Supplementary Material Chapter 4

1. Dataset description.

Birth records were obtained from the Netherlands Perinatal Registry (Perined),[1] which provides individual-level information on maternal characteristics and birth outcomes, along the four digit postcode of the mother's place of residency at delivery. The registry covers 97% of all births in the Netherlands.[1] Statistics Netherlands (CBS) performed individual-level linkage of Perined records to CBS sociodemographic data national registries. Due to stillbirths being non-linkable, records available for analysis consisted of live births only. Further details on the linkage procedure are available elsewhere.[2] Information on ethnicity, educational level, and household income was extracted from CBS registries.

The Netherlands Institute for Social Research (SCP) Status Score was used as measure of neighbourhood SES.[3] The SCP Status Score is a relative measure of neighbourhood SES available for four-digit postcodes areas. The score is calculated using factor analysis by aggregating the information of all neighbourhood inhabitants to summarise: 1) percentage of inhabitants with a low income, 2) percentage of inhabitants without a paid job, and 3) percentage of inhabitants with a low education level. For this work we used the scores for the reporting years between 2010 and 2017. The SCP Status Scores have been previously used in health inequalities research in the Netherlands.[2,4] The Status Scores for each reporting year were categorized into advantaged and disadvantaged SES. First, quintiles were calculated and the areas in the lowest quintile were set to the disadvantaged category and the areas in the remaining quintiles were set to the advantage category. Then, the categories were assigned to the birth records based on postcode and year of birth, e.g., the Status Score category for reporting year 2017 were assigned to births occurred during 2017.

The neighbourhood social environment score was obtained from the assessment of neighbourhood liveability index ("Leefbaarometer") by the Netherlands Ministry of the Interior.[5] The social environment score is one of the dimensions of the Leefbarometer and is based on the following indicators: residential stability (number of relocations), life stage diversity of households (e.g., single, couples, family households), population density per 100 m2, and perceived social cohesion. The social cohesion component is based on a representative national survey and summarises information on the frequency and quality of social contact among neighbourhood inhabitants, sentiment of belonging, and perceived social support.[6] The social environment scores have been released for the years 2014 and 2017. Similar to the SCP Status Score, for each reporting year available, quintiles were calculated and categories (disadvantaged vs non-disadvantaged) were created. Then, the social environment categories were assigned to the birth records based on postcode and year of birth. Births occurred between 2010 and 2014 received the category for reporting year 2014 and births occurred between 2015 and 2017 received the category for reporting year 2017.

Additional neighbourhood characteristics were obtained from the postcode-level data collected yearly by CBS, which is calculated by aggregating the information of all residents from each area (average of 4,000 inhabitants).[6]

Due to privacy concerns, neither the SCP Status Scores nor the Leefbarometer measure on social environment are calculated for areas with less than 100 households. Thus, this measures could not be assigned to births from mothers living in such areas.

Figure S1 shows the population flow diagram. Table S1 contains a summary of the population characteristics.

