

Introducing the PRIDaL model for linking identifiable self-reported patient data with routinely collected data

Mark-Rhys Kingston, Helen Snooks, Martin Heaven, Alan Watkins, Swansea University Medical School, Wales, UK
Contact: prismatic@swansea.ac.uk

Introduction

The rise in availability, quality and use of routine health data has resulted in well-developed methods for anonymised linkage of data from multiple sources. However, methods for combining identifiable data (e.g. patient self-reported questionnaires) with routine anonymised data are not yet tried-and-tested. Linking these data presents opportunities to improve the feasibility and effectiveness of observational and experimental studies, but emerging data linkage processes must balance data security and usability [1].

As part of PRISMATIC – a mixed methods stepped wedge cluster randomised trial of the effectiveness of an emergency admission risk tool – we needed to efficiently link routine hospital data and patient-reported quality of life and quality of care questionnaire data, while satisfying information governance requirements. *Key issues included, how to:*

- account for the roles of multiple data sources;
- ensure patients that dissent are excluded from the data linkage;
- explain and illustrate data processes.

Our objective was to develop an efficient privacy-protecting model for linking routine and identifiable patient-reported research data.

Methodology

The Process for Routine and Identifiable Data Linkage (PRIDaL) was conceived and developed by the PRISMATIC study team, alongside specialists in e-trials, health informatics, information governance and process mapping from the Centre for the Improvement of Population Health through E-records Research (CIPHER) at Swansea University. As a group we managed the steps for linking databases as summarised by Bradley et al [2]:

- 1) identify data sources and owners;
- 2) obtain approvals;
- 3) select variables used to link databases;
- 4) determine methods for linking databases;
- 5) evaluate quality of linked data.

We used process mapping to help plot, trace and visualise data flows, including the interconnections between processes, information systems and data sources or organisations [3]. We captured and refined the process using swim lane diagrams, which have been infrequently applied in health research.

Results

Table 1 summarises the sources and types of data we wished to link.

Source	Dataset	Data Items	Type
SAIL (via NWIS)	Emergency Department Data	Emergency department attendance	Routine
SAIL (via NWIS)	Patient Episode Database for Wales	Hospital admission history	Routine
SAIL (via GP practices)	Primary Care GP dataset	Event history (e.g. visits)	Routine
GP Practices	Prism risk model	Risk of emergency admission within 12 months	Routine
NWIS	Patient questionnaire sample	Study IDs	Sample
Patients	Self reported questionnaires	Quality of life (SF-12), Client Service Receipt Inventory, Quality of Care Monitor	Sample

