BMC Med Inform Decis Mak

https://doi.org/10.1186/s12911-026-03449-6

Article in Press

# Integrating data science for the analysis of early years health inequalities: a Birmingham, UK case study

Received: 6 October 2023

Accepted: 16 March 2026

Published online: 30 April 2026

**Cite this article as: Haidar D., Tawil A.H., Vilas J. *et al.* Integrating data science for the analysis of early years health inequalities: a Birmingham, UK case study. *BMC Med Inform Decis Mak* (2026). https://doi.org/10.1186/s12911-026-03449-6**

**Diana Haidar, Abdel-Rahman H. Tawil, Julia Vilas, Konstantinos Vlachos, Maria Clark, Edlira Vakaj & Nigel Sharratt**

# Integrating Data Science for the Analysis of Early Years Health Inequalities: A Birmingham, UK Case Study

Diana Haidar[1], Abdel-Rahman H. Tawil[1*], Julia Vilas[2], Konstantinos Vlachos[1], Maria Clark[3], Edlira Vakaj[1], Nigel Sharratt[4]

[1*] College of Computing, Birmingham City University, Belmont Row, Birmingham, B4 7RQ, UK .
[2] ISIS Engineering School, INU Champollion, Castres, France .
[3] School of Health Sciences, University of Nottingham, Nottingham, UK .
[4] Information Technology & Digital Services, Birmingham City Council, Birmingham, UK .

*Corresponding author(s). E-mail(s): abdel-rahman.tawil@bcu.ac.uk;
Contributing authors: diana.haidar@bcu.ac.uk; konstantinos.vlachos@bcu.ac.uk; maria.clark@nottingham.ac.uk; edlira.vakaj@bcu.ac.uk;



**Abstract**

**Background:** Early years health and development analysis is a key component of current government strategies, as it represents one of the main determinants for early intervention. A substantial body of research has applied data science methods—such as global cognitive scores, multilevel linear regression, stepwise linear regression, and logistic regression—to model risk factors for individual children. However, limited research has explored how socioeconomic status affects young children's health and development, or how this correlation can be accurately measured.

**Methods:** In collaboration with Birmingham City Council's Public Health, Education, and Skills directorates, Birmingham Community Healthcare NHS Foundation Trust, and the University of Nottingham's School of Health Sciences, we conducted a comprehensive review and analysis of early interventions across Birmingham communities. The aim was to compare health and social deprivation indices to identify the data required to understand correlations between indicators from both domains, supporting evidence-based intervention planning. Using Pearson and Spearman correlations, factorial analysis was performed to identify deprivation variables most relevant to studying the relationship between socioeconomic deprivation and early health and developmental outcomes.

**Results:** The analysis revealed significant correlations between socioeconomic factors—such as income deprivation, universal credit, and unemployment claimants—and early developmental outcomes, including obesity and the proportion of children performing below expected levels in communication, fine motor, gross motor, personal-social, and problem-solving skills.

**Conclusions:** The findings highlight concerning public health patterns: a weak but notable correlation between socioeconomic indices and indicators of child and family poverty, linked to wider determinants influencing childhood obesity and poor school readiness at ages 2 to 2.5 years. These results underscore the need for stronger emphasis on understanding and addressing the relationship between local environmental conditions and public health outcomes.

**Keywords:** Children Needs, Socioeconomic Status, Deprivation, Correlation Analysis

## 1 Background

Giving every child the best start in life is a public health and policy imperative [1]. Addressing health disparities at both the family and community level can have a profound impact on improving children's health outcomes and establishing a strong foundation for good health throughout their lives [2]. In England, the Public Health Outcomes Framework (PHOF) sets out a vision for public health, including a focus on the 'wider determinants' of health. Local authorities are tasked with reducing health inequalities for infants and young children - working together with statutory health visiting services and early years providers to identify childhood adversities and reduce the long-term health impacts for individuals and communities [1]. Spencer et al. [3] showed how adverse health outcomes would be reduced by as much as 18% to 59% if all children were as healthy as the most socially advantaged. Webb [4] recently demonstrated how an intersectional analysis of interventions in child social care can help build critical research on the social gradient measure as an indicator of local inequities. A range of data analysis tools are provided to assist local authority commissioners and service providers to identify local priorities and work together towards an integrated child health review at 2-2.5 years. Two related priority areas for child health development in Birmingham in the West Midlands region are school readiness and obesity prevention. These two priorities are more readily measurable here and reflect significantly worse than average public health outcomes for children under 5 years of age, when compared with children across the country [1].

According to the 2018 Health profile for England report, there is an inequality in life expectancy between the most deprived and the least deprived areas of England [5]. The roots of health inequalities can be traced back to early childhood, and areas with limited resources tend to have lower scores on various indicators of child health, such as low birth weight, infant mortality, dental health, school readiness, and child obesity. These findings are highlighted in the annual reports of Public Health England [6, 7]. Since a child's well-being generally continues into adulthood, there is a need to ensure the best start from an early age [8]. Socioeconomic factors, including parental age, marital status, employment, and level of education, as well as parents' ethnicity, and experience of stigma, all have an impact on an individual's quality of life [9]. In cases such as the one presented in this study, where social determinants of health are examined instead of the medical causes of a particular disease, the local government takes the lead in public health services by investigating non-medical factors that impact health outcomes [10, 11]. Local government has an impact on individual and population health improvement through tackling the social determinants of health, and commissioning specialist services, such as health visitors and school nurses, for delivering the Healthy Child Programme (HCP) for children aged 0 to 19 [7, 12–14]. The world health organization suggests that social determinants accounts for over 45% of health outcomes with an estimated population that exceeds the contribution from the health sector [15]. Exploring the correlation between socioeconomic and demographic factors and the needs of children can aid local governments in making informed decisions about interventions. Taking appropriate measures to address these factors is crucial for enhancing children's health and mitigating long-standing health inequities. However, achieving this goal necessitates the involvement of all sectors and civil society,including the public health workforce [7].

While national studies have documented the impact of socioeconomic inequality on child health and development across England, our contribution is distinct in focusing on Birmingham as a detailed city-level case study. By integrating multiple datasets (ASQ-3, NCMP, IMD, IDACI, Universal Credit, and unemployment claimants), we provide a multi-dimensional picture of early years outcomes within one of the UK's largest and most diverse urban centres. This approach allows us to demonstrate how place-based data integration can uncover localised patterns that may be masked in national analyses, thereby generating insights of direct relevance for local policy and service delivery.

This paper identifies the socioeconomic factors that are most influential on children's needs. We focus on Birmingham in the UK as a local case study. Birmingham has one of the most diverse populations in Europe and faces significantly worse than average outcomes in child obesity and school readiness compared to the national average. These characteristics make it a critical illustrative case for exploring the interaction between socioeconomic deprivation and early years health. While our findings are not intended to be generalizable to all UK regions, they provide transferable insights that may inform early years interventions in other urban areas with similar demographic and deprivation profiles. As another important dimension in the analysis, family demographics are also included. The study is focused on two themes: school readiness and obesity. A review of obesity in this area showed interventions designed to reduce the risk of overweight/obesity delivered antenatally or during the first 2 years of life, with outcomes reported from birth to 7 years of age [16].

School readiness is evaluated by means of the Early Development Instrument (EDI) total score, which measures five developmental areas, including physical health and well-being, social competence, emotional maturity, language cognitive development and communication skills, and general knowledge.

In Section 2 we investigate Related Works. In Section 3 we design a conceptual framework modeling the relationships between health and development outcome indicators and cover data collection and processing. Further, Section 4 covers in depth the technical methods, the results achieved and a discussion of how these results can be helpful to the end users i.e. decision makers and more importantly the community. Section 6 presents our conclusion and findings.

## 2 Related work

The first few years of a child's life are critical for later health and development [17]. The chapter reports that the health of children in early years has been improving in recent years as seen across many indicators of health and development outcomes including child obesity, and school readiness. There exists a fair body of research that addresses school readiness or child obesity themes for early years children. Table 1 lists and briefly compares some of the related work articles that tackle these themes based on KPI, socioeconomic factor(s), and correlation technique used. We next discuss these studies and report any examined associations.



