Building trust in real–world data: lessons from INSIGHT, the UK's health data research hub for eye health and oculomics

Alastair K. Denniston[a,b,c], *Aditya U. Kale*[a,b], *Wen Hwa Lee*[a,d], *Susan P. Mollan*[a,c,e], *and Pearse A. Keane*[a,f]

Purpose of review
In this review, we consider the challenges of creating a trusted resource for real-world data in ophthalmology, based on our experience of establishing INSIGHT, the UK's Health Data Research Hub for Eye Health and Oculomics.

Recent findings
The INSIGHT Health Data Research Hub maximizes the benefits and impact of historical, patient-level UK National Health Service (NHS) electronic health record data, including images, through making it research-ready including curation and anonymisation. It is built around a shared 'north star' of enabling research for patient benefit. INSIGHT has worked to establish patient and public trust in the concept and delivery of INSIGHT, with efficient and robust governance processes that support safe and secure access to data for researchers. By linking to systemic data, there is an opportunity for discovery of novel ophthalmic biomarkers of systemic diseases ('oculomics'). Datasets that provide a representation of the whole population are an important tool to address the increasingly recognized threat of health data poverty.

Summary
Enabling efficient, safe access to routinely collected clinical data is a substantial undertaking, especially when this includes imaging modalities, but provides an exceptional resource for research. Research and innovation built on inclusive real-world data is an important tool in ensuring that discoveries and technologies of the future may not only favour selected groups, but also work for all patients.

Keywords
artificial intelligence, eye health, Health Data Research UK, INSIGHT, oculomics, patient voice, real-world data, vision

INTRODUCTION

At some point in the future, we will look back on our current 'information profligacy' with puzzlement. We will wonder why we let precious information sit in disconnected silos: dots unjoined, lessons unlearned and tomorrow's patients served with the same level of ignorance as those seen today. There is, however, hope of change. Widespread adoption of electronic health records connected to digital platforms for imaging and other data should allow us to finally harness the 'everyday encounters' between patient and almost any part of the health system [1]. Systematic capture through these systems supported by curation, aggregation, sharing and analysis can accelerate the translation of *data* into *knowledge*, and *knowledge* into improved *practice* [2▪▪].

[a]INSIGHT Health Data Research hub for Eye Health, [b]Academic Unit of Ophthalmology, Institute of Inflammation & Ageing, College of Medical and Dental Sciences, University of Birmingham, [c]Ophthalmology Department, University Hospitals Birmingham NHS Foundation Trust, Birmingham, [d]Action Against Age-Related Macular Degeneration, London, [e]Institute of Metabolism and Systems Research, College of Medical and Dental Sciences, University of Birmingham and [f]NIHR Biomedical Research Centre At Moorfields Eye Hospital NHS Foundation Trust, UCL Institute of Ophthalmology, London, UK

Correspondence to Alastair K. Denniston, Ophthalmology Department, University Hospitals Birmingham NHS Foundation Trust, Birmingham B15 2GW, UK. E-mail: alastair.denniston@insight.hdrhub.org

KEY POINTS

- The use of routinely collected health data for research should be built on patient and public trust, and directed towards research that will lead to patient benefit.

- A safe approach to the use of health data for research can be supported by a 'Five Safes' approach, which considers projects, people, settings, data and outputs.

- To maximize relevance and impact, data resources should be as contemporary as possible without compromising quality; this can be achieved through a quality-assured 'pipeline' approach that continuously updates to provide near-real time data.

- Inclusion and optimal scale should be considered at the design stage of any data resource to ensure that the technical structure and tools support efficient onboarding of a diverse range of data institutions.

- INSIGHT, a cross-sector partnership within the UK's National Health Service, provides an example of how routinely collected eye health data (including more than 25 million images) can be safely made accessible to researchers to support discovery, innovation and care in eye health.

Real-world data (RWD), such as described here, does need to be treated with respect. The scale of datasets can be dazzling, but is only of value if the data itself can be trusted. As with any other source of scientific data, the researcher needs to have confidence in the provenance of the data, and be assured of the end-to-end process from original observation to the data point (or image) provided to them. Provided this can be achieved, then RWD does have a powerful part to play in research from discovery to validation, from epidemiology to the creation of artificial intelligence health technologies [3▪▪,4,5▪].

