorić, M., Bovet, P., Miranda, J. J., Bhutta, Z., Stevens, G. A., Laxmaiah, A., Kengne, A.-P., & Bentham, J. (2019). The epidemiological burden of obesity in childhood: A worldwide epidemic requiring urgent action. *BMC Medicine*, *17*(1), 212. https://doi.org/10.1186/s12916-019-1449-8

Dong, Q., Elliott, M. R., & Raghunathan, T. E. (2014). A nonparametric method to generate synthetic populations to adjust for complex sampling design features. *Survey Methodology*, *40*(1), 26–46.

Dornan, P., & Woodhead, M. (2015). How Inequalities Develop through Childhood: Life course evidence from the Young Lives cohort study. *Office of Research DiscussionPaper*(February).

Dubois, L., & Girard, M. (2006). Early determinants of overweight at 4.5 years in a population-based longitudinal study. *International Journal of Obesity (2005)*, *30*(4), 610–617. https://doi.org/10.1038/sj.ijo.0803141

Dubois, L., & Girard, M. (2007). Accuracy of maternal reports of pre-schoolers' weights and heights as estimates of BMI values. *International Journal of Epidemiology*, *36*(1), 132–138. https://doi.org/10.1093/ije/dyl281

Epstein, L. H., & Anzman-Frasca, S. (2017). The Promise of Early Childhood Self-Regulation for Obesity Prevention. *PEDIATRICS*, *139*(5). https://doi.org/10.1542/peds.2017-0389

Fong, C., Hazlett, C., & Imai, K. (2018). Covariate balancing propensity score for a continuous treatment: Application to the efficacy of political advertisements. *The Annals of Applied Statistics*, *12*(1), 156–177. https://doi.org/10.1214/17-AOAS1101

Forest, V., Lanthier, C., Nelissen, M., & Roy, J. (2007). *Meeting Early Childhood Needs: Québec's educational program for childcare services*. Québec, Montréal. Ministère de la famille et des aînés.

Fortin, P. (2018). Quebec's Childcare Program at 20: How it has done, and what the rest of Canada can learn. *Inroads*(42). http://inroadsjournal.ca/quebecs-childcare-program-20-2/

Francis, L. A., & Susman, E. J. (2009). Self-regulation and rapid weight gain in children from age 3 to 12 years. *Archives of Pediatrics & Adolescent Medicine*, *163*(4), 297–302. https://doi.org/10.1001/archpediatrics.2008.579

Freedman, D. S., & Sherry, B. (2009). The validity of BMI as an indicator of body fatness and risk among children. *PEDIATRICS*, *124 Suppl 1*, S23-34. https://doi.org/10.1542/peds.2008-3586E

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, *33*(1), 1–22. http://www.jstatsoft.org/v33/i01/

Friendly, M., Beach, J., & Turiano, M. (2002). *Early childhood education and care in Canada 2001* (5th ed.). Childcare Resource and Research Unit, Centre for Urban and Community Studies, University of Toronto.

Friendly, M., & Childcare Resource and Research Unit. (1997). *Child Care in Canada: Provinces and Territories, 1995*. Toronto. Centre for Urban and Community Studies, University of Toronto.

Friendly, M., Larsen, E., & Feltham, L. (2018). *Early Childhood Education and Care in Canada 2016*. Childcare Resource and Research Unit.

Fuemmeler, B. F., Østbye, T., Yang, C., McClernon, F. J., & Kollins, S. H. (2011). Association between attention-deficit/hyperactivity disorder symptoms and obesity and hypertension in early adulthood: A population-based study. *International Journal of Obesity (2005)*, *35*(6), 852–862. https://doi.org/10.1038/ijo.2010.214

Gelman, A., & Pardoe, I. (2006). Bayesian Measures of Explained Variance and Pooling in Multilevel (Hierarchical) Models. *Technometrics*, *48*(2), 241–251. https://doi.org/10.1198/004017005000000517

Gelman, A., & Pardoe, I. (2007). Average Predictive Comparisons for Models with Nonlinearity, Interactions, and Variance Components. *Sociological Methodology*, *37*(1), 23–51. https://doi.org/10.1111/j.1467-9531.2007.00181.x

Geoffroy, M.-C., Power, C., Touchette, E., Dubois, L., Boivin, M., Séguin, J. R., Tremblay, R. E., & Côté, S. M. (2013). Childcare and overweight or obesity over 10 years of follow-up. *The Journal of Pediatrics*, *162*(4), 753-758.e1. https://doi.org/10.1016/j.jpeds.2012.09.026

