

BMJ Open Association of socioeconomic position and childhood obesity in Finland: a registry-based study

Laura Paalanen ¹, Esko Levälahti,¹ Päivi Mäki,¹ Hanna Tolonen,¹ Franco Sassi,² Majid Ezzati,³ Tiina Laatikainen^{1,4}

To cite: Paalanen L, Levälahti E, Mäki P, *et al.* Association of socioeconomic position and childhood obesity in Finland: a registry-based study. *BMJ Open* 2022;**12**:e068748. doi:10.1136/bmjopen-2022-068748

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2022-068748>).

LP and EL are joint first authors.

Received 28 September 2022
Accepted 15 December 2022



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Department of Public Health and Welfare, Finnish Institute for Health and Welfare, Helsinki, Finland

²Centre for Health Economics and Policy Innovation, Imperial College Business School, London, UK

³MRC Centre for Environment and Health, School of Public Health, Imperial College London, London, UK

⁴Institute of Public Health and Clinical Nutrition, Faculty of Health Sciences, University of Eastern Finland, Joensuu, Finland

Correspondence to

Dr Laura Paalanen;
laura.paalanen@thl.fi

ABSTRACT

Objective To identify what dimensions of socioeconomic position (SEP) are most closely associated with childhood obesity in Finland, leveraging population-wide data among the whole child population aged 2–17 years in Finland.

Design Registry-based study.

Setting Data from several administrative registries linked on individual level covering the whole of Finland were used. Data on height and weight measurements in 2018 were obtained from the Register of Primary Health Care visits and data on sociodemographic and socioeconomic indicators (2014–2018) from Statistics Finland.

Participants Children aged 2–17 years with valid height and weight measurements performed at the child health clinic or school healthcare in 2018 (final n=194 423).

Main outcome measures Obesity was defined according to WHO Growth Reference curves. Sociodemographic and socioeconomic indicators were linked on individual level for adults (both parents) who lived in the same household (42 predictors). Boosted regression model was used to analyse the contribution of SEP to obesity.

Results From socioeconomic indicators, annual household income (12.6%) and mother and father's educational level (12.6% and 8.1%, respectively) had the highest relative influence on obesity risk. The relative influence of a child's sex was 7.7%.

Conclusions The parents' SEP was inversely associated with obesity among the offspring. A remarkable number of objective SEP indicators were analysed with parents' education and household income finally being the indicators most strongly associated with obesity among children. In future research, more attention should be paid to reliable and objective ways of measuring educational status and income rather than on developing new SEP indicators. Administrative registries with information on both healthcare and socioeconomic indicators can in future provide better opportunities to assess the influence of SEP on various health risks.

INTRODUCTION

Obesity is a globally growing challenge even among children.^{1 2} Like most health risk factors, obesity is not evenly distributed in the populations, with socioeconomic status being one of the key determinants influencing its risk. It is observed that in low-income countries high socioeconomic position (SEP)

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ The study is based on objective data from standardised height and weight measurements of children and adolescents aged 2–17 years (n=194 423), and on a large set of socioeconomic indicators from population-based Finnish administrative registries.
- ⇒ Height and weight data from the Register of Primary Health Care visits and socioeconomic indicators obtained from Statistics Finland were linked on individual level using the personal identity code.
- ⇒ A large number of objective indicators of socioeconomic position were analysed; altogether, 42 predictors were included in a boosted regression model.
- ⇒ Because of the study setting, self-report bias and selection bias could be virtually avoided.
- ⇒ The imperfect coverage of height and weight data among children aged 2–17 years (40%) is due to challenges in data transfer from different software used in primary healthcare across the country to the national healthcare register.

mostly increases the risk of obesity while in high-income countries, the risk is often increased among those with low SEP, with some exceptions.^{3–8}

As SEP is a descriptive term for the position of a person in society, it is dependent on various societal factors. Even though it is commonly classified based on occupational, economic and educational criteria,⁹ also other factors such as ethnicity, literacy and cultural characteristics are related to SEP.¹⁰ Thus, determining SEP and comparisons between populations are challenging.^{11 12} Even more, regarding children, SEP needs to be designated based on parents or family's characteristics.^{12 13} This complicates the operationalisation of childhood SEP as the indicators can be created in various ways using either information related to one or both parents and/or the household as a whole.^{7 14}

Researchers have also identified challenges in collecting reliable socioeconomic data. Socioeconomic data can be collected from

different data sources and using different methodology. Information from administrative registries, from questionnaires having self-reported data and information relying on neighbourhood statistics may often be discordant.¹⁵ In addition, even the key socioeconomic indicators such as education, occupation and income are not interchangeable and are differently affected by other factors, such as culture, and thus do not necessarily measure the same exposure.^{11 12 16 17}

Income data have been seen to be especially affected by different reporting biases, even more so in low-income settings.¹⁸ For income, three main problems with self-reported survey information have been identified: (1) many forms of income exist, and people may not count all of them, which leads to misreporting; (2) income may be considered as sensitive information which may lead to high proportion of item non-response in surveys; and (3) people may be prone to over-report their income (social desirability bias).^{12 15 19–21}

The wide-ranging administrative, population-based registries in Finland and the possibility of linking data on individual level from several registry sources using the personal identity code give possibility to use objective data sources for examining the associations of sociodemographic and socioeconomic characteristics of children and their families and the risk of childhood obesity. Furthermore, the use of administrative registers allows including several SEP indicators in the analyses simultaneously to examine their relative influence. The aim of this study was to analyse the impact of a large set of registry-based indicators of SEP on childhood obesity, based on measured height and weight, among the whole child population aged 2–17 years in Finland. This study was carried out within the framework of the STOP project (<http://www.stopchildobesity.eu/>).

