Review into the use of health data for research and analysis
Response from the British Heart Foundation
March 2021

About the BHF

The British Heart Foundation (BHF) is the largest independent funder of research into heart and circulatory disease and the third largest charitable funder of medical research in the UK. Each year, thanks to the generosity of our supporters, we are able to fund around £100 million of new research across the UK and in all four nations.

We support an extensive portfolio of projects focused on the use of health and care data, which has grown substantially in recent years. For example, in 2019 we announced the launch of a £10 million BHF Data Science Centre (BHF DSC) in partnership with Health Data Research UK (HDR UK). The Centre was established to work in partnership with patients, the public, the NHS, researchers and clinicians to promote the safe and ethical use of data for research into the causes, prevention and treatment of all diseases of the heart and circulation. The Centre has already risen to the many challenges of the Covid-19 pandemic and is leading several largescale data projects that aim to better understand the relationship between Covid-19 and cardiovascular diseases. In April 2020, the BHF and the BHF DSC supported the development of a paper that fed into SAGE decision-making,[1] which proposed a way to share data among researchers to rapidly explore the answers to Covid-19 related research questions, across the four nations, using cardiovascular disease as an exemplar.

The BHF also provides significant core funding to both HDR UK and the UK Biobank (currently £2 million and £3.2 million over five years, respectively), and has supported 12 BHF/Alan Turing Institute Cardiovascular Data Science Awards to date, at a value of over £0.5 million. This is a joint funding scheme to support collaborative research between cardiovascular investigators and data scientists seeking to generate data science solutions to key cardiovascular problems. In addition to these strategic investments, as of November 2019 we were supporting approximately £16m across other ongoing data science, machine learning and artificial intelligence research activities.

Introduction

The BHF welcomes the opportunity to respond to the invitation to share our thoughts and experience of using health data for research, in support of the review currently underway within Government. We work to drive forward the effective use of health data in the fight for improved heart and circulatory disease research, treatments, care and prevention, and recognise the importance of its safe and ethical use informed by, and communicated to, patients and the public. It is vital that regulators and law-making bodies act promptly, ensuring that laws keep pace with technological developments, and that effective oversight and trustworthiness stand at the heart of how data is governed and regulated. We are committed to realising the potential that comprehensive, nationwide, representative data has in this

---

[1] British Heart Foundation et al. A national health data research capability to support COVID-19 research questions (14 April 2020).

exciting era of personalised medicine and digital innovation through our considerable investment in data science programmes.

We are privileged to work with a world-class community of researchers who utilise the considerable health data asset of the UK on a day to day basis. Our expert patient data panel ensures that our data and data science work is always supported by a strong patient voice, encouraging, and challenging us on matters of engagement and transparency.

This response outlines four key priorities which the BHF would like to emphasise to the review, with the primary purpose of supporting and enabling our research community to utilise the remarkable health data assets available in the UK to best effect, whilst ensuring that processes around data stewardship are ethical and transparent.

Bring the researcher to the data

A standardised model of access which brings the researcher to the data

We would strongly advocate for a 'researcher to data' model whereby researchers access health data held in a secure environment where access can be controlled by the relevant data controllers. Many current models undertake the opposite approach, bringing 'data to the researcher' through data request processes and bespoke data extracts, which leads to significant process and governance burdens for both researcher and data provider, potentially delaying important research progress. A 'researcher to data' model would also further enhance opportunities for researchers to undertake discovery and exploration of data sources to maximise their value, as opposed to limiting requirements in advance of request. There are some good examples where such a model has supported rapid research during the pandemic, including NHS Digital's Trusts Research Environment[2], and the OpenSafely[3] analytics platform. We would encourage this model to be the default for all health data access for research.