Geserick, M., Vogel, M., Gausche, R., Lipek, T., Spielau, U., Keller, E., Pfäffle, R., Kiess, W., & Körner, A. (2018). Acceleration of BMI in Early Childhood and Risk of Sustained Obesity. *The New England Journal of Medicine*, *379*(14), 1303–1312. https://doi.org/10.1056/NEJMoa1803527

Gialamas, A., Mittinty, M. N., Sawyer, M. G., Zubrick, S. R., & Lynch, J. (2015). Time spent in different types of childcare and children's development at school entry: An Australian longitudinal study. *Archives of Disease in Childhood*, *100*(3), 226–232. https://doi.org/10.1136/archdischild-2014-306626

Gomajee, R., El-Khoury, F., Côté, S., van der Waerden, J., Pryor, L., & Melchior, M. (2018). Early childcare type predicts children's emotional and behavioural trajectories into middle childhood. Data from the EDEN mother-child cohort study. *Journal of Epidemiology & Community Health*, *72*(11), 1033–1043. https://doi.org/10.1136/jech-2017-210393

Gorter, R., Fox, J.-P., & Twisk, J. W. R. (2015). Why item response theory should be used for longitudinal questionnaire data analysis in medical research. *BMC Medical Research Methodology*, *15*, Article 55, 1–12. https://doi.org/10.1186/s12874-015-0050-x

Gortmaker, S. L., Wang, Y. C., Long, M. W., Giles, C. M., Ward, Z. J., Barrett, J. L., Kenney, E. L., Sonneville, K. R., Afzal, A. S., Resch, S. C., & Cradock, A. L. (2015). Three Interventions That Reduce Childhood Obesity Are Projected To Save More Than They Cost To Implement. *Health Affairs (Project Hope)*, *34*(11), 1932–1939. https://doi.org/10.1377/hlthaff.2015.0631

Greenland, S. (2006). Bayesian perspectives for epidemiological research: I. Foundations and basic methods. *International Journal of Epidemiology*, *35*(3), 765–775. https://doi.org/10.1093/ije/dyi312

Haeck, C., Lebihan, L., & Merrigan, P. (2018). Universal Child Care and Long-Term Effects on Child Well-Being: Evidence from Canada. *Journal of Human Capital*, *12*(1), 38–98. https://doi.org/10.1086/696702

Harms, T. (Ed.). (1993). *Collection Ressources et petite enfance*. *Un univers à découvrir: Grille d'évaluation des services de garde en milieu familial* (2e éd. rev. et corr). Gouvernement du Québec, Office des services de garde à l'enfance.

Harms, T., Clifford, R. M., & Cryer, D. (1998). *Early childhood environment rating scale* (Rev ed.). Teachers College Press.

Harrell, F. E. (2015). *Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis / Frank E. Harrell, Jr*. *Springer series in statistics, 0172-7397*. Springer.

Hernán, M. A. (2018). The C-Word: Scientific Euphemisms Do Not Improve Causal Inference From Observational Data. *American Journal of Public Health*, *108*(5), 616–619. https://doi.org/10.2105/AJPH.2018.304337

Hernán, M. A., & Robins, J. M. (2020). *Causal Inference: What If*. Chapman & Hall/CRC. https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/

Holland, P. W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association*, *81*(396), 945–960. https://doi.org/10.1080/01621459.1986.10478354

Howard, S. J., & Williams, K. E. (2018). Early Self-Regulation, Early Self-Regulatory Change, and Their Longitudinal Relations to Adolescents' Academic, Health, and Mental Well-Being Outcomes. *Journal of Developmental and Behavioral Pediatrics : JDBP*, *39*(6), 489–496. https://doi.org/10.1097/DBP.0000000000000578

Imai, K., & Ratkovic, M. (2014). Covariate balancing propensity score. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *76*(1), 243–263. https://doi.org/10.1111/rssb.12027

Institut de la statistique du Québec. *Enquête québécoise sur la qualité des services de garde éducatifs: Grandir en qualité, 2003*. Québec Québec. Institut de la statistique du Québec.

