Drug Discovery Today • Volume 00, Number 00 • December 2015 PERSPECTIVE

feature

Designing and incorporating a real world data approach to international drug development and use: what the UK offers

Andrew Bate[1,2], andrew.bate@pfizer.com, Jane Juniper[3,4], Andy M. Lawton[5] and Rob M.A. Thwaites[6]

Assessments of the safety, efficacy and appropriate use of new medicines lie at the heart of treatment development and subsequent adoption in clinical practice. Highly controlled randomised clinical trials routinely inform decisions on the approval, coverage and use of a medicine. Researchers and decision makers have become increasingly aware that these experimental data alone are insufficient to address those decisions fully. Real world data recorded from routine healthcare delivery by healthcare professionals and patients help provide a more complete picture of care. The UK, with its connectivity and rich longitudinal patient records, accumulated research and informatics experience and National Health Service, provides an exemplar of how real world data address a wide range of challenges across drug development.

Introduction

Real world data (RWD) have been defined simply as 'data used for decision-making that are not collected in conventional randomised clinical trials (RCTs)' [1]. RWD are characterised by very large (relative to RCTs), complex, intricately structured datasets, with several years of data on millions of patients. The National Health Service (NHS) is a near monopoly provider to UK residents and visitors of primary healthcare, through a network of contracted general practitioners, and secondary healthcare, provided by wholly state-owned hospitals. General practitioners are the gatekeepers of the healthcare system because they are the routine first point of entry into the healthcare system, referring patients as needed for more-specialist treatment; and electronic medical records (EMRs) provide a detailed record of primary care interactions as well as insights into referrals to and from other healthcare delivery systems. As software increasingly were introduced in surgeries for medical record storage in electronic format the same software systems meant that with appropriate safeguards and caveats large numbers of inter-surgery EMRs could be stored in databases, and used in anonymised format by external parties for research. For example, one such dataset in the UK, the Clinical Practice Research Datalink (CPRD) [2], formerly known as the General Practice Research Datalink (GPRD) [3], has data now on over 15 million patients contributing over 82 million years of prospectively collected longitudinal data – more than 5 years of history per patient on average with more than 3.5 million patients who have more than 10 years of history (Williams, T., personal communication). Copies of real-world healthcare data from other sources, also originally collected for the clinical care of patients, are stored and linked in the Health and Social Care Information Centre (HSCIC) in England, and equivalent bodies in the rest of the UK. This data confederation serves as a focus for many epidemiological research projects in the UK. The near-universal existence of an NHS number, unique per patient, also facilitates data linkage across the healthcare systems. Further information on CPRD and other examples of UK RWD sources are included in Table 1. We consider a RWD approach as one that systematically (i) considers which data are required to answer decision maker questions, (ii) determines whether those data are available for secondary analysis for

http://dx.doi.org/10.1016/j.drudis.2015.12.002

Please cite this article in press as: Bate, A. et al. Designing and incorporating a real world data approach to international drug development and use: what the UK offers, Drug Discov Today (2015), http://dx.doi.org/10.1016/j.drudis.2015.12.002

TABLE 1

Examples of the many hundreds of health data sources in the UK, each distinct in terms of varying size and the type of data that are captured.

Database	Characteristic	Population size
Clinical Practice Research Datalink (CPRD; formerly GPRD)	General practitioner (GP) primary care database	12+ million patients [3]
The Health Improvement Network (THIN)	GP primary care database	10.5 million patients [44]
Hospital Episode Statistics (HES)	All patients treated in an English National Health Service (NHS) hospital. Clinical data and demographics including hospital in-patient diagnoses and procedures from the hospitals [45]	Over 125 million patient, outpatient and accident and emergency records each year [46]
Cancer screening: National Breast Screening Programme data	Data on all English women undergoing a mammography screening, aged between 50 and 70, all of whom are invited every third year, and other and self referrals [47]	Data on the 1–2 million women screened per year [48], 1.94 million women aged 45 and over screened in the year 2011–2012 [49]
Cancer screening: National Bowel Screening Programme data	Bowel Cancer Screening System has clinical and demographic data for English patients screened at over 59 for bowel cancer, their episode history, results of their Faecal Occult Blood Testing (FOBT) tests and diagnostic tests and histology information [50]	11.2+ million people of age greater than 59 [50]
The National Cancer Waiting Times (CWT) Monitoring Data Set	Patients diagnosed with suspected cancer in England in an NHS hospital setting [51]	865 494 '2 week wait referrals' and 224 984 cancers (from 8049 English practices) [52]
National Cancer Data Repository (NCDR)	Information about every patient with cancer diagnosed in England, obtained from routine NHS data sources. Created by linking cancer registry and HES data [22]	Data on 114 155 patients underwent major resection for a colorectal tumour over a 5 year period, and data on other cancer patients [53]