### 2.1 School Readiness - Child Development Level

Jones et al. [18] provided a cross-sectional study to explore the association between the child's Socioeconomic status (SES), using Index of multiple Deprivation (IMD) 2019 ward scores and each of good level of development and school-readiness scores (Early Years Foundation Stage Profile (EYFSP)) using Pearson bivariate correlations. In

**Table 1** Comparing related work articles on school readiness or child obesity themes based on KPI, Socioeconomic Factor(s), and Correlation Technique used

| Paper | Year | Location | KPI | Socioeconomic Factor(s) | Correlation Technique |
|---|---|---|---|---|---|
| [18] | 2021 | England | good level of development and school readiness scores | IMD 2019 scores - ward level | Pearson |
| [19] | 2020 | England | ASQ-3 domains | IDACI scores - postcode level | Spearman |
| [20] | 2020 | Valencia | global cognitive scores | parental socioeconomic characteristics | Linear regression |
| [21] | 2016 | Hong Kong | school readiness scores for 5 developmental domains | family SES index | Multilevel linear regressions |
| [21] | 2016 | Hong Kong | school readiness scores for 5 developmental domains | gender | Chi-squared test |
| [22] | 2012 | United States | pre-academic knowledge scores (cognitive skills) | SES index | Bivariate correlations |
| [23] | 2018 | United States | child overweight status | child risk factors | Logistic regression |
| [24] | 2014 | England | prevalence of overweight and obesity in Reception NCMP for 2007/8 and 2009/10 | IDACI 2010 scores - MSOA level | Stepwise linear regression |
| [25] | 2012 | England | prevalence of overweight and obesity in Reception NCMP from 2007/8 to 2009/10 | IMD 2010 scores - LA level | Pearson |

their study, data was collected from 326 children aged 4-5 years from the northeast of England. The analysis showed no significant effect of SES on the level of development of children. On the other hand, lower school-readiness scores was significantly associated with children from more deprived areas ($r = 0.12, p < 0.05$).

The relationships between Income Deprivation Affecting Children Index (IDACI) score taken from the family home postcode and ASQ-3 (Ages & Stages Questionnaires) domains scores was investigated by Law et al. [19] using Spearman correlation. The data was collected from 894 children using the Early Language Identification Measure – Extended (ELIM-E) across five sites in England (Derbyshire, Middlesbrough, Newham, Wakefield, and Wiltshire). The results showed that poorer performance was associated with more deprived postcodes (i.e. higher IDACI score) for the ASQ-3 communication, problem solving, and personal-social skills domains, however those associations were weak.

The study by González et al. [20] assessed the relationship between child cognitive development and each of maternal/paternal socioeconomic characteristics (social class, education, and employment status). The study was conducted on 525 children aged 5-6 years in a Valencia Birth cohort and measured children cognitive development using the Global Cognitive Score (GCS) from McCarthy Scales of Children's Abilities. Linear regression models were used to evaluate the associations where a strong gradient was found between GCS and maternal/paternal social class ($\beta = -10.0$ and $\beta = -11.5$ for class $IV + V$ (lowest) for mothers and fathers respectively). Regarding educational level, mothers who were with university degree had children with a 15.4 points higher GCS than those with primary school ($\beta = 11.0$ for fathers). Also, unemployment for both mother and father had a strong relationship with child's GCS, being stronger for long-term unemployed fathers ($\beta = -7.8$).

In the study on Chinese preschool children in Hong Kong, Multilevel linear regression models were used to examine the effect of family SES on school readiness [21]. The study measured the school readiness scores of 567 children using the Chinese version of the Early Development Instrument (EDI) for 5 developmental domains: physical health and well-being, social competence, emotional maturity, language and cognitive development, and communication skills & general knowledge. The family SES index was also determined from parental educational attainment, occupation, monthly income adjusted for household size, and asset score. The estimates of correlation between SES and EDI scores were significant for each developmental domain excluding emotional maturity, and the Intraclass correlation coefficients (ICC) indicate poor reliability with $ICC = 0.24$ for total EDI score (all domains together). Furthermore, the study performed Chi-squared test to examine the relationship between gender and school readiness revealing a significant gender effect ($\chi^2 = 11.79, p = .008$).

In the longitudinal study of child health and development by Dotterer et al. [22], the link between school readiness and SocioEconomic Status (SES) index (created by standardizing income-to-needs ratio and maternal education level) was examined. The study invited participants from an urban community via fliers, postings, mailings, and phone contact and included mother-child dyads from African American and European American families. Results show that negative/intrusive parenting connect the link between socioeconomic status and children's pre-academic knowledge scores. Furthermore, socioeconomic status was positively correlated to pre-academic knowledge especially among African Americans.

## 2.2 Child Obesity

The data analysis study by [23] investigated the correlation between parental and child risk factors (such as socioeconomic status, race, birth weight, parental smoking, and lack of family meals) and childhood obesity. The study was conducted on the 2001 United States Early Childhood Longitudinal Birth Cohort data sample using logistic regression to model the child overweight status from each individual risk factor and evaluates socioeconomic status as interaction term. The results showed that parental and child risk factors were associated to children being overweight/obese. In addition, Black or Hispanic children were 60% more likely to be overweight/obese than white children in the USA. Moreover, children belonging to the lowest socioeconomic status quintile exhibited a 70% higher likelihood of being overweight or obese in comparison to children from the highest quintile.

A further large cross sectional study [24] was conducted on combined National Child Measurement Programme (NCMP) data for the years 2007/8 and 2009/10 to examine the relationship between weight status of children in Reception (4-5 years) and socioeconomic deprivation (as represented by IDACI 2010 scores) at Middle Layer Super Output Areas (MSOA) level in England. Using step-wise linear regression, the authors identified a positive association between IDACI scores and the prevalence of overweight or obesity (slope $B = 9.511, p < .001$ for overweight or obese status, and $B = 7.254, p < .001$ for obese status).

Earlier in 2012 [25] studied the association between childhood overweight and obesity in Reception and IMD 2010 scores using NCMP data from 2007/8 to 2009/10 at local authority level in England. The authors carried out Pearson's correlation analysis to measure the strength of the correlation and found that the prevalence of obesity in Year 0 and IMD scores are strongly correlated ($r = 0.539$, $r = 0.612$, and $r = 0.625$, $p < 0.001$) in 2007/8, 2008/9, and 2009/10 respectively.



## 3 Materials and Methods

In our preparation for early years review analysis, we design a conceptual framework modeling the relationships between health and development outcome indicators with socioeconomic and demographic factors. After that, we present and discuss the data sets used in our study and the preprocessing steps performed to prepare the data for analysis.

## 3.1 Conceptual Framework

Fig. 1 presents the conceptual framework guiding the early years health and development review analysis. Blue rectangles represent the themes used in our study, orange hexagons represent the socioeconomic factors, green hexagons represent the demographic factors, and the arrows (links) represent reference or potential relationships/correlations between the themes and socioeconomic/demographic factors. The arrows having a list of paper references enclosed in brackets refer to papers already reviewed in Section 2 which addresses the links-related relationships.

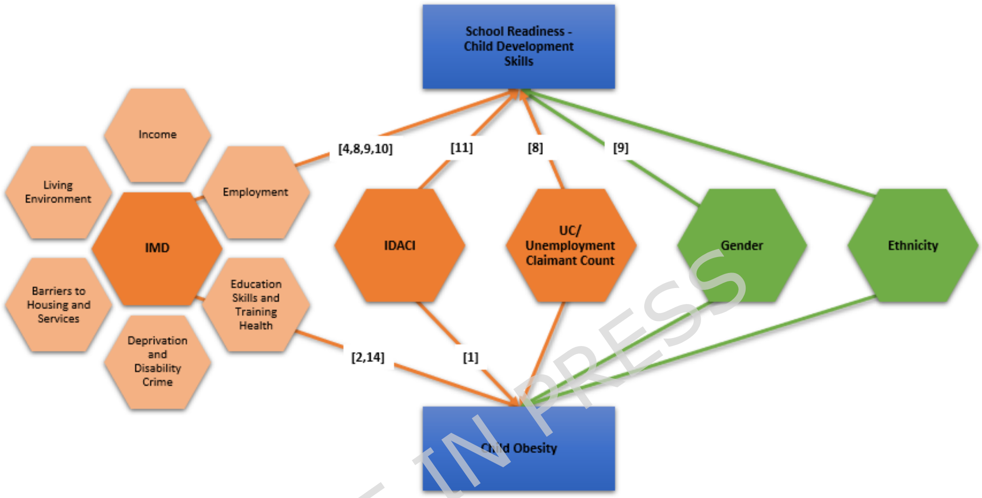


**Fig. 1** Conceptual framework guiding our analysis.

As illustrated in the conceptual diagram, the themes tackled in our work are: School Readiness - Child Development Level, and Child Obesity. Children development level in early years is an important marker in a child's life course. Monitoring child development at a young age ensures that children meet their developmental milestones and are physically, emotionally, and socially ready to enter school and learn. It also helps to identify if a child has any developmental delays and the opportunity of early intervention to develop his competency and ensure better readiness.