In this review, we consider the challenges of creating a trusted resource for RWD in ophthalmology, based on our experience of establishing INSIGHT, the UK's Health Data Research Hub for Eye Health and Oculomics [6].

INSIGHT IN BRIEF

INSIGHT forms part of Health Data Research UK (HDR UK), the United Kingdom (UK) National Institute for Health Data Science, which aims to 'unite the UK's health data to enable discoveries that improve people's lives' [7]. Initial funding was awarded through the UK's Industrial Strategy Challenge Fund in 2019, as one of the six initial health data hubs covering different specialist areas (such as Acute Care or Cancer), and with further specialist hubs established since then [8].

Our aim in establishing INSIGHT was to create an exceptional, trusted bioresource for eye health and oculomics [9▪] based on routinely collected imaging and linked clinical data within the UK's National Health System. Although the value of imaging to eye health is a well recognized need, oculomics is an emerging discipline that uses the 'eye as a window' to provide insights into systemic health and well being, an ancient concept enhanced for the modern era by the combination of high-resolution quantitative digital imaging, artificial intelligence and large-scale datasets.

INSIGHT's founding partners include two National Health Service (NHS) Trusts (Moorfields Eye Hospital NHS Foundation Trust, MEH; University Hospitals Birmingham NHS Foundation Trust, UHB), academic and charity partners (University of Birmingham, UK; Action Against Age-related Macular Degeneration, AAAMD), and two industry partners (Google and Roche). This cross-sector partnership brought complementary expertise, including in the management of health service data, industry-grade processes for secure data architecture, data curation, quality assurance and quality control, such as would be expected for real-world evidence forming part of regulatory submissions.

PRINCIPLES

At the time of writing, INSIGHT is in its third year since its launch in 2019. Building from our north star of 'patient benefit', we have established a number of principles, which reflect the priorities and ethos of our work at INSIGHT. We share these here in the hope that they may be helpful to those who wish to undertake similar work, or wish to understand the challenges and opportunities in greater depth.

(1) Build trust through patient involvement, openness and transparency
(2) Ensure safety through 'Five Safes' approach to data security
(3) Promote equity and shared benefit through inclusivity and data diversity
(4) Create data 'pipelines' that ensure efficient access to continually updating, near-real time data
(5) Design structures and tools that support scaling
(6) Invest for long-term sustainability and impact

Principle 1: Build trust through involvement, openness and transparency

Trust is a critical foundation for all clinical research, and it should be recognized that the use of data
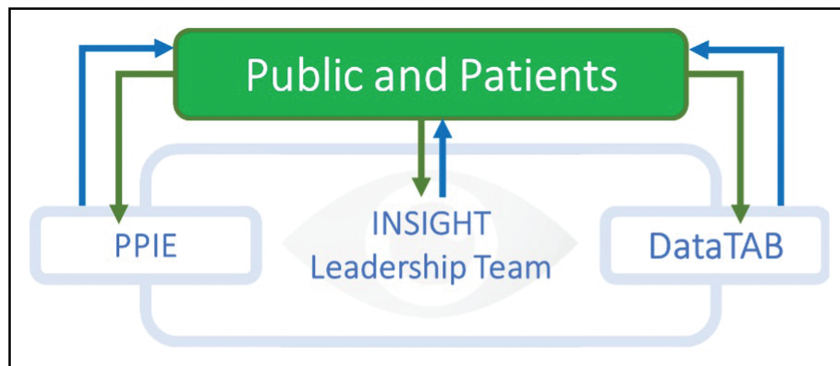
FIGURE 1. The public and patients are involved at multiple levels within the strategy, function and governance of INSIGHT, forming the Data Trust Advisory Board, being represented on the leadership team, and being involved through a wide-ranging patient and public involvement and engagement (PPIE) programme.

(health or otherwise) is an area where trust may be in short supply. Patients are at the heart of the 'why' of INSIGHT (its purpose) and the 'how' of INSIGHT (its strategy, function and governance). INSIGHT has three major ways in which patients and the public contribute to its strategy, function and governance: patient representation within the INSIGHT leadership team; a Data Trust Advisory Board comprising patients and public representatives which provides independent advice on all data access requests to INSIGHT; and our Patient and Public Involvement and Engagement (PPIE) programme, which works with the communities we serve, and informs what we do (Fig. 1).

