

Extracting Family Connections from Administrative Data

Vasquez, SA¹, Palfy, M¹, and Garcia, AF¹

¹SA NT DataLink

Introduction

Record linkage has existed for decades, regarded as an activity where an entity is ascertained across different data sources. Initially medically driven to determine further outcomes. Developments in technology led to cheaper computing and storage, which ushered capture of information beyond the primary entities such as carer's details.

Objectives and Approach

Administrative datasets that capture an event involving a child would more often than not, capture information about their parents or carers including relationship to the primary entity within the record.

The main objective was to exploit the connections across the entities captured in a record and have that as an additional layer of information that may contribute to context. All entities extracted from one record were clearly defined as related. Information was considered true at the point of capture. Then record linkage for all the individual entities extracted were performed.

Results

Family structures evolve and a family structure during birth may change by the time the child attends school where carer information was recorded during enrolment.

Each data source inevitably becomes a "source of truth" and understanding the different aspects of a dataset affects its credibility. Authority for collection dictates if the recording of the event is mandatory or optional and directly influences the coverage. The process of data capture and storage is also important; manual vs digital input and if any validation was present.

Application of family connection to specific types of research will dictate the relevant "source of truth". If one is to use family connections in the context of hereditary diseases, pregnancy outcomes would be more suitable than a birth registry.

Conclusion/Implications

Family unit as defined in the traditional sense have changed and all variations are reflected in the data.

Assessing the impact of social connections to ones health may now be possible. Extracting these connections from administrative datasets can serve as additional intelligence that may inform if not shape longitudinal studies.