Institut de la statistique du Québec. (2016). *Data Collection Tools.* Gouvernement du Québec. I am, I'll be (QLSCD).

https://www.jesuisjeserai.stat.gouv.qc.ca/informations_chercheurs/outils_collecte/outils_collecte_an.html

Jacobson, L. A., Pritchard, A. E., Koriakin, T. A., Jones, K. E., & Mahone, E. M. (2016). Initial Examination of the BRIEF2 in Clinically Referred Children With and Without ADHD Symptoms. *Journal of Attention Disorders.* Advance online publication. https://doi.org/10.1177/1087054716663632

Japel, C., Tremblay, R. E., & Côté, S. M. (2005). Quality counts! Assessing the Quality of Daycare Services Based on the Quebec Longitudinal Study of Child Development. *Choices*, *11*(5), 4–42.

Jenson, J. (2002). Against the Current: Child Care and Family Policy in Quebec. In S. Michel & R. Mahon (Eds.), *Child care policy at the crossroads : gender and welfare state restructuring*.

Jester, J. M., Nigg, J. T., Adams, K., Fitzgerald, H. E., Puttler, L. I., Wong, M. M., & Zucker, R. A. (2005). Inattention/hyperactivity and aggression from early childhood to adolescence: Heterogeneity of trajectories and differential influence of family environment characteristics. *Development and Psychopathology*, *17*(1), 99–125. https://doi.org/10.1017/S0954579405050066

Jetté, M. (2002). *Survey Description and Methodology – Part I: Logistics and Longitudinal Data Collections* (Québec Longitudinal Study of Child Development (QLSCD 1998-2002) – From Birth to 29 Months Vol. 2, No. 1). Québec. Institut de la statistique du Québec. http://www.jesuisjeserai.stat.gouv.qc.ca/publications/bebe_v2no1_an.html

Kakinami, L., Séguin, L., Lambert, M., Gauvin, L., Nikiéma, B., & Paradis, G. (2014). Poverty's latent effect on adiposity during childhood: Evidence from a Québec birth cohort. *Journal of Epidemiology & Community Health*, *68*(3), 239–245. https://doi.org/10.1136/jech-2012-201881

Kaufman, J. S. (2019). Commentary: Causal Inference for Social Exposures. *Annual Review of Public Health*, *40*, 7–21. https://doi.org/10.1146/annurev-publhealth-040218-043735

Kautz, T., Heckman, J. J., Diris, R., Weel, B. ter, & Borghans, L. (2014). *Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success* (NBER working paper No. 20749). OECD. https://www.nber.org/papers/w20749

Keil, A. P., Daza, E. J., Engel, S. M., Buckley, J. P., & Edwards, J. K. (2018). A Bayesian approach to the g-formula. *Statistical Methods in Medical Research*, *27*(10), 3183–3204. https://doi.org/10.1177/0962280217694665

Knowlden, A. P., Sharma, M., Cottrell, R. R., Wilson, B. R. A., & Johnson, M. L. (2015). Impact evaluation of Enabling Mothers to Prevent Pediatric Obesity through Web-Based Education and Reciprocal Determinism (EMPOWER) Randomized Control Trial. *Health Education & Behavior : The Official Publication of the Society for Public Health Education*, *42*(2), 171–184. https://doi.org/10.1177/1090198114547816

Kottelenberg, M. J., & Lehrer, S. F. (2013). New Evidence on the Impacts of Access to and Attending Universal Child-Care in Canada. *Canadian Public Policy / Analyse De Politiques*, *39*(2), 263–285.

Kottelenberg, M. J., & Lehrer, S. F. (2014). Do the Perils of Universal Childcare Depend on the Child's Age? *CESifo Economic Studies*, *60*(2), 338–365. https://doi.org/10.1093/cesifo/ifu006

Lang, K. M., & Little, T. D. (2016, May 24). *Don't be Fancy. Impute your Dependent Variables!* 6th Annual Modern Modeling Methods M3 Conference, Storrs, CT.