As well, overweight and obesity at a younger age increases the risk of poor health in adulthood including diabetes, cardiovascular disease, and risk of premature death. Assessing the child's weight in early years helps to identify children in need of early help and treatment if needed (for instance, support in meal planning and shopping, medical investigations, investigating potential social and environmental factors contributing, etc.).

The socioeconomic factors covered include: (1) Index of Multiple Deprivation (IMD), (2) Income Deprivation Affecting Children Index (IDACI), and (3) Universal Credit (UC)/ Unemployment Claimant Counts. Indeed, in this study, we considered the IMD factor as well as each of its subdomains: Income Deprivation; Employment Deprivation; Education Deprivation, Skills and Training Deprivation; Health Deprivation and Disability; Crime; Barriers to Housing and Services; and Living Environment Deprivation. A detailed description of each of the socioeconomic factors is provided in Section 3.2.2.

In addition, we covered the demographic factors including: (1) gender (male, female), and (2) ethnicity (Asian, Black. White, Mixed, Other).

## 3.2 Data Collection and Processing

### 3.2.1 Themes Data Sets

In this study, we collected themes data sets from various data sources.

***2-2½ year reviews ASQ-3 Data Set***

For the development skills theme, we inquired the 2-2½ year reviews data set that uses Ages and Stages Questionnaire (ASQ-3) in the Healthy Child Programme development review from Birmingham Community HealthCare (BCHC) NHS Foundation Trust. The ASQ-3 provides an objective measure of child development outcomes around their second birthday over several domains of development including communication, gross motor, fine motor, problem solving, and personal social skills. It helps to identify children below the expected level of development and support them before they enter school to achieve their full potential.

The 2-2½ year reviews ASQ-3 data set is provided at Birmingham postcode geographical level and reported monthly over the years 2019/20, 2020/21, and 2021/22. It records a measure score for each domain of development per child, associated to the following attributes: gender, ethnicity, age in months at point of contact, postcode, report month, and report financial year.

The Healthy Child Programme development review uses three types of ASQ-3 questionnaires: 24 months, 27 months, and 30 months questionnaire. For each questionnaire type, a threshold score for expected level of development is set per domain. Table 2 presents the full list of threshold scores relevant to the questionnaire type and domain of development.

**Table 2** Threshold score for expected level of development per questionnaire type and domain of development.

| | Domain of development | | | | |
|---|---|---|---|---|---|
| ASQ-3 questionnaire type | Communication | Fine motor | Gross motor | Personal social | Problem solving |
| 24 month | 25.17 | 35.16 | 38.07 | 31.54 | 29.78 |
| 27 month | 24.02 | 18.42 | 28.01 | 25.31 | 27.62 |
| 30 month | 33.30 | 19.25 | 36.14 | 32.01 | 27.08 |

In order to asses a child's level of development, it is required to compare the measure scores for each domain of development to the relevant threshold scores, then identify children below expected levels of development, as well as children at or above expected levels of development. In this work, we are interested to target children below expected level of development for early intervention. Thus, we used the percentage of children below expected levels for each domain of development as well as for all domains to be the Key Performance Indicators (KPIs) in our study.

***NCMP Data Set***

For the child obesity theme, we inquired the National Child Measurement Programme (NCMP) data set from the Public Health Division at Birmingham City Council (BCC). NCMP is a national public health programme that provides indicators of the children excess weight as part of the 'Tackling Obesity' Strategy of the UK government. The NCMP annually measures the weights and heights of children in Year 0 (a.k.a Reception year) and Year 6 across eligible state-funded schools in England.

The NCMP data set is provided at Birmingham LSOA[1] geographical level for the period 2010/11 to 2019/20 as a whole together. It records the number of pupils measured with respect to the following attributes: LSOA, school year (*Year 0, Year 6*), Body Mass Index (BMI) category (*Normal, Obese, Overweight*), and ethnicity. In our study, we only focus on Year 0 children.

We should note the collection of the NCMP 2020/21 data was carried out as a sample [26] (11% sample for Birmingham) due to the impact of the Covid-19 pandemic. However, according to the dataset publisher [26], statistical weighting was applied to the data and a meticulous sampling methodology was followed for the data collection process to ensure that the data will be representative of the entire pupil population and a data quality investigation proved that the data are comparable to the previous collection periods.

On the other hand, the number of children measured at Reception level in Birmingham in the earlier NCMP 2019/20 data was around 55% of the annual average over the previous three years (around 8,700 compared to almost 16,000 children) due to the school closures in March 2020 in response to the Covid-19 pandemic. The Public Health Division at BCC advised that NCMP 2019/20 data is representable and can

[1]LSOAs are small areas designed to be of a similar population size, with an average of approximately 1,500 residents or 650 households. There were 32,844 LSOAs in England for Census 2011. They are produced by the Office for National Statistics based on Census data for the reporting of small area statistics.

be used in our analysis despite the lower than usual number of measurements, thus it was included in our analysis.

The dataset includes '*' values for the numbers of pupils measured in some records (that is for the number of pupils inside an LSOA for a specific combination of gender, ethnicity, and BMI category); these values refer to numbers below 5 that were substituted with an asterisk by the Public Health Division at BCC to preserve the confidentiality of data and protect pupils from being identified. We replaced these '*' values with the value 2 as an estimation for the average number of pupils measured in a record.

In this work, our objective is to help children with excess weight (i.e. overweight or obese) and ensure their healthy growth. Thus, we used the prevalence (percentage) of children in Year 0 who are (1) overweight, (2) obese, (3) overweight or obese to be the KPIs in our study.

### 3.2.2 Socioeconomic Factors Data Sets

For the socioeconomic factors data sets, we collected the relevant releases of data available with the aim to correlate to the theme data sets. As our focus is on Birmingham community, we filtered these data sets to only include Birmingham LSOA geographical level in our study.

***Index of Multiple Deprivation***

The Index of Multiple Deprivation (IMD) data set [27] is available online from GOV.UK[2] at England Lower-layer Super Output Areas (LSOA) geographical level and was last reported in 2019 to update the previous English indices of deprivation of 2015.

[2]GOV.UK is a United Kingdom public sector information website, created by the Government Digital Service to provide a single point of access to Her Majesty's (HM) Government services.

The IMD is an overall composite measure of relative deprivation for LSOAs across England. It is created by summing weighted relative indices of deprivation for the seven domains: Income Deprivation; Employment Deprivation; Education Deprivation, Skills and Training Deprivation; Health Deprivation and Disability; Crime; Barriers to Housing and Services; and Living Environment Deprivation.

The IMD data set records the ranks and deciles at LSOA level for the IMD and for each of the indices of the deprivation seven domains. The ranks are calculated for the 32,844 LSOAs in England where a rank of 1 is given to the most deprived LSOA, and a rank of 32,844 is given to the least deprived. The deciles are calculated by allocating the LSOAs into 10 equal deciles (groups) based on their calculated ranks where the most deprived 10% of LSOAs nationally are allocated to decile 1, the least deprived 10% of LSOAs nationally are allocated to decile 10, and so on.

In our analysis, we used the deciles at LSOA level for the overall composite measure IMD as well as each of the indices of deprivations for seven domains.

### *Income Deprivation Affecting Children Index*

The Income Deprivation Affecting Children Index (IDACI) data set [27] is also available online from GOV.UK at England LSOA geographical level and was last reported in 2019.

The IDACI is a measure of the rate of children aged below 16 living in income deprived families in LSOAs across England. Income deprived families include people who are out-of-work or people who are in work but have low earnings. In a similar format to the IMD data set, the IDACI records the ranks and deciles at LSOA level for the IDACI, as well as recording the IDACI rates. In our analysis, we used the deciles at LSOA level for the IDACI.

### Universal Credit Claimants

The Universal Credit (UC) Claimants data set [28] is available from Stat-Xplore [3] at LSOA geographical level and was last reported in June 2022. The UC Claimants is a measure of the number of people on Universal Credit for whom no closure of their claim have been recorded up to the release date (second Thursday in each month). The UC claimants include unemployed people and employed people on a low income.

