Real World Evidence (RWE) is clinical data which has been collected outside of a conventional clinical trial.

In Part I of the series we introduced and explained the importance of RWE. In this article we move on to the next step which is data collection:

There are three routes available to you.

Which one will you take?

Read on to make your choice…

This is the second part of a regular series available at svmpharma.com
The Big Data Route
Taking this path you will access vast and comprehensive datasets to collect data at a scale far beyond what would be feasible in a conventional clinical trial. Using powerful data tools, the filtering and selection of data within millions of records can be performed efficiently. Using search and filtering tools and via combining datasets you can zoom into specific data to answer your questions and discover new opportunities.

The RWE Trial Route
If you choose this path you will find the established and accomplished group of study designs known as observational studies. For decades these studies have shed light on the real-world impact of treatments and services. Alongside these studies you will find their newly established neighbour, pragmatic trials. Find out how pragmatic trials adapt the existing methodology of RCTs to add fresh insight to real-world effectiveness.

The Custom Data Collection Route
Heading in this direction, you will uncover a unique approach to collecting RWE. You will find a custom-built online treatment evaluation programme, designed and developed alongside an expert group. Healthcare professionals (HCPs) across selected centres enter data retrospectively from medical records and data can be prospectively entered as it becomes available. You will discover the positive outcomes derived from this method and the additional benefits which it brings.
1) The Big Data Route

Big Data is a collection of large and complex digital datasets which typically require non-standard computational facilities for storage, management and analysis. In recent years Big Data has become a buzzword across many industries; this is driven by the increased collection of data throughout our daily lives, the universal digitisation of information and the increased technological capacity to handle this data (1, 2).

The UK is renowned for a number of comprehensive healthcare-based datasets, here we will focus on four key RWE data sources: Hospital Episode Statistics (HES), Clinical Practice Research Datalink (CPRD), GP Prescribing Data and the Quality Outcomes Framework (QOF). It is important to note that the majority of these datasets cover England only (with the exception of CPRD); the devolved nations have their own equivalents (e.g. Scottish Morbidity Record (SMR) or Patient Episode Database Wales (PEDW)).

HES and CPRD have been used for research for over 25 years, and have supplied the data for hundreds of publications. However, it is much more recently that the technology has been available to realise the full potential of these datasets; you can now search, sort and interpret this data with increasing speed and efficiency. There is a continuing effort to increase data linkage; CPRD can be linked with HES and other national datasets including the Cancer Registry and Cardiovascular Outcomes (3, 4).

Recently, GP Prescribing Data has been used alongside incidence rates from QOF data to map spending on metformin and methylphenidate. Maps have been created to a high
spatial resolution and combined with demographic and geographic data. This data can be analysed to distinguish between chance fluctuations and genuine differences in prescribing rates, and users can accurately identify where action is required (5).

The table below shows the types of data available in the four key datasets, where to find it, and how you can access it.

### A GUIDE TO THE KEY SOURCES OF BIG DATA

<table>
<thead>
<tr>
<th>Hospital Episode Statistics (HES)</th>
<th>GP Prescribing Data</th>
<th>Clinical Practice Research Datalink (CPRD)</th>
<th>Quality Outcomes Framework (QOF)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Held by</strong></td>
<td>Health and Social Care Information Centre (HSCIC)</td>
<td>Health and Social Care Information Centre (HSCIC)</td>
<td>Medicines and Healthcare Products Regulatory Agency (MHRA)</td>
</tr>
<tr>
<td><strong>Contents</strong></td>
<td>Hospital inpatient, outpatient and A&amp;E activity with up to 300 fields per record. Includes Patient Reported Outcome Measures (PROMS). Data submitted to the Commissioning Data Set (CDS) for monitoring and payment purposes and this data is input and coded from patient notes and hospital records.</td>
<td>Practice level prescription information for dispensed items including BNF code, cost, quantity and geographic information of where it was prescribed.</td>
<td>Longitudinal primary care data and includes diagnoses and symptoms irrespective of hospitalisation in addition to drug prescriptions, vaccinations, blood test results and risk factors. The data is recorded using the Read Coded Clinical Terms system of nearly 100,000 codes (4).</td>
</tr>
<tr>
<td><strong>Areas Covered</strong></td>
<td>England (devolved nations have their own equivalents)</td>
<td>England (+Wales since 2011 and Northern Ireland since 2013)</td>
<td>UK</td>
</tr>
<tr>
<td><strong>Years Covered</strong></td>
<td>1987-present</td>
<td>April 2010-present</td>
<td>1987-present (previously known as GP Research Database GPRD)</td>
</tr>
<tr>
<td><strong>Size</strong></td>
<td>125 million records added yearly, where one record is a period of care/ finished consultant episode</td>
<td>Approximately 4 million rows of data every month, where a row is a specific drug preparation in combination with a specific GP Practice</td>
<td>5 million active patients (13 million overall)</td>
</tr>
<tr>
<td><strong>Access</strong></td>
<td>Direct from HSCIC via license application or via intermediary</td>
<td>Freely available</td>
<td>Application only. In-house researchers work on behalf of applicant, no direct access to data</td>
</tr>
</tbody>
</table>

