[Title]

Preparation

- Github link *[Optional]*
- Number of words: ***
- Runtime: *** hours (*Memory 10 GB, CPU Intel i7-10700 CPU @2.90GHz*)
- Coding environment: SDS Docker
- License: this notebook is made available under the Creative Commons Attribution license (or other license that you like).
- Additional library *[libraries not included in SDS Docker or not used in this module]*:
  - **watermark**: A Jupyter Notebook extension for printing timestamps, version numbers, and hardware information.
  - To install required package:

```
!pip install linearmodels
!pip install nbconvert weasyprint
```

Table of contents

1. Introduction
2. Literature Review
3. Research questions
4. Data
5. Methodology
6. Results and discussion
7. Conclusion
8. References

1. Introduction

[go back to the top]

Childhood obesity remains a critical public health challenge in London, adversely affecting children's health and causing significant economic costs. To address this, the UK government's *Childhood Obesity: A Plan for Action* (Department of Health and Social Care, 2018) substantially expanded investments in physical activity infrastructure, including funding for school sports and community facilities. Public Health England's *Local*

Health Profiles report increased availability of parks, playgrounds, and sports centres across London wards (Public Health England, 2023). Nonetheless, reductions in childhood obesity have been limited and uneven, especially among younger children, underscoring obesity's multifactorial nature and insufficient impact from infrastructure expansion alone (Wyszyńska et al., 2020; Lieb, 2009).

This study aims to quantitatively evaluate how different facility types influence childhood obesity rates across age groups in London wards from 2009/10 to 2013/14. It also examines the moderating roles of socioeconomic status (SES), green space, and public transport accessibility (PTAL). Leveraging spatial and temporal fixed effects, this research seeks to uncover spatial heterogeneity and construct a novel risk-intervention two-dimensional quadrant framework combining spatial fixed effects and residuals to guide targeted, equitable public health interventions.

Requirements to Run the Analysis

An overview of packages used to run the analysis with brief explanation of their role.

```
!pip install linearmodels
!pip install nbconvert weasyprint
```

```
import os
import requests
import zipfile
from functools import reduce

import pandas as pd
import geopandas as gpd
from shapely.geometry import Point

import numpy as np
import scipy.stats as stats

import matplotlib.pyplot as plt
from matplotlib.gridspec import GridSpec

import seaborn as sns

from sklearn.ensemble import RandomForestRegressor
from sklearn.preprocessing import StandardScaler

import statsmodels.api as sm
import statsmodels.formula.api as smf
from statsmodels.iolib.summary2 import summary_col
from statsmodels.stats.outliers_influence import variance_inflation_factor

from linearmodels.panel import PanelOLS

import nbformat
import nbconvert
from nbconvert.exporters import HTMLExporter
from nbconvert.preprocessors import Preprocessor
from weasyprint import HTML, CSS
```

2. Literature Review

[go back to the top]

Increasing evidence challenges the notion that expanding physical activity facilities alone suffices to reduce childhood obesity rates. While infrastructure is necessary, obesity results from complex interactions among behaviours, socioeconomic factors, and environmental contexts (Danielli et al., 2021). Research shows age influences intervention responses: older children engage more independently with their surroundings, whereas younger children's behaviours are more shaped by familial and cultural factors (Wyszyńska et al., 2020; Ayala et al., 2021), indicating the need for age-specific approaches.

Socioeconomic disparities exacerbate risk, as children from lower SES backgrounds often face poor access to quality facilities and supportive environments (Lieb et al., 2009; Yuan, 2024). Environmental factors such as green space and public transport accessibility further modify physical activity opportunities and obesity outcomes (Danielli et al., 2021). Additionally, spatial heterogeneity within London wards manifests uneven obesity prevalence and resources, warranting geographically targeted interventions (Titis et al., 2023).

Notably, Nau et al. (2019) used panel data integrating socioeconomic and built environment factors to study youth BMI trajectories across diverse communities, emphasizing the importance of accounting for spatial and temporal heterogeneity and dynamic interactions over time in obesity research.

Building on these foundations, the present study integrates facility types, age groups, SES, and environmental factors within a spatial-temporal fixed effects framework. This approach aims to better capture how facility impacts vary by age and SES while precisely modeling spatial heterogeneity in childhood obesity patterns, thereby informing more nuanced, targeted interventions.