The UC Claimants data set records monthly releases of UC claimant count at Birmingham LSOA level over the fiscal years 2019/20, 2020/21, and 2021/22 (the fiscal years have been selected to match the years available for the themes data sets). The data is extracted in monthly format and transformed into the format of annual average claimant count per fiscal year at Birmingham LSOA level. For the 2-2½ year reviews ASQ-3 data set, the relevant annual average claimant count has been selected to match the ASQ-3 KPIs per fiscal year at LSOA level. On the other hand, for the NCMP data set, as the data was available for the period 2010/11 to 2019/20 as a whole, the annual average claimant count for 2019/20 has been selected to match the NCMP KPIs at LSOA level.

### Unemployment Claimants

The Unemployment Claimants data set [29] is available from Nomis[4] at LSOA geographical level and was last reported in May 2022.

The Unemployment Claimants is a measure of the number of people who claim Jobseeker's Allowance (JSA) (i.e. people claiming unemployment related benefits while looking for work) in addition to out of work people who claim Universal Credit. Unlike JSA, Universal Credit requires a wider range of claimants to look for work. The data

[3] Stat-Xplore is a browser based client to explore benefit data administered by the Department for Work and Pensions.

[4] Nomis is a service provided by Office for National Statistics (ONS). On this website, ONS publishes statistics related to population, society and the labour market at national, regional and local levels.

set records monthly releases of Unemployment claimant count at Birmingham LSOA level over the fiscal years 2019/20, 2020/21, and 2021/22. Similar to the UC Claimants data set, the Unemployment Claimants data sets has been processed to match the themes data sets.

**Table 3** Summary of data sources and collection years

| **Dataset** | **Source** | **Years covered** |
|---|---|---|
| ASQ-3 (2–$2\frac{1}{2}$ year reviews) | Birmingham Community Healthcare NHS Foundation Trust | 2019/20–2021/22 |
| NCMP (Reception year) | Public Health Division, Birmingham City Council | 2010/11–2019/20 (2020/21 sample 11% due to COVID) |
| Index of Multiple Deprivation (IMD) | GOV.UK (ONS) | 2019 |
| IDACI | GOV.UK (ONS) | 2019 |
| Universal Credit Claimants | Stat-Xplore (DWP) | 2019/20–2021/22 |
| Unemployment Claimants (JSA + UC) | NOMIS (ONS) | 2019/20–2021/22 |



To improve transparency, we have added Table 3, which summarises the datasets used, their sources, years of collection, and the degree of temporal overlap. While the ASQ-3, Universal Credit, and Unemployment datasets align closely over 2019–2022, the NCMP data spans a longer period (2010–2020). The IMD and IDACI indices are only available as 2019 releases. These temporal inconsistencies, while typical of routinely collected public health datasets, limit full integration across all measures. We address the implications of this limitation in the Discussion. It should be noted that the study period coincided with the COVID-19 pandemic, which affected both health and socioeconomic data collection. For example, the NCMP dataset for 2020/21 represents only 11% of the expected sample due to school closures. In addition, some small-area records were suppressed to preserve confidentiality; in these cases, imputation methods such as aggregation or mean substitution were applied. Socioeconomic indicators, including Universal Credit and unemployment claimant counts, were also strongly influenced by pandemic-related economic shocks.

### 3.3 Statistical Analysis

Correlation analyses were conducted to explore associations between socioeconomic indicators and child health/development outcomes. Pearson's correlation coefficient was applied when both variables were normally distributed (assessed using Shapiro–Wilk test), while Spearman's rank correlation coefficient was used for non-normally distributed variables or ordinal data.

All analyses were performed in R (version 4.3.1) using the functions `cor.test()` from the `stats` package and `rcorr()` from the `Hmisc` package. Visualisations of the correlation results (scatterplots and fitted trend lines, presented in Section (4.2) were produced using the `ggplot2` package in R. To address the issue of multiple comparisons, we applied the Benjamini–Hochberg False Discovery Rate (FDR) correction using the function `p.adjust(method = "BH")` in R. Both uncorrected and FDR-corrected p-values (q-values) are reported. The workflow consisted of the following steps: (i) data cleaning and preparation; (ii) assessment of variable distributions (Shapiro–Wilk test); (iii) selection of correlation method (Pearson or Spearman); (iv) calculation of correlation coefficients and associated p-values; (v) application of Benjamini–Hochberg FDR correction; and (vi) generation of scatterplots and fitted trend lines for visualisation.

## 4 Data Analysis

### 4.1 Data Visualisation and Analysis

We utilised the Power BI Business Intelligence (BI) tool (Version: 2.102.683.0 64-bit) to create visuals in dashboards that helped in understanding the data and provided insights about the prevalence of children below the expected level of development and the prevalence of children in Year zero who are overweight/obese. Power BI was selected as a software tool because of its rich feature set that allows importing different datasets, applying data transformation steps, and creating high quality interactive visualizations [30]. In addition, Power BI provides the ability to link the different datasets using table relationships to generate an underlying data model consisting of different data tables. This model can then be used as a basis to create reports with visuals (graphs, charts, tables) that interact with each other to slice and cross-filter the displayed data, thus simplifying the data interrogation and analysis process and facilitating insight extraction.

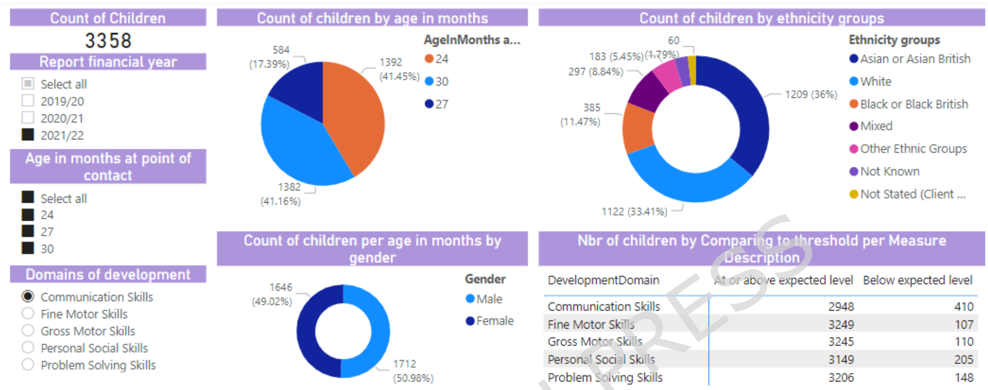


**Fig. 2** Overview on 2-2½ year reviews ASQ-3 data.

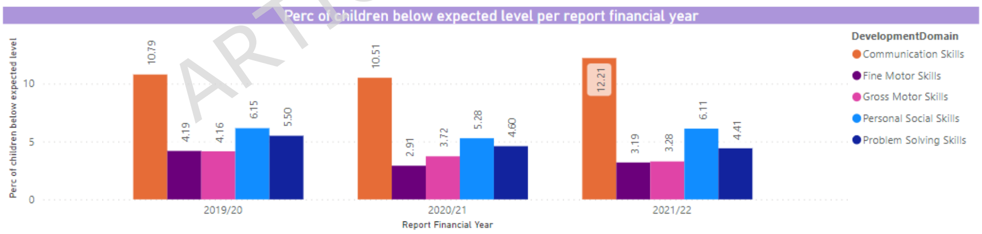


**Fig. 3** Prevalence of children below expected level of development per report financial year.

In Figure 2, we give an overview on the 2-2½ year reviews ASQ-3 data. For instance, in 2021/22, the 2-2½ year reviews for Communication Skills domain cover 3358 children of ages (24, 27, and 30 months) where 41.45% are of age 24 months, 41.16% are of age 30 months, followed by 17.39% of age 27 months. 50.98% of children are males and most of children are from Asian or Asian British ethnic group (36%), followed by 33.41% of White ethnic group. Later in this work, we study the correlation of children development level to the demographic factors (gender, ethnicity).

With respect to all domains of development, the majority of children measure scores are At or above expected level compared to the relevant threshold scores. For instance, 2948 of children are At or above expected level in Communication Skills domain, while around around 410 are Below expected level. As the aim of this work is to target children below expected level, the number of children to help is quite high, and this is only for the financial year 2021/22. Filtering the report financial year to cover the 3 years together reveals 1068 children who have been in need for early help to develop their communication skills and improve their competence before entering school.

Figure 3 also presents the percentages of children below expected level of development per report financial year. The colored bars/columns represent the percentage in each of the 5 domains of development. We observe that Communication Skills domain has the highest percentage of children below expected level over the 3 financial years (10.79% in 2019/20, 10.51% in 2020/2021, and 12.21% in 2021/22) followed by Personal Social Skills domain (6.15% in 2019/20, 5.28% in 2020/2021, and 6.11% in 2021/22).