Patient and public advisers' representation in leadership

INSIGHT is supported by a patient and public advisory group of 15 patients, which is supported by the PPIE team of the Moorfields Eye Hospital NHS Foundation Trust (specifically the NIHR Biomedical Research Centre of Moorfields and UCL). Two members of the group were appointed through open invitation to sit on the main INSIGHT Leadership Group, and fully participate in all activities relating to progress, function and strategy. These members help ensure that the 'patient voice' [10] is well represented within INSIGHT, supporting the priorities and concerns of patients.

Data trust advisory board

Data access requests to INSIGHT undergo three main levels of assessment. The NHS Trust remains the legal data controller and has ultimate responsibility for providing data access.

In stage 1, the INSIGHT team undertake a series of checks, which include: due diligence assessment of the applicant and institution (evidence that this would be a 'safe user'); assessment of the technical feasibility of satisfying the request (nature and volume of data enquired after); assessment of whether the description of the project and public benefit is described adequately and appropriately for a lay audience. Data access applications are welcomed from any sector provided that there is a clear path to patient benefit; in addition to the NHS, applicants have included UK and international applicants from academia, small and medium enterprises, pharmaceutical industry, healthtech industry and charities.

In stage 2, the application is reviewed by a Data Trust Advisory Board (DataTAB), which provides an independent view on the application. The DataTAB comprises an independent Chair, and members of diverse background and experience. The DataTAB reviews the applications against open criteria [11], which include the need to demonstrate public benefit and safe use of data aligned to the 'Five Safes' [12▪▪,13]. Evidence provided to the DataTAB includes the full application, list of data fields requested and applicants' credentials. The DataTAB provides a consensus-based recommendation to INSIGHT and the NHS Trusts, which may be recommend approval; recommend approval subject to conditions; recommend reject; or recommend pause pending further clarification.

In stage 3, the NHS Trust considers the evidence provided to it from stages 1 and 2, including the DataTAB recommendation. If the NHS Trust's senior legal officer for data (the 'Data Protection Officer') is satisfied, then the application is approved, and it can proceed to contract discussions. Any outcomes from Stage 3 and any research outputs from approved projects are shared back to the DataTAB to help refine their assessment criteria.

Principle 2: Ensure safety through `five safes' approach to data security

The use of health data for research should not only comply with all legal requirements, but should also seek to align to best practice. INSIGHT is wholly committed to promoting the protection of privacy and data security in line with the OECD Recommendation of the Council on Health Data Governance, and in compliance with both GDPR and national law, specifically the UK's Data Protection Act. In practical terms, our approach to data access is guided by the 'Five Safes' model, which is summarized in the following section [12▪▪]. INSIGHT recognizes the model's key feature that the five dimensions 'severally and jointly' contribute to the safety (or risk) around data access.

Safe setting

INSIGHT provides a safe setting through technical and physical security, education and culture, and contractual safeguards. Data sit within the NHS data framework, with each NHS Trust retaining data controllership of data relating to patients under their care. Data are pseudonymized at the earliest possible step with a secret one-way HASH and further anonymized (irreversibly deidentified) prior to researcher access.

Researcher access is provided within an approved Trusted Research Environment (TRE), or a safe environment which provides equivalent security requirements [14]. INSIGHT itself can provide access to data within a TRE at UHB, and a second is under construction at MEH, enabling access in multiple cloud environments. Contractual safeguards provide additional protection, and are particularly important if data are to be held in research environments not under the direct control of the two NHS Trusts.

Safe data

INSIGHT operates a data minimization principle, ensuring that access is only provided to data that are needed to address the specific research question, and considering both its sensitivity and risk of reidentification. When evaluating access requests, INSIGHT requires justification of every datafield requested, level of granularity requested (e.g. age to the nearest month, year, decade) and volume of data requested. Data within INSIGHT are held in a pseudonymized form, but with irreversible deidentification at the point of request for access to provide an anonymized dataset to the researcher.

Risk is managed proportionately with regard to providing access to any data that might alone or through combination lead to identification of an individual. A general principle of INSIGHT is that data made accessible should be necessary and proportionate to the purposes required, that is there is data minimization. When requesting access to a dataset, an applicant must justify the inclusion of each datafield. INSIGHT reserves the right to refuse an application or limit the data fields available or their granularity based on concerns around possible identification. Contractual safeguards provide additional protection.