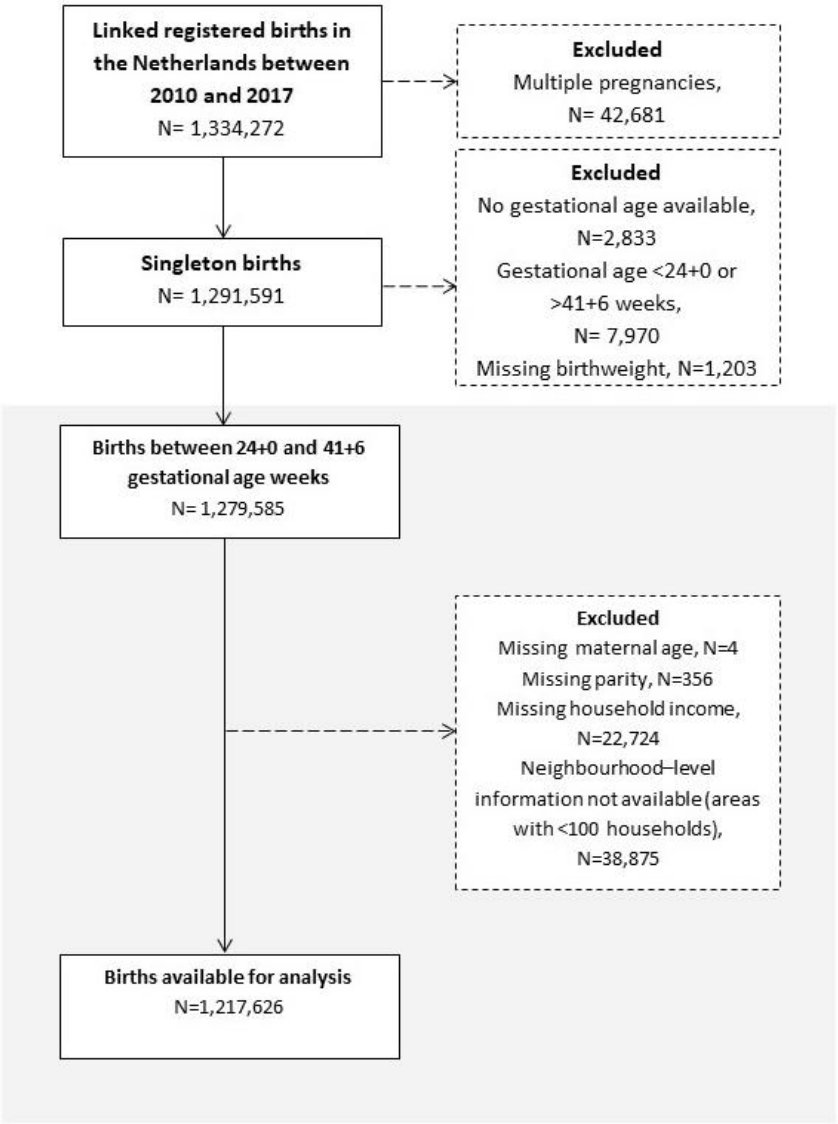

Figure S1. Population flow diagram.

Table S1. Summary of population characteristics by neighbourhood socioeconomic status (SES) category

	Neighbourhood SES category		
Characteristics	**Total**	**Disadvantaged**	**Non-disadvantaged**
N, %	1,217,626	881,738 (72.4)	335,888 (27.6)
Maternal age mean (SD)	30.6 (4.8)	30.0 (5.1)	30.8 (4.7)
Primiparous, N , %	545,904 (44.8)	155,230 (46.2)	390,674 (44.3)
Ethnicity, N, %			
Dutch	994,595 (81.7)	235,071 (70.0)	759,524 (86.1)
Morocco	30,759 (2.6)	18,911 (5.6)	11,848 (1.3)
Turkey	20,881 (1.7)	13,490 (4.0)	7,391 (0.8)
Suriname	11,568 (1.0)	6,388 (1.9)	5,180 (0.6)
Antilles	9,550 (0.8)	5,515 (1.6)	4,035 (0.5)
Others non western	76,372 (6.3)	31,796 (9.5)	44,576 (5.1)
Others western	73,901 (6.0)	24,717 (7.4)	49,184 (5.6)
Education, N , %			
Low	106,093 (8.7)	49,838 (14.8)	56,255 (6.4)
Medium	497,265 (40.8)	142,987 (42.6)	354,278 (40.2)
High	456,387 (37.5)	104,047 (31.0)	352,340 (40.0)
Unknown	157,881 (13.0)	39,016 (11.6)	118,865 (13.5)
Standardized disposable Household income Median in euros(IQR)	26,259 (15,743)	22,197 (15,697)	27,587 (15,127)
Small-for-gestational-age, N , %	131,309 (10.8)	43,272 (12.9)	88,037 (10.0)
Neighbourhood charateristics			
Percentage non-western migrants, mean (SD)	13.67 (15.04)	27.81 (19.99)	8.28 (7.41)
Average home value (1000 euros), mean (SD)	243.05 (83.6)	179.76 (52.3)	267.16 (80.7)
Disadvantaged Social Context, N, %	243,526 (20.0)	179,122 (53.3)	64,404 (7.3)

