BMJ Oncology

Making administrative healthcare systems clinical data the future of clinical trials: lessons from BladderPath

Harriet Paige Mintz ,[1,2] Amandeep Raj Singh Dosanjh ,[3] Helen Parsons ,[4] Matthew Sydes ,[2,5] Richard T Bryan ,[6] Nicholas D James ,[7,8] Prashant Patel [9,10]

To cite: Mintz HP, Dosanjh ARS, Parsons H, *et al*. Making administrative healthcare systems clinical data the future of clinical trials: lessons from BladderPath. *BMJ Oncology* 2023;**2**:e000038. doi:10.1136/bmjonc-2023-000038

For numbered affiliations see end of article.

Correspondence to
Prashant Patel;
p.patel@bham.ac.uk

Clinical trials in recent years have become more challenging and costly to run. COVID-19 has revealed the power of huge driving forces behind research studies and how new and alternative ways of working are major agents behind this success. For over a decade, we have been investigating the use of routinely collected hospital administrative data in randomised controlled trials (RCTs). In this editorial, we share our practical experiences with such data for our bladder cancer trial (BladderPath). We aim to transparently report our journey for the benefit of future trialists working with such data, to aid future studies in successfully acquiring and using this data goldmine.

Conventional processes of collecting RCT data for outcomes are expensive, laborious and sometimes inefficient, using invaluable National Health Service (NHS) resources. This can limit follow-up data collection and important outcomes can be unreportable or unreliable. Hence, using routinely collected data to identify outcomes may permit more efficient RCT designs, be less burdensome for patients and healthcare providers, be more cost-effective, enable the collection of longer-term outcomes and divert scarce resources to priority needs. This is long awaited, and many groups[1 2] are investigating the use of these vast data which are described in detail elsewhere.[3–5] One example, the Hospital Episode Statistics (HES), code every inpatient, outpatient and emergency visit to NHS England hospitals[6] with fields including diagnoses and procedure codes.

We previously described our BladderPath team's development and validation of novel methods to solely use routinely collected data to replace conventional follow-up to populate case report forms (CRFs).[3 4] BladderPath (Image Directed Redesign of Bladder Cancer Treatment Pathway, ISRCTN35296862) compares two diagnostic pathways for bladder cancer.[7] Our rationale included: data providers periodically sending the trial team data for processing, these extracts would automatically prepopulate CRFs to be sent for site verification and upload into the central trial database.

Literature highlights administrative data benefits and pitfalls, including accuracy.[8–10] However, it is also recognised that traditional data collection already yields imperfect outcome data.[11] For our validation, we compared NHS routinely collected data to reference clinical datasets and identified substantial sensitivity improvements (example events: surgery, radiotherapy and chemotherapy).[3] In 2011, 41/117 (35%) events were detected, compared with 104/109 (95%) in 2017, with 95% (657/692) sensitivity over the last five data years.[3] Despite being a single site validation, remuneration is driving central and local initiatives which we hypothesise is driving further improvements in coding accuracy nationally.[3] We proposed manually querying all administrative data derived events against local clinical notes for further validation in the BladderPath trial.[3] With this approach, we intended to set up a framework with data providers to continually return quality measures to enhance these data, removing future need for data queries. We proposed using multiple datasets to reduce missingness and algorithm rules to capture miscoded events.[3] We considered our design would address pre-empted data missingness, outcome availability, governance, data retention, privacy and security concerns.[8] Using BladderPath as a case study we set out to implement our schema and below we share our experiences.

During BladderPath setup, two data providers were contacted: NHS Digital (NHSD) (April 2017) and the National Cancer Registration and Analysis Service (NCRAS) (July 2017). NCRAS was previously run by Public Health England (PHE) which became part of NHSD in October 2021 and is now known as NHS England due to merging in February 2023. We sought monthly participant linked HES, mortality and diagnostic imaging data from NHSD and radiotherapy, chemotherapy and cancer registration data[6] from PHE.

Initially, we requested monthly HES and mortality data for assessment of rapid outcomes (at that time unaware that NHSD provided data with a 2-month lag). A data linkage fee for every data drop was quoted (£2060) in addition to set up fees, costing £4910 per month; £58 920 per year and £589 200 for 10-year study follow-up. This substantially exceeded trial budgets and did not include the necessary central staff time for data processing.

