Researchers need access to human population data in order to model everything from transport patterns and spread of disease to consumer behaviour and household energy use. Privacy considerations mean that real-world datasets for these types of applications are often hard to come by, but researchers can now use ‘synthetic data’ instead. This anonymised, artificially generated data retains the original data’s statistical properties, but contains no sensitive information linked to real people.

A new Turing-developed tool – the Synthetic Population Catalyst (SPC) – aims to make it vastly easier for researchers to access synthetic population data. Using a computational process that typically takes just seconds to run, it combines real-life data from a variety of sources – including UK census data and health surveys – and outputs a synthetic dataset for any user-specified county in England or Wales. The dataset zooms right down to the individual level, including variables related to personal health, household type, employment status and social interactions. If researchers need a specific variable that is not covered by the basic version of the SPC, they can also work with the SPC team to incorporate it into the output – they just need the necessary data source.

As a fully documented, open-source tool, the SPC can be readily adapted across multiple domains. One Turing project is using synthetic populations generated by the tool to feed a model for analysing the individual health impacts of climate change-related extreme heat. Another project, currently based at the Technical University of Denmark, is using SPC-generated data to find out why obese people are more susceptible to COVID-19 and other viral diseases, with the aim of identifying underlying socio-economic factors that could be targeted to improve population health.

Read more: The Synthetic Population Catalyst on GitHub

“My research into the links between obesity, health and poverty wouldn’t be possible without the data provided by the Synthetic Population Catalyst. The Turing’s cross-disciplinary approach to developing this tool means that it will be useful across so many different domains.”

Karyn Morrissey
Professor of Applied Economics at Technical University of Denmark and former Turing Fellow