

Institution: University of Sussex		
Unit of Assessment: 11 – Computer Science and Informatics		
Title of case study: Improving efficiency and effectiveness in patient identification for clinical trials		
Period when the underpinning research was undertaken: September 2010 – May 2015		
Details of staff conducting the underpinning research from the submitting unit:		
Name(s):	Role(s) (e.g. job title):	Period(s) employed by submitting HEI:
Dr Anne Rosemary Tate Dr Natalia (Natasha) Beloff	Senior Research Fellow Senior Lecturer	June 2007 – May 2015 October 2000 – present
Period when the claimed impact occurred: August 2013 – December 2020		
Is this case study continued from a case study submitted in 2014? N		
1. Summary of the impact <p>Significant increases in the efficiency and effectiveness of patient selection for randomised clinical trials have been achieved through the development and successful deployment of TrialViz, a Big Data analytic platform, on the largest Primary Care Electronic Healthcare Records (EHR) Database in the world, CPRD Gold. Underpinned by a breakthrough rapid search engine and innovative visualisation tools developed by the University of Sussex in collaboration with industrial partners Dataline and CPRD, TrialViz allows for an unprecedented near-real-time identification and visualisation of patient cohorts suitable for any given clinical trial. This reduces cohort identification time from 2-3 weeks to 1-6 minutes, resulting in significantly reduced costs, greater efficiency in clinical trial recruitment and improved patient outcomes.</p> <p>TrialViz is now a core component of the IQVIA E360™ Cohort Builder. As of January 2021, E360™ had over 600 million unique recorded patient journeys with over 150 billion medical events across nearly 30 distinct datasets, with numbers growing rapidly.</p>		
2. Underpinning research <p>Randomised Clinical Trials (RCTs) are one of the key elements of the drug discovery and adoption process. Prior to this research project, the routine procedure for RCT-suitable cohort identification was time-consuming and often a semi-manual process, carried out by highly-trained clinical professionals in order to satisfy multiple and complex inclusion / exclusion criteria matching a particular study protocol. According to research collaborator Tim Williams, Head of Interventional Research at CPRD, 50% of clinical trials were failing to recruit to time, resulting in significant cost implications and delays [S6].</p> <p>In 2010, Beloff and colleagues in the Department of Informatics at Sussex (INF Sussex) proposed to combine their academic expertise in Big Data analytics and Data Quality assessment with the skills of industrial software developers from Dataline Software Ltd (Dataline), and Electronic Healthcare Records experts / commercial data providers from Clinical Practice Research Datalink UK (CPRD), to investigate and address the need for a robust and efficient patient selection procedure for clinical trials.</p> <p>This resulted in an Innovate UK-funded project [G1], with the aim to develop a system (now known as TrialViz) for automatically generating RCT-suitable cohorts, thus maximising the UK's participation in international drug discovery efforts and speeding up the adoption process of newly available treatments. Research by INF Sussex in collaboration with project partners focused on three main themes:</p> <p>(1) Data abstraction / extraction: the design of the complex protocols needed to create cohorts of patients that match all of the clinical trials' multiple criteria. The key innovation was to leverage the high input/output speed of solid-state drives to create an optimised SQL-based query-processing tool to operate on each of the 3 databases making up TrialViz. The optimization was based on a divide-and-execute approach, which, in conjunction with the database model and underlying architecture, reduced the search time by several orders of magnitude [R1]. Additional speed gains were obtained through judicious use of caches and the development of data quality</p>		

procedures [R2, R3] that allowed for the timely elimination of possible errors in patient selection. Implemented using the emerging HADOOP framework, the resulting system allowed a fully automated search through over 6 million patient records in a matter of minutes, compared to the 2-3 weeks required by the previously used semi-manual procedures [S1, S6]. The work published in [R1] was the basis of the patent by Dataline (Patent Number: US-9672268-B2), as stated on p. 295 (penultimate paragraph) of [R1].

(2) Data visualisation: A novel visualisation approach was pioneered based on an intuitive “card selection” interface, with special “card stacks” to help define inclusion/exclusion criteria (fig. 2 from [R1] below). This allowed for clear geographical representations of available patient cohorts, together with their density and suitability levels. In the words of Tim Williams, CPRD, the practical importance of such visualisation was “ground-breaking” [S6]. This same visualisation interface is still in use in the current IQVIA E360™ Cohort Builder [S2].