We subsequently approached PHE regarding HES access, in addition to radiotherapy, chemotherapy and cancer registration data. NCRAS through the Office for Data Release were able to provide these datasets affordably without charging multiple linkage fees. However, discussions revealed that neither monthly nor quarterly HES releases were feasible for technical and operational reasons, instead offering us 6 monthly provision (minimum). This was unfortunately insufficient due to our requirement for rapid access to outcomes.

Our experience of working with healthcare systems data in England was currently too cumbersome and expensive for practical and economic implementation into RCT methodology[12]: such data have to be made more affordable with timely access. In the meantime, trialists should be aware of the need to allocate substantial budgets for data on grant application and to be satisfied that real-time access to outcomes data is not yet possible. However, the future may hold more promise. An alternative new and more accessible dataset, the Rapid Cancer Registration Dataset,[6] is now available. This can signal specific treatment events from January 2018 but with a smaller set of variables. Data providers have also been undergoing reform, with the NHSD DigiTrials service set up to support trialists.[6] In addition, the Secretary of State for Health and Social Care were tasked to find ways to 'deliver better, broader, safer use of NHS data for analysis and research' suggesting the use of Trusted Research Environments or Secure Data Environments.[13] We remain hopeful that these initiatives improve access and timeliness.

Our negative experiences of acquisition of these data have been offset by a positive experience for one-off retrospective data collection for our prostate cancer trial (STAMPEDE),[14] despite not yet translating into a simple approach for long-term data.[4] NCRAS delivered a service with allocated, supportive and helpful team members. We recommend that future trialists initiate such dialogue before the grant application stage, to ensure that the lengthy data application processes can be completed prior to recruitment. A new combined application system, such as that recently adopted by the Integrated Research Application System (IRAS), or even better, enabling application through the IRAS system when setting up a clinical trial, may enhance and simplify the application process. This would enable a single application for all datasets.

The success of COVID-19 trials like RECOVERY,[15] who used such data, emphasises the huge public value of well-conducted research and the power of NHS trials. Globally, these lessons are broadly applicable by using equivalent datasets. Nordic datasets are extensive, whereby a unique personal identity number is assigned at birth/immigration which tracks healthcare and other interactions. Due to these extensive datasets the methods that we outline above may be strengthened. A unique assigned number would potentially be of huge research value in England.

Our experiences show that administrative data can be repurposed to collect trial outcomes, however, the caveats described above need to be considered. BladderPath has now closed recruitment and on agreement with NHSD, we plan to use these data for a follow-up data sweep, bypassing the costs and timeliness concerns seen with real-time acquisition. With the newly available services and reforms, we are optimistic that this huge resource can be more widely used to benefit future research. We continue working to drive trial conduct to the forefront of technology.

Author affiliations
[1]University of Birmingham Medical School, University of Birmingham, Birmingham, UK
[2]MRC Clinical Trials Unit, University College London, London, UK
[3]Institute of Cancer & Genomic Sciences, University of Birmingham, Birmingham, UK
[4]Clinical Trials Unit, University of Warwick Warwick Medical School, Coventry, UK
[5]Health Data Research, London, UK
[6]Bladder Cancer Research Centre, Institute of Cancer & Genomic Sciences, University of Birmingham, Birmingham, UK
[7]The Institute of Cancer Research, London, UK
[8]Royal Marsden NHS Foundation Trust, London, UK
[9]Cancer Research Clinical Trials Unit, Institute of Cancer & Genomic Sciences, University of Birmingham, Birmingham, UK
[10]Department of Urology, University College London Hospitals NHS Foundation Trust, London, UK

Twitter Harriet Paige Mintz @HattieMintz, Matthew Sydes @mattsydes and Nicholas D James @Prof_Nick_James

Acknowledgements Ana Hughes and Ann Pope: Involved with management of the trial and data curation, critical revision of the work and final approval of the version to be published. Alicia Jakeman: Input into the manuscript and critical revision of the work and final approval of the version to be published. Wenyu Liu and Sarah Pirrie: BladderPath trial statisticians, provided input into the manuscript and critical revision of the work and final approval of the version to be published. Rachael Brannan: Advice and guidance on data stewardship of confidential patient information contributing to the research programme, provided input into the manuscript and critical revision of the work and final approval of the version to be published. Luke Hounsome: Agreement for NDRS data access and provision of the data, critically revised the work and given final approval of the version to the published. Haiyan Wu: Providing advice and guidance on data stewardship of confidential patient information, facilitating agreement for data access from PHE and the provision of the data to the study, provided input into the manuscript and critical revision of the work and final approval of the version to be published. James Carpenter: Co-leads research into using routinely collected data for outcome measures in randomised clinical trials, contributed to the conception, writing and editing of the manuscript and approved the version to be published.