MATERIALS AND METHODS

Study population

Data of children were achieved from the Register of Primary Health Care visits (Avohilmo) for all children aged 2–17 years who had visited child health clinic or school healthcare between 1 January and 31 December 2018. This registry includes real-time data on healthcare visits and treatments collected from primary healthcare in the public sector. Healthcare visits are regular (at least yearly among 2–15 year-olds and at least once among 16–17 year-olds in high school or vocational school), cover the vast majority of children and include assessment of health and development of children including height and weight measurements.^{22 23} The measurements at child health clinics and at school healthcare are carried out by trained healthcare professionals using standardised measurement protocols.²⁴ From local and regional electronic health records data on height and weight are transferred electronically to the Register of Primary Health Care visits. Since 2011, the data collection has covered all health centres and municipalities in

Finland. Notwithstanding, the coverage of data collection of height and weight of children was approximately 40% in 2018.²⁵

Information extracted for the current study included sex, date of birth and all available height and weight measurements, with measurement date for children aged 2–17 years with a visit in 2018 (n=397 047). To check the measured height and weight values extracted from the Register of Primary Health Care visits, we calculated deviation statistics for height and weight using Finnish growth standard as reference values (relative to weight by sex and height, height by sex and age, body mass index (BMI) by sex and age).²⁶ Height and weight measurements with deviation values outside [−4,4] boundary were excluded. Also, sex and age-specific weight index (ISO-BMI) was calculated based on Finnish growth standard. Height and weight measurements resulting in ISO-BMI≥50 were also excluded. Altogether 1537 children were excluded because no valid height and weight measurements from 2018 were not available. As a result, the number of children who were 2–17 years of age in 2018 and had at least one valid height and weight measurement in 2018 was 395 510.

Socioeconomic indicators

Socioeconomic indicators were obtained from Statistics Finland for adults (both parents) who lived in the same household with a child covering years 2014–2018, and linked on individual level to the height and weight data of children using deterministic record linkage and personal identification code. An extensive amount of information related to SEP was extracted from the registries including parents' age, native language, country of birth, marital status, classification of socioeconomic group, educational level of highest qualification/degree, educational field of highest qualification/degree, occupational status, employed/unemployed, code of occupation, size of family, size of household-dwelling unit, number of children in the family and information whether there are children under 3, 7, 14 and 18 years old in the family, family type, municipality group of municipality of domicile according to the 2016 regional division, region according to the 2016 regional division, mode of housing, living space, annual household's disposable money income (without equalising according to household size), earned total income in state taxation, total capital income, housing benefits and debts in total. Data on SEP of parents were used from year 2017. If this information was missing, it was imputed using data from years 2014–2016 and/or from year 2018. In addition, for adolescents aged ≥16 years, the type of their current education was obtained: (a) adolescent in high school and (b) adolescent in vocational school.

Many variables related to SEP had several categories. Based on logistic regression analyses on binary obesity outcome, some parental multinomial categorical variables were classified into fewer classes: region (from 19 to 4 categories), occupational status (from 8 to 3 categories), educational field of highest qualification/degree

(from 12 to 4 categories for fathers and to 3 categories for mothers) and classification of socioeconomic group for mother/father (from 19 to 5 categories for fathers and to 4 categories for mothers, occupational code from 11 to 5 categories) (see online supplemental table S1 for categorisation of variables included in the final model). Variables like age, income, size of family, etc were kept in the analyses as continuous variables. Additional dichotomous indicators informing whether each parent was living in the same household as the child were created using the addresses of a child and parents: (1) proportion of households with mother and (2) proportion of households with father.

Children with missing data on SEP from Statistics Finland were excluded ($n=7887$). Also, siblings and children having either same mother or father as well as children with two male or two female adults in the household were excluded ($n=193\,200$), so each parent was only once in the data as well as only one child of each family to guarantee that each adult-child pair was independent of each other. The number of families with two male or two female adults was 374. The final data included 194 423 children. The used variables and their distributions are presented in table 1.

Definition of obesity

For the analyses, obesity was defined according to WHO Growth Reference curves.^{27 28} For children under 5 years of age the definition for obesity is weight for height greater than 3 SDs above WHO Child Growth Standards median, and for children over 5 years greater than 2 SDs above the WHO Growth Reference median.

Statistical analyses

Most of the categorical predictor variables on SEP were imputed using values from previous years. In case no previous individual data were available, data from 2018 were used. Adolescents' own education was imputed using values from the previous year. For continuous predictors, individual trends were fitted and used for prediction of missing values. For continuous predictor data with only one observation, one trend line was fitted and used for prediction of missing values.

The analyses were carried out using randomly selected modelling data: training data ($n=155\,479$) and validation data ($n=38\,944$; 20% of the training data).

A machine learning model, boosted regression model, was used to analyse the contribution of parents' and family's socioeconomic characteristics to children's obesity defined as a binary variable.²⁹ The analyses were carried out with R-package Generalized Boosted Regression Models (GBM). GBM was run using 10-fold cross-validation, that is, automatically selecting test data for every 10-fold which was specified as 20% random sample of the training data set. Finally, overfitting was avoided by using the so-called early stopping rule, which means that an optimal number of ensemble models were chosen by the GBM program.