User-oriented data environments

We have found that data access processes and analytical environments are often designed with the needs of the data provider put first. We would like to see this shift towards a more user-oriented approach. Designing processes and environments with the needs of users in mind, could involve providing clear and accessible training and professional expertise to support researchers use health data assets effectively and appropriately. Consideration of the use of a 'librarian' role within data providers who can advise researchers on, for example, known data quality issues, would be particularly welcome. This is well exemplified in NHS Digital's Data Access Environment which, in partnership with the BHF Data Science Centre, makes dataset expertise and analytical support available through a dedicated data scientist, specific support channels for analysts and researchers and comprehensive documentation and 'how to' guides.

Tools and methods:

Data analysis environments should support a variety of tools and methods to further support the democratisation of data availability and analysis to as wide a range of researchers as possible. Which tools and methods are supported should be informed by the research

---

[2] The TRE service provides approved researchers from trusted organisations with timely and secure access to health and care data. https://digital.nhs.uk/coronavirus/coronavirus-data-services-updates/trusted-research-environment-service-for-england

[3] OpenSAFELY is a secure analytics platform for electronic health records in the NHS, created to deliver urgent results during the global COVID-19 emergency. It accesses 58 million patients' full pseudonymised primary care NHS records. All analytic software is open for security review, scientific review, and re-use. https://opensafely.org/

community and open source approaches should be encouraged wherever possible in order to support wider dissemination and learning (see below).

Streamline data access

Reduce variation in data access processes

We would like to see a simplification and standardisation of data access processes across all four UK nations. A joined-up approach across the UK will result in improved representation within data used for research projects both in terms of patient populations and the differences inherent in system implementation, which will benefit research through increased power and impact. Currently, the processes and requirements for access to health data varies considerably between nations, adding to administrative and time burdens for researchers and resulting in a barrier to whole UK analysis. A single, UK-wide, agreed data access pathway and approvals process for all data assets would remove this.

Individual researchers as the 'unit of trust'

Current processes for health data access in some parts of the UK, primarily England, rely on individual researchers belonging to organisations with specific accreditation or credentials, for example the Data Security and Protection Toolkit required by NHS Digital in England. Such requirements can add a significant burden to individual organisations whose primary focus is not research, for example health charities. We believe individual researchers who can demonstrate training and appropriate accreditation under a valid scheme (e.g. ONS's Research Accreditation Service) for using health data appropriately should be able to be associated with approved projects irrespective of their organisational affiliation. This can be a key blocker in some instances to health data access, particularly in England, and limits the ability of researchers and research groups to bring in additional expertise without recourse to 'work-arounds' such as honorary contracts. In addition, linking accreditation to the individual supports professionalisation and has the potential to increase rigour and accountability in terms of access to data.

Multiple uses principle

Where extracts of data are currently required, the requirement for 'one extract, one use' adds to administrative burdens for researchers and can stifle data-driven research and innovation as data is required to be destroyed at the end of a project rather than being used to further develop or explore emerging research questions. Teams will often need to reapply for the same dataset, which results in repetition, additional time and administrative burden for research teams and data controllers. Appropriately accredited researchers should be able to undertake multiple analyses on data for which they have approved access on the proviso that this is made explicit and transparent to the data provider, approved as soon as practicable prior to further analysis being undertaken, and made transparent to the public. A move to a 'researcher to data' model of data access (see above) would negate this issue, but in the short term at least it remains a reality for access to many datasets.

Invest in wider engagement and involvement

Patient and public engagement:

Much of UK health data is derived from publicly funded health and care services; patients and the public rightly expect their data to be managed securely and be utilised for the benefit of services and the nation's health. Effective and representative public engagement is vital in understanding public concerns and views relating to the use of health and care data in research. Recent reports, including by Understanding Patient Data (UPD)[4] and OneLondon,[5] have highlighted just how valuable patient views are in developing data policies, as well as how important it is to educate the public regarding the use, and impact, of health data in research. The BHF also seeks to be led by patients on data related issues and has established a Patient Data Panel alongside Cancer Research UK. The panel was initially created to ensure patients were represented in our response to the NHS national data opt-out consultation in 2018, and has since been consulted by Public Health England, NHSX and UPD, as well helping to inform plans for the BHF Data Science Centre, among others.