2. Underlying model for the outcome, assessment of potential confounding and sensitivity analyses.

The exchangeability identification assumption (described in Table 2, main text) requires that the underlying models used in the g-formula are adjusted for factors that confound the exposure-outcome relationship, i.e., covariates that are expected to be common precursors of the exposure and the outcome.[7] Additionally, the decomposition of the Total Effect into Direct and Indirect effects assumes no unmeasured (and uncontrolled) confounding in the mediator-outcome and exposure-outcome relationships.[8,9]

At the individual level, the models included a first set of variables: maternal age in categories (≤19, 20-34, ≥35 years), parity (nulliparous vs multiparous), maternal ethnicity as registered in CBS (Dutch, Turkish, Moroccan, Surinamese, Antillean, others western, other non-western),[10] maternal educational level in categories as defined by CBS (low, medium, high, unknown),[11] and equivalised disposable household income (quintiles). Household income is often preferred over individual-level income in inequalities research as it might be a more useful indicator of SES, particularly for women, who may not be the main earners in the household.[12] At the neighbourhood level, neighbourhood average home value (quintiles), and percentage of non-western migrants (quintiles) were included. A measure of urbanity was not included as the measure for neighbourhood social environment already accounts for this. Additionally, models were adjusted for year of birth to account for cohort effects. Table S2 shows the odds ratios for the underlying model for the outcome used in the g-formula.

At the neighbourhood level, there might be features that are related to both neighbourhood social environment (mediator) and SGA (outcome), i.e., a potential mediator-outcome confounder. This is the case for green space, which can influence both, social environment (e.g., by facilitating social interaction) and SGA (e.g., by promoting physical activity). [13,14] It must be noted that greenness has been highlighted in the literature as a mediator for the relationship between neighbourhood SES and birth outcomes.[15] If necessary, the g-formula approach would enable analyses involving confounders affected by the exposure (exposure-dependent).[16] To assess the relevance of the aforementioned factors we used information from the Statistics Netherlands Neighbourhood Characteristics.[17] We fitted models that were additionally adjusted for a measure of neighbourhood greenness, i.e., the percentage of neighbourhood land dedicated to green space and recreational areas.

Our dataset, unfortunately, only consists of live-births. This selection on live-births would result in a violation of the exchangeability condition. The underlying models were adjusted for known common causes of stillbirths and SGA, i.e., maternal age, parity, education, and income. Adjustment for common causes of the collider and the outcome is a well-known strategy to reduce this bias (see Table 2, main text). However, we also have access to additional information on lifestyle factors (alcohol consumption, drug use, and smoking) and pre-existent conditions (hypertension and diabetes) which are also common causes of stillbirths and SGA. These variables, particularly life style factors, are known to suffer from

underreporting in the perinatal registry. Thus, they have been included in the models in a sensitivity analysis.

Last, as a sensitivity analysis, we assessed the impact of women moving to another neighbourhood during (or shortly prior) their pregnancy, by restricting the underlying models used in the g-formula to women who have been living in the same residential address for at least two years at the time of delivery. Results from this analysis can be found in table S3 (Model III).

Table S3 shows the g-formula total and mediated effects (point estimates) for the sensitivity analyses. Model I, used the underlying model adjusted for the first set of variables (individual and neighbourhood characteristics). Model II, was additionally adjusted for neighbourhood greenness. In Model III we accounted for maternal lifestyle factors and pre-existent conditions (addressing selection on live-births). Last, Model IV was restricted to women who have been living in the same residential address for at least two years at the time of delivery. We observed that the adjustment for the additional variables resulted in little to no change in the point estimates of the TE, NIE, and NDE. Based on these results, Model I (most parsimonious model) was used for the g-formula underlying models.