3. Research questions

[go back to the top]

Building on the above, and given that facility effects vary by children’s development and socioeconomic access, while environmental factors (SES, greenspace, PTAL) may modify these effects through space and time, this study addresses:

- **RQ1 |** How do different facility types affect childhood obesity across age and socioeconomic groups, after accounting for spatial and temporal factors?
- **RQ2 |** How do greenspace and public transport accessibility modify the effects of facilities on obesity?

Recognizing that obesity prevalence and environmental influences vary across geographic areas and over time, a comprehensive spatial-temporal analysis is necessary. Consequently:

- **RQ3 |** How can spatial and temporal fixed effects reveal heterogeneous patterns to inform targeted public health interventions?

4. Data Collection

[go back to the top]

This study focuses on London wards at the 2011 level (London Data Store, 2011), offering the finest publicly available spatial resolution. The research covers the academic years from 2010/2011 to 2013/2014, spanning four academic years in total.

Table 3.1: Variable Descriptions and Data Sources for Urban Childhood Obesity Analysis

Variable Category	Variable Name	Description	Time Coverage	Data Source
Obesity Rates	`obese_recp`	Reception children obesity (ages 4-5)	2010/2011–2013/2014	Ward Atlas-NCMP
	`obese_y6`	Year 6 children obesity (ages 10-11)	2010/2011–2013/2014	Ward Atlas-NCMP
Facility Types	`facility_edu`	Education facilities per 10,000 pop	2010–2014	Active Places, ONS
	`facility_comm`	Commercial facilities per 10,000 pop	2010–2014	Active Places, ONS
	`facility_local`	Local authority facilities per 10,000 pop	2010–2014	Active Places, ONS
Socioeconomic	`unemp`	Unemployment rate	2010–2014	Ward Profiles
Environmental	`greenspace_pct`	Percentage of greenspace	2010, 2012*	Ward Profiles
	`ptal`	Public Transport Accessibility Level	2010, 2011, 2012, 2014*	Ward Profiles

[Note: Missing years for greenspace and PTAL were estimated via interpolation as described below.]

The key independent variables are densities of physical activity facilities, with the proportion of children from unemployed households and children's age groups used as heterogeneity stratifiers; greenspace coverage and public transport accessibility (PTAL) are included as control variables.

1. Facilities were filtered to include only those operational before 2014 by considering their opening and closing dates. Population estimates for wards each year were taken from the Office for National Statistics (ONS, 2023). Facility density per 10,000 population was calculated using the formula:

$$\textbf{Facility density} = \textbf{(Number of operational facilities in ward / Ward population)} \times \textbf{10,000}$$

   This calculation standardises facility counts by population size, enabling meaningful comparison across wards with differing population numbers.

```
# Count active facilities per ward per year
def count_facilities_by_ward(gdf_fac_london, years, facility_types):
    wards_fac_count = wards[ward_cols].copy()
    for yr in years:
        # Filter active facilities for the year
        mask_active = (
            (gdf_fac_london['year_built'] <= yr) &
            (gdf_fac_london['closed_date'].isna() | (gdf_fac_london['closed_date'].dt.year
>= yr))
        )
```

```
        active = gdf_fac_london[mask_active]

        # Count facilities by type for each ward
        for ftype in facility_types:
            temp = active[active['mgmt_group'] == ftype]
            by_ward = temp.groupby('GSS_CODE').size().rename(f'facility_{ftype.lower()}
_{yr}').reset_index()
            wards_fac_count = wards_fac_count.merge(by_ward, on='GSS_CODE', how='left')

    # Fill missing values with 0
    for ftype in facility_types:
        for yr in years:
            col = f'facility_{ftype.lower()}_{yr}'
            if col in wards_fac_count.columns:
                wards_fac_count[col] = wards_fac_count[col].fillna(0)

    return wards_fac_count
```

```
# Compute facility density per 10,000 population
def compute_facility_density(wards_fac_count, df_pop, pop_years, facility_types,
ward_cols):
    wards_fac_pop = wards_fac_count.merge(df_pop, on='GSS_CODE', how='left')

    for yr in pop_years:
        for ftype in facility_types:
            fac_col = f'facility_{ftype.lower()}_{yr}'
            wards_fac_pop[f'facility_{ftype.lower()}_per_10000pop_{yr}'] = (
                wards_fac_pop[fac_col] / wards_fac_pop[f'pop_{yr}'] * 10000
            )

    to_keep = ward_cols.copy()
    for yr in pop_years:
        for ftype in facility_types:
            to_keep.append(f'facility_{ftype.lower()}_per_10000pop_{yr}')

    return wards_fac_pop[to_keep]

# Main execution
wards_fac_count = count_facilities_by_ward(gdf_fac_london, years, facility_types)
df_pop = load_population_data(pop_years)
wards_fac_pop = compute_facility_density(wards_fac_count, df_pop, pop_years,
facility_types, ward_cols)
```