Lang, K. M., & Little, T. D. (2018). Principled Missing Data Treatments. *Prevention Science : The Official Journal of the Society for Prevention Research*, *19*(3), 284–294. https://doi.org/10.1007/s11121-016-0644-5

Laurin, I., Guay, D., Fournier, M., Bigras, N., & Solis, A. (2016). La fréquentation d'un service éducatif préscolaire : un facteur de protection pour le développement des enfants de familles à faible revenu? [Not Available]. *Canadian Journal of Public Health / Revue Canadienne de Sante'e Publique*, *106*(7 Suppl 2), eS14-20. https://doi.org/10.17269/cjph.106.4825

Laurin, I., Lavoie, S., Guay, D., Boucheron, L., Durand, D., & Goulet, N. (2012). Enquête sur le développement des enfants montréalais à leur entrée à l'école. *Santé Publique*, *24*(1), 7. https://doi.org/10.3917/spub.121.0007

Lawlor, D. A., Benfield, L., Logue, J., Tilling, K., Howe, L. D., Fraser, A., Cherry, L., Watt, P., Ness, A. R., Davey Smith, G., & Sattar, N. (2010). Association between general and central adiposity in childhood, and change in these, with cardiovascular risk factors in adolescence: Prospective cohort study. *BMJ (Clinical Research Ed.)*, *341*, c6224. https://doi.org/10.1136/bmj.c6224

Lefebvre, P., Merrigan, P., & Roy-Desrosiers, F. (2011). *Quebec's Childcare Universal Low Fees Policy 10 Years After: Effects, Costs, and Benefits* (Working Paper 11-01). Montreal. CIRPEE: Centre Interuniversitaire sur le Risque, les Politiques Economiques et l »emploi.

Lemoine, N. P. (2019). Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. *Oikos*, *128*(7), 912–928. https://doi.org/10.1111/oik.05985

Lenth, R. (2020). *emmeans* (Version 1.4.6) [Computer software]. R Foundation for Statistical Computing. https://CRAN.R-project.org/package=emmeans

Lin, B., Liew, J., & Perez, M. (2019). Measurement of self-regulation in early childhood: Relations between laboratory and performance-based measures of effortful control and executive functioning. *Early Childhood Research Quarterly*, *47*, 1–8. https://doi.org/10.1016/j.ecresq.2018.10.004

Loeffler-Wirth, H., Vogel, M., Kirsten, T., Glock, F., Poulain, T., Körner, A., Loeffler, M., Kiess, W., & Binder, H. (2018). Longitudinal anthropometry of children and adolescents using 3D-body scanning. *PLOS ONE*, *13*(9), e0203628. https://doi.org/10.1371/journal.pone.0203628

Lucie and André Chagnon Foundation. (2019). *Preschool programs*. https://naitreetgrandir.com/en/feature/kindergarten/preschool-programs/

Lumeng, J. C., Miller, A. L., Horodynski, M. A., Brophy-Herb, H. E., Contreras, D., Lee, H., Sturza, J., Kaciroti, N., & Peterson, K. E. (2017). Improving Self-Regulation for Obesity Prevention in Head Start: A Randomized Controlled Trial. *PEDIATRICS*, *139*(5). https://doi.org/10.1542/peds.2016-2047

Mazza, J. R. S. E., Lambert, J., Zunzunegui, M. V., Tremblay, R. E., Boivin, M., & Côté, S. M. (2017). Early adolescence behavior problems and timing of poverty during childhood: A comparison of lifecourse models. *Social Science & Medicine*, *177*, 35–42. https://doi.org/10.1016/j.socscimed.2017.01.039

McCain, M. N., Mustard, J. M., & Shanker, S. (2007). *Early Years Study 2: Putting Science into Action*. Toronto. Council for Early Child Development.

McCartney, K., Burchinal, M., Clarke-Stewart, A., Bub, K. L., Owen, M. T., & Belsky, J. (2010). Testing a series of causal propositions relating time in child care to children's externalizing behavior. *Developmental Psychology*, *46*(1), 1–17. https://doi.org/10.1037/a0017886

McClelland, M., Geldhof, J., Morrison, F., Gestsdóttir, S., Cameron, C [Claire], Bowers, E., Duckworth, A., Little, T., & Grammer, J. (2018). Self-Regulation. In N. Halfon, C. B. Forrest, R. M. Lerner, & E. M. Faustman (Eds.), *Handbook of Life Course Health Development* (pp. 275–298). Springer International Publishing. https://doi.org/10.1007/978-3-319-47143-3_12

McCoy, D. C. (2019). Measuring Young Children's Executive Function and Self-Regulation in Classrooms and Other Real-World Settings. *Clinical Child and Family Psychology Review*, *22*(1), 63–74. https://doi.org/10.1007/s10567-019-00285-1

McLaren, L., & McIntyre, L. (2014). Conceptualizing child care as a population health intervention: can a strong case be made for a universal approach in Canada, a liberal welfare regime? *Critical Public Health*, *24*(4), 418–428. https://doi.org/10.1080/09581596.2013.803035

McLaren, L., Zarrabi, M., Dutton, D. J., Auld, M. C., & Emery, J. C. H. (2012). Child care: Implications for overweight / obesity in Canadian children? *Chronic Diseases and Injuries in Canada*, *33*(1), 1–11.