Table S2. Odds ratios (95% CI) from underlying logistic regression for the relationship between neighbourhood SES and SGA

Disadvantaged SES	1.06 (1.04,1.07)
Disadvantaged Social Context	1.03 (1.01,1.05)
2011	1.03 (1.01,1.06)
2012	1.03 (1.01,1.05)
2013	1.06 (1.03,1.08)
2014	1.04 (1.02,1.07)
2015	1.04 (1.01,1.06)
2016	1.05 (1.03,1.07)
2017	1.04 (1.01,1.06)
Multipara	0.46 (0.46,0.47)
Maternal ethnicity	
Morocco	0.73 (0.71,0.76)
Turkey	0.89 (0.85,0.93)
Suriname	2.13 (2.03,2.23)
Antilles	1.28 (1.21,1.35)
Others non western	1.17 (1.14,1.20)
Others western	0.95 (0.93,0.98)
Equivalized disposable household income	
Quintile 2	0.86 (0.84,0.87)
Quintile 3	0.75 (0.74,0.77)
Quintile 4	0.68 (0.67,0.69)
Quintile 5 (high)	0.68 (0.67,0.69)
Mother's education	
Medium	1.30 (1.28,1.32)
Low	1.70 (1.66,1.74)

Unknown	1.29 (1.27,1.32)
Maternal age	
19-35	1.05 (1.00,1.10)
Above 35	1.26 (1.20,1.32)
Neighbourhood percentage of non-western migrants	
Quintile 2	1.03 (0.98,1.07)
Quintile 3	1.08 (1.04,1.13)
Quintile 4	1.18 (1.13,1.22)
Quintile 5 (high)	1.29 (1.24,1.34)
Average neighbourhood house value	
Quintile 2	0.96 (0.95,0.98)
Quintile 3	0.94 (0.93,0.96)
Quintile 4	0.96 (0.94,0.98)
Quintile 5 (high)	0.99 (0.97,1.01)

SGA, small-for-gestational-age.
Omitted categories correspond to reference categories.
Underlying model used in the g-formula, which includes exposure, mediator variables and confounders (maternal age, parity, maternal ethnicity, maternal educational level, equivalised-household income, neighbourhood average home value, and percentage of non-western migrants).

Table S3. G-formula total and mediated effects (point estimates) for model assessing potential confounding and sensitivity analyses.

	Model I	**Model II**	**Model III**	**Model IV**
TE, on SGA	-0.0070	-0.0072	-0.0071	-0.0072
NIE, on SGA	-0.0008	-0.0008	-0.0009	-0.0009
NDE, on SGA	-0.0063	-0.0064	-0.0062	-0.0063

SGA, small-for-gestational-age.
Model I, adjusted for individual and neighbourhood characteristics.
Model II, was additionally adjusted for neighbourhood greenness.
In Model III we accounted for maternal lifestyle factors and pre-existent conditions.
Model IV was restricted to women who have been living in the same residential address for at least two years at the time of delivery.

3. Mean values from simulated scenarios and absolute effect sizes.

Absolute effect sizes for total effect (TE), total direct effect (NDE) and total indirect effect (NIE) are shown in Table S4. Table S5 shows the mediator and outcome means for the simulated scenarios. To calculate the relative TE the absolute TE value is divided by the mean of the CF *ses*=1 and multiplied by 100. This is interpreted as the percentage change attributed to the hypothetical improvement in Neighbourhood SES from disadvantaged (*ses*=1) to non-disadvantaged (*ses*=0). The same calculation is done for the NDE and the NIE. The proportion mediated was calculated as *PM*=*NIE*/*TE*, i.e., -0.0008/-0.0070 = 0.114. Thus, the NIE accounted for 11.4% of the TE.

Table S5 also shows the comparison of observed means vs natural course scenario means along with the comparison between counterfactual scenario means. The comparison between observed and natural course outcome means is used as a check against gross misspecification.

Table S4. Absolute effect sizes (point estimates and 95% confidence interval).

TOTAL EFFECT (TE)			
	Mean	**95% CI**	
TE, on SGA	-0.0070	-0.0085	-0.0070
TE, on Social environment	-0.1647	-0.1800	-0.1495
NATURAL INDIRECT EFFECT (NIE)			
NIE, on SGA	-0.0008	-0.0010	-0.0007
NATURAL DIRECT EFFECT (NDE)			
NDE, on SGA	-0.0063	-0.0078	-0.0058

SGA, small-for-gestational-age.

Table S5. Mediator and outcome means.