In Figure 4, we give an overview on the NCMP data. The number of pupils measured in Year 0 (reception) over the period 2010/11 to 2019/20 was 138830 pupils, where 51.21% are males. We observe that the majority of pupils were from White ethnicity (32.47%) followed by 27.16% from Asian ethnicity. Most of the pupils were classified to Normal BMI category (76.4%), while the percentage of children classified overweight or obese was approximately 23.6% (12.28% for only overweight and 11.31% for obese).

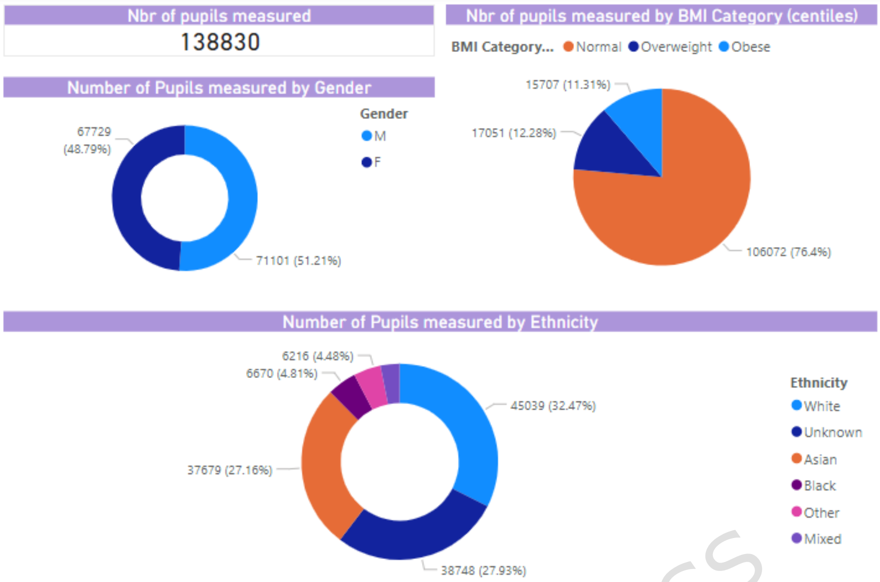


**Fig. 4** Overview on NCMP data.

Before moving into the correlation analysis, we give below some examples that illustrate the relationships between the themes and the economic factors (for instance, IMD).

Figure 5 shows the prevalence of children below expected level of development by IMD decile. The Power BI visuals are filtered to cover the years 2019/20 to 2021/22 for the Personal Social Skills and Problem Solving Skills domain. In these domains, We observe that the percentages of children below expected level are the highest in lower IMD deciles. For Personal Social Skills, the percentages are 6.8% and 6.4% in IMD deciles 1 and 2 compared to 2.4% and 1.4% in IMD deciles 9 and 10 respectively. In general, the average percentage of children below expected level is 6.34% in the IMD deciles below 5 compared to 3% in the IMD deciles at or above 5.

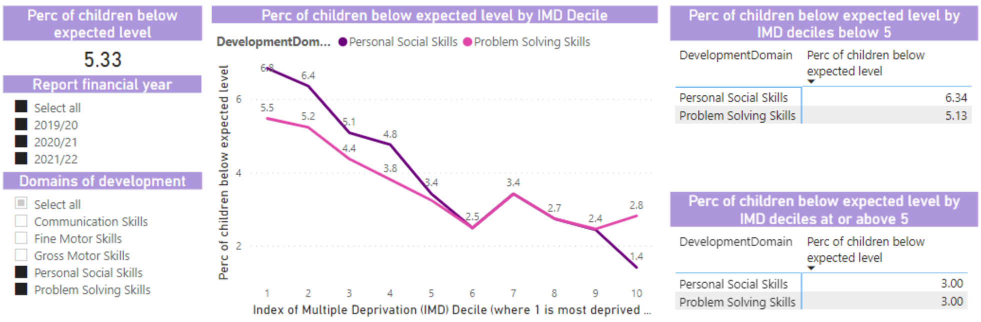


**Fig. 5** Prevalence of children below expected level by IMD decile.

Similarly, Figure 6 shows the prevalence of children who are overweight/obese by IMD decile. The percentage of children classified obese is the highest in lower IMD deciles (25% and 23.45% in IMD deciles 1 and 2 compared to 16.17% and 13.31% in IMD deciles 9 and 10 respectively. The same is also observed for the overall percentage of children classified overweight or obese. On the other hand, the percentage of children classified as overweight is almost the same over all IMD deciles. In general, the average percentage of children classified overweight or obese in the IMD deciles below 5 (24.31%) is higher than that in the IMD deciles at or above 5 (19.17%). Similarly, the average percentage of children classified obese is 11.92% compared to 7.53%. For the overweight category, the percentages are almost the same.

## 4.2 Correlation Analysis

### 4.2.1 Socioeconomic Factors Correlation

The rest of this section addresses the correlation between the themes and the socioeconomic factors, in addition to the demographic factors. Recall the conceptual framework presented earlier in Figure 1. Our aim here is to study the links/relationships between each of the themes and factors. We start with scatter plots to examine the links

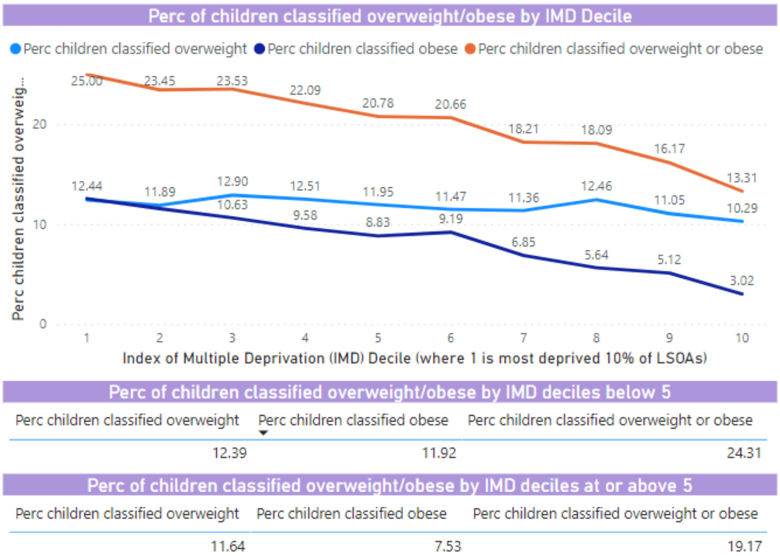


**Fig. 6** Prevalence of children classified overweight/obese by IMD decile.

between each of themes KPIs and IMD to get a flavor of the linearity/non-linearity of the relationships, and decide on using Pearson[5] or Spearman[6] correlations.

In Figure 7, we present scatter plots of prevalence of children below expected level in Communication Skills, Personal Social Skills, and Problem Solving Skills domains by IMD Rank using PowerBI Craydec Regression Chart[7]. Here, we chose IMD Ranks instead of IMD deciles to better present the spread of data points. It is evident that the data the data points are sparse and linear regression performs poorly ($R2=0$ for Communication Skills, $R2=0.03$ for Personal Social Skills, and $R2=0.3$ for Problem Solving Skills). Hence, for the prevalence of children below expected level KPI, we lean toward using Spearman Correlation as will be detailed later.

Figure 8 also present scatter plots of prevalence of children classified overweight/obese by IMD Rank using PowerBI Craydec Regression Chart. We notice that



[5]Pearson's correlation measures the linear relationship between two continuous variables. A relationship is said to be linear when a change in one variable is associated with a proportional change in the other variable.

[6]Spearman's correlation measures the strength and direction of monotonic association between two variables. Monotonicity is "less restrictive" than that of a linear relationship.

[7]Craydec Regression Chart is a scatter plot with a simple linear regression. The visual calculates Pearson's correlation coefficient, R2 value, and it draws the correlation equation as abline on the chart.