Table 1 Subject characteristics

	n*	% (95% CI)/mean (SD)
Obesity (%)	194 423	8.6 (8.4 to 8.7)
Child's age, mean	194 423	9.6 (4.6)
Child's sex, proportion of girls (%)	194 423	48.5 (48.2 to 48.7)
Annual household disposable money income (€), mean	194 282	60 814 (63 618)
Proportion of households with children <14 years of age (%)	180 108	54.7 (54.5 to 54.9)
Child: municipality class (%)	194 423	
City		73.8 (73.6 to 74.0)
Urban		15.5 (15.3 to 15.7)
Rural		10.7 (10.6 to 10.9)
Proportion of households with mother (%)	194 423	97.7 (97.6 to 97.8)
Proportion of households with father (%)	194 423	81.6 (81.5 to 81.8)
Father's age, mean	168 156	41.6 (8.1)
Mother's age, mean	190 958	39.2 (7.4)
Father: degree of highest education (%)	169 498	
No qualification		13.3 (13.2 to 13.5)
Vocational		43.3 (43.1 to 43.6)
Specialised vocational		1.9 (1.8 to 2.0)
Lowest tertiary		6.8 (6.6 to 6.9)
Bachelor's degree		16.2 (16.1 to 16.4)
Master's degree		16.3 (16.1 to 16.5)
Doctor's degree		2.2 (2.1 to 2.2)
Mother: degree of highest education (%)	190 967	
No qualification		9.4 (9.3 to 9.5)
Vocational		36.0 (35.8 to 36.2)
Specialised vocational		1.1 (1.0 to 1.1)
Lowest tertiary		8.7 (8.5 to 8.8)
Bachelor's degree		22.1 (21.9 to 22.3)
Master's degree		21.0 (20.8 to 21.2)
Doctor's degree		1.8 (1.7 to 1.8)
Proportion of adolescents (aged ≥16 years) in high school (%)	21 205	41.8 (41.1 to 42.5)
Proportion of adolescents (aged ≥16 years) in vocational school (%)	21 205	23.9 (23.3 to 24.5)
*Total number available for analyses. CI, confidence interval; SD, standard deviation.		

For model performance evaluation, deviance and relative influences of predictors were used, and area under the receiver operating characteristic curve (AUC) and percentage of model deviance explained (pseudo- R^2) were calculated. Based on cross-validation, AUC and deviance values were expected to have very small differences between the training and validation data sets. Also, model performance of the GBM model in the excluded data set was evaluated.

Logistic regression approximation of boosted regression prediction was used to calculate contributions of predictors as differences from mean predicted obesity prevalence.

Patient and public involvement

No patients were involved in setting the research question, study design, outcome measures or the conduct of the study. This study is solely based on data from Finnish administrative registries.

RESULTS

Boosted regression modelling and model accuracy evaluation

The full model had 42 predictors included as 73 variables due to splitting of the original multinomial categorical variables into three or more indicator variables. None of the predictors had zero influence. The optimal number of ensemble models was 1530 selected out of the prespecified 2000 models by using a cross-validation method. The final model had 12 predictors as 12 variables, of which none had zero influence and the optimal number of ensemble models was 1237 selected. The AUC of the full model was 0.724 in the validation data set which was 0.025 lower than in the training data set and 0.018 lower than in the excluded data set. In the final model of the validation data set, the AUC was 0.718, which was 0.018 lower than in the training data set (0.003 lower than in the excluded data set). The difference of AUC between the full model and the final model in validation data set was 0.006. The difference between the full and final model AUCs was tested separately in 16 age groups (for ages 2–17) and none of the tests were significant at $p=0.001$ level.

The difference in AUC by sex was not significant ($p=0.108$). The AUC was higher for children aged 2–6 years (0.782) than for children aged 7–17 years (0.655) ($p<0.0001$).

Predicted risk and observed prevalence of childhood obesity

Figure 1 illustrates the predicted risk and the observed prevalence of obesity based on the training data set. The modelling resulted in practically same obesity estimates with the observed prevalences. The risk of being obese was about 1% among children under 5 years of age but increased to 5%–11% among children between 5 and 7 years of age. The risk of being obese was highest among children before puberty being 9.8% among girls and

Table 2 The relative influence of sociodemographic and socioeconomic indicators on childhood obesity

	Relative influence	Direction
Child's age	39.2	2–11: +, 12–15: –, 16–17: +
Mother: degree of highest education	12.6	–*
Annual household disposable money income	12.6	–
Father: degree of highest education	8.1	–
Child's sex	7.7	Girls: –
Father's age	5.3	+
Any children <14 years of age	5.2	If yes: –
Mother's age	4.6	+
Municipality class	3.4	+ (for a more rural municipality)
Adolescent (aged ≥16 years) in high school	0.9	If yes: –
Mother lives in the same household	0.2	If yes: –
Father lives in the same household	0.1	If yes: –

*Special vocational education is almost equal to vocational education.

17.4% among boys. During and after prepuberty and related growth spurt, the risk of being obese declined to about 8% in girls and 12% in boys.

