Harvesting research outcomes from clinical databases - demonstrating the potential of TPP SystmOne.

Richard Gillott, Cardiovascular Database Developer, LTHT

Rick Jones, Yorkshire Centre of Health Informatics, University of Leeds
Background

• Clinical research in the UK is suffering from the high cost of obtaining and analyzing patient data, especially when large cohorts are needed to satisfy research questions.
• It’s hypothesized that leveraging the data already contained within General Practice records, through automated mechanisms, would vastly reduce costs and help to restore the UK as one of the leaders in this field.
• Currently, researchers experience difficulties obtaining the data they require for analysis, and linking datasets is extremely time consuming, which makes it difficult to gain a complete view of patient care.
• The majority of clinical trials in the UK rely upon a labour intensive paper based methodology, inconvenient for both trials and practices.
The UK can significantly enhance its clinical research capability by using, strictly within the bounds of patient confidentiality, the electronic patient data that the UK’s National Programmes for IT in the NHS have the potential to allow. This will have enormous benefits for all types of clinical, public health and health services research and for many aspects of patient care.
Context: Simulations

- **4.1 Surveillance (Pharmacovigilance)**
  - The vision for an ideal surveillance system is of a nationwide “active” system for tracking patients’ responses to medical interventions (POMS, immunisations and OTC drugs as well as Devices) and of disease and other incidents requiring reporting.

- **4.2 Clinical Trials**
  - Within the range of activities involved in running a successful clinical trial in the future, there will be a need to access and process data from electronic records at a number of stages before, during and after the trial.

- **4.3 Prospective tracking of a known cohort**
  - This simulation team concluded that in order for UK Biobank (which is a resource for prospective studies) to be able to provide benefits access to data will have to be at patient level data (identifiable) both coded and textual.

- **4.4 Observational epidemiology**
  - The construction of retrospective observational epidemiological studies based on routine data sources requires access to data from a very wide range of electronic records, both within and without the health services.
OSCHR

Strategic oversight of research-related activities

OSCHR e-Health Research Records Board
- Advisory group to OSCHR on e-Health Research Records
- External Reference Group for NHS CfH Research Capability Programme
- Forum for developing activities jointly funded with non-government stakeholders-UKCRC

NIHR Information Systems Programme

DH RDD

NHS National Programme for IT Programme Board

NHS Connecting for Health Research Capability Programme
- Manages infrastructure programme enabling research and analysis to improve quality and safety of care
- Ensures research input to SUS, SIP, NHS NP, ISB

Information Standards Board

National Information Governance Board

OSCHR

NHS CfH Comms & Stakeholder Engagement

Chief Clinical Officer and National Clinical Leads

Service Implementation

Secondary Uses Service

NHS Number Programme
Context: The NHS Research Capability Programme

- Six work streams:
  - Technical Architecture
  - Functional Scope
  - Data Quality, Standards and Linkage
  - Information Governance and Threat Assessment
  - Infrastructure
  - Communications and Stakeholder Engagement
Previous Findings
State of Research Data Infrastructure
Previous Findings
Confirmation of Knowledge Gap

"You did say you wanted HL7 format?"
Part 1: The current research
Aims & objectives

Based on the work of the RCP could a practical trial be carried out as proof of concept to:

1. Prove the feasibility of extracting identifiable patient data from GP systems for use in research

2. Construct an architecture to enable the rapid, repeatable, and secure query & collection of data.
Pilot trial

• A small trial was planned with the aid of 2 research groups based in Leeds

• The pilot aimed to demonstrate the value of the information contained in the patient record, and prove whether the data was sufficient in its coverage of the population and its completeness
Methods

Research Unit
- Contact 3rd parties for advice on ethics and consent
- Trial picked and consent verified
  - Relevant fields discussed and confirmed
  - Clinical Trial data manager generate NHS numbers on consented patients. Add a secondary inhouse identifier.
  - Clinical trials data manager re-adds NHS number and stores data securely
  - Clinical trials PI re-adds NHS number and stores data securely
  - A anonymised view of the data is made available to RG. NHS number, postcode etc removed
- Analysis subjected to expert review by trial PI's

TPP
- TPP release dummy extract of all available fields
  - TPP construct SQL query to extract agreed data set from records
  - TPP removes NHS number from extract and sends to research unit
- Relevant fields noted

RG
- Analysis: Proportion of patients found
  - Rates per case of:
    - Relevant admin, referral codes
    - Relevant diagnosis codes
    - Relevant visit codes
    - Relevant test, drug, procedures
- Results compiled and documented

End
Results: Yields of Records

- 81% Cardiovascular
- 66% Oncology
Results: GP Practice Coverage

Of the 231 patients with practice codes in the C.V. data - 197 are unique

Of the 4727 patients with practice codes in the oncology data - 548 are unique
Results: Data Quality

- 2.5% of patients noted as deceased in SystmOne are recorded as alive in PPM
- 26.5% of those recorded as alive in SystmOne are marked as deceased in PPM
Results: Data Quality

Figure 12: Frequency of drugs for cardiology patients

Figure 13: Frequency of diary entries for cardiology patients
Part 2: The consent problem
Trial deliberately avoided access to detailed identifiable data because:

- Strength of patient consent in trials not adequate
- Issues of depth and breadth of data available in full record going beyond trial interest
- Potential insecurity of free text entries in record
- Whole issue subject to investigation in current EPSRC workshops as follow-up to this study and Wellcome Grant on NHS/research database integration.

www.ychi.leeds.ac.uk/eprresearch
• Are consent forms from conventional trails sufficiently detailed when considering electronic access?

