

International Journal of Population Data Science

Journal Website: www.ijpds.org



Kids' Environment and Health Cohort

Tolu Adedire^{1,2}, Esther Lewis^{1,2}, Dilly Stephenson^{1,2}, Nathan O'Connor^{1,2}, Leah Quinn^{1,2}, and Hannah Goode^{1,2}

¹Office for National Statistics

²Administrative Data Research UK (ADR UK)

This presentation discusses Kids' Environment and Health Cohort (KEHC), a bespoke Administrative Data Research UK (ADR UK) funded linkage project. KEHC will be a comprehensive and yearly updated dataset for children born in England from 2006 onwards, integrating health, education, pregnancy, and environmental data. This integrates data from different administrative systems including the Office for National Statistics, the National Health Service, and the Department for Education as well as open environmental data.

This presentation will outline the deterministic and probabilistic linkage methods utilised and the quality assurance undertaken through clerical review and bias analysis. Additionally, we will delve into the challenges encountered in the process of data ingest, linkage, and delivery due to the diverse origins and purposes of the collected data, as well as temporal disparities. In navigating the intricacies of multiple linkages, we will discuss our unique methods, including plans to use an indexing first approach.

The first cohort of approximately 10 million children has been created to form the spine and further linkages to the 2011 and 2021 UK census data have been concluded. Linkages between address histories and open environmental data is ongoing. Linkage results have been promising, with high match rate and excellent quality.

KEHC is a powerful example of how integrated data can be used for research for the public good.

KEHC will provide a unique opportunity for academic researchers investigating the longitudinal effects of local environments on children's health and educational outcomes and will become one of ADR UK's flagship Research Ready Datasets.



Temporary page!

$\LaTeX$ was unable to guess the total number of pages correctly. As there was some unprocessed data that should have been added to the final page this extra page has been added to receive it.

If you rerun the document (without altering it) this surplus page will go away, because $\LaTeX$ now knows how many pages to expect for this document.