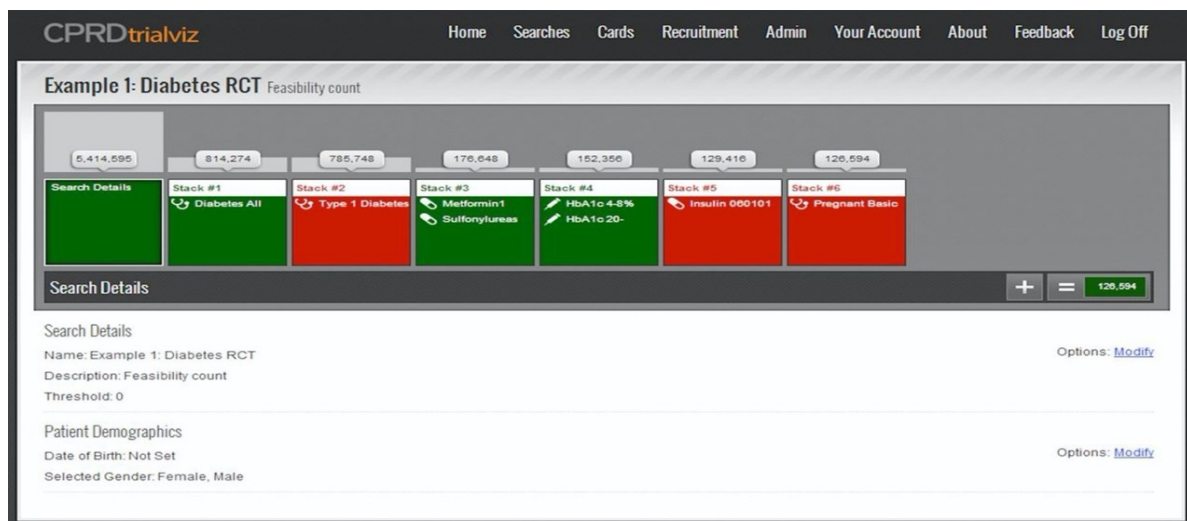


Figure 2 from [R1]: Stack and card interface showing the results of a search.

(3) User testing. Extensive user-testing sessions for the TrialViz prototype were designed and run with end-user focus groups comprising key stakeholders: (i) pharmaceutical industry representatives from AstraZeneca, GlaxoSmithKline, and Novo Nordisk, (ii) Primary Care Research Network representatives (iii) General Practitioners as primary generators of Electronic Health Records and as the first contacts for patient recruitment for clinical trials, and (iv) patient interest groups [R1-R3].

Finally, INF Sussex was instrumental in garnering peer-review recognition in this highly competitive field [R1-R5], stipulating the technical innovation of TrialViz in algorithmic search and in visualisation mechanisms, as well as demonstrating its applicability in a number of settings for RCTs and epidemiological studies.

3. References to the research

[R1] A. Rosemary Tate, Natalia Beloff, Balques Al-Radwan, Joss Wickson, Shivani Puri, Timothy Williams, Tijeed Van Staa, and Adrian Bleach (2014). *Exploiting the potential of large databases of electronic health records for research using rapid search algorithms and an intuitive query interface*. Journal of the American Medical Informatics Association 21(2):292–298. <https://doi.org/10.1136/amiainl-2013-001847>

[R2] A. Rosemary Tate, Natalia Beloff, Timothy Williams, Shivani Puri, and Tjeerd van Staa (2012). *Developing quality scores for electronic health records for clinical research*. Pharmacoepidemiology and Drug Safety, 21(S3):654, p. 305. <https://doi.org/10.1002/pds.3324>

[R3] A. Rosemary Tate, Tim Williams, Shivani Puri, Natalia Beloff, and Tjeerd Van Staa (2011). *Developing quality scores for electronic health records for clinical research: a study using the General Practice Research Database*. In ACM Proceedings of the first International Workshop

on Managing Interoperability and Complexity in Health Systems (MIXHS'11), pp. 35–42.
<https://doi.org/10.1145/2064747.2064756>

[R4] Sheena Dungey, Natalia Beloff, Rachael Williams, Tim Williams, Shivani Puri, and A. Rosemary Tate (2015). *Characterisation of data quality in electronic healthcare records*. In Briassouli, Benois-Pineau, Hauptmann, Alexander (Eds.) *Health Monitoring and Personalized Feedback Using Multimedia Data*, pp. 115-135, Springer. https://doi.org/10.1007/978-3-319-17963-6_7 (PDF available on request).