Relative influence of predictors on childhood obesity

Table 2 shows the relative influence of the final 12 predictors. The age of the child had the highest relative influence (39.1%) on the risk of being obese. From socioeconomic indicators, annual household income (12.6%) and mother's and father's educational level of highest degree (12.6% and 8.1%, respectively) had the highest relative influence on obesity risk, with higher risk among the lower household's disposable money income and educational level groups. Relative influence of child's sex was 7.7%, the risk being higher among boys. In addition to factors related to education and income, the number of children under 14 years of age in the family, mother's and father's age, child in high school and municipality

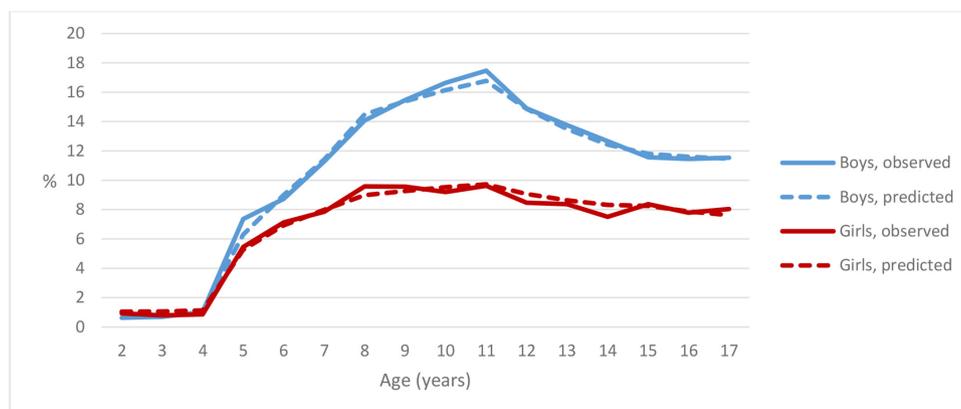


Figure 1 The predicted risk and the observed prevalence of obesity (%) among children according to WHO criteria by sex and age in the large register-based data from Finland in 2018.

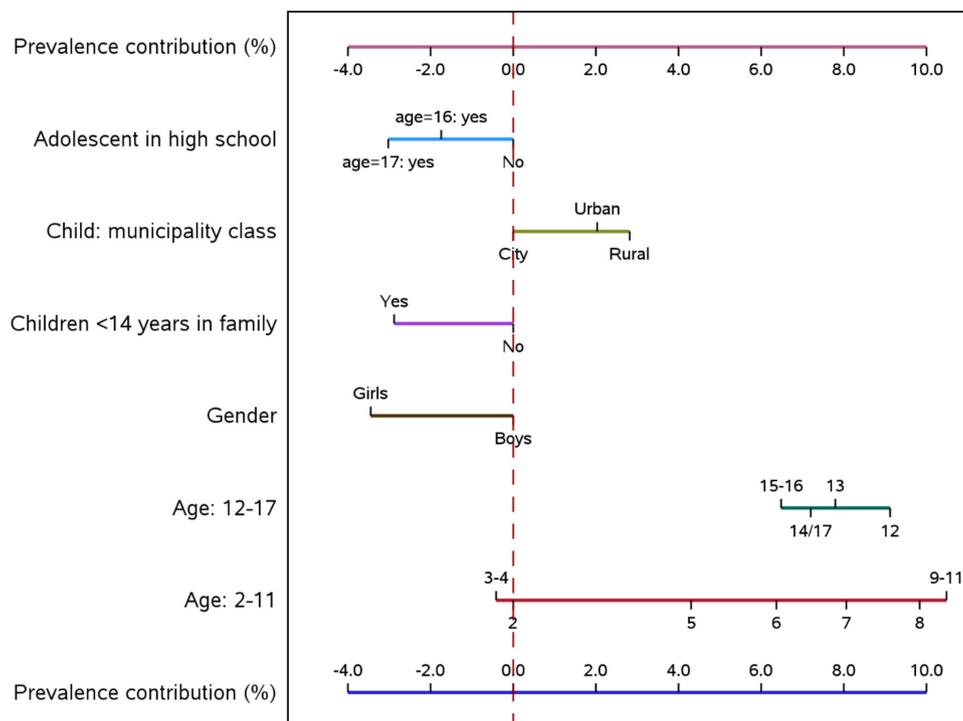


Figure 2 Contributions of child and family predictors to obesity prevalence (mean prevalence=8.6%).

class (relative influence 0.9%–5.3%) influenced obesity risk. Two additional variables were included in the model, indicators of mother or father living in the same address, but both had a very low influence on obesity risk (<0.3%).

Based on multivariate logistic approximation, there were three significant interactions when entered separately to the model with all main effects: age of father*annual household income, age of mother*annual household income and age of child*degree of highest education of father. Only one interaction was significant when entered in full final model with main effects and interactions: age of father*annual household income.

Boosted regression model effects were approximated using multiple logistic regression model for child and family predictor main effects (figure 2) and parental main effects and interaction of father's age and annual household income (figure 3) excluding variables for mother or father living in the same household.

Continuous variables such as annual household income, mother and father's age and child's age were included as categorical (see online supplemental table S1 for categorisation). Some of the category effects which were similar on obesity prevalence were combined for graphical representation. Based on logistic regression model, the prevalence was between 0.3% and 34.0% in the used training data set. In the original boosted regression model, the prevalence was between 0.7% and 66.7%. The prevalence >34.0% (516 predictions) can be interpreted as outliers in prevalence prediction distribution.