1. Linear interpolation was applied to address missing years in greenspace and PTAL data. The method assumes that changes between two observed values occur in a steady, linear fashion over time and estimates missing values accordingly by proportionate weighting between neighbouring data points.

   This approach is reliable because environmental variables like greenspace and PTAL typically evolve gradually over short periods, and the brief interpolation spans (one to two years) minimise potential inaccuracies. Linear interpolation is a well-established technique in spatial-temporal data analysis, supported by comprehensive reviews such as Li et al., 2014.

```python
# Interpolate missing years
df_green["greenspace_pct_2011"] = df_green["greenspace_pct_2012"]
df_green["greenspace_pct_2013"] = (df_green["greenspace_pct_2012"] +
df_green["greenspace_pct_2014"]) / 2
df_ptal["ptal_2013"] = (df_ptal["ptal_2012"] + df_ptal["ptal_2014"]) / 2
```

After cleaning and interpolation, all datasets were merged on the standard geographic identifier (ward_code) to produce a comprehensive dataset for analysis.

5. Methodology

[go back to the top]

Following Nau et al. (2019), this study analyses childhood obesity across London wards from 2010 to 2014 using panel data integrating socioeconomic and built environment variables. We develop age-specific fixed effects regression models at a fine ward-level scale, incorporating detailed facility types, age groups (Reception and Year 6), and their interactions with socioeconomic status (SES). This framework enables comprehensive control over spatial and temporal heterogeneity, while also accounting for residual variation and inherent spatial risk (see Figure 1).

5.1 Multicollinearity Testing

Multicollinearity is first assessed using Variance Inflation Factor (VIF) and Pearson correlation matrix to avoid high correlations among independent variables that may bias model estimation (Hair et al., 2019):

- Variables with VIF > 5 are removed or transformed to ensure predictor independence.
- Variables with absolute Pearson correlation coefficients > 0.7 are excluded.

This procedure ensures stability and interpretability of subsequent panel regression models.

5.2 Cross-sectional OLS Regression

We conduct annual cross-sectional Ordinary Least Squares (OLS) regressions—a standard method estimating linear relationships—to assess baseline spatial associations between childhood obesity and variables. Treating years independently captures spatial variation but omits time-invariant unobserved factors.

This approach provides important guidance on variable selection and model set-up for panel fixed effects models that address temporal dynamics and unobserved heterogeneity.

5.3 Panel Fixed Effects Regression Model

Based on the screened facility and socioeconomic variables, a panel fixed effects regression model is constructed using ward-level data from 2010 to 2014 to analyse spatial and temporal variations of childhood obesity rates. The fixed effects model effectively controls for time-invariant regional characteristics and annual uniform impacts, reducing omitted variable bias and improving causal inference robustness.

The baseline model is expressed as follows:

$$\textbf{Obesity}_{i,t} = \alpha + \sum_k \beta_k \textbf{ Facility}_{k,i,t} + \gamma_1 \textbf{ Unemployment}_{i,t} + \gamma_2 \textbf{ Greenspace}_{i,t} + \gamma_3 \textbf{ PTAL}_{i,t} + \mu_i + \tau_t + \varepsilon_{i,t}$$

where:

- $\text{Obesity}_{i,t}$: Childhood obesity rate of ward i in year t
- $\text{Facility}_{k,i,t}$: Density of facility type k in ward i and year t
- $\text{Unemp}_{i,t}$, $\text{Greenspace}_{i,t}$, $\text{PTAL}_{i,t}$: Control variables for unemployment rate, greenspace coverage, and public transport accessibility level respectively
- μ_i: Ward fixed effects controlling for time-invariant regional characteristics
- τ_t: Year fixed effects controlling for common annual shocks
- $\varepsilon_{i,t}$: Error term

Models are estimated separately for Reception (ages 4-5) and Year 6 (ages 10-11) groups to examine heterogeneous effects across developmental stages.