Safe outputs

It is important that researchers publish their findings, and with sufficient detail to maximize the value of the study. However, the way that data are presented, particularly in tables, may provide sufficient detail for inadvertent disclosure at individual level. INSIGHT requires authors to ensure that this is avoided through an 'output statistical disclosure control' in which they evaluate all statistical output for risk of disclosure [15]. A common example is for tables wherein any cells may have less than five units. In such cases, authors may be asked to either consider collapsing categories if possible; or replace the cell count with '<5'. In some cases, an alternative threshold such as 10 units may be used.

Safe people

'Safe People' reviews the knowledge, skills and incentives of the users to use the data appropriately. One of the essential criteria by which INSIGHT evaluates all applications is whether the applicant (both individual and institution) is deemed to be appropriate, and this is undertaken through the INSIGHT Due Diligence Process. Applicants are also required to evidence that they have completed appropriate training in the safe use of health data for research.

Safe projects

Although this consideration is less directly related to the issue of ensuring confidentiality, it is included here for completeness as part of the assessment within the 'Five Safes'. Under 'Safe Projects', INSIGHT considers the legal, moral and ethical considerations surrounding use of the anonymized data, including the likelihood of patient benefit. It is important to recognize that although the 'Five Safes' are often considered in terms of assessing and minimizing risk related to proceeding with a data project, INSIGHT also considers the risk of 'loss to public benefit' through *not* doing the project.

Principle 3: Promote equity and impact through data diversity and inclusivity

Health data poverty is the inability for individuals, groups or populations to benefit from a discovery or

innovation due to a scarcity of data that are adequately representative [16■]. It is of particular concern in the context of artificial intelligence health technologies, which utilize machine learning models that are highly tuned to the data that the model has been trained on. When exposed to a new population, such models may be 'brittle', with deterioration of performance due to poor generalizability [17■■]. Unless this is taken seriously, harm may occur through *ignorance, acquiescence* or *exclusion.*

Where the level of artificial intelligence bias arising from unrepresentative health data is not recognized (*ignorance*) or simply accepted (*acquiescence*), then technologies may cause harm through systematic, wide-scale under-performance in certain groups. To combat this, reporting guidelines for artificial intelligence health technologies ask that analysis of errors includes stratification by demographic subgroups, and this is under active consideration by regulators. It also highlights the need for such technologies to have robust postmarket surveillance measures, which include consideration of errors, adverse events and harms by demographic group, including ethnicity, age and sex [18,19].

Where under-performance is recognized but dealt with by denying access to that technology for specific groups (*exclusion*), a two-tier health system is likely to occur in which majority populations are likely to benefit from the most advanced digital healthcare, whereas minority populations are required to continue to use increasingly outdated and disinvested 'analogue' systems. It is important to note that this form of digital health divide is based entirely on access to data, and not on access to technology. It is under-recognized, but is a serious threat to equitable digital healthcare, and therefore to health itself [16■,20].

How big a problem might this be in ophthalmology? In 2020, a global survey of publicly available datasets of ophthalmic images identified and analysed 94 datasets containing 507 724 clinical images and 125 videos of eyes gathered from at least 122 364 people. Much of the world was found to be unrepresented: for example, there was only one dataset from sub-Saharan Africa and two datasets from South America, whereas most images came from populations in Asia and Europe, with very few datasets from large parts of the world such as sub-Saharan Africa (one dataset) and South America (two datasets). Critically, reporting of demographic information was generally poor with age, sex and ethnicity being missing in more than 80% of datasets. Given that the vast majority of artificial intelligence health technologies are developed on these types of easily accessible datasets, there is a need to create well characterized, diverse datasets that

address health data poverty and better represent the populations who could benefit from these technologies [21].

INSIGHT is exceptionally well placed to address this issue. First, the NHS itself is a health service for the whole population of the UK, being free at the point of need. The profile of NHS users is therefore highly representative of the whole, diverse UK population. Second, alongside other similar initiatives in the UK [22■], it is based on routinely collected healthcare data and is approved to be undertaken on an 'opt out' basis. Numerous studies have shown that the majority of the UK population supports the use of their health data for research [23,24■■], with fewer than 6% of the UK population having opted out after considerable media coverage of the issue and public awareness [25]. Finally, the two founding NHS Trusts serve large, diverse populations in London and Birmingham, and this is reflected within the datasets created.

The diversity within the population that INSIGHT serves is reflected in the leadership of INSIGHT, which is primarily drawn from the NHS Trusts and the communities they serve. In addition, a major focus of the PPIE theme is to engage with groups that are 'seldom heard', and to ensure that INSIGHT recognizes their views and communicates effectively.