Figures 2 and 3 represent the contribution of model predictors to obesity prevalence. For example, the mean effect of being a girl in the multivariate model on obesity prevalence is about 3.5% points less than the overall

mean obesity prevalence (8.6%) (figure 2). Having a mother aged >39 years has a mean effect, which is about 2% points higher than the overall mean prevalence (figure 3). Five-year-old boys with no younger siblings and parents <30 years of age, living in an urban area, and annual household income over €60 000 with mother's highest educational qualification of bachelor's degree and father's highest educational qualification of master's degree have 1.5% (=4.5+0+0+0+2–0.5–4–3.5) lower obesity prevalence than an average child. The wider the line for the predictor, the greater the multivariate contribution of a predictor.

The same applies to interaction for father's age*annual household income, which was categorised to six age groups and to three to six income groups. Children with fathers over 49 years of age have the highest annual household income contribution and children with fathers under 40 years of age (separated into three groups) have the lowest annual household income contribution. Child's age, mother's highest educational qualification and annual household income with father's age over 45 years (separated into two groups) have the highest contribution to children's obesity prevalence.

DISCUSSION

Our study based on objective data from Finnish administrative registry sources showed that child's age and sex and his/her parents' SEP are strong predictors of obesity in childhood. Altogether, 42 registry-based predictors were included in the analyses. Mother's educational level and annual household income were the two socioeconomic indicators with the highest relative influence on

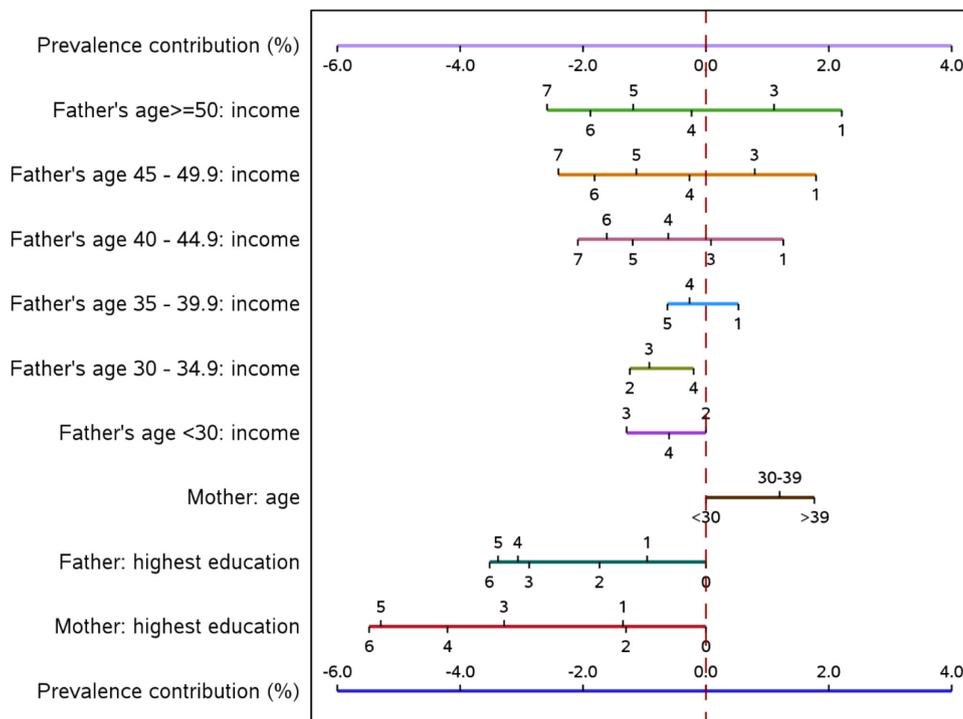


Figure 3 Contributions of parental predictors and interaction of father's age and annual household income (mean prevalence=8.6%).

Parental education categories: 0=no qualification, 1=vocational education, 2=special vocational education, 3=lowest tertiary, 4=bachelor's degree, 5=master's degree, 6=doctoral education. **Annual household income categories (€):** 1=<30 000; 2=30 000–39 999.99; 3=40 000–59 999.99; 4=60 000–69 999.99; 5=70 000–79 999.99; 6=80 000–99 999.99; 7= \geq 100 000. When 'father's age 35–39', categories 2 and 3 were recoded into 4, and categories 6 and 7 were combined into 5. When 'father's age 30–34' or 'father's age <30', category 1 was recoded to 2 and categories 5–7 were recoded into 4. When 'father's age 40–44', 'father's age 45–49' or 'father's age >49', category 2 was recoded into 3.

the risk of obesity among the offspring. Childhood obesity was inversely associated with the included SEP indicators, that is, the risk of obesity was higher among children with lower parental SEP. The number of children in our analyses was remarkable ($n=194\,423$) and enabled sophisticated utilisation of statistical methods such as machine learning models.

Finland has a comprehensive public healthcare system and almost all families with children use these services.^{22–23} The provision of health services, such as health check-ups, is mandatory for municipalities, but services are voluntary and free of charge for families. The coverage is extremely high among children from all socioeconomic groups eliminating the non-response bias commonly seen in survey data.²² In addition to providing an excellent data source, using registry data from health monitoring of children is cost-effective because there are no additional data collection costs. On the other hand, a disadvantage of relying on registry data is that information on behavioural factors, such as on diet or physical activity, is not available. Furthermore, in our study, the data did not allow for accurate inference on the relation of a child and the adults living with the child. As examples, whether the adults of the household were biological parents of a child or whether the family was reconstituted was not known.