healthcare applications and, if so, (iii) looks to find an effective way to collect, access and analyse those data. We consider the conduct of pure primary data collection studies out of scope for this manuscript and instead focus on analyses of routinely collected healthcare data. We present case studies from the UK that show how a RWD approach can and is being adopted and how it might transform decision maker evaluations.

Although we focus here on the UK, internationally there are many RWD sources, each dataset with its own strengths and weaknesses associated with the underlying healthcare system. For example, the fragmented nature of the US healthcare system means it can be difficult to follow patients across care providers. In addition, data are captured primarily for billing purposes rather than healthcare provision which means important research data such as laboratory test results are often not captured. Databases in Scandinavia and Scotland are rich and link reliably across the healthcare system (the ubiquitous existence and use of a personal ID number in, for example, Sweden facilitates this) but can be too small for some research questions, because the population of these countries is relatively small compared with England and the USA.

In recent years an increasing number of studies have been performed synthesising RWD from different healthcare organisations or countries, often motivated by the need to improve statistical power [4]. To facilitate the RWD approach in this context, there are several within, and across, country initiatives: examples include the US Sentinel Initiative and Observational Medical Outcomes Partnership (OMOP); the Canadian CNODES; the pan-Asian network Asian Pharmacoepidemiology Network (ASPEN); and European initiatives such as Innovative Medicines Initiative's Pharmacoepidemiological Research on Outcomes of Therapeutics (IMI PROTECT) and European Union-Adverse Drug Reaction (EU-ADR) [5–10]. It is beyond the scope of this article to provide a comprehensive list of databases, database networks or inter-database comparative analyses; for more information about database selection see for example [11,12]. We merely argue that the UK, with the NHS and its existing data and capabilities, is an ideal location for a RWD approach and UK RWD can and should play an even more extensive part in global drug development [13,14].

Case study: psoriasis and the risk of chronic kidney disease – the association between the diseases and the implications for therapy development and healthcare policy

A group of US-based researchers (Wan *et al.*) [15] describe a study examining the risk of chronic kidney disease in patients with psoriasis. A key challenge for this study was to characterise the level of severity of the psoriasis accurately. Body surface area direct measurement had previously been shown to be a good estimate of psoriasis severity, but is not a measure routinely captured in records. EMRs sometimes cannot adequately capture severity directly nor its fluctuations accurately, often relying on proxy measurement, see for example [16,17]. The authors prospectively collected data from general practitioners (GPs) on a subset of patients in a UK EMR database called The Health Improvement Network (THIN). This cohort, called the Incident Health Outcomes and Psoriasis Events (iHOPE24) study, had disease severity for 10 500 patients aged 25–64, based on extent of body surface area affected by psoriasis.

The iHOPE24 cohort was then used to show that treatment patterns correlated with body surface area psoriasis coverage, and that treatment patterns could therefore be used for estimating psoriasis severity. Treatment patterns were then used as a proxy of disease severity for patients across the full THIN dataset, where data on body surface area coverage were normally absent, so that the full power and generalisability of THIN data could still be used for inference. The authors reported an increased risk of incident chronic kidney disease in psoriasis sufferers but in particular in the severe psoriasis

Please cite this article in press as: Bate, A. et al. Designing and incorporating a real world data approach to international drug development and use: what the UK offers, Drug Discov Today (2015), http://dx.doi.org/10.1016/j.drudis.2015.12.002