Principle 4: Create data `pipelines' that ensure efficient access to continually updating, near-real time data

When creating datasets for research from routinely collected health data, a key strategic decision is whether to create a *static* dataset or to create a *dynamic* structure. INSIGHT is designed to be an efficient ongoing data resource with the ability to create data extracts on demand, and to recreate these at any future point. The emphasis is on 'getting the pipeline right', such that curation processes occur automatically, and are supported by built-in quality assurance and quality control processes. Requests for new datasets that use all or a majority of existing data-fields can therefore be serviced rapidly, and more cost-efficiently.

The dynamic pipeline approach also means that the research database can be much more responsive to changes in the health environment. The Covid-19 pandemic has highlighted the value to healthcare providers in being able to rapidly collect data relating to new conditions within days, and to meet unexpected changes in health demand [26]. INSIGHT datasets refresh on a daily basis from the live NHS systems, with this frequency being increased if required.

Principle 5: Design structures and tools that support replication and scaling

Design should take into account whether there is a need to replicate and scale. This is seen both in technical design choices and human factors. For INSIGHT, the aim has always been to 'build and share', enabling other NHS sites to join and be supported to make their data accessible to researchers efficiently and safely. Our 'cloud first' strategy – aligned to NHS policy – facilitates this, as our infrastructure and tools can be replicated in other NHS cloud environments [27]. We have adopted an 'infrastructure as code' approach using *Terraform* (Hashicorp) suitable for a 'lift and drop' approach in other sites.

One of the biggest challenges for sites who wish to make their data available for research is the need to ensure that imaging data are both anonymized and in a suitable format for research. INSIGHT has addressed this for ophthalmic imaging through the creation of Moorfields Librarian, a tool that was first created for the Moorfields-DeepMind collaboration, and which enables a pipeline approach to the anonymization and subsequent 'DICOMisation' of OCT and other retinal images [28].

In addition to technical considerations, there is a need to consider human factors, including the need for complete, accessible documentation; the need for training in new tools, standard operating procedures and data dictionaries; the need to consider nonstandard integration, including the use of 'on premises' solutions; and the need to ensure shared understanding of how information governance is supported end-to-end.

Principle 6: Invest for long-term sustainability and impact

Building an infrastructure that supports efficient, safe access to research-ready data is a significant task. It requires a team with wide-ranging expertise (including technical, clinical, information governance) who need to be backed by significant commitment from the host institutions. Rather like building a biobank, there is considerable upfront investment and effort before public research outputs are seen. For INSIGHT, it was exactly 2 years after its launch that the first complete research study using data made available through INSIGHT was published [29▪]. That study combined data from MEH and UHB with preexisting trial data to enhance predictive modelling in neovascular Age-related Macular Degeneration, and evaluate the potential detrimental effects of prolonged treatment delay due to the pandemic as a way of informing care strategies within the NHS and beyond. Since that first application, INSIGHT has supported a wide range of applications from the most common sight-threatening conditions to novel retinal diagnostics for systemic health, and from epidemiological studies to requests to build new artificial intelligence tools.

A resource like INSIGHT is a precious opportunity for the eye research community, and the patients we serve. From the outset, INSIGHT has been designed to be sustainable, ensuring that it would continue to support eye health and oculomic researchers from around the world.

CONCLUSION

Everyday encounters between patients and the health system are probably the most valuable resource we have to better understand disease, and to progress development of new diagnostics, therapeutics and prognostics. Initiatives such as INSIGHT show how we can unlock this potential through building ethical processes that capture, curate and provide safe access to high-quality RWD that enables research that improves people's lives.

Conflicts of interest

Professor Keane has received speaker fees from Heidelberg Engineering, Topcon, Carl Zeiss Meditec, Haag-Streit, Allergan, Novartis and Bayer. He has served on advisory boards for Novartis and Bayer and has been an external consultant for DeepMind and Optos.

Professor Mollan reports consulting fees from Invex Therapeutics and Neurodiem; advisory board compensation from Janssen, Santhera, GenSight Biologics and Chugai-Roche Ltd; and speaker fees from Chiesi, Heidelberg engineering, Allergan, Chugai-Roche Ltd, Santen, Santhera, Roche and Teva. All conflicts of interest are outside the submitted work.

