

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

Mark Rhys Kingston, Hayley Hutchings, Martin Heaven, Alan Watkins, Helen Snooks

College of Medicine, Swansea University, Wales, UK



Swansea University
Prifysgol Abertawe

BACKGROUND

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 [1]. 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 address the appropriate balance between data security and usability [2].

As part of PRISMATIC – a mixed methods cluster randomised trial of the effectiveness of an emergency admission risk tool in 32 general practices [3] – 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 owners, including patients, health services and data warehouses
- ensure patients that dissent from study participation are excluded from the data linkage
- explain and illustrate processes so that they can be understood by reviewers

Objective

To present an efficient privacy-protecting model for linking routine and identifiable patient-reported research data.

METHODS

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. We managed the steps for linking databases as summarised by Bradley et al: 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 [4].

We used process mapping to help plot, trace and visualise data flows, including the interconnections between processes, information systems and data sources or organisations [5]. We captured and refined the process using swim lane diagrams, which are common in organisational research but have been infrequently applied to research processes.

Disclaimer. This project was funded by the National Institute for Health Research HS&DR programme (project number 09/1801/1054). The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the HS&DR programme, NIHR, NHS or the Department of Health.

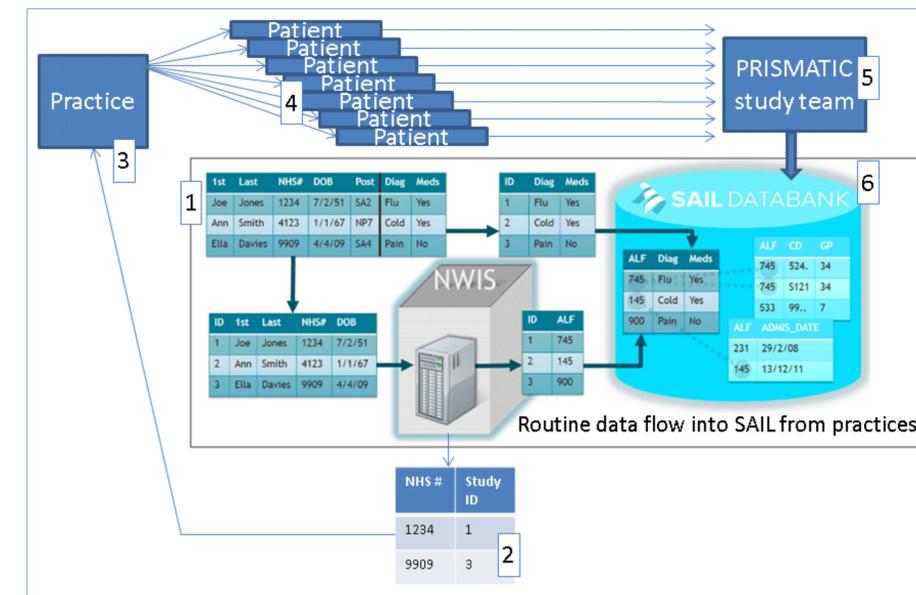
RESULTS

Table 1 (below) summarises the sources and types of data we wished to link in our study dataset

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

The PRIDaL model (figures 1 and 2) was part of our successful application to the independent Information Governance Review Panel for approval to use the SAIL databank at Swansea University. SAIL brings together person-based health care data for the population of Wales, UK [6]. It uses a split file approach to anonymisation (figure 1) to overcome issues of confidentiality and disclosure. SAIL links with the NHS Wales Informatics Service (NWIS) as a Trusted Third Party (TTP) for person-level encrypted identifiers (Anonymised Linking Fields or ALFs) to merge data from multiple sources.

Figure 1: PRIDaL model with SAIL split file methodology



Key

1. Normal anonymisation process into SAIL
2. NWIS generate sample for questionnaires and feed back to Practices
3. Practices invite patients to participate
4. Patients return questionnaires directly to Prismatic Study team, or decline to participate
5. Prismatic team provide dissenter ID to SAIL
6. SAIL link questionnaires and eliminate dissenter information from all flows