Further, interaction terms between facility variables and socioeconomic status (SES) are incorporated:

$$\textbf{Obesity}_{i,t} = \alpha + \sum_k \beta_k \textbf{ Facility}_{k,i,t} + \sum_k \delta_k (\textbf{Facility}_{k,i,t} \times \textbf{SES}_{i,t}) + \gamma_1 \textbf{ Unemployment}_{i,t} + \gamma_2 \textbf{ Greenspace}_{i,t} + \gamma_3 \textbf{ PTAL}_{i,t} + \mu_i + \tau_t + \varepsilon_{i,t}$$

which explores how SES moderates the impact of different facility types on childhood obesity, enhancing the policy relevance and scientific rigour of recommendations.

5.4 Residuals and Spatial Heterogeneity Analysis

Inspired by approaches combining spatial risk assessment with local variation analysis (Walker et al., 2016), we leverage spatial fixed effects and residuals to capture inherent spatial risk across wards and unexplained local deviations under current government interventions. To apply this insight, we construct a two-dimensional quadrant plot classifying wards by spatial risk level and residual sign.

This classification identifies:

- **High spatial risk & positive residuals**: Persistent high-risk wards where obesity remains above expected levels, suggesting current interventions are insufficient.
- **High spatial risk & negative residuals**: Wards with high inherent risk but better-than-expected outcomes, indicating effective local measures.
- **Low spatial risk & positive residuals**: Wards usually at low risk but showing unexpected obesity increases, signalling emerging problems.
- **Low spatial risk & negative residuals**: Wards with low risk and favourable outcomes, representing stable, well-managed areas.

By combining spatial risk and residual deviations, this approach highlights where current strategies succeed or fail, guiding more precise resource allocation and multi-level public health actions.

