

International Journal of Population Data Science

Journal Website: www.ijpds.org



Investigating cystic fibrosis in Wales using linked routine genetic data

Robert Maddison¹, Rebecca Cannings-John¹, Fiona Lugg-Widger¹, Karen Reed¹, and Andrew Fry¹

¹Cardiff University, Cardiff, United Kingdom

Objectives

Routinely collected clinical genetic test records are under-utilised for research. We aimed to explore the utility of such records for population genetics and epidemiology by linking cystic fibrosis (CF) genetic test reports to administrative datasets in a trusted research environment (TRE).

Conclusions

We have demonstrated that linkage of routine clinical genetic test records can facilitate population genetics, epidemiological and healthcare research. Our experience has highlighted limitations of using linked routine clinical genetic data in TREs which have particular implications for rare disease research.

Methods

We manually curated 16,181 NHS CFTR gene test reports for anonymisation and linkage. Cohorts of people with CF (pwCF) and CF carriers were validated using the UK CF Registry. We investigated the population genetics of CF in Wales longitudinally, using demographic information to control for related individuals in the anonymised data, and geographically, using spatial mapping and linear regression. The CF carrier phenotype was broadly characterised using multiple healthcare datasets to conduct a phenome wide association study. In line with CF community research priorities, we investigated the long term effects of CFTR modulator therapies on pwCF in Wales.

Results

Simulation modelling suggests we detected first-degree relatives in the anonymised data with approximately 70% sensitivity. Local authorities in South Wales exhibited significantly higher rates of CF births, carrier status, testing, and CFTR allelic diversity compared with other regions. Higher rates of negative CF tests correlate with deprivation indicators. No statistically significant differences in allele frequencies were observed between 1993 and 2023. Frequencies for rare alleles relating to small counts could not be reported safely due to disclosure risk. CF carriers exhibit an intermediate phenotype relative to controls and pwCF in respect of physiological measurements and risk of developing CF-related conditions. The age at which pwCF are first prescribed modulator therapy may affect age-related comorbidity risks in later life.



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.