The data for the parents and families' sociodemographic and socioeconomic indicators were obtained from Statistics Finland and included information from several Finnish registries. Data from all included registries were linked on individual level using the personal identity code. Again, such data are much more objective compared with self-reported data and do not suffer from awareness or social desirability bias. Furthermore, individual and household-level SEP indicators used in this study are generally seen as more reliable and less tricky to interpret than area-based indicators, which may underestimate the association between SEP and a health outcome.^{30–31}

Although theoretically the whole child population aged 2–17 years was included, the coverage of height and weight data on Finnish children was only around 40% in 2018.²⁵ The imperfect coverage is due to challenges in data transfer from different software used in primary healthcare across the country to the national healthcare register. Therefore, despite the incomplete geographical coverage, the data were highly representative for different socioeconomic groups and the selection bias is minimal.

The age of a child was the most significant predictor of childhood obesity in our study. A remarkable shift in the prevalence of obesity at the age of 5 was seen. The shift was related to the inherent characteristics of WHO

definition of childhood obesity, which is different for children under and over 5 years of age.^{27–28} The risk of being obese was only about 1% among children under 5 years of age but increased to 5%–11% among children between 5 and 7 years of age. Previous studies have shown differences between various childhood obesity definitions pointing to higher prevalence rates with WHO definition compared with the International Obesity Task Force, and national French, Italian as well as Finnish references.^{32–34} Therefore, the selection of childhood obesity definition directly affects the resulting prevalences, and the differences between the definitions are important to acknowledge when comparing results from different studies or countries.

Previous studies have suggested that in high-income countries, the parents' lower SEP is associated with higher risk of childhood obesity, whereas the direction of the association is the opposite in low or middle-income countries.³⁴⁷ Our results from Finland, a high-income country located in northern Europe, support the earlier research literature with consistent inverse association of parents' SEP and obesity among the offspring. In our study, both parents' educational level had an influence on the risk of a child to be obese. However, the association was somewhat stronger for mother's educational level. In general, mother's education has been more often used as a SEP indicator in previous studies than that of father's.⁷ However, previous studies also point to slightly stronger association between mothers' educational level and obesity among children compared with that of fathers'.⁷ In our data, the proportion of missing values for educational level was higher among fathers than mothers, which may have resulted in somewhat less accurate or attenuated estimates regarding the association of fathers' educational level and obesity among the children.

In our data among Finnish children aged 2–17 years, also the annual household income after taxes was a significant indicator of childhood obesity. In previous literature, wealth and family income indicators have been observed to be associated with other health-related outcomes as well, such as mortality.³⁵ Whenever feasible to measure, these indicators would be valuable in studies examining the association of SEP and health behaviour or other health-related outcomes. Most often, however, education is used as the main SEP indicator, as it has proven to be more straightforward to assess. However, the indicators of education and income are not interchangeable, which justifies the use of separate indicators for the different dimensions of SEP such as education, occupation and household income.³⁶ The information on income is regarded as rather sensitive information and therefore is challenging to inquire in surveys.^{12 37} In surveys, income questions tend to have relatively high item non-response rates.²¹ In a large study on the socioeconomic differences in overweight of children in 24 countries from the WHO European Region, self-reported data on SEP indicators were used, and the authors discussed that reporting bias may have occurred particularly for family-perceived

wealth.³ In our study, on the contrary, the information on income, namely the annual household income, was obtained from reliable administrative registry sources, and thereby the challenges related to misreporting or item non-reporting were overcome.

The magnitude of data available for the analyses of the current study was remarkable. Using the unique personal identity code given to every Finnish resident, we were able to link each child's height and weight data to registry data of their parents and household. We obtained data including information on the family composition, household disposable money income and parents' educational level, among others from the abundant registries of Statistics Finland. The influence of a multitude of SEP indicators on childhood obesity was analysed. However, we observed that finally, the indicators which most strongly predicted obesity among children and were selected to the final model were those which quite commonly have been reported in previous studies, namely the parents' educational level and household income. This is probable to apply to many other high-income countries as well. The results of our study thereby support the use of these indicators also in future studies.

To conclude, the commonly used indicators of family SEP, education and income, were most strongly associated with childhood obesity. Thus, in future research, more attention should be paid to ensure standardised, reliable and objective measures of educational status and household income rather than putting effort on developing new SEP indicators. Furthermore, development of administrative registries including information on both healthcare and socioeconomic indicators can provide excellent opportunities to assess the influence of SEP on various health risks in future. This is also a cost-effective way of data collection. However, comprehensive data sources, possibilities to link data on individual level and high-capacity and secure cloud computing platforms to analyse the data following the data protection rules are needed for efficient use of registry data. European Union proposal on European Health Data Space regulation is a step towards this.

Acknowledgements We would like to thank the members of the STOP management team for their valuable advice and support in all phases of the study.

Contributors TL, PM, EL, HT, FS and ME designed the study. EL had access to all the data and performed all statistical analyses. EL is the guarantor and takes responsibility for the content of the study. EL, PM and TL contributed to data acquisition and checking the data. FS and ME provided methodological guidance for the study. TL wrote the first draft of the manuscript. LP was responsible for writing the later version as well as the final version of the manuscript. All authors evaluated and interpreted the results of the analyses; read and commented on the manuscript; and gave final approval for the version to be submitted. LP and EL contributed equally to this paper.

