Description of DNBC urban exposome data generated at ISGlobal as part of the Athlete project

Brief summary of the process generating the urb ex data

We sent the complete home address history of geocoded addresses and the dates of changing addresses, DNBC births and interviews of DNBC participants to ISGlobal (the full address history of the mothers a year before the delivery of the index child participating in the DNBC and for the children after birth until 2017 as children may not live with the mothers after birth for the participants with available addresses, maternal interview 1 data and liveborn child) for use within the Lifecycle and Athlete projects. ISGlobal used the geocoded addresses to generate the urban exposures for a subset of the DNBC population; Within the Lifecycle project, the DNBC study population with urban exposome assessment was restricted to those within Copenhagen and exposure until 12 years of child's age. For Athlete, the urban environment subset of DNBC includes 28936 children living in the 4 larger cities: Copenhagen, Aarhus, Odense and Aalborg (the social context data, e.g. area SES information has been generated only for Copenhagen area as part of Lifecycle) and exposures were estimated for ages after 12 years of age when possible and a wider range of exposures were estimated.

The number of observations for the urban exposome variables is decreasing with age of the child (i.e. numbers of exposures that include older ages are fewer than those for pregnancy) which is due to loss of follow-up and that DNBC families' movements in other areas than the 4 large cities for which urban environment can be assessed. At child age 18 years, the number of observations is lower because not all the children reached the age of 18 at the time of collecting information about addresses. Because in DNBC we have the full address history of each participant, it means that for some participants we even have more than one address within the same period. For air pollution, the time-weighted means considering the time registered at each address have been calculated. For categorical variables, the address with the longest time during each period had been used. For example, if a Subject 1 during pregnancy (year 0) lives in an URBAN address, but a few months later during pregnancy (still in year 0) Subject 1 moves to SUBURBAN address. So, during the same temporal period (year 0) the same subject has two measures. For those cases and to exclude multiple addresses (and thus different exposures) of the same participants at the same temporal period, we selected the address where they lived longer during each period. For subjects who lived 50% and 50% with the same duration of time at two different addresses, one of the two addresses were randomly selected.

Remarks about certain urb exposome variables in DNBC

Ambient air pollution data

Note: For time-dependent variables such as air pollutants and meteorological data the exposure at birth was not calculated because it will represent calculating only 1 single exposure day. At the contrary, non-time dependent variables were calculated also at the birth time period so they don't suffer any temporal adjustment.

Food environment data

Note: food environment data lacks proper quality validation.

Variable: ath_fesupermarket300
All values are 0 for this variable as most of the geocodes don’t have a large supermarket nearby.

Variable: ath_fesupermarket800
All values are 0 for this variable as most of the geocodes don’t have a large supermarket nearby.

Meteorological data

Note: For time-dependent variables such as air pollutants and meteorological data the exposure at birth was not calculated because it will represent calculating only 1 single exposure day. At the contrary, non-time dependent variables were calculated also at the birth time period so they don't suffer any temporal adjustment.

Variable: ath_uvddc
We have between 84-99% missing values in DNBC at pregnancy, all trimesters, and y1. Also, there’s 20-65% missings for y2-y5. The source data (TEMIS) wasn’t available until 2002 and most of the DNBC children were born before 2003, which explains why we have a lot of missing data in the younger ages.

Variable: ath_uvdec
We have between 84-99% missing values in DNBC at pregnancy, all trimesters, and y1. Also, there’s 20-65% missings for y2-y5. The source data (TEMIS) wasn’t available until 2002 and most of the DNBC children were born before 2003, which explains why we have a lot of missing data in the younger ages.

Variable: ath_uvdvc
We have between 84-99% missing values in DNBC at pregnancy, all trimesters, and y1. Also, there’s 20-65% missings for y2-y5. The source data (TEMIS) wasn’t available until 2002 and most of the DNBC children were born before 2003, which explains why we have a lot of missing data in the younger ages.

Social context data

Note: Social context data in DNBC is based on a subset of the cohort living in Copenhagen area.