International Journal of Population Data Science

Journal Website: www.ijpds.org

Challenges in Data Linkage – Experiences from An Upper Gastrointestinal Cancer Data Linkage Study

Khan, NN^{1*}, Ioannou, L¹, Pilgrim, J², Earnest, H¹, Maharaj, A¹, Croagh, D³, Liew, D¹, Atwood, D⁴, Holland, J¹, Philip, J⁵, Emery, J⁵, Ijzerman, M⁵, Brown, W², Zalcborg, J¹, and Evans, S¹

¹Monash University

²Alfred Health

³Monash Health

⁴Consumer Advocate

⁵The Victorian Comprehensive Cancer Centre

⁶The University of Melbourne

Introduction

Linked, population-level data is valuable for mapping patterns of care and evaluating health service utilisation, particularly in difficult-to-reach populations. Upper gastrointestinal (UGI) cancers have a dismal prognosis, creating difficulties engaging patients in research. The utility of a linked dataset in this population is of high value.

Objectives and Approach

Key objectives included identifying the operational and feasibility issues associated with linking Australian state-based administrative and registry data for understanding health service utilisation in UGI cancers. Datasets pertained to hospital admissions, radiotherapy, community health, primary care, palliative care, Medicare and Pharmaceutical Benefits Schedule's and UGI cancers.

Results

From a logistical perspective, data access request approval processes varied, with some requiring consent to be sought from individual services contributing data. The availability of unique person-level identifying information varied widely. Additionally, the time period of data capture differed between and within datasets, limiting the quality of the linked data. Significant costs were associated with linking with primary care and Medicare and Pharmaceutical Benefits Schedule's. Federal dataset linkage required at least a one-year waiting period.

Conclusion / Implications

Whilst in theory data linkage is a powerful mechanism for obtaining population-level data, in reality, there are many logistical and financial barriers to linking multiple datasets. Consequently, critical data, which has the potential to inform policy and improve patient outcomes, cannot be procured.

*Corresponding Author:

Email Address: nadia.khan@monash.edu (NN Khan)