### THERE IS GREAT POTENTIAL TEMPERED BY A NUMBER OF CONCERNS

You will find that Big Data in healthcare is a highly discussed topic where there is great potential tempered by a number of concerns regarding confidentiality. The use of these large datasets in healthcare faces more obstacles and scrutiny than ever before, which has led to ongoing changes in the structure and availability of the data (6-8).

In 2013 the care.data programme was developed which would add pseudonymised GP records through the General Practice Extraction Service (GPES) to HES data. This is currently undergoing a pathfinder phase after initial delays (9, 10).
2) The RWE Trial Route

Designing and conducting a trial is an important and robust way of collecting RWE with a significant amount of flexibility. RWE trials can be divided into two main groups, observational studies and pragmatic trials. Both types use existing or adapted study designs but shift the emphasis towards external validity in order to gain better understanding of real world practice.

Observational Studies

In observational studies the assignment of subjects into a treatment group and control group is outside the control of the invigilator and these studies *draw inferences* about the possible effects of treatment on a subject. Observational research has a long and successful history in generating RWE, with advocates of observational studies promoting its ability to assess real world practice and outcomes (11). Both retrospective (case-control) and prospective (cohort) observational studies have a role in obtaining RWE due their external validity (applicability to real world practice). However observational studies can be slow to setup and inefficient in delivering results (12, 13).

Pragmatic Trials

Pragmatic RCTs (pRCTs) have become increasingly common in recent years with a tenfold increase in publications over the last decade. Pragmatic RCTs are randomised and follow a similar methodology to a traditional RCT. However they differ on a number of counts and use broad eligibility criteria to ensure the inclusion of subjects who are representative of real world clinical settings.

The medical management within these trials are consistent with actual clinical care protocols and they measure outcomes that are important to patients and decision makers (including functional status, quality of life and costs). Pragmatic RCTs can be more expensive to run than traditional RCTs, due to the broader scope and increased range of outcomes measured (14-17).
3) The Custom Data Collection Route

The approach outlined here uses retrospective or prospective data from 50-200 patient health records entered by health professionals into a web-based tool or Treatment Evaluator™. Data entry is guided and validated; complex patient histories are streamlined and reduced to the relevant information.

**The Service Evaluation Approach**

Clinical data collection programmes can be classed as research, audit or service evaluation. Research programmes derive generalisable new knowledge and generate and test hypotheses. An audit is designed and conducted to produce information to inform delivery of best care. An audit asks ‘Does this service reach a pre-determined standard?’

A service evaluation programme considers a service’s effectiveness or efficiency through systematic assessment of its aims, objectives, activities, outputs, outcomes and costs.

This approach can be used to compare a new treatment or service with an existing one, but cannot be used for measurement against standards. Service evaluation generates evidence of the effectiveness of a service which may lead to service redesign and is fully endorsed by the NHS (18, 19).

Using a service evaluation approach usually does not require a sponsoring organisation, R&D approval or ethics approval and avoids the judgemental nature of an audit. Unlike clinical research, this approach does not allow generalisability of the result, but the aims of the project, the selection of the centres and the parameters of the data collection can be adapted to ensure that the conclusions are clear and persuasive (20).

**Scoping Phase**

This begins with the assembly of an expert group which consists of doctors, other healthcare professionals (HCPs), pharmaceutical company representatives, and commissioners. Group and one-to-one meetings are vital in determining the aims and scope of the project and its viability. Assumptions are developed and agreed at this stage.