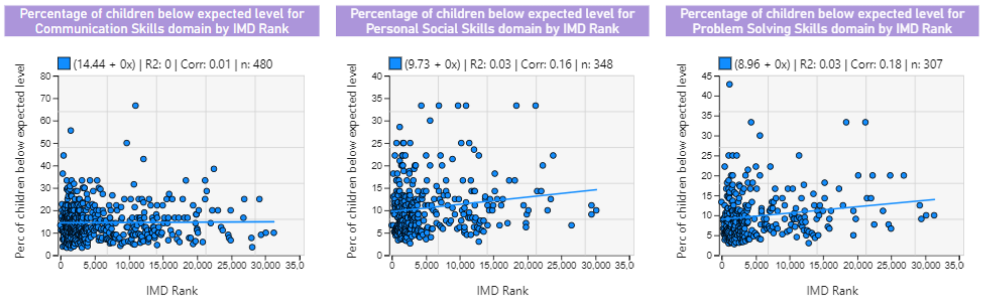


**Fig. 7** Scatter plot of prevalence of children below expected level by IMD Rank.

linear regression performs better here as there exists linear relationships between IMD Rank and each of the percentage of children classified obese (Corr=-0.55, R2=0.31) and the percentage of children classified overweight or obese ($Corr = -0.47, R2 = 0.23$). On the other hand, linear regression performs poorly for the overweight category. Hence, as our interest is in the prevalence of children class overweight or obese, we lean toward using Pearson correlation.

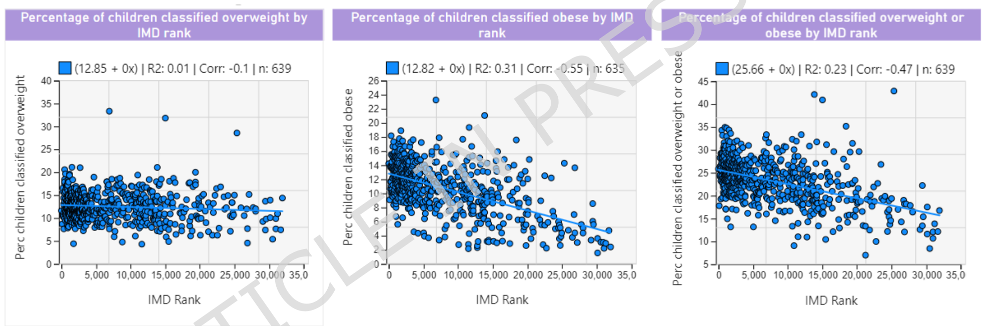


**Fig. 8** Scatter plot of prevalence of children classified overweight/obese by IMD Rank.



In what follows, we discuss the Pearson/Spearman correlation analysis performed for each of the themes and the socioeconomic factors.

Table 4 displays the Spearman correlation coefficients of each of the socioeconomic factors and the prevalence of children below expected level for each domain of development. we observe that the highest correlation coefficients are mostly for

Personal Social skills followed by Communication skills. For instance, the correlation between IMD decile and percentage of children below expected level for Personal Social skills domain is $\rho = -0.27, p < 0.0001$ ($\rho$ is Spearman's rank correlation coefficient); there exists a negative correlation with statistical significance. Thus, as the IMD decile decreases (i.e. more deprived), the percentage of children below expected level in Personal Social skills increases. We also notice that the correlation with annual average UC claimant count is the highest for both Personal Social and Communication skills domains($\rho = 0.24, p < 0.0001$); there exists a positive correlation with statistic significance. Thus, as the annual average UC claimant count increases, the percentage of children below expected level in these domain increases too. However, the socioeconomic factors are weakly correlated to the domains of development.

Table 5 displays the Pearson correlation coefficients of each of the socioeconomic factors and the prevalence of children below expected level for each domain of development for the overweight, and overweight or obese categories. It is evident that for the overweight, and overweight or obese categories the correlation of the prevalence of children classified obese is $r = -0.55$ ($r$ is Pearson's rank correlation coefficient) with IMD decile, $r = -0.56$ with IDACI decile, $r = 0.46$ with annual average claimant count, $r = 0.43$ and annual unemployment claimant count ($p < 0.001$ in all cases). Hence, the socioeconomic factors are moderately correlated with statistical significance to the children obesity.

Our results can be compared to that of the study [25] that investigates the associations between the IMD scores and the prevalence of overweight or obesity using Pearson's correlation analysis. Their study was conducted on children in Year 0 and Year 6 of local authority districts in England over the years 2007/8, 2008/9, and 2009/10. Even though the data sets cover other areas and academic years, the overall correlation assessment is similar to ours. The authors [25] found that the prevalence of overweight in Year 0 and IMD scores are weakly correlated ($r = 0.111$ and

**Table 4** Spearman's correlation coefficients of each of the socioeconomic factors and the prevalence of children below expected level for each domain of development measured.

| | Domain of development | | | | |
|---|---|---|---|---|---|
| Socioeconomic factors | Communication | Fine motor | Gross motor | Personal social | Problem solving |
| IMD Decile | -0.26*** | -0.2*** | -0.18*** | **-0.27*** ** | -0.22*** |
| Income Decile | -0.24*** | -0.21*** | -0.18*** | **-0.27*** ** | -0.23*** |
| Employment Decile | -0.24*** | -0.18** | -0.2*** | **-0.26*** ** | -0.23*** |
| Education Skills and Training Decile | -0.24*** | -0.2*** | -0.18*** | **-0.26*** ** | -0.21*** |
| Health and Disability Decile | -0.22*** | -0.17*** | -0.2*** | **-0.28*** ** | -0.21*** |
| Crime Decile | **-0.14*** ** | -0.1* | -0.06 | -0.13*** | -0.08* |
| Barriers to Housing and Services Decile | **-0.19*** ** | -0.02 | -0.02 | -0.16*** | -0.14*** |
| Living Environment Decile | 0 | **-0.11* ** | -0.02 | -0.05 | -0.02 |
| IDACI Decile | **-0.26*** ** | -0.18*** | -0.17*** | **-0.26*** ** | -0.23*** |
| Average UC claimant count | **0.24*** ** | 0.17*** | 0.13*** | **0.24*** ** | 0.21*** |
| Average Unemployment claimant count | **0.23*** ** | 0.15*** | 0.11** | **0.23*** ** | 0.2*** |

$*\ p < 0.05, **\ p < 0.01, ***\ p < 0.001$; Highest correlation coefficients with statistical significance in **bold**.

**Table 5** Pearson's correlation coefficients of each of the socioeconomic factors and the prevalence of children classified overweight/obese.

| | BMI Category | | |
|---|---|---|---|
| Socioeconomic factors | Overweight | Obese | Overweight or Obese |
| IMD Decile | -0.09* | **-0.55*****| -0.46*** |
| Income Decile | -0.06 | **-0.51***** | -0.41*** |
| Employment Decile | -0.06 | **-0.46***** | -0.38*** |
| Education Skills and Training Decile | -0.1* | **-0.58***** | -0.49*** |
| Health and Disability Decile | -0.19*** | -0.43*** | **-0.45***** |
| Crime Decile | -0.12** | **-0.3***** | -0.29*** |
| Barriers to Housing and Services Decile | 0.06 | **-0.44***** | -0.28*** |
| Living Environment Decile | 0.09* | **-0.22***** | -0.1* |
| IDACI Decile | -0.15*** | **-0.56***** | -0.51*** |
| Average UC claimant count | -0.01 | **0.46***** | 0.33*** |
| Average Unemployment claimant count | -0.05 | **0.43***** | 0.29*** |

$* \ p < 0.05, ** \ p < 0.01, *** \ p < 0.001$; Highest correlation coefficients with statistical significance in **bold**.

$r = 0.114, p < 0.05$) in 2007/8 and 2008/9 respectively and not significantly correlated in 2008/10. In our study, the results were similar where $r = -0.09, p < 0.05$ (the difference in correlation sign $(+/-)$is because we are using IMD decile not IMD score). On the other hand, the paper [25] reveals that the prevalence of obesity in Year 0 and IMD scores are strongly correlated ($r = 0.539$, $r = 0.612$, and $r = 0.625$, $p < 0.001$) in 2007/8, 2008/9, and 2009/10 respectively. Likewise, we found that the correlation is $r = -0.55, p < 0.001$.

#### 4.2.2 Demographic Factors Correlation

Here, we perform Chi-squared test[8] to investigate the links between the themes and categorical variables gender/ethnicity). Table 6 presents Chi-Squared values of the dependence of development level and BMI category on each of the demographic factors gender/ethnicity. For the development skills theme, the dependence of children development level (at or above expected level, below expected level) to gender (male, female) is $\chi^2 = 365.83, p = 1.52E-81$ and to ethnicity (Asian, Black, Mixed, White, Other) is $\chi^2 = 6.57, p = .16$.

[8]Chi-squared test (also $\chi^2$ test) is a statistical test applied to sets of categorical data to evaluate how likely it is that any observed difference between the sets arose by chance.