Medvedyuk, S., Ali, A., & Raphael, D. (2018). Ideology, obesity and the social determinants of health: a critical analysis of the obesity and health relationship. *Critical Public Health*, *28*(5), 573–585. https://doi.org/10.1080/09581596.2017.1356910

Melhuish, E., Ereky-Stevens, K., Petrogiannis, K., Ariescu, A., Penderi, E., Rentzou, K., Tawell, A., Leseman, P., & Broekhuisen, M. (2015). *A review of research on the effects of early childhood education and care (ECEC) on child development* (Early childhood education and care: promoting quality for individual, social and economic benefits 613318; FP7-SSH-2013-2). CARE: Curriculum Quality Analysis and Impact Review of European Early Childhood Education and Care (ECEC). http://ecec-care.org/fileadmin/careproject/Publications/reports/summaries/D4__1_EcecutiveSummary.pdf

Michel, S., & Mahon, R. (Eds.). (2002). *Child care policy at the crossroads : gender and welfare state restructuring.*

Miller, A. L., Gearhardt, A. N., Fredericks, E. M., Katz, B., Shapiro, L. F., Holden, K., Kaciroti, N., Gonzalez, R., Hunter, C., & Lumeng, J. C. (2018). Targeting self-regulation to promote health behaviors in children. *Behaviour Research and Therapy*, *101*, 71–81. https://doi.org/10.1016/j.brat.2017.09.008

Miller, E. B., Farkas, G., & Duncan, G. J. (2016). Does Head Start differentially benefit children with risks targeted by the program's service model? *Early Childhood Research Quarterly*, *34*, 1–12. https://doi.org/10.1016/j.ecresq.2015.08.001

Mischel, W., Shoda, Y., & Rodriguez, M. L. (1989). Delay of gratification in children. *Science*, *244*, 933–938.

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., Houts, R., Poulton, R., Roberts, B. W., Ross, S., Sears, M. R., Thomson, W. M., & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(7), 2693–2698. https://doi.org/10.1073/pnas.1010076108

Morinis, J., Maguire, J., Khovratovich, M., McCrindle, B. W., Parkin, P. C., & Birken, C. S. (2012). Paediatric obesity research in early childhood and the primary care setting: The TARGet Kids! Research network. *International Journal of Environmental Research and Public Health*, *9*(4), 1343–1354. https://doi.org/10.3390/ijerph9041343

Muller, C. J., & MacLehose, R. F. (2014). Estimating predicted probabilities from logistic regression: Different methods correspond to different target populations. *International Journal of Epidemiology*, *43*(3), 962–970. https://doi.org/10.1093/ije/dyu029

Nigg, J. T. (2017). Annual Research Review: On the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk-taking, and inhibition for developmental psychopathology. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *58*(4), 361–383. https://doi.org/10.1111/jcpp.12675

O'Loughlin, J., Gray-Donald, K., Paradis, G., & Meshefedjian, G. (2000). One- and two-year predictors of excess weight gain among elementary schoolchildren in multiethnic, low-income, inner-city neighborhoods. *American Journal of Epidemiology*, *152*(8), 739–746. https://doi.org/10.1093/aje/152.8.739

Ontario Agency for Health Protection and Promotion, Singh, T., & Tyler, I. (2015). *Evidence Brief: Comparing self/proxy report versus direct measures of height, weight and BMI in 0–18 year olds* (Evidence Briefs). Toronto. Public Health Ontario.