The group can select a practical primary outcome, with test results and other disease indicators as secondary

---

**The methodology of a custom data collection programme**

- **Scoping Phase**
  - 1-to-1 Meetings
  - Identification & Recruitment of Expert Group
  - Collection Parameters

- **Go or No-Go Decision**

- **Development Phase**
  - Online Database & Development Sign-Off

- **Data Collection Phase**
  - Ongoing Centre Management
  - Real-Time Analysis

- **Review Phase**
  - Expert Group Reconvenes
  - In-Depth Analysis
  - Feedback
  - Future Actions Planned

- **Outcomes**
  - Journals
  - National & International Conferences
  - Re-submissions to NICE/SMC
  - Service Redesign

© 2015 SVMPharma Ltd. All rights reserved.
outcomes. This is of particular interest to commissioning groups and other payors.

Practical Outcomes:
- Total patient-NHS interactions
- Number of procedures
- Number of outpatient appointments
- Number of bed days
- Emergency department visits.

Development Phase
The online data entry tool must be standardised across centres and tested. The internal logic can be designed to check for incomplete fields and inconsistency, validating the data. Intelligent design at this stage greatly shortens the collection and analysis phase.

Data collection phase
Patient data can be entered quickly following online training sessions with each of the participating centres. Real-time analysis allows monitoring of the process and centres can visualise their outcomes.

This online data entry tool will allow HCPs across selected centres enter data retrospectively from patient records.

Review Phase
The expert group reconvenes and the findings across the centres are presented and discussed. The objectives of the programme and the data collected determines the next steps.

Outcomes
RWE generated using this methodology has a proven record of contributing to meaningful outcomes (21, 22):
- Abstract presentation at national and international conferences
- Publications
- Successful re-submission to advisory bodies e.g. Scottish Medicines Consortium (SMC)
- Presentations to specialist commissioning groups
- Case Studies
Destination: RWE

Each of the three routes leads to RWE, providing data which can drive clinical outcomes and enhance the value of a product. However you will see that each route has strengths and weaknesses, did you make the right choice?

Big Data is important in identifying trends and relationships and inspiring new directions for marketing and growth. The scope and volume of data can be persuasive and powerful, however accessibility and completeness varies.

Pragmatic RCTs and prospective observational studies require a considerable investment in time and money but can produce robust and powerful RWE.

Custom data collection, via the approach outlined, allows data to be collected, analysed and presented within a period of several months. This is invaluable when faced with an approaching deadline, for instance, submissions or re-submissions to NICE, SMC or specialist commissioners. This data can differentiate a product in a competitive marketplace and the data collection process encourages engagement with the centres and key prescribers.

### A COMPARISON OF RWE DATA SOURCES ALONGSIDE THE TRADITIONAL RCT

<table>
<thead>
<tr>
<th></th>
<th>Retrospective</th>
<th>Prospective</th>
<th>On-label</th>
<th>Off-label</th>
<th>HCP Engagement</th>
<th>Cost</th>
<th>Speed</th>
<th>Level of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Low</td>
<td>High</td>
<td>Slow</td>
<td>High</td>
</tr>
<tr>
<td>RWE Trial</td>
<td>✔️</td>
<td>✔️</td>
<td></td>
<td></td>
<td>Low</td>
<td>High</td>
<td>Slow</td>
<td>Medium-High</td>
</tr>
<tr>
<td>Big Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
</tr>
<tr>
<td>Custom Data Collection</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>High</td>
<td>Low</td>
<td>Fast</td>
<td>Medium</td>
</tr>
</tbody>
</table>

The table compares the key attributes of these RWE data sources in addition to the traditional RCT. There is no single way to collect RWE, and each of the routes offer benefits which may align with your objectives. Of the three routes, custom data collection offers an excellent balance: it is a low-cost and expedient programme, enabling both retrospective and prospective data collection and offering the bonus of HCP engagement. Adding this programme to your strategy can bridge the gaps in clinical trial data and offers a clear path to success.

**WHICH ROUTE WILL YOU TAKE FOR YOUR NEXT REAL WORLD EVIDENCE PROJECT?**

NEXT ON THE RWE SERIES JOIN US FOR PART III IN WHICH WE LOOK AT RWE DATA ANALYSIS. VISIT [SVMPHARMA.COM](http://SVMPHARMA.COM) & FOLLOW US [@SVMPHARMA](https://twitter.com/SVMPHARMA)

2. Bollier D, Firestone CM. The promise and peril of big data: Aspen Institute, Communications and Society Program Washington, DC, USA; 2010.


12. Black N. Why we need observational studies to evaluate the effectiveness of health care1996 1996-05-11 07:00:00. 1215-8 p.