[R5] A. Rosemary Tate, Sheena Dungey, Simon Glew, Natalia Beloff, Rachael Williams, and Tim Williams (2017). *Quality of recording of diabetes in the UK: how does the GP's method of coding clinical data affect incidence estimates? Cross-sectional study using the CPRD database*. *British Medical Journal Open* 7(1):e012905.
<https://bmjopen.bmj.com/content/7/1/e012905>

Evidence of research quality. Funding includes:

[G1] PI: J. Cuthbert (Dataline); Co-I: Dr N. Beloff, "Visualisation of the Quality of Electronic Health Records for Clinical Research", Innovate UK (TSB), 02/04/2011- 31/10/2012, £340,000 (INF Sussex share £86,383).

[G2] PI: Dr N. Beloff, "Developing measures to assess the quality of GPRD data for research purposes", CPRD, 01/02/12 - 30/09/12, £42,638.

[G3] PI: Dr N. Beloff, "Data Quality in the Clinical Practice Research Datalink", CPRD, 11/03/2013 – 10/04/2014, £120,408 to Sussex.

[G4] PI: Dr N. Beloff, "Methods to characterise and monitor data quality in the Clinical Practice Research Datalink", CPRD, 11/03/2014 – 31/05/2015, £71,666 to Sussex.

4. Details of the impact

1. Improved efficiency of patient identification in Randomised Clinical Trials worldwide

Following the success of the TrialViz prototype, significant subsequent funding from CPRD was obtained by INF Sussex to continue its development **[G2-G4]**. Tim Williams (CPRD) emphasises that the immediate success and potential of the prototype delivered by the INF Sussex-Dataline-CPRD collaboration was self-evident, and explicitly recognises the importance of the contributions by INF Sussex within this collaborative effort:

"Whilst the implementation of the technology was being undertaken by Dataline, a lot of the innovative thinking and the ideas behind this were being generated through Natasha [Beloff, INF Sussex] who has a good understanding and capability in data management and is familiar with CPRD data ... along with Adrian Bleach who was also keen on innovation, I would say that together they brought about the technological innovation which led to the success of the project." **[S6]**

TrialViz became a core component of the patient cohort selection process for RCTs in the IQVIA E360™ Cohort Builder **[S1-S2]**. This happened in two stages; firstly in 2013 through the acquisition of Dataline by IMS Health Holdings Inc. (a multinational US company specialising in providing and analysing medical data from many different countries including France, Germany, UK and the USA), and secondly in 2016 with the merger of IMS Health with Quintiles, forming IQVIA (a leading global company, providing information, services and technology for the healthcare industry worldwide) **[S1]**. Importantly, the same basic algorithm and architecture from the original prototype are still being used by the E360™ Cohort Builder to deploy and research multi-national datasets, accommodating a variety of coding systems whilst still fulfilling the promise of near real-time performance **[S6, S1]**. Through E360™, the impact of the INF Sussex-industry collaboration has achieved global reach. It supports multiple geographies (including Europe, America, Canada and Australia) and allows research and pharmaceutical organisations access to the world's largest single source of pseudo-anonymised longitudinal medical records, claims and registry data in a single accessible application. According to Joss Wickson (Senior Principal, Technology) at IQVIA, at time of writing (January 2021), "E360™ has over 600 million unique recorded patient journeys with over 150 billion medical events across nearly 50 distinct

datasets, and this number is growing rapidly” [S1-S5]. Commenting on the significance of the E360™ suite for IQVIA, he states:

“The positive impact to IQVIA’s internal research teams is overwhelming, it has cut patient population identification for Recruitment, HEOR, Commercial, R&D and Data science teams from weeks to hours, saving hundreds of millions of pounds. External use of the tools has grown rapidly amongst the major pharmaceuticals and the tools have found a market in government organisation use with the FDA, CDC, MHRA and EMA all active members of the user base. 26 research and pharmaceutical organisations globally use the platform, and this number is steadily growing” [S1].