Pandey, A [Anuja], Hale, D., Das, S., Goddings, A.-L., Blakemore, S.-J., & Viner, R. M. (2018). Effectiveness of Universal Self-regulation-Based Interventions in Children and Adolescents: A Systematic Review and Meta-analysis. *JAMA Pediatrics*, *172*(6), 566–575. https://doi.org/10.1001/jamapediatrics.2018.0232

Pearce, A., Li, L., Abbas, J., Ferguson, B., Graham, H., Law, C., & Millenium Cohort Study Child Health Group (2010). Is childcare associated with the risk of overweight and obesity in the early years? Findings from the UK Millennium Cohort Study. *International Journal of Obesity*, *34*(7), 1160–1168. https://doi.org/10.1038/ijo.2010.15

Pearce, A., Scalzi, D., Lynch, J., & Smithers, L. G. (2016). Do thin, overweight and obese children have poorer development than their healthy-weight peers at the start of school? Findings from a South Australian data linkage study. *Early Childhood Research Quarterly*, *35*, 85–94. https://doi.org/10.1016/j.ecresq.2015.10.007

Peterson, C. M., Su, H., Thomas, D. M., Heo, M., Golnabi, A. H., Pietrobelli, A., & Heymsfield, S. B. (2017). Tri-Ponderal Mass Index vs Body Mass Index in Estimating Body Fat During Adolescence. *JAMA Pediatrics*, *171*(7), 629–636. https://doi.org/10.1001/jamapediatrics.2017.0460

Peth-Pierce, R. (1998). *The NICHD Study of Early Child Care* (NIH-Pub-98-4318). Bethesda, MD. National Inst. of Child Health and Human Development (NIH). https://www.readingrockets.org/article/nichd-study-early-child-care

Piche, G., Fitzpatrick, C., & Pagani, L. S. (2012). Kindergarten Self-Regulation As a Predictor of Body Mass Index and Sports Participation in Fourth Grade Students. *Mind, Brain, and Education*, *6*(1), 19–26.

Public Health Agency of Canada. (2011). *Curbing Childhood Obesity: A Federal, Provincial and Territorial Framework for Action to Promote Healthy Weights*. Ottawa. Public Health Agency of Canada.

Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and longitudinal modeling using Stata: Volume I: Continuous Responses* (3rd ed., Vol. 1). Stata Press Publication.

Rabe-Hesketh, S., & Skrondal, A. (2016). Generalized Linear Latent and Mixed Modeling. In W. J. van der Linden (Ed.), *Chapman & Hall/CRC statistics in the social and behavioral sciences series: Vol. 1. Handbook of item response theory. Volume 1, Models* (pp. 503–526). Chapman & Hall/CRC.

Rademacher, A., & Koglin, U. (2018). The concept of self-regulation and preschoolers' social-emotional development: a systematic review. *Early Child Development and Care*, *3*, 1–19. https://doi.org/10.1080/03004430.2018.1450251

Raghunathan, T. E. (2015). *Missing data analysis in practice* (1st). *Chapman & Hall/CRC interdisciplinary statistics series*. Chapman & Hall/CRC.

Raghunathan, T. E., Solenberger, P., Berglund, P. A., & Hoewyk, J. van. (2016). *IVEware* (Version 0.3) [Computer software]. The Regents of the University of Michigan. https://www.src.isr.umich.edu/software/iveware-documentation/

Rao, D. P., Kropac, E., Do, M. T., Roberts, K. C., & Jayaraman, G. C. (2016). Tendances en matière d'embonpoint et d'obésité chez les enfants au Canada [Childhood overweight and obesity trends in Canada]. *Health Promotion and Chronic Disease Prevention in Canada : Research, Policy and Practice*, *36*(9), 194–198. https://doi.org/10.24095/hpcdp.36.9.03

Renk, K. (2005). Cross-Informant Ratings of the Behavior of Children and Adolescents: The "Gold Standard". *Journal of Child and Family Studies*, *14*(4), 457–468. https://doi.org/10.1007/s10826-005-7182-2

Roberts, K. C., Shields, M [Margot], de Groh, M., Aziz, A., & Gilbert, J.-A. (2012). Overweight and obesity in children and adolescents: Results from the 2009 to 2011 Canadian health measures survey. *Health Reports*, *23*(3), 37–41.

Rose, G. (1985). Sick individuals and sick populations. *International Journal of Epidemiology*, *14*(1), 32–38.

Roth, R. M., Isquith, P. K., & Gioia, G. A. (2014). Assessment of Executive Functioning Using the Behavior Rating Inventory of Executive Function (BRIEF). In S. Goldstein & J. A. Naglieri (Eds.), *Handbook of Executive Functioning* (pp. 301–331). Springer New York. https://doi.org/10.1007/978-1-4614-8106-5_18

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. *Wiley series in probability and mathematical statistics. Applied probability and statistics, 0271-6232*. Wiley.