For the child obesity theme, the dependence of children BMI category (Normal, Overweight, Obese) to gender is $\chi^2 = 0.004, p = 1.00$ and to ethnicity is $\chi^2 = 74.73, p = 6.33E - 10$.

**Table 6** Chi-Squared values of the dependence of development level and BMI category on each of the demographic factors gender/ethnicity

| Demographic factors | Development Level | BMI Category |
|---|---|---|
| Gender | $\boldsymbol{\chi^2 = 365.83, p = 1.52E - 81}$ | $\chi^2 = 0.004, p = 1.00$ |
| Ethnicity | $\chi^2 = 6.57, p = .16$ | $\boldsymbol{\chi^2 = 74.73, p = 6.33E - 10}$ |

Our results can be compared to that of the study[21] who investigated the relationship between school readiness, socioeconomic status, and gender. School readiness is measured using the Early Development Instrument (EDI) total score which assesses the following five developmental domains: physical health and well-being, social competence, emotional maturity, language and cognitive development, and communication skills and general knowledge. The dependence of EDI total score and socioeconomic status to gender was $\chi^2 = 11.79, p = 0.008$.

## 5 Discussion

Our analysis underscores the significant correlation between socioeconomic status (SES) and early childhood development outcomes, specifically in areas such as obesity and school readiness. It is important to emphasise that this study is framed as a Birmingham case study. While Birmingham provides an important example due to its size, diversity, and persistent inequalities, the results should be understood as illustrative rather than nationally representative. Nonetheless, the findings highlight patterns and dynamics that are highly relevant to other UK cities facing similar social and health challenges. Lower SES levels in the Birmingham locality highlight prevalent issues of early childhood deprivation, as also supported by research linking poverty in early years to increased disease risk in adulthood [31, 32].

The findings reflect established literature on health inequalities, where children from deprived and socioeconomically challenged communities demonstrate higher morbidity indicators, including early childhood obesity. These outcomes support previous studies that link SES with adverse health and developmental results, showing an urgent need for targeted early interventions. However, the current data lacks granularity on whether early interventions directly influence long-term health outcomes for these children and their families. Future research could benefit from a more in-depth analysis, potentially integrating frontline health service data, such as frequency of health assessments or direct contact measures, to identify optimal support levels.

In terms of methodology, while Power BI provided a robust platform for data interrogation and visualization, there are inherent limitations in our visual analysis. A deeper justification for using specific visualization tools and regression models is necessary to clarify their role in refining the insights derived. We recommend further work to justify these tools and consider other data modeling advancements to address the nuanced relationships within our data.

It is crucial to critically reflect on the structural inequities inherent in governmental policies aimed at addressing inequalities. These policies often risk stereotyping and victimizing low-SES population groups by framing them solely through the lens of deficits or what they "lack" rather than recognizing their potential capabilities and resilience [33]. For example, policies like the 'two-child benefit cap' and chronic unemployment responses may inadvertently blame the affected communities for their circumstances rather than addressing the structural shortcomings of the system itself. This deficit-focused narrative undermines the agency and abilities of these communities, effectively excluding them from being active contributors to solutions.

Future studies should adopt a strengths-based approach that emphasizes the capabilities, resources, and resilience of affected communities. By integrating the lived experiences of these groups, research can shift from deficit-focused narratives to approaches that highlight community strengths and co-created solutions. This approach aligns with broader calls for participatory research methodologies that empower communities and amplify their voices.

The impact of the COVID-19 pandemic is another critical factor to consider, as it likely exacerbated existing inequalities by limiting access to public health services, physical activities, and outdoor spaces. For instance, the collection of the NCMP dataset was disrupted, and unemployment rates surged during this period, reflecting broader socioeconomic challenges. While this study addresses these impacts briefly, future research should take an intersectional approach to examine how the pandemic influenced health outcomes and social determinants, particularly in socioeconomically challenged communities. Moreover, investigating how community strengths and adaptive strategies mediated resilience and recovery during the pandemic could provide valuable insights into fostering sustainable interventions. This includes exploring environmental factors such as access to green spaces and their role in promoting health and well-being among children in deprived areas. These pandemic-related disruptions and imputation procedures introduce potential biases and limit comparability across years, which should be borne in mind when interpreting the results.

## 5.1 Limitations and Future Research

A notable limitation of this study is the temporal inconsistency between datasets. While the ASQ-3, Universal Credit, and Unemployment Claimants datasets align closely over the period 2019–2022, the NCMP data covers a longer span (2010–2020), and the IMD and IDACI indices are available only for 2019. These differences restrict the extent to which fully integrated longitudinal analyses can be conducted. This limitation is common when using routinely collected public health data, but it highlights the importance of future work that seeks to harmonise data collection cycles across domains to enable more robust inferential analyses. Our study is also limited by the availability and structure of public datasets, which might not capture the full picture of early years health interventions. Future research should explore more sophisticated data models and include qualitative measures to address gaps in our current data. Incorporating community feedback and using mixed-method approaches can provide more nuanced insights into how SES affects childhood health and development. Another limitation is that our analyses rely primarily on bivariate correlations. While these provide useful exploratory insights, they cannot account for potential confounding factors or establish causal relationships. Accordingly, the associations observed here should be regarded as hypothesis-generating rather than causal. Future work should employ multivariable regression models and causal inference approaches (e.g., longitudinal analyses, structural equation modelling) to more robustly assess the drivers of early years outcomes once harmonised datasets become available.

The correlation between socioeconomic deprivation and adverse health outcomes in early years is evident, with implications for both public health policies and community interventions. Addressing these issues will require multi-sectoral involvement and an intersectional research approach that considers both the structural inequities faced by communities experiencing socioeconomic disadvantage and their inherent strengths. By focusing on community-driven solutions and resilience, research can provide actionable insights to mitigate health inequalities and support child development.

## 6 Conclusions

Our findings showed a weak correlation between socioeconomic factors and wider determinants that influence childhood obesity and poor school readiness at age 2-2.5 years (Tables 4, 5). Computer modelling analysis showed concerning public health trends. Reinforcing the public health outcomes data from 2018/2019 there is again an increasing trend in childhood obesity and poor school readiness at this early years' child health developmental assessment point. Child poverty is suggestive. Although we are unable to give a robust explanation for this tentative link, the health assessment at 2 – 2.5 years is crucial to determining the relevance and appropriateness of early years public health intervention. How health visiting and early years providers of child health services worked together to identify and address wider determinants suggested by the social gradient is unknown. We measured the data routinely captured (NCMP) and so were limited by less access to whole service level data, perhaps partially due to social restrictions during the pandemic and local reorganisation of child health services. The impact of the pandemic on children and family capacity to undertake and sustain health enhancing activities is underexplored. Physical activity, access to outdoor green space and nutritional interventions (such as vitamin supplementation and breast feeding supports) are mediating public health determinants of healthy weight [16] and learning attainment. Future research could focus on the comparative analysis of 'wider' determinants, such as locality access to green spaces, and their impact on weight and nutritional aspects within local 'ward' data. Paying closer attention to the intersection between the social gradient of child health outcomes and applied collaborative interventions to reduce gross child health inequities in communities.

Finally, we note that this study is context-specific to Birmingham and should be interpreted as an exploratory case analysis. Its contribution lies in providing localised insights that can inform targeted early years interventions, while also offering transferable lessons for policymakers in other urban contexts across the UK facing similar inequalities.



# Declarations

## Ethics approval and consent to participate

We diligently sought informed consent from the data owners, providing each participant data provider with clear and detailed information about the study, its objectives, procedures, potential risks, and benefits. We took every measure to protect confidentiality, anonymising the dataset by removing all identifiable information. The project 'Data Science Collaboration 3' (Project ID: Tawil /#10052 /sub4 /R(A) /2022 /Apr /CEBE FAEC) received ethics approval from the Computing, Engineering and the Built Environment Faculty Academic Ethics Committee at Birmingham City University on April 25, 2022. This approval covers the attached paper, ensuring ethical conduct in alignment with the approved project parameters, with Professor Mark Josephs overseeing the process.

## Consent for publication

Not applicable

## Availability of data and materials

The socioeconomic factors data sets are publicly available online. The Index of Multiple Deprivation (IMD) data set and the The Income Deprivation Affecting Children Index (IDACI) data set are available online from GOV.UK [27] at England LSOA geographical level and were last reported in 2019. The Universal Credit (UC) Claimants data set is available from Stat-Xplore [28] at LSOA geographical level and was last reported in June 2022. The Unemployment Claimants data set is available from Nomis [29] at LSOA geographical level and was last reported in May 2022.