2. Building up capability for pragmatic epidemiological studies in CPRD

The project also had a significant impact on CPRD by prompting and enabling new uses for the UK medical data it holds as part of the UK Department for Health. Previously, CPRD data were used solely for observational studies and not for feasibility in clinical trials. The development of Trialviz prompted new directions for CPRD:

“TrialViz as a tool was essentially a ‘first in class’ development. Subsequent development and work around it has enabled CPRD to run data enabled clinical trials within a Primary Care setting ...the work undertaken was original and it set a precedent in terms of use of this data in clinical trials...the groundswell of understanding that this is the way forward for clinical trials is very much in the wake of Trivalviz’s success” [S6].

The TrialViz research project enabled CPRD to build its capacity for conducting pragmatic epidemiological studies i.e. using existing Primary Care patient records rather than RCT setups. Because CPRD retained a licence to use Trialviz within the NHS for UK-based data sets, they have been able to run longitudinal epidemiological studies in realistic settings [S6].

Another direct benefit from the Trialviz research project has been the development of the Clinical Practice Research Datalink (CPRD) Assisted Patient Recruitment tool (as part of the UK Government’s Innovative Licensing and Access Pathway (ILAP) for medicines, launched December 2020) [S7]. The Medicines and Healthcare products Regulatory Agency promotes the tool as a “fast, targeted and efficient approach to patient recruitment across the UK”, attesting that the “quality of CPRD’s data is globally renown[ed] and is used extensively by industry, regulators and academia worldwide” [S7]. Williams confirms that “this capability is predicated upon use of the CPRD primary care EHRs for clinical trial feasibility which directly relates to the project we did with Natasha several years ago” [S6].

In his assessment of the significance of the TrialViz research project, Tim Williams, CPRD, stated: “the tool itself is a number of years old and enhanced and revised versions are now in place as part of the clinical trial capability but there is a direct link back to the TrialViz research project. From an academic perspective I see the work we did as an original and innovative piece of research with real world application” [S6]. Indeed, the Trialviz project has had both national and international impact through the deployment of E360™ and its legacy continues through innovation and development initiatives within CPRD that are directly inspired by it.

5. Sources to corroborate the impact

[S1] Testimonials (original from June 2018, updated with recent numbers from January 2021) from Jonathan (Joss) Wickson, Principal, Technology, Real World Analytic Services, IQVIA, direct collaborator with INF Sussex on TrialViz project.

[S2] E360™ Cohort Builder promo video presented by IMS Health, published on 24 April 2016, before IQVIA bought out IMS Health. <https://youtu.be/kTG5mPuW5q4>. The original TrialViz interface is clearly visible (1:52 - 3:02) (fig. 2 of [R1], also reproduced in Section 2).

[S3] Reference article from Braid Equity Research demonstrating how important E360™ is becoming to the fabric of IQVIA as a 21 Billion USD market capital company.

[S4] IQVIA video “Solutions for the UK NHS”, published 22 July 2019
<https://youtu.be/98ieMJpyxhs> Head of the Healthcare IQVIA UK and Ireland, Peter Lane, describes the impact IQVIA UK team has delivered (0:40 - 1:22).

[S5] IQVIA video by Clinical Project Manager Dr Priyanka Narayan
<https://youtu.be/ggXvuHivAXc> where she discusses what impact IQVIA clinical trial cohort builder has on the success of the trials (0:10 - 1:55). Published 1 March 2019

[S6] Testimony from Dr Tim Williams, Head of Interventional Research, CPRD, part of Medicines and Healthcare products Regulatory Agency (MHRA) <https://www.cprd.com/>; direct collaborator with INF Sussex on TrialViz project.

[S7] CPRD Tool for Assisted Patient Recruitment, listed on GOV.UK site for Innovative Licensing and Access Pathway (ILAP) for medicines. Published 22 December 2020.
<https://www.gov.uk/government/publications/innovative-licensing-and-access-pathway-ilap-for-medicines/application-tools#tool-clinical-practice-research-datalink-cprd-assisted-patient-recruitment>