We would therefore advocate for greater investment in education and engagement activities to raise awareness of how communities' health data is used in the medical research context, liaising with charities and patient representation groups to support increasing the profile of high quality research driven by UK health data and its impact and benefits. There should be a purposeful focus on representativeness to ensure research reflects the populations and communities it ultimately serves, capitalising on growing recognition of the importance of diversity and inclusion both in research datasets and research itself.[6,7]

Widen opportunities for additional datasets:

In the UK, health is defined and influenced by more than clinical data; developing a health data ecosystem that can receive, link and allow researchers to access a diverse range of data sets beyond the health sphere would add richness and extend opportunities for investigation including other publicly funded datasets from, for example, Genomics England, education and social services. This should extend, assuming principles of privacy, transparency and openness could be guaranteed, to datasets from private industry, for example in relation to wearable technology and consumer trends.

Community development:

The research community itself is an invaluable asset to the use, promotion and long-term value of the UK's health data. The provision of a common platform to support the development of a community of researchers and expert patients to share work and drive improvements would not only support peer to peer learning and development, but also give data providers the means of engaging meaningfully, at scale, with their user base to help plan and deliver improvements in the areas outlined above. Individual data providers have good examples of such community focussed sites, delivered through platforms such as GitHub. However, central hub would allow for better awareness and improved communication between providers and researchers, particularly in the case where researchers are engaged with multiple providers on the same project.

Adopt open principles

Transparency:

We are committed to the principle that UK health data derived from the NHS and other services funded through taxation is a public resource, and as such the outputs from its use should be made publicly and freely available where legally permissible. This should include publication of relevant analytic code and other supporting information under licensing

---

[4] Understanding Patient Data Accountability, transparency and public participation must be established for third-party use of NHS data (2 March 2020)
[5] OneLondon, Ipsos MORI, The King's Fund Public deliberation in the use of health and care data (30 June 2020)
[6] Manolio, T. (2019) Using the Data We Have: Improving Diversity in Genomic Research AJHG
[7] Wellcome Clinical Trials Policy – Including people from under-served groups (September 2020)

agreements to support credited reuse, not for commercial gain. Recent projects such as OpenSafely and work from the Health Foundation[8], among many others, exemplify commitment to transparent approaches. Such examples support the creation of a 'commons of knowledge' and continues to rapidly build a collaborative culture to foster innovation and capacity building. Furthermore, such transparency supports public trust and fosters greater understanding of benefit. These principles are well understood and discussed extensively in the context of health data provision and research; however, transparent approaches are not the default. There are ongoing issues with variability in the timeliness of publication and ensuring publications are accessible and usable, for example in stable, machine readable format and clearly linked from searchable web pages. Such issues can have the result of 'transparency' being achieved in theory only, with researchers frustrated by ever changing publication locations, definitions and formats.

Open standards:

We would support any agreement reached between data providers on common and consistent models of open data access. We believe this would promote common understanding and a shared vision to enable greater interoperability between data providers and researchers. There has been good progress in this area, with the work of the Open Data Institute and efforts from a wide range of data providers to make anonymised data freely available to the public. The pandemic period has exemplified the benefits of promoting this way of working. However, this is not always the 'default' for datasets and there is considerable variation across the UK. We would support an 'open-by-default' approach, with appropriate safeguards in place, applied at least to data from publicly funded services. This will ultimately support the development of a health data ecosystem that serves medical research rather than being viewed as a blocker to research progress, and engender ever greater public trust.

---

[8] E.g. Watt et al, (2020) Use of primary care during the COVID-19 pandemic, The Health Foundation, 17/03/2021, https://www.health.org.uk/news-and-comment/charts-and-infographics/use-of-primary-care-during-the-covid-19-pandemic