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

Mark Rhys Kingston, Hayley Hutchings, Martin Heaven, Alan Watkins, Helen Snooks

College of Medicine, Swansea University, Wales, UK



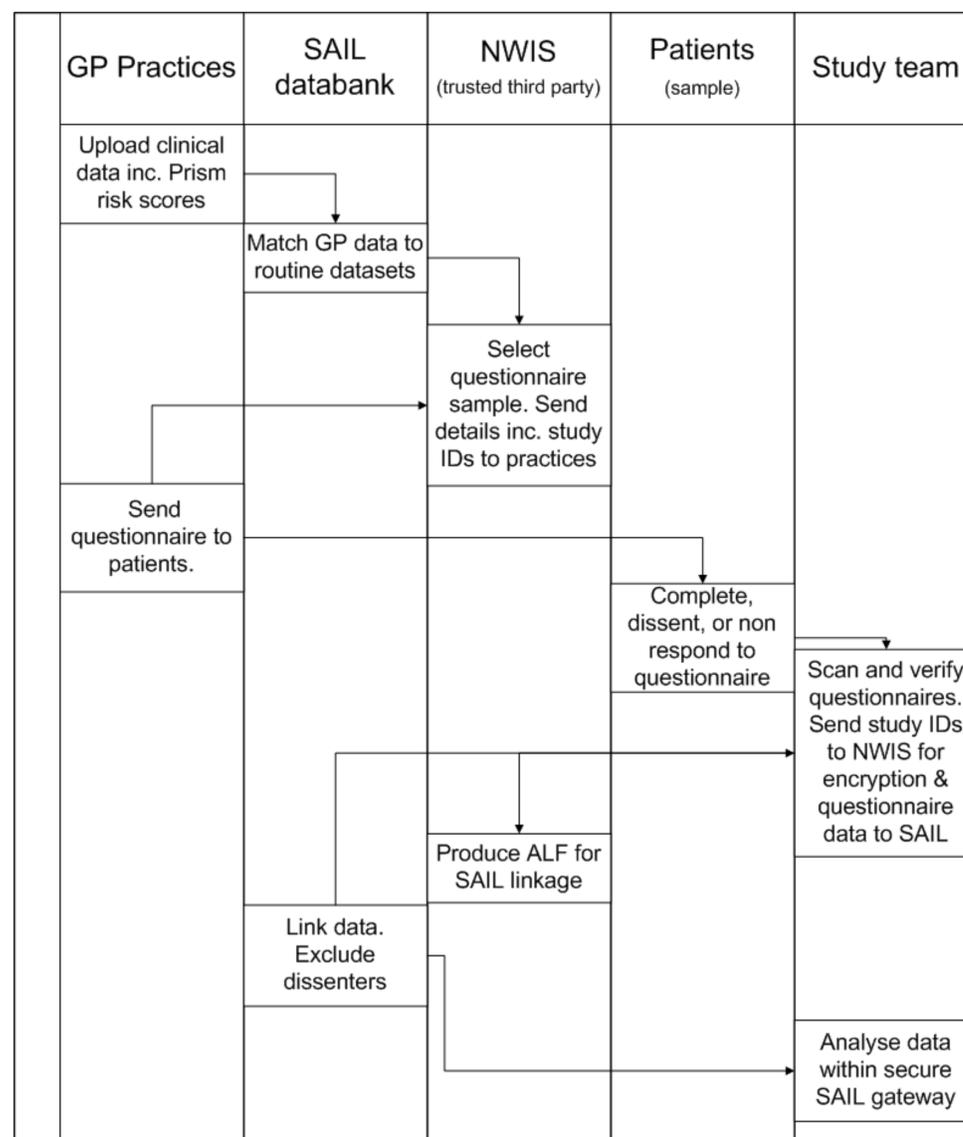
RESULTS continued

Our process for data linkage was refined and captured in the PRIDaL swim lanes flowchart (figure 2). The flowchart defines the path of our cohort and routine data, plus the architecture connecting the activities and organisations along the path [7]. The model developed through our discussions with the IGRP review board, and alongside the process map captured in figure 1, became an important visualisation of our data linkage process incorporating privacy protection.

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

Contact for further information.
m.r.kingston@swansea.ac.uk

Figure 2: PRIDaL model swim lanes flowchart



DISCUSSION & 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 [1].

We found that 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 along the path. Our swim lanes diagram was an important factor in gaining information governance approval.

The implementation of PRIDaL had a number of dependencies:

- the use of complex systems for data management to protect security of data (building on best practice in data linkage research through the SAIL databank)
- good relationships with data owners – including study sites (GP practices), and Trusted Third Party (NWIS).
- an administrative burden on study sites based on their access to patient address details for questionnaire distribution. To help, we offered administration costs reimbursement.
- all patients confirming that they did not want to participate in study had to be removed from anonymised follow up (less than 0.5% of invited sample).

Conclusion

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, as a potential solution.

REFERENCES

- 1: Hye-Chung K, Ashok K, Ashwin M, Michael KR, Stanley A. Privacy preserving interactive record linkage (PPiRL). *Journal of the American Medical Informatics Association* 2013.
2. 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/>
3. Hutchings HA, Evans BA, Fitzsimmons D, Harrison J, Heaven M, Huxley P, 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>
- 4 Bradley et al , Health Services Research and Data Linkages: Issues, Methods, and Directions for the Future. *Health Serv. Res.* 2010;45(5p2):1468-88.
- 5: http://www.mindtools.com/pages/article/newTMC_89.htm
- 6: 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.
- 7 Damelio R. *The Basics of Process Mapping*, 2nd Edition, 2011