All authors' employing institutions are partners in the INSIGHT award. No authors have any other interests to declare.

REFERENCES AND RECOMMENDED READING

Papers of particular interest, published within the annual period of review, have been highlighted as:

- ■ of special interest
- ■■ of outstanding interest

1. Denniston AK. How using data from 'everyday encounters' could be the most powerful tool we have to help improve people's lives [Internet]. HDR UK. 2019 [cited 23 March 2022]. https://www.hdruk.ac.uk/news/how-using-data-from-everyday-encounters-could-be-the-most-powerful-tool-we-have-to-help-improve-peoples-lives/. [Accessed 28 February 2022]

2. Foley T, Horwitz L, Zahran R. Realising the potential of learning health
■■ systems. Newcastle University; May 2021 May. https://learninghealthcare-project.org/wp-content/uploads/2021/05/LHS2021report.pdf. [Accessed 28 February 2022]

A helpful introduction to Learning Health Systems, with a focus on how more effective use of data can be translated into knowledge, change in practice and improved outcomes for patients.

3. Writing Committee for the UK Age-Related Macular Degeneration EMR Users
■■ Group. The neovascular age-related macular degeneration database: multi-center study of 92 976 ranibizumab injections: report 1: visual acuity. Ophthalmology 2014; 121:1092–1101.

An outstanding early example of the power of big data in eye health research enabled by a shared electronic health record platform and collaborative effort across health institutions; this, the first of a series of articles, is not only focused on treatment outcomes in age-related macular degeneration, but also has relevant discussion of the opportunities and challenges for RWD in an eye health context.

4. Denniston AK, Lee AY, Lee CS, *et al.,* UK DR EMR Users Group. United Kingdom Diabetic Retinopathy Electronic Medical Record (UK DR EMR) Users Group: report 4, real-world data on the impact of deprivation on the presentation of diabetic eye disease at hospital services. Br J Ophthalmol 2019; 103:837–843.

5. Lee CS, Owen JP, Yanagihara RT, *et al.* Smoking is associated with higher
■ intraocular pressure regardless of glaucoma: a retrospective study of 12.5 million patients using the Intelligent Research in Sight (IRIS®) Registry. Ophthalmol Glaucoma 2020; 3:253–261.

An important example of the use of large-scale routinely collected health data from the USA, in this case focused on the association of smoking and intraocular pressure.

6. INSIGHT. INSIGHT, the health data research hub for eye health [Internet]. [cited 2 November 2021]. https://www.insight.hdrhub.org/. [Accessed 28 February 2022]

7. Health Data Research UK [Internet]. 2019 [cited 23 March 2022]. http://www.hdruk.ac.uk. [Accessed 28 February 2022]

8. Health Data Research UK. Health Data Research Hubs [Internet]. HDR UK. 2020 [cited 23 March 2022]. https://www.hdruk.ac.uk/helping-with-health-data/health-data-research-hubs/. [Accessed 28 February 2022]

9. Wagner SK, Fu DJ, Faes L, *et al.* Insights into systemic disease through retinal
■ imaging-based oculomics. Transl Vis Sci Technol 2020; 9:6.

An introduction to 'oculomics', noting how the concept of the 'eye as a window' is being accelerated by high-resolution imaging, large-scale datasets and artificial intelligence to reveal novel ophthalmic signatures of systemic disease.

10. Dean S, Mathers JM, Calvert M, *et al.* 'The patient is speaking': discovering the patient voice in ophthalmology. Br J Ophthalmol 2017; 101:700–708.

11. Data trust advisory board [Internet]. INSIGHT. [cited 23 March 2022]. https://www.insight.hdrhub.org/data-trust-advisory-board. [Accessed 28 February 2022]

12. Desai T, Ritchie F, Welpton R. Five safes: designing data access for research.
■■ 2016. Bristol Business School Working Papers in Economics. https://www2.uwe.ac.uk/faculties/BBS/Documents/1601.pdf. [Accessed 28 February 2022]

A foundational reference introducing the helpful 'five safes' concept of data management: projects, people, settings, data and outputs.

