Why adding GP data to UK Biobank will be transformational for research

Professor Martin Rutter is Deputy Chief Scientist at UK Biobank. Here, Martin sets out why adding de-identified primary care data to UK Biobank's database will transform the research potential overnight.

Last October, Wes Streeting announced a hugely exciting development for healthcare research. The Secretary of State for Health and Social Care declared that NHS England (NHSE) will take responsibility for GP data of NHSE patients in consented cohorts like UK Biobank. This was incredibly welcome news to all of us at UK Biobank as it means that soon, we can add the coded GP data of our 500,000 volunteers to our database. These data are eagerly anticipated by UK Biobank's global research community (currently standing at over 20,000 researchers) and the volunteers who generously consented to share this information when they joined UK Biobank over 15 years ago.

A treasure trove of data for research

For those who have not heard of UK Biobank, we are the world's most comprehensive source of medical data available for health research in the public interest. We have collected biological, health and lifestyle information from half a million UK volunteers, which is stripped of identifying information before being made available to vetted scientists around the world.

The beauty of UK Biobank is that we are always adding more information from our willing participants, who recognise that their contributions will benefit future generations. The growing dataset which approved researchers can access includes hospital admissions data, the world's largest set of whole genome sequencing data, imaging data from an incredible 90,000 participants and the world's largest study looking at proteins circulating in the blood.

Studies using UK Biobank data are translating into real world health impacts, including new ways to predict, diagnose and treat diseases. For example, a team of scientists have found protein patterns that predict the onset of dementia more than a decade before formal diagnosis, bringing a blood test for this condition a step closer. UK Biobank data has also been used to create a simple, genetic score to detect those at higher risk of conditions including breast and prostate cancer, heart disease and strokes, enabling earlier and more effective intervention and prevention. Another finding has even changed the way doctors diagnose and treat diabetes, as researchers found that the incidence rate of Type 1 diabetes remains about the same up to at least age 60 years – previously thought to be only a disease developed in childhood.

The power of patient data

Although UK Biobank's depth of data has already proved transformational for public health research, the GP data for our participants is a long-awaited missing piece of the puzzle. Our GPs are the first port of call for many of us when we feel unwell and lots of conditions are only treated by our GP, not by a hospital. Take diabetes, asthma, migraines and mental health conditions – most of these are picked up and managed by our GPs.

Adding the GP data will roughly double the cases of dementia and depression seen within UK Biobank, as well as increase the number of early-stage disease cases. Having these data will allow researchers to examine the full spectrum of disease severity, bringing new diagnostics and treatments closer.

GP data has already been used successfully for research, for example, during the COVID-19 pandemic. Emergency legislation allowed us to access UK Biobank participant’s GP data for COVID-19 research. Access to these data allowed rapid discoveries to be made, from identifying risk factors for severe infection, to understanding how infection impacts our wider organs like the brain. GP data access had a powerful impact on the trajectory of COVID-19 treatment, and it’s possible for us to see the same impact across other diseases too.

Looking ahead

Since Wes Streeting's announcement last October, we’ve been working closely with NHS England to ensure we are ready to receive the data. The data we will receive will be coded, meaning no letters of other free text will be included to ensure patient confidentiality.

The work of UK Biobank and use MY data highlights how responsible data sharing improves health outcomes and ultimately saves lives. The National Patient Data Day conference is a brilliant platform for discussions about how - when done safely, securely and with principals of transparency – sharing healthcare data can revolutionise public health.