Macey Murray: Co-leading research into using routinely collected health data for outcome measures in randomised clinical trials, contributed to the writing and editing of the manuscript and approved the version to be published. Saiam Ahmed: Involved in the STAMPEDE study, provided input into the manuscript and critical revision of the work and final approval of the version to be published. Veronica Nanton: Co-investigator of BladderPath qualitative research substudy, provided input into the manuscript and critical revision of the work and final approval of the version to be published. Kieran Jefferson: Contributed to conception and design of BladderPath, Principal Investigator at UHCW and provided input into the manuscript and critical revision of the work and final approval of the version to be published. James Catto: Contributed to the conception and design of BladderPath, Principal Investigator at the Royal Hallamshire Hospital and was involved in the revision of the manuscript and final approval of the version to be published. Jean Gallagher: Contributed to the patient perspective throughout, critical revision of the work and final approval of the version to be published. Allen Knight: Contributed on all matters of patient and public interest, most notably helping in the development of patient facing documents, provided input into the manuscript and critical revision of the work and final approval of the version to be published.

Collaborators The BladderPath trial management group.

Contributors HM: conducted the validation and primary author. AD: involved in development and curation of reference datasets for validation, provided input into manuscript, critical revision and final approval of the version to be published. HP: contributed to the design of BladderPath and helped write and critically appraise the manuscript and final approval of the version to be published. MS: provided input into the manuscript, discussion of key points and approved the version to the published. RB: coinvestigator of BladderPath, substantially contributed to the design of the work, involved in the revision of the manuscript and final approval of the version to be published. He also obtained funding for and led the diagnostic urinary biomarker BladderPath substudy. NJ: Chief Investigator of BladderPath, substantially contributed to the conception, design and running of the trial and routine data substudy and was involved in writing the manuscript, critical revision and approval of the version to be published. PP: corresponding author for this article. Substantially contributed to the conception, design and running of the trial and routine data substudy and the validation database. Principal Investigator at University Birmingham Hospitals. Involved in writing the manuscript, critical revision and had final approval of the version to be published.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests MS is supported by UKRI (MRC) MC_UU_00004/08. All support for the present manuscript (eg, funding, provision of study materials, medical writing, article processing charges) in past 36 months: Astellas (Research grants and drug costs or biomarker testing costs, all to institution and all active in past 36 months but on research outside of this research); Clovis Oncology (Research grants and biomarker testing costs, all to institution and all active in past 36 months but on research outside of this research); Janseen (Research grants and drug costs or biomarker testing costs, all to institution and all active in past 36 months but on research outside of this research); Pfizer (Research grants and biomarker testing costs, all to institution and all active in past 36 months but on research outside of this research); Novartis (Research grants and biomarker testing costs, all to institution and all active in past 36 months but on research outside of this research); Sanofi-Aventis (Research grants and biomarker testing costs, all to institution and all active in past 36 months but on research outside of this research). Payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events; Lilly Oncology—Speaker fees at clinical trial statistics training meeting for clinicians (no discussion of particular drugs); Janssen—Speaker fees at clinical trial statistics training meeting for clinicians (no discussion of particular drugs); Eisai, tutorial.Participation on a Data Safety Monitoring Board or Advisory Board—Independent member of many Independent Data Monitoring Committees but all for academic sponsors and none paid. NDJ: All support for the present manuscript (eg, funding, provision of study materials, medical writing, article processing charges)—National Institute of Health and Care Research Health Technology Assessment programme, Project funding to the Cancer Research UK Clinical Trial Unit, University of Birmingham (Project no. 14/08/60). RTB: All support for the present manuscript (eg, funding, provision of study materials, medical writing, article processing charges)—BladderPath research funding to University of Birmingham, UK. Grants or contracts from any entity (all in list research funding to university of Birmingham, UK)—Cancer Research UK (Early Detection & Diagnosis); Cancer Research UK (Data Innovation Award); Janssen; University Hospitals Birmingham Charity, UK; Cancer Research UK (Biospecimen Collection); QED Therapeutics, USA; UroGen Pharma, USA; Cancer Research UK (Biomarker Project Award); Cancer Research UK (Early Detection Spark Award. Royalties or licenses—Nonacus Limited, UK—Diagnostic urinary biomarker royalties to University of Birmingham, UK. Payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events—The Karolinska Institute, Sweden, Personal honorarium as PhD opponent. Patents planned, issued or pending—International patent application (PCT/GB2019/052776)—University of Oxford, UK and University of Birmingham, UK. Leadership or fiduciary role in other board, society, committee or advocacy group, paid or unpaid—Action Bladder Cancer UK—unpaid trustee.