Rubin, D. B. (2007). The design versus the analysis of observational studies for causal effects: Parallels with the design of randomized trials. *Statistics in Medicine*, *26*(1), 20–36. https://doi.org/10.1002/sim.2739

Saarela, O., Stephens, D. A., Moodie, E. E. M., & Klein, M. B. (2015). On Bayesian estimation of marginal structural models. *Biometrics*, *71*(2), 279–288. https://doi.org/10.1111/biom.12269

Schisterman, E. F., Cole, S. R., & Platt, R. W. (2009). Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology (Cambridge, Mass.)*, *20*(4), 488–495. https://doi.org/10.1097/EDE.0b013e3181a819a1

Sezgin, E., & Demiriz, S. (2019). Effect of play-based educational programme on behavioral self-regulation skills of 48-60 month-old children. *Early Child Development and Care*, *189*(7), 1100–1113. https://doi.org/10.1080/03004430.2017.1369972

Shields, M [Margot], & Tremblay, M. S. (2010). Canadian childhood obesity estimates based on WHO, IOTF and CDC cut-points. *International Journal of Pediatric Obesity : IJPO : An Official Journal of the International Association for the Study of Obesity*, *5*(3), 265–273. https://doi.org/10.3109/17477160903268282

Shpitser, I., VanderWeele, T. J., & Robins, J. M. (2012). On the Validity of Covariate Adjustment for Estimating Causal Effects. In *26th Conference on Uncertainty in Artificial Intelligence, UAI 2010*, Catalina Island, CA. http://arxiv.org/abs/1203.3515

Singh, A. S., Mulder, C., Twisk, J. W. R., van Mechelen, W., & Chinapaw, M. J. M. (2008). Tracking of childhood overweight into adulthood: A systematic review of the literature. *Obesity Reviews : An Official Journal of the International Association for the Study of Obesity*, *9*(5), 474–488. https://doi.org/10.1111/j.1467-789X.2008.00475.x

Smithers, L. G., Sawyer, A. C. P., Chittleborough, C. R., Davies, N., Davey Smith, G., & Lynch, J. (2018). A systematic review and meta-analysis of effects of early life non-cognitive skills on academic, psychosocial, cognitive and health outcomes. *Nature Human Behaviour*, *2*(11), 867–880. https://doi.org/10.1038/s41562-018-0461-x

Snowden, J. M., Rose, S., & Mortimer, K. M. (2011). Implementation of G-computation on a simulated data set: Demonstration of a causal inference technique. *American Journal of Epidemiology*, *173*(7), 731–738. https://doi.org/10.1093/aje/kwq472

Solheim, E., Wichstrøm, L., Belsky, J., & Berg-Nielsen, T. S. (2013). Do time in child care and peer group exposure predict poor socioemotional adjustment in Norway? *Child Development*, *84*(5), 1701–1715. https://doi.org/10.1111/cdev.12071

Stan Development Team. (2018). *RStan: the R interface to Stan* (Version 2.18) [Computer software]. http://mc-stan.org/

Statistics Canada. (2018). *Live births, by place of residence of mother.* https://doi.org/10.25318/1310041401-ENG

Stein, A., Malmberg, L.-E., Leach, P., Barnes, J., & Sylva, K. (2013). The influence of different forms of early childcare on children's emotional and behavioural development at school entry. *Child: Care, Health and Development*, *39*(5), 676–687. https://doi.org/10.1111/j.1365-2214.2012.01421.x

Swyden, K., Sisson, S. B., Lora, K., Castle, S., & Copeland, K. A. (2017). Association of childcare arrangement with overweight and obesity in preschool-aged children: A narrative review of literature. *International Journal of Obesity (2005)*, *41*(1), 1–12. https://doi.org/10.1038/ijo.2016.198

Tremblay, R. E., Loeber, R., Gagnon, C., Charlebois, P., Larivée, S., & LeBlanc, M. (1991). Disruptive boys with stable and unstable high fighting behavior patterns during junior elementary school. *Journal of Abnormal Child Psychology*, *19*(3), 285–300.

U.S Department of Housing and Urban Development. (2007). *Measuring Overcrowding in Housing* (017-002). Office of Policy Development and Research.

U.S. Department of Health and Human Services, Administration for Children and Families. (2010). *Head Start Impact Study: Final Report*. Washington, DC. Administration for Children and Families, U.S. Department of Health and Human Services.