Funding The STOP project (<http://www.stopchildobesity.eu/>) received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement number 774548. The STOP consortium is coordinated by Imperial College London and includes 24 organisations across Europe, the USA and New Zealand.

Disclaimer The content of this publication reflects only the views of the authors, and the European Commission is not liable for any use that may be made of the information it contains.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Laura Paalanen <http://orcid.org/0000-0003-2738-2654>

REFERENCES

- NCD Risk Factor Collaboration (NCD-RisC). 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. *Lancet* 2017;390:2627–42.
- World Health Organization. *WHO European regional obesity report 2022*. Copenhagen: WHO Regional Office for Europe, 2022. <https://apps.who.int/iris/bitstream/handle/10665/353747/9789289057738-eng.pdf>
- Buonocristiano M, Williams J, Simmonds P, et al. Socioeconomic inequalities in overweight and obesity among 6- to 9-year-old children in 24 countries from the World Health Organization European region. *Obes Rev* 2021;22 Suppl 6:e13213.
- Chung A, Backholer K, Wong E, et al. Trends in child and adolescent obesity prevalence in economically advanced countries according to socioeconomic position: a systematic review. *Obes Rev* 2016;17:276–95.
- Brisbois TD, Farmer AP, McCargar LJ. Early markers of adult obesity: a review. *Obes Rev* 2012;13:347–67.
- Wang Y, Lim H. The global childhood obesity epidemic and the association between socio-economic status and childhood obesity. *Int Rev Psychiatry* 2012;24:176–88.
- Sares-Jäske L, Grönqvist A, Mäki P, et al. Family socioeconomic status and childhood adiposity in Europe - A scoping review. *Prev Med* 2022;160:107095.
- Muthuri SK, Onyera VO, Tremblay MS, et al. Relationships between parental education and overweight with childhood overweight and physical activity in 9-11 year old children: results from a 12-country study. *PLoS One* 2016;11:e0147746.
- Pollack CE, Chideya S, Cubbin C, et al. Should health studies measure wealth? A systematic review. *Am J Prev Med* 2007;33:250–64.
- Scott J. *Oxford dictionary of sociology*. 4th ed. Oxford University Press, 2014.
- Braveman PA, Cubbin C, Egerter S, et al. Socioeconomic status in health research: one size does not fit all. *JAMA* 2005;294:2879–88.
- Galobardes B, Shaw M, Lawlor DA, et al. Indicators of socioeconomic position (Part 1). *J Epidemiol Community Health* 2006;60:7–12.
- Lien N, Friestad C, Klepp KI. Adolescents' proxy reports of parents' socioeconomic status: how valid are they? *J Epidemiol Community Health* 2001;55:731–7.
- Chittleborough CR, Baum FE, Taylor AW, et al. A life-course approach to measuring socioeconomic position in population health surveillance systems. *J Epidemiol Community Health* 2006;60:981–92.
- Hariri JG, Lassen DD. Income and outcomes: social desirability bias distorts measurements of the relationship between income and political behavior. *Public Opin Q* 2017;81:564–76 <https://doi.org/10.1093/poq/nfw044>
- Psaki SR, Seidman JC, Miller M, et al. Measuring socioeconomic status in multicountry studies: results from the eight-country MAL-ED study. *Popul Health Metr* 2014;12:8.
- Macintyre S, McKay L, Der G, et al. Socio-economic position and health: what you observe depends on how you measure it. *J Public Health Med* 2003;25:288–94.
- O'Donnell O, van Doorslaer E, Wagstaff A. Analysing health equity using household survey data: a guide to techniques and their implementation. World Bank, 2008. Available: <https://openknowledge.worldbank.org/handle/10986/6896> [Accessed 15 Sep 2022].
- Moore JC, Stinson LL, Welniak EJJ. Income measurement error in surveys: a review. *J Off Stat* 2000;16:331–61.
- Moore JC, Loomis L. Reducing income nonresponse in a Topic-Based interview. Res Rep Ser. statistical research division, U.S. Bureau of the census, 2002. Available: <https://www.census.gov/content/dam/Census/library/working-papers/2002/adrm/rsm2002-06.pdf> [Accessed 04 Feb 2022].
- Kuhn I. Measurement of income in surveys. FORS guide No. 02, version 1.0. Swiss centre of expertise in the social sciences FORS, 2019. Available: https://serval.unil.ch/resource/serval:BIB_AFED4FF73C.P001/REF [Accessed 15 Sep 2022].
- Mölläri K, Hakulinen T, Hietanen-Peltola M. Terveystarkastusten ja muiden käyntien toteumat äitiys- ja lastenneuvolassa 2018 sekä kouluterveydenhuollossa lukuvuonna 2018-19 (in Finnish). Statistical report 46/2019. Finnish Institute for Health and Welfare, 2019. Available: <https://urn.fi/URN:NBN:fi-fe2019121848672> [Accessed 24 Aug 2022].
- Government decree 338/2011 on maternity and child health clinic services, school and student health services and preventive oral health services for children and youth. issued in Helsinki on 2011 Apr 6. Available: <https://www.finlex.fi/fi/laki/kaannokset/2011/en20110338.pdf> [Accessed 13 Jun 2022].
- Mäki P, Wikström K, Hakulinen T. Terveystarkastukset lastenneuvolassa ja kouluterveydenhuollossa. Menetelmäkäsikirja (in Finnish). In: *National Institute for health and welfare*. 4th, 2017. <https://urn.fi/URN:ISBN:978-952-302-964-4>
- Lundqvist A, Jääskeläinen S. Lasten ja nuorten ylipaino ja lihavuus 2018 (in Finnish). Statistical report 9/2019. Finnish Institute for Health and Welfare, 2019. Available: <https://urn.fi/URN:NBN:fi-fe2019042413181> [Accessed 13 Jun 2022].
- Saari A, Sankilampi U, Hannila M-L, et al. New Finnish growth references for children and adolescents aged 0 to 20 years: Length/height-for-age, weight-for-length/height, and body mass index-for-age. *Ann Med* 2011;43:235–48.
- WHO Multicentre Growth Reference Study Group. WHO child growth standards based on length/height, weight and age. *Acta Paediatr Suppl* 2006;450:76–85.
- de Onis M, Onyango AW, Borghi E, et al. Development of a WHO growth reference for school-aged children and adolescents. *Bull World Health Organ* 2007;85:660–7.
- Friedman JH. Stochastic gradient boosting. *Comput Stat Data Anal* 2002;38:367–78.
- Galobardes B. Socioeconomic inequalities in health: individual or area level; does it matter? *BMC Public Health* 2012;12:171.
- Geronimus AT. Invited commentary: using area-based socioeconomic Measures—Think Conceptually, act cautiously. *Am J Epidemiol* 2006;164:835–40.
- Kéké LM, Samouda H, Jacobs J, et al. Body mass index and childhood obesity classification systems: a comparison of the French, International Obesity Task Force (IOTF) and World Health Organization (WHO) references. *Rev Epidemiol Sante Publique* 2015;63:173–82.
- Valerio G, Balsamo A, Baroni MG, et al. Childhood obesity classification systems and cardiometabolic risk factors: a comparison of the Italian, World Health Organization and International Obesity Task Force references. *Ital J Pediatr* 2017;43:19.
- Sarkkola C, Viljakainen J, de Oliveira Figueiredo RA, et al. Prevalence of thinness, overweight, obesity, and central obesity in Finnish school-aged children: a comparison of national and international reference values. *Obes Facts* 2022;15:240–7.