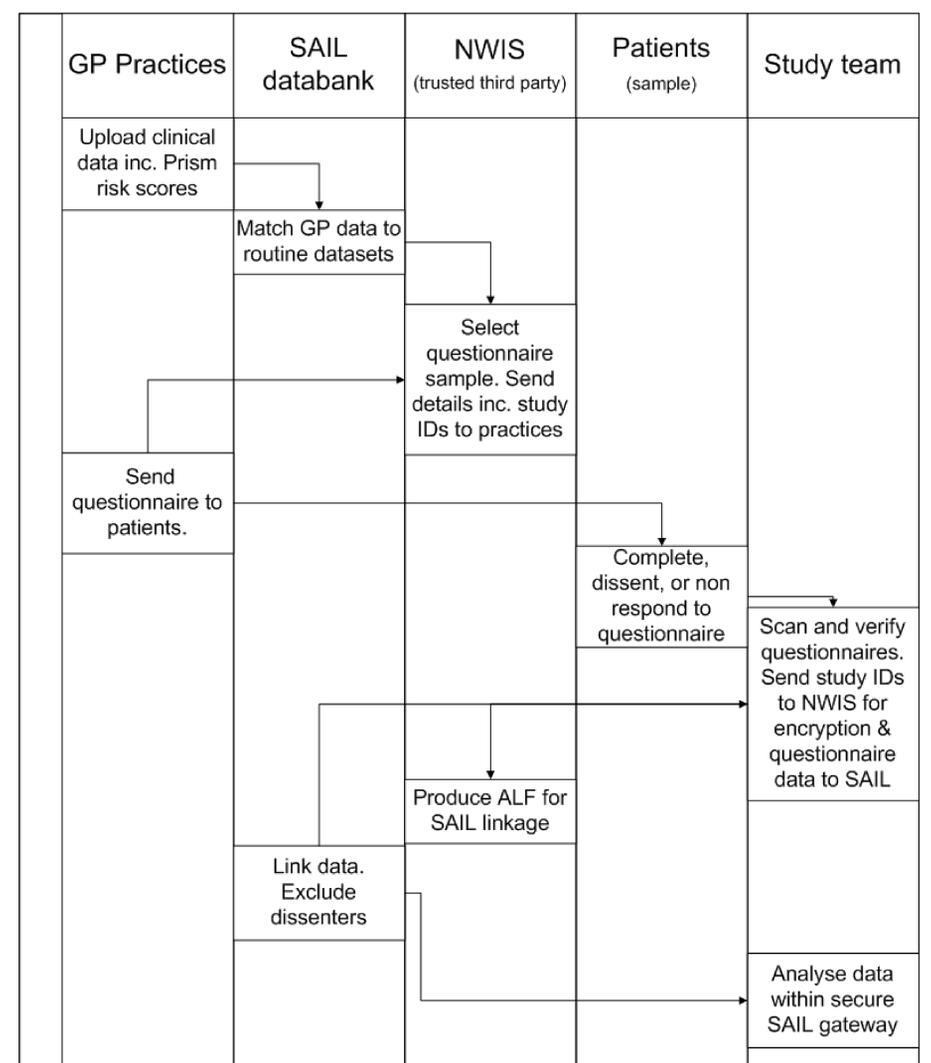
References: 1) Dusetzina SB, Tyree S, Meyer AM, et al (2014) *Linking Data for Health Services Research: A Framework and Instructional Guide*. <http://www.ncbi.nlm.nih.gov/books/NBK253313/>
2) Bradley et al, Health Services Research and Data Linkages: Issues, Methods, and Directions for the Future. *Health Serv. Res.* 2010;45(5p2):1468-88 3) http://www.mindtools.com/pages/article/newTMC_89.htm 4.) Lyons RA, Jones KH, John G, et al The SAIL databank: linking multiple health and social care datasets. *BMC Medical Informatics and Decision Making*, 2009, 9: 3 5) Hutchings HA, Evans BA, Fitzsimmons D, et al. Predictive risk stratification model: a progressive cluster-randomised trial in chronic conditions management (PRISMATIC) research protocol. *Trials* 2013;14(1):301. <http://www.trialsjournal.com/content/14/1/301>

Our process for data linkage was refined and captured in the PRIDaL swim lanes flowchart (figure 1) -part of our successful application for approval to use the SAIL databank. SAIL brings together person-based health care data for Wales, UK, using split file anonymisation to address issues of confidentiality and disclosure and linking with a Trusted Third Party (TTP) for person-level encrypted identifiers – called ALFs - to merge data from multiple sources [4].

The flowchart defines the path of our cohort and routine data, and the architecture connecting the activities and organisations.

A sample of over 269,000 unique patients have been included in our analysis, each with emergency admission risk scores linked to primary and secondary care data in SAIL (99.9% match rate). 2368 questionnaires were included from 1374 patients (100% match with SAIL. Response rate 40%). 90 patients dissented (<0.5% of sample) and were excluded.

Figure 1: PRIDaL swim lanes flowchart



Conclusions

We developed a data linkage model for the PRISMATIC study, that balanced usability and security by determining the minimum information required for acceptable linkage, and developing a model to reveal only that information. Convening an expert group and using swim lanes flowcharts helped to define the path of data and the associated architecture connecting the activities, people, and information systems. Our swim lanes diagram was important in gaining information governance approval. The linking of routine health and patient self-reported data presents valuable research opportunities, but clear, replicable models are needed to support ethical and practical data linkage.

We present PRIDaL, developed through expert input and process mapping, and tested through the PRISMATIC study [5], as a potential solution.