## Competing interests

We would like to confirm that there are no competing interests to declare in relation to the submitted manuscript. Our research was conducted with the objectivity and integrity, and no financial, personal, or professional interests exist that could be perceived as influencing the work or its conclusions. We are dedicated to upholding the highest ethical standards in research and publication and have adhered to all relevant guidelines and regulations.

## Funding

Not applicable

## Authors' contributions

Haidar D. and Tawil A.R. devised the project, the main conceptual ideas and proof outline. They also initiated the study, overseeing its overall direction and planning. Haidar D. designed the conceptual diagram and conducted data collection. In addition, she shaped the manuscript and wrote the sections: Related Work, Materials and Methods, and Data Analysis and Results. Vilas J. conducted data processing, performed the experiments, derived the results and presented in charts, tables, and other visual aids with the guidance of Haidar D. and Tawil A.R. Haidar D., Tawil A.R., Vilas J., Vlachos K., and N.S. collectively analysed the data and results contributing to data science aspects. Clark M., Tawil A.R., and Vlachos K. wrote the manuscript sections: Abstract, Background, and Conclusion. Additionally, Clark M. provided valuable insights related to health. Vakaj E. reviewed the paper for consistency and evaluated the overall contribution of each section to ensure cohesion in the final manuscript.

## Acknowledgements

The completion of this paper was made possible by key people from Public Health, Education and Skills Divisions at Birmingham City Council and from Birmingham Community Healthcare NHS Foundation who facilitated data collection. We would like to acknowledge all and we are very grateful for their assistance.

An appendix contains supplementary information that is not an essential part of the text itself but which may be helpful in providing a more comprehensive understanding of the research problem or it is information that is too cumbersome to be included in the body of the paper.

ARTICLE IN PRESS

# References


[1] Scally G.: England's new Office for Health Improvement and Disparities. British Medical Journal Publishing Group.

[2] Marmot M. Fair Society, Health Lives: Strategic Review of Health Inequalities in England Post-2010-Executive Summary; 2010.

[3] Spencer N, Raman S, O'Hare B, Tamburlini G. Addressing inequities in child health and development: towards social justice. BMJ Paediatrics Open. 2019;3(1).

[4] Webb C, Bywaters P, Scourfield J, McCartan C, Bunting L, Davidson G, et al. Untangling child welfare inequalities and the 'inverse intervention law'in England. Children and Youth Services Review. 2020;111:104849.

[5] Straatmann VS, Lai E, Law C, Whitehead M, Strandberg-Larsen K, Taylor-Robinson D. How do early-life adverse childhood experiences mediate the relationship between childhood socioeconomic conditions and adolescent health outcomes in the UK? J Epidemiol Community Health. 2020;74(11):969–975.

[6] Public Health England.: Chapter 5: inequalities in health. Last accessed 05.08.2022. Available from: https://www.gov.uk/government/publications/health-profile-for-england-2018/chapter-5-inequalities-in-health.

[7] Public Health England.: Health Profile for England 2021. Last accessed 05.08.2022. Available from: https://fingertips.phe.org.uk/static-reports/health-profile-for-england/hpfe_report.html.

[8] Khambati N, Mahedy L, Heron J, Emond A. Educational and emotional health outcomes in adolescence following maltreatment in early childhood: a population-based study of protective factors. Child abuse & neglect. 2018;81:343–353.

[9] National Research Council (US) Panel on Race, Ethnicity, and Health in Later Life; Anderson, N B , Bulatao, R A , Cohen, B , editors. Critical perspectives on racial and ethnic differences in health in late life. Washington (DC): National Academies Press (US); 2004.

[10] Maggio AB, Martin XE, Saunders Gasser C, Gal-Duding C, Beghetti M, Farpour-Lambert NJ, et al. Medical and non-medical complications among children and adolescents with excessive body weight. BMC pediatrics. 2014;14(1):1–9.

[11] Van de Mheen H, Stronks K, Looman C, Mackenbach J. Does childhood socioeconomic status influence adult health through behavioural factors? International journal of epidemiology. 1998;27(3):431–437.

[12] Marmot M. Social determinants of health inequalities. The lancet. 2005;365(9464):1099–1104.

[13] Marmot M. Health equity in England: the Marmot review 10 years on. Bmj. 2020;368.

[14] WHO.: Public health, environmental and social determinants of health (PHE). World Health Organization.

[15] Marmot M, Allen JJ.: Social determinants of health equity. American Public Health Association.

[16] Redsell SA, Edmonds B, Swift JA, Siriwardena AN, Weng S, Nathan D, et al. Systematic review of randomised controlled trials of interventions that aim to reduce the risk, either directly or indirectly, of overweight and obesity in infancy and early childhood. Maternal & child nutrition. 2016;12(1):24–38.

[17] Public Health England.: Chapter 4: health of children in the early years. Last accessed 05.08.2022. Available from: https://www.gov.uk/government/publications/health-profile-for-england-2018/chapter-4-health-of-children-in-the-early-years.

[18] Jones D, Innerd A, Giles EL, Azevedo LB. The Association between Physical Activity, Motor Skills and School Readiness in 4–5-Year-Old Children in the Northeast of England. International Journal of Environmental Research and Public Health. 2021;18(22):11931.

[19] Law J, Charlton J, McKean C, Watson R, Roulstone S, Holme C, et al.: Identifying and supporting children's early language needs. Available from: https://myimpact.ncl.ac.uk/UploadFiles/270761/7d1567da-dbc1-460b-94ac-53f2e62f42f7.pdf.

[20] González L, Cortés-Sancho R, Murcia M, Ballester F, Rebagliato M, Rodríguez-Bernal CL. The role of parental social class, education and unemployment on child cognitive development. Gaceta sanitaria. 2020;34:51–60.

[21] Ip P, Rao N, Bacon-Shone J, Li SL, Ho FKw, Chow Cb, et al. Socioeconomic gradients in school readiness of Chinese preschool children: The mediating role of family processes and kindergarten quality. Early Childhood Research Quarterly. 2016;35:111–123.

[22] Dotterer AM, Iruka IU, Pungello E. Parenting, race, and socioeconomic status: Links to school readiness. Family Relations. 2012;61(4):657–670.

[23] Williams AS, Ge B, Petroski G, Kruse RL, McElroy JA, Koopman RJ. Socioeconomic status and other factors associated with childhood obesity. The Journal of the American Board of Family Medicine. 2018;31(4):514–521.

[24] Cetateanu A, Jones A. Understanding the relationship between food environments, deprivation and childhood overweight and obesity: evidence from a cross sectional England-wide study. Health & place. 2014;27:68–76.

[25] Conrad D, Capewell S. Associations between deprivation and rates of childhood overweight and obesity in England, 2007–2010: an ecological study. BMJ open. 2012;2(2):e000463.

[26] NHS Digital.: National Child Measurement Programme, England 2020/21 School Year. Last accessed 01.06.2022. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/national-child-measurement-programme/2020-21-school-year.

[27] Ministry of Housing CLG.: English indices of deprivation 2019. Last accessed 25.05.2022. Available from: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019.

[28] Stat-Xplore.: People on Universal Credit. Last accessed 01/07/2022. Available from: https://stat-xplore.dwp.gov.uk/webapi/jsf/dataCatalogueExplorer.xhtml.

[29] Stat-Xplore.: Jobseeker's Allowance with rates and proportions. Last accessed 12/04/2023. Available from: https://www.nomisweb.co.uk/query/construct/summary.asp?reset=yes&mode=construct&dataset=1&version=0&anal=1&initsel=#.

[30] Gonçalves CT, Gonçalves MJA, Campante MI. Developing Integrated Performance Dashboards Visualisations Using Power BI as a Platform. Information. 2023;14(11):614.

[31] Kivimäki M, Batty GD, Pentti J, Shipley MJ, Sipilä PN, Nyberg ST, et al. Association between socioeconomic status and the development of mental and

physical health conditions in adulthood: a multi-cohort study. The Lancet Public Health. 2020;5(3):e140–e149.

[32] Khanolkar AR, Patalay P. Socioeconomic inequalities in co-morbidity of overweight, obesity and mental ill-health from adolescence to mid-adulthood in two national birth cohort studies. The Lancet Regional Health–Europe. 2021;6.

[33] Parbery-Clark C, McSweeney L, Lally J, Sowden S. How can health systems approach reducing health inequalities? An in-depth qualitative case study in the UK. BMC Public Health. 2024;24(1):2168.