International Journal of Population Data Science

Journal Website: www.ijpds.org

Using Free Text From Medical Notes To Enrich a Longitudinal Cohort Study

Humphries, K¹, Davies, A¹, Wren, Y¹, and Sandy, J¹

¹University of Bristol

Introduction

The Cleft Collective Cohort study is the world's largest multi-disciplinary cleft lip/palate research programme. Despite being one of the most common birth anomalies, the causes of clefting are unknown. Treatment involves a considerable burden of care from birth onwards, together with a variety of social and psychological challenges.

The richness of these data allows for more in depth phenotypic analyses.

Objectives and Approach

Our aim is to create the infrastructure and resources necessary to gain important new knowledge that will advance our understanding of the causes of cleft lip/palate, inform treatment and improve the lives of people born with cleft; data linkage is a key aspect of this. There are challenges associated with linking to multiple data sources (NHS Digital, Cleft National Registry). However, using the consent obtained, we are also able to link directly to the participant's medical notes held by their NHS cleft team; enabling us to access a variety of data, including free text.

Results

The study already collects social and demographic data via questionnaires and genetic data from biological samples. Data linkage enriches this but also enables us to validate and address missing data problems. However, linkage to external sources brings many challenges, including, governance, costs, access issues. By gaining direct access to cleft team medical notes these issues are significantly reduced and provide us with rich phenotype data that cannot be obtained elsewhere, for example, via 'read codes' within electronic medical records. By tailoring our own data collection tool we can collect specific cleft data to enhance the resource and allow for subtype analyses. This process can be repeated throughout the duration of this longitudinal study for subsequent data.

Conclusion/Implications

Data linkage is a valuable resource but comes with many challenges. One route to overcome many of these issues is by accessing free text data directly from participants medical notes.