			Counterfactual scenarios	
	Observed data	**Natural course**	**Disadvantaged (*ses*=1)**	**Non-disadvantaged (ses=0)**
Social environment	0.2000	0.2004	0.2866	0.1187
SGA	0.1078	0.1078	0.1129	0.1059

SGA, small-for-gestational-age; CF, counterfactual scenario.

4. Link to R code used in illustrative case

The R code used in the illustrative case is available in the GitHub repository

References Supplementary Material Chapter 4

1. Perined. Perinatale Zorg in Nederland 2015 [Internet]. 2015. Available from: https://assets.perined.nl/docs/980021f9-6364-4dc1-9147-d976d6f4af8c.pdf

2. Burgos Ochoa L, Bertens LCM, Garcia-Gomez P, Ourti T van, Steegers EAP, Been J v. Association of neighbourhood socioeconomic trajectories with preterm birth and small-for-gestational-age in the Netherlands: a nationwide population-based study. The Lancet Regional Health-Europe. Elsevier; 2021;10:100205.

3. Social and Cultural Planning Office (SCP). Socio-Economic Status by postcode area [Internet]. 2019. Available from: https://bronnen.zorggegevens.nl/Bron?naam=Sociaal-Economische-Status-per-postcodegebied

4. Tromp M, Eskes M, Reitsma JB, et al. Regional perinatal mortality differences in the Netherlands; care is the question. BMC Public Health. BioMed Central; 2009;9(1):1–9.

5. Mandemakers J, Leidelmeijer K, Burema F, Halbersma R, Middeldorp M, Veldkamp J. Leefbaarometer 3.0 Instrumentontwikkeling [Internet]. 2021. Available from: https://www.leefbaarometer.nl/resources/LBM3Instrumentontwikkeling.pdf

6. Central Bureau for Statistics. Statistiche gegevens per vierkant postcode [Internet]. 2020. Available from: https://www.cbs.nl/-/media/cbs/dossiers/nederland-regionaal/postcode/statistische-gegevens-per-vierkant-en-postcode.pdf

7. Pearl J. Causality: Models, reasoning and inference cambridge university press. Cambridge, MA, USA. 2000;9:10–11.

8. Wang A, Arah OA. G-computation demonstration in causal mediation analysis. Eur J Epidemiol. Springer; 2015;30(10):1119–1127.

9. Greenland S, Robins JM. Identifiability, exchangeability and confounding revisited. Epidemiologic Perspectives & Innovations. Springer; 2009;6(1):1–9.

10. Central Bureau for Statistics. Migration background [Internet]. 2019. Available from: https://www.cbs.nl/en-gb/our-services/innovation/project/cbs-experimenting-with-dot-maps/migration-background

11. CBS. Education-level [Internet]. Available from: https://www.cbs.nl/en-gb/news/2018/20/well-being-not-distributed-equally/education-level

12. Galobardes B, Shaw M, Lawlor DA, Lynch JW, Smith GD. Indicators of socioeconomic position (part 1). J Epidemiol Community Health (1978). BMJ Publishing Group Ltd; 2006;60(1):7–12.

13. Lee KJ, Moon H, Yun HR, et al. Greenness, civil environment, and pregnancy outcomes: perspectives with a systematic review and meta-analysis. Environmental Health [Internet]. 2020;19(1):91. Available from: https://doi.org/10.1186/s12940-020-00649-z

14. Jennings V, Bamkole O. The Relationship between Social Cohesion and Urban Green Space: An Avenue for Health Promotion. Int J Environ Res Public Health. 2019 Feb;16(3).

15. Ham M van, Manley D, Bailey N, Simpson L, Maclennan D. Understanding neighbourhood dynamics: New insights for neighbourhood effects research. Understanding neighbourhood dynamics. Springer; 2012. p. 1–21.

16. Daniel RM, Cousens SN, Stavola BL de, Kenward MG, Sterne JAC. Methods for dealing with time-dependent confounding. Stat Med. Wiley Online Library; 2013;32(9):1584–1618.

17. Central Bureau for Statistics. Wijk en buurt statistieken. 2020; Available from: https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/wijk-en-buurtstatistieken