Verbeken, S., Braet, C., Naets, T., Houben, K., & Boendermaker, W. (2018). Computer training of attention and inhibition for youngsters with obesity: A pilot study. *Appetite*, *123*, 439–447. https://doi.org/10.1016/j.appet.2017.12.029

Vogel, M. (2019). *childsds: Data and Methods Around Reference Values in Pediatrics* (Version 0.7.4) [Computer software]. CRAN. https://CRAN.R-project.org/package=childsds

Voigt, K., Nicholls, S. G., & Williams, G. (2014). *Childhood Obesity: Ethical and Policy Issues* (1st edition). Oxford University Press. https://www.amazon.com/dp/B00IOMQ9PQ/ref=r_soa_w_d

Volger, S., Rigassio Radler, D., & Rothpletz-Puglia, P. (2018). Early childhood obesity prevention efforts through a life course health development perspective: A scoping review. *PLOS ONE*, *13*(12), e0209787. https://doi.org/10.1371/journal.pone.0209787

Wang, Y., & Chen, H.-J. (2012). Use of Percentiles and Z-Scores in Anthropometry. In V. R. Preedy (Ed.), *Handbook of Anthropometry* (pp. 29–48). Springer New York. https://doi.org/10.1007/978-1-4419-1788-1_2

Westreich, D., Edwards, J. K., Lesko, C. R., Cole, S. R., & Stuart, E. A. (2019). Target Validity and the Hierarchy of Study Designs. *American Journal of Epidemiology*, *188*(2), 438–443. https://doi.org/10.1093/aje/kwy228

WHO. (2019). *Growth reference 5-19 years: BMI-for-age (5-19 years)*. WHO. https://www.who.int/growthref/who2007_bmi_for_age/en/

WHO Multicentre Growth Reference Study Group (2006). Who Child Growth Standards based on length/height, weight and age. *Acta Paediatrica (Oslo, Norway : 1992). Supplement*, *450*, 76–85. https://doi.org/10.1080/08035320500495548

Wickins-Drazilova, D., & Williams, G. (2011). Ethics and Public Policy. In L. A. Moreno, I. Pigeot, & W. Ahrens (Eds.), *Epidemiology of Obesity in Children and Adolescents* (pp. 7–20). Springer New York. https://doi.org/10.1007/978-1-4419-6039-9_2

Willms, J. D., & Shields, M [M.]. (1996). *A Measure of socioeconomic status for the National Longitudinal Study of Children.: Report prepared for Human Resource Development Canada*. Human Resource Development Canada.

Wolfenden, L., Barnes, C., Jones, J., Finch, M., Wyse, R. J., Kingsland, M., Tzelepis, F., Grady, A., Hodder, R. K., Booth, D., & Yoong, S. L. (2020). Strategies to improve the implementation of healthy eating, physical activity and obesity prevention policies, practices or programmes within childcare services. *The Cochrane Database of Systematic Reviews*, *2*, CD011779. https://doi.org/10.1002/14651858.CD011779.pub3

Yang, S., Charters, T., Weinstock, D., & Nandi, A. *Impacts of Universal Childcare on Behavioral Development: Evidence from a Birth Cohort in Quebec, Canada*.

Zheng, H., & Little, R. J. A. (2003). Penalized Spline Model-Based Estimation of the Finite Populations Total from Probability-Proportional-to-Size Samples. *Journal of Official Statistics*, *19*(2), 99–117.

Zhou, H., Elliott, M. R., & Raghunathan, T. E. (2016). A two-step semiparametric method to accommodate sampling weights in multiple imputation. *Biometrics*, *72*(1), 242–252. https://doi.org/10.1111/biom.12413

Zhou, T., Elliott, M. R., & Little, R. J. A. (2019). Penalized Spline of Propensity Methods for Treatment Comparison. *Journal of the American Statistical Association*, *114*(525), 1–19. https://doi.org/10.1080/01621459.2018.1518234

Zhou, X., & Reiter, J. P. (2010). A Note on Bayesian Inference After Multiple Imputation. *The American Statistician*, *64*(2), 159–163. https://doi.org/10.1198/tast.2010.09109

Zoritch, B., Roberts, I., & Oakley, A. (2009). Day care for pre-school children. In B. Zoritch (Ed.), *Cochrane Database of Systematic Reviews*. John Wiley & Sons, Ltd. https://doi.org/10.1002/14651858.CD000564