13. Karrar N, Khan SK, Manohar S, *et al.* Improving transparency in the use of health data for research: recommendations for a data use register standard [Internet]. 2022. https://zenodo.org/record/5902743. [Accessed 28 February 2022]

14. UK Health Data Research Alliance, NHSX. Building trusted research environments: principles and best practices; Towards TRE ecosystems [Internet]. 2021. https://zenodo.org/record/5767586. [Accessed 28 February 2022]

15. Griffiths E, Greci C, Kotrotsios Y, *et al.* Handbook on statistical disclosure control for outputs. The Health Foundation; 2019. https://ukdataservice.ac.uk/app/uploads/thf_datareport_aw_web.pdf. [Accessed 28 February 2022]

16. Ibrahim H, Liu X, Zariffa N, *et al.* Health data poverty: an assailable barrier to
■ equitable digital healthcare. Lancet Digit Health 2021; 3:e260–e265.

An introduction to the concept of health data poverty, and its risk to equity in digital health.

17. Kaushal A, Altman R, Langlotz C. Geographic distribution of US cohorts used
■■ to train deep learning algorithms. JAMA 2020; 324:1212–1213.

In clinical applications of deep learning across multiple disciplines, algorithms trained on U.S. patient data were disproportionately trained on cohorts from California, Massachusetts and New York, with little to no representation from the remaining 47 states.

18. Liu X, Cruz Rivera S, Moher D, *et al.,* SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nat Med 2020; 26:1364–1374.

19. Liu X, Glocker B, McCradden MM, *et al.* The medical algorithmic audit. Lancet Digit Health 2022; 4:e384–e397.

20. Burton MJ, Ramke J, Marques AP, *et al.* The Lancet Global Health Commission on Global Eye Health: vision beyond 2020. Lancet Glob Health 2021; 9:e489–e551.

21. Khan SM, Liu X, Nath S, *et al.* A global review of publicly available datasets for ophthalmological imaging: barriers to access, usability, and generalisability. Lancet Digit Health 2021; 3:e51–e66.

22. Gallier S, Price G, Pandya H, *et al.* Infrastructure and operating processes of
■ PIONEER, the HDR-UK Data Hub in Acute Care and the workings of the Data Trust Committee: a protocol paper. BMJ Healthcare Inform [Internet] 2021; 28:; http://dx.doi.org/10.1136/bmjhci-2020-100294. [Accessed 28 February 2022]

A detailed description of a similar data resource to INSIGHT but focused on Acute Care; this provides a helpful detailed description of the operation of the hub, and in particular the patient and public involvement in decisions around data access.

23. Atkin C, Crosby B, Dunn K, *et al.* Perceptions of anonymised data use and awareness of the NHS data opt-out amongst patients, carers and healthcare staff. Res Involv Engagem 2021; 7:40.

24. Understanding patient data. Public attitudes to the use of patient data. 2021
■■ [Internet]. https://understandingpatientdata.org.uk/sites/default/files/2021-08/Public%20attitudes%202018-2021.pdf. [Accessed 28 February 2022]

This resource summarizes (and provides links to) recent studies in the UK regarding patient and public attitudes to the use of their health data. Although undertaken in a UK context, the public-facing resources of Understanding Patient Data provide a helpful model for all researchers thinking about how they involve and communicate with public and patients in the context of health data resources.

25. NHS Digital. National data opt-out, September 2021 [Internet]. 2021 [cited 23 March 2022]. https://digital.nhs.uk/data-and-information/publications/statistical/national-data-opt-out/september-2021. [Accessed 28 February 2022]

26. Kabarriti R, Brodin NP, Maron MI, *et al.* Association of race and ethnicity with comorbidities and survival among patients with COVID-19 at an urban medical center in New York. JAMA Netw Open 2020; 3:e2019795.

27. NHS Digital. NHS architecture principles [Internet]. 2020 [cited 23 March 2022]. https://digital.nhs.uk/about-nhs-digital/our-work/nhs-digital-architecture/principles. [Accessed 28 February 2022]

28. De Fauw J, Ledsam JR, Romera-Paredes B, *et al.* Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat Med 2018; 24:1342–1350.

29. Mollan SP, Fu DJ, Chuo C-Y, *et al.* Predicting the immediate impact of national
■ lockdown on neovascular age-related macular degeneration and associated visual morbidity: an INSIGHT Health Data Research Hub for Eye Health report. Br J Ophthalmol [Internet] 2021; http://dx.doi.org/10.1136/bjophthalmol-2021-319383. [Accessed 28 February 2022]
The first research study published from the INSIGHT database measured impact of the UK's national lockdown on patients' treatment and visual outcomes in neovascular age-related macular degeneration.