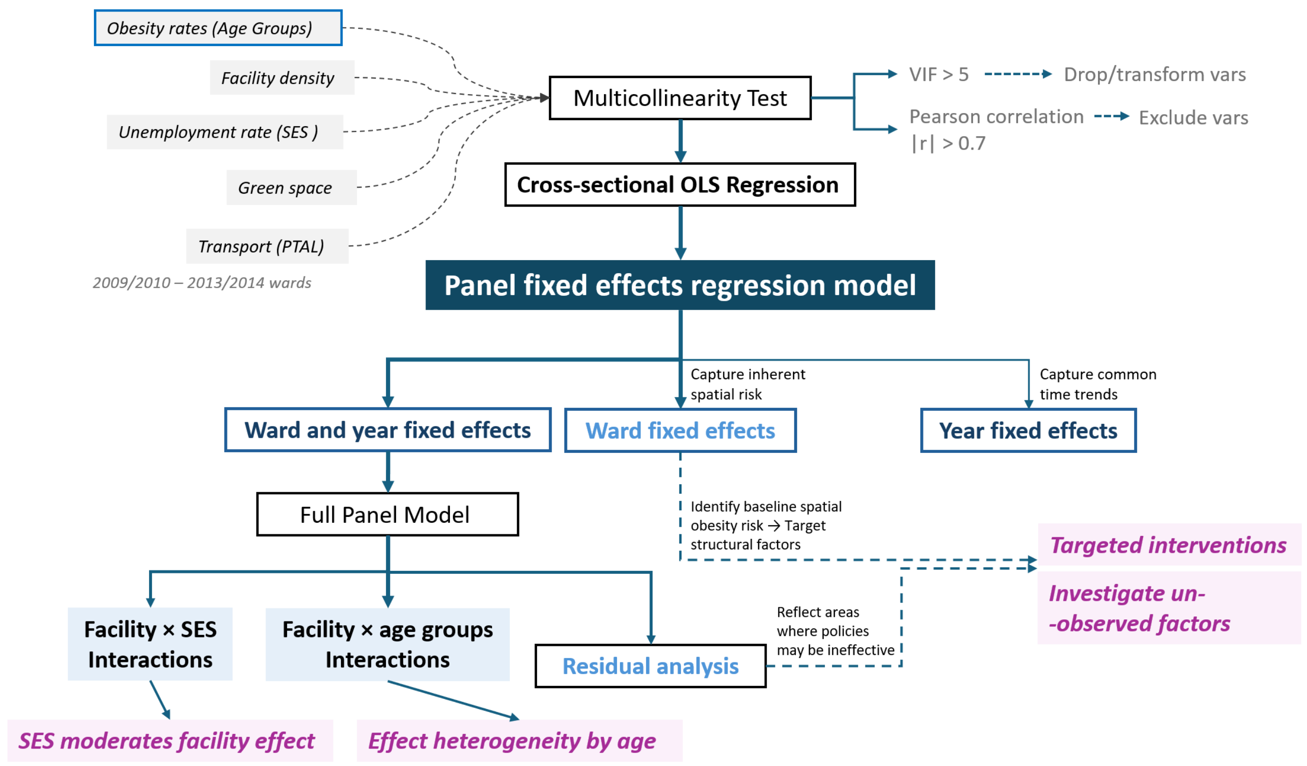

Figure 1: Methodology flow chart.

6. Results and discussion

[go back to the top]

7. Conclusion

[go back to the top]

References

[go back to the top]

Ayala, G.X., et al., 2021. The social environment and childhood obesity. *International Journal of Environmental Research and Public Health*. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8365653/ [Accessed 20 Apr 2025].

Danielli, A., Parker, L. and Swinburn, B., 2021. Systematic review into city interventions to address obesity. *International Journal of Environmental Research and Public Health*, 18(12). Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7910670/ [Accessed 20 Apr 2025].

Lieb, D.C., 2009. Socioeconomic Factors in the Development of Childhood Obesity. *Current Obesity Reports*. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5596457/ [Accessed 20 Apr 2025].

Public Health England, 2023. London’s Local Health Profiles. London: Public Health England. Available at: https://fingertips.phe.org.uk/profile/health-profiles/data#page/4/gid/8000073/pat/15/par/E92000001/ati/6/are/E12000007/iid/90323/age/201/sex/4/cat/-1/ctp/-1/yrr/1/cid/4/tbm/1/page-options/car-do-0_tre-ao-0 [Accessed 20 Apr 2025].

Titis, E., 2023. Socio-economic correlates of childhood obesity in urban and rural England. *Public Health Nutrition*. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10478054/ [Accessed 20 Apr 2025].

Wyszyńska, J., et al., 2020. Physical Activity in the Prevention of Childhood Obesity. *Frontiers in Pediatrics*. Available at: https://www.frontiersin.org/articles/10.3389/fped.2020.535705/full [Accessed 20 Apr 2025].

Yuan, Y., 2024. Public health interventions against childhood obesity in China and globally. *Preventive Medicine Reports*, 29. Available at: https://www.sciencedirect.com/science/article/pii/S2468266724002457 [Accessed 20 Apr 2025].

London Data Store (2011). *Statistical GIS Boundary Files for London*. Available at: https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london (Accessed: 16 April 2025).

London Data Store (2023). *Ward Profiles and Atlas*. Available at: https://data.london.gov.uk/dataset/ward-profiles-and-atlas (Accessed: 16 April 2025).

Office for National Statistics (ONS) (2023). *Ward Level Mid Year Population Estimates (Experimental)*. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/wardlevelmidyearpopulationestimatesexperimental (Accessed: 16 April 2025).

Active Places Power (2023). *Active Places Dataset*. Available at: https://www.activeplacespower.com/datasets/1e189022e7064e0e9b1bdaadef6d3b78/explore (Accessed: 16 April 2025).

Li, J., Heap, A. D., Potter, A., & Daniell, J. (2014). Spatial interpolation methods applied in the environmental sciences: A review. *Environmental Modelling & Software*, 53, pp.182–190. https://doi.org/10.1016/j.envsoft.2013.12.008

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis* (8th Edition). Cengage Learning.

Nau, C., Ellis, H., Huang, H., et al. (2019). Multidimensional Socioeconomic and Built Environment Factors and Their Associations with Youth Body Mass Index Trajectories across Geographically Heterogeneous Communities. *Preventive Medicine Reports*, 15, 100939. https://doi.org/10.1016/j.pmedr.2019.100939

Walker, P. G. T., et al. (2016). Estimating the most efficient allocation of interventions to reduce malaria burden and transmission. *The Lancet Global Health*, 4(7), e474-e484. Available at: https://www.thelancet.com/journals/langlo/article/PIIS2214-109X(16)30073-0/fulltext30073-0/fulltext) [Accessed 27 April 2025].