Patient and public involvement Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

Patient consent for publication Not applicable.

Ethics approval The BladderPath study protocol and subsequent amendments were approved by the the London Bridge Research Ethics Committee (REC 17/LO/1819) and local institutional review boards in accordance with the ethical principles of the Declaration of Helsinki and Good Clinical Practice Guidelines. Signed informed consent was obtained from all participants before registration in to the study.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data sharing not applicable as no datasets generated and/or analysed for this study.

ORCID iDs
Harriet Paige Mintz http://orcid.org/0000-0003-1583-6268
Amandeep Raj Singh Dosanjh http://orcid.org/0000-0002-4522-7722
Helen Parsons http://orcid.org/0000-0002-2765-3728
Matthew Sydes http://orcid.org/0000-0002-9323-1371
Richard T Bryan http://orcid.org/0000-0003-2853-4293
Nicholas D James http://orcid.org/0000-0002-7314-8204
Prashant Patel http://orcid.org/0000-0002-2882-5990

REFERENCES

1 Health Informatics Working Group. Available: https://www.methodologyhubs.mrc.ac.uk/about/working-groups/health-informaticswg/ [Accessed Jul 2022].
2 Health Data Research UK (HDR UK). Available: https://www.hdruk.ac.uk/ [Accessed 5 Sep 2022].
3 Mintz HP, Dosanjh A, Parsons HM, *et al*. Development and validation of a follow-up methodology for a randomised controlled trial, Utilising routine clinical data as an alternative to traditional designs: a pilot study to assess the feasibility of use for the Bladderpath trial. *Pilot Feasibility Stud* 2020;6:165.
4 Mintz HP. *Can routinely collected data be used to inform randomised controlled trial outcomes in oncology*. University of Warwick, 2019.
5 Lensen S, Macnair A, Love SB, *et al*. Access to routinely collected health data for clinical trials - review of successful data requests to UK registries. *Trials* 2020;21:398.
6 Dataset catalogue. NHS Digital. Available: https://digital.nhs.uk/services/data-access-request-service-dars/dars-products-and-services/data-set-catalogue#h [Accessed 16 May 2023].
7 Bryan RT, Liu W, Pirrie SJ, *et al*. Comparing an imaging-guided pathway with the standard pathway for staging muscle-invasive bladder cancer: preliminary data from the Bladderpath study. *Eur Urol* 2021;80:12–5.
8 Mc Cord KA, Al-Shahi Salman R, Treweek S, *et al*. Routinely collected data for randomized trials: promises, barriers, and implications. *Trials* 2018;19:29.
9 Sydes MR, Barbachano Y, Bowman L, *et al*. Realising the full potential of data-enabled trials in the UK: a call for action. *BMJ Open* 2021;11:e043906.
10 Murray ML, Love SB, Carpenter JR, *et al*. Data provenance and integrity of health-care systems data for clinical trials. *Lancet Digit Health* 2022;4:e567–8.

11 Little RJ, D'Agostino R, Cohen ML, *et al*. The prevention and treatment of missing data in clinical trials. *N Engl J Med* 2012;367:1355–60.
12 Macnair A, Love SB, Murray ML, *et al*. Accessing routinely collected health data to improve clinical trials: recent experience of access. *Trials* 2021;22:340.
13 Goldacre B, Morley J. *Better, broader, safer: using health data for research and analysis. A Review Commissioned by the Secretary of State for Health and Social Care*. 2022.
14 Systemic therapy in advancing or metastatic prostate cancer: evaluation of drug efficacy, A multi-arm multi-stage randomised controlled trial [The protocol v23]. Available: http://www.stampedetrial.org/media/2646/stampede-protocol_v23-05-oct-2021-2.pdf [Accessed 04 Jul 2022].
15 McCall B. Data, data all around. *Lancet Digit Health* 2021;3:e284–5.