• Do patient information leaflets need revision when considering this type of access?

• How can the system cope with multiple trial requests when patients may be in several trials?

• Is there a need for a consent register?
• The GPRD takes the opinion that patients should contact their GP to opt out of their records being accessed, however the RCP believes that opt in model should be utilized.
• GPES, takes an alternative view, stating that it may extract data without GP or patient consent, if a representative board is consulted on their behalf.
For linked records...

We need:

• Granulated consent – all or part of record
• A model to allow selective searching by:
  – Clinical relevance
  – Administrative components
  – Therapeutics
  – Diagnostics
• An understanding of when and how frequently to conduct searches
• A business model to reimburse GP systems suppliers / data guardians for their search time
Part 3: Architecture
Background

- Building upon the work undertaken by the RCP
  - Proposed a number of high level architectures
    - Centralized vs federated data
    - Concentrates on Secondary Care
  - Honest broker

- The RCPs work and its database of documents was reviewed.

- The aim was not to provide a conclusive model for implementation, but to discuss some of the important considerations that need to be made.
Centralized vs distributed

Linkage layer
- Federated query manager
- Meta data

Data layer
- Index
- Data marts (DM1, DM2, DMn)
- Small central repository

Mediation layer
- Push channel
- Pull channel

Tools layer
- Analysis tools
- Cleansing tools
- Mining tools
- Commercial interface

Key:
- Data sent
- Data link

Analyst
Reseacher
General Practice Extraction Service (GPES)

- Provides a collection of methods to enable the extraction of data from NHS GP datasets
- SystmOne incorporates this functionality
- Does not provide functionality for updating information
- Based upon a pull model
- IC expected to take a central role of managing requests
About the model

• Uses the Service Orientated Architecture (SOA) paradigm, in a format that ties in with the work being conducted by the Research Capability Program to link data from Primary, Secondary and Tertiary sites across the NHS.

• Adopts the hybrid model of centralized & federated data storage and processing

• Recognises the mechanisms necessary for the querying of data are already available in some systems, such as SystmOne, thanks to the GPES initiative. However, this scheme will need to be extended, to support both the push and the pull methods required by a scalable solution.
SPINE services utilized

Enterprise services
- Personal demographics service (PDS)
- Personal spine information service (PSIS)
- Gazeteer Service
- Electronic prescription service (EPS)
- Electronic booking service
- Choose and book
- Secondary data uses service (SUS)
- Reference data service

Service management services
- Enhanced reporting service
- Spine operational support service

Integration infrastructure services
- Spine portal
- Transaction messaging service (TMS)
- Service and endpoint directory
- Local identity agent (IA)
- Spine security broker (SSB)
- Certificate management system (CMS)
- Spine user directory
- Role based access control
- NHS organisation structure
- Sealed envelope service
- Patient consent / permissions
- Legitimate relationship service

Key:
- Spine services to be utilized

Directory service

Access control service

Security infrastructure services

Electronic Health Record

- Patient accessible
- Supporting 24-hour care
- Routine patient care
- Aggregated anonymised subsets
  - Developing Health Improvement Programmes
  - Clinical Governance
  - Epidemiological Research

Treatment and care

Primary care EHR

NHS Trust EPR

Social care records

Knowledge for

Public
Patients
Healthcare professionals
Managers

Analysis

Public health
Clinical Governance
Health Improvement Programme
Performance management
Vision for Future Research Informatics

Interconnected Interoperable Clinical Research Community

Research Based Organizations

Clinical Trials Networks

Biomedical Informatics Networks (e.g. BIRN, caBIG, etc.)

Exchange of standardized, de-identified data

Research Enterprise

Inpatient and Ambulatory Clinical Care Data
External labs, Retail Pharmacy
Medical & Pharmacy Claims
Tissue Banks/Genomic Data
Clinical Trials Data

Exchange of standardized, de-identified data

From CSCA – FDA Briefing - 2007
Conclusions

• Research has demonstrated potential of linking data from trials to records in a regional primary care data system.
• Consent remains a major issue
• By adopting SOA architecture it may be possible to enable secure linkage between systems to support:
  – Consent management
  – Automated data retrieval

• To take part in the next stage you are welcome to attend the forthcoming EPSRC funded workshops
  – March 11th - 2 – 4 Trial planning and consent
  – March 23rd -10 – 12 Technological and architectural analysis
  – March 31st - 10 – 12 Feedback and synthesis

Email – r.g.jones@leeds.ac.uk to book a place