- 35 Duncan GJ, Daly MC, McDonough P. Optimal indicators of socioeconomic status for health research. *Am J Public Health* 2002;92:1151–7.
- 36 Turrell G, Hewitt B, Patterson C, *et al*. Measuring socio-economic position in dietary research: is choice of socio-economic indicator important? *Public Health Nutr* 2003;6:191–200.
- 37 Turrell G. Income non-reporting: implications for health inequalities research. *J Epidemiol Community Health* 2000;54:207–14.

S1 Supplemental table. Description and categorization of variables

Variable	Source	Variable description	Original categories	Coding in the analyses
Categorical variables				
Municipality class	¹	Municipality group of municipality of domicile according to the 2016 regional division	1=city 2=urban 3=rural	1=city 2=urban 3=rural
Adolescent (age≥16 years) in high school	¹	Type of education	2=high school 3=vocational school 4=higher (third-level) education 5=university of applied sciences 6=vocational school	Yes: 2 No: other
Mother lives in the same household	¹	Status in family	1=head of the family 2=spouse 3=child 4=head of the cohabiting family 5=spouse of cohabiting family 9=unknown 0=not member of family	Yes: 1, 2, 5 (1 among single mothers) No: other
Father lives in the same household	¹	Status in family	1=head of the family 2=spouse 3=child 4=head of the cohabiting family 5=spouse of cohabiting family 9=unknown 0=not member of family	Yes: 1, 4 No: other
Mother/father: degree of highest education	¹	Educational level of highest qualification/degree	1= Vocational 2= Specialised vocational 3= Lowest tertiary 4= Bachelor's degree 5= Master's degree 6= Doctor's degree	0= No qualification (missing values coded as 0) 1= Vocational 2= Specialised vocational 3= Lowest tertiary 4= Bachelor's degree 5= Master's degree 6= Doctor's degree
Continuous variables				
Child's age (with decimal numerals)	²	Child's age when weight and height was measured		age= (date of measurement - date of birth)/365.25 Figure 1: integer value of age was calculated Figure 2: age in 12 categories: 2, 3-4, 5, 6, 7, 8, 9-11, 12, 13, 14, 15-16, 17

Any children <14 years of age	¹	Number of children aged under 14 in the family		Yes: ≥1 No: other
Annual household's disposable money income (euros)	¹	Household's disposable money income without equalizing according to household's size		Figure 3: 1= <30 000 2= 30 000–39 999.99 3= 40 000–59 999.99 4= 60 000–69 999.99 5= 70 000–79 999.99 6= 80 000–99 999.99 7= ≥ 100 000 When “father's age 35–39”, categories 2 and 3 were recoded into 4, and categories 6 and 7 were combined into 5. When “father's age 30–34” or “father's age <30”, category 1 was recoded to 2 and categories 5–7 were recoded into 4. When “father's age 40–44”, “father's age 45–49” or “father's age >49”, category 2 was recoded into 3.
Mother's age (years, integer value)	¹	Person's age at last day of year		Figure 3: 1= <30 2= 30-39 3= >39
Father's age (years, integer value)	¹	Person's age at last day of year		Figure 3: 1= <30 2= 30-34 3= 35-39 4= 40-44 5= 45-49 6= >49

¹Statistics Finland²Register of Primary Health Care visits