MRC/CSO Social and Public Health Sciences Unit Consultation Response

Title of consultation
Goldacre Review
Name of the consulting body
Department for Health and Social Care
Link to consultation
https://www.goldacrereview.org/
Why did the MRC/CSO Social and Public Health Sciences Unit contribute to this consultation?
The SPHSU uses a wide range of health-relevant data in its research and therefore has a keen interest in the questions highlighted in the review's Terms of Reference, in order to facilitate access to data to support efficient and high-quality research relevant to population health and inequalities.
Our consultation response
This response is submitted on behalf of the *Institute of Health and Wellbeing*, University of Glasgow and was prepared by members of the *MRC/CSO Social and Public Health Sciences Unit*. Our response to the consultation is informed by views from across the Institute. Correspondence address: Emily Tweed emily.tweed@glasgow.ac.uk **About us** The *Institute of Health and Wellbeing*, University of Glasgow is a global leader in research into population health and health inequalities. We comprise academic leaders from public health, medicine, epidemiology, psychology and health data science to address some of the world's greatest health challenges. Our research is guided by three research themes which support multidisciplinary approaches with a shared vision for health and healthcare equality – determinants of health and health inequalities, solutions focused research and data science. It is delivered across seven research groups. Further information about the Institute is available from our website. The *MRC/CSO Social and Public Health Sciences Unit* is one of the research groups within the Institute of Health and Wellbeing. The Unit is an interdisciplinary group of sociologists, anthropologists, psychologists, epidemiologists, geographers, political scientists, public health physicians, statisticians, information scientists, trial managers and others. The Unit receives core-funding from the Medical Research Council and the Scottish Government Chief Scientist Office, as well as grant funding for specific projects from a range of sources. We conduct research to understand the determinants of population health and health inequalities, and to develop and test interventions to improve health and reduce inequalities, using a wide variety of methods including qualitative research; evidence synthesis; the collection, linkage, and analysis

of social survey and routinely collected data; natural experimental studies; and randomised controlled trials. Further information about the Unit is available from our website.

Together, we use a wide range of data to underpin our research (population-wide surveys, administrative data, healthcare data including hospitalisation and GP records) from across the UK and internationally, which we access in a variety of ways (virtual safe havens, physical trusted research environments, end user-licence agreements and open access data). Additionally, we have expertise across a range of disciplines and career profiles and are therefore well-placed to comment on the review into use of health data for research and analysis. Although health is a devolved matter, there is an opportunity for wider learning from across the four UK nations which will increase the benefits to patients and society as a whole: our responses below therefore draw on examples from across the UK, with a particular focus on research in which we have been closely involved. Our response focuses on quantitative data, in keeping with the questions outlined in the Terms of Reference, though we note (especially in Section 1) the enormous value and richness of qualitative data in health research.

Summary of response

We have organised our response into the following eight themes, linked to the questions set out in the review's Terms of Reference

1	**Good health research needs more than just healthcare data**	ToR 1, 3, 4, 8
2	**Governance and approvals processes require profound redesign if the benefits of health data research are to be realised, and the risks minimised**	1, 3, 4, 5
3	**A strategic approach to data acquisition and curation should be adopted, matched to priority research areas**	7, 10, 12
4	**Technical factors, including analytic platforms, trusted research environments, and data flows, are important but not the major barrier**	2, 3
5	**The public must play a central role in guiding the use of data for health research, at every stage of the process**	1, 6, 9
6	**Greater value must be placed on open science in approvals processes and reward structures for research**	6, 11
7	**Staffing models should recognise and reward the specialist, skilled work of data management and access**	7, 11
8	**Our experiences highlight both successes and challenges**	4

Response

1. Good health research needs more than just healthcare data

Good health among the population is created by a much wider range of factors than just healthcare, or even health and social care. These 'upstream' social, environmental, commercial, and political determinants of health are fundamental: they underlie the stalling and decreases in life expectancy observed prior to the pandemic, whilst Covid has brought many of them - particularly housing, occupation, social protections (such as sick pay and welfare benefits), and structural racism – into even sharper relief (Aldridge, Lewer et al. 2020, Douglas, Katikireddi et al. 2020, Mutambudzi, Niedwiedz et al. 2020, Niedzwiedz, O'Donnell et al. 2020, Walsh, McCartney et al. 2020). Yet most health data research continues to exclusively use data from healthcare sources, which might better be described as ill-health data.

Profound health inequalities persist in the UK, and on many metrics, are widening (Brown, Allik et al. 2019, Marmot, Allen et al. 2020). This carries a huge human and economic cost (Marmot, Allen et al. 2020). These inequalities have come into sharper focus during the pandemic, and contributed to the UK's unacceptably high and unequal death rate from Covid-19 (Marmot, Allen et al. 2020). Researchers from our Institute estimated that, over ten years, the life expectancy impact of inequality in the UK is greater than six unmitigated Covid-19 pandemics (McCartney, Leyland et al. 2021). Since inequalities in health arise from inequalities in the distribution of the social, environmental, commercial, and political determinants of health, healthcare data alone is insufficient if we are to understand and address them effectively.

It is therefore essential that data for health research encompass data from sectors other than healthcare: without this, we are greatly limited in our understanding of what health is and how it can be improved, as well as our ability to address inequalities. This is not a new idea: it was a central theme of the Academy of Medical Science's Health of the Public 2040 report, published in 2016 (Academy of Medical Sciences 2016). For instance, the observational and quasi-experimental study designs required to address many population health questions not amenable to randomised controlled trials require additional data from non-healthcare sources (such as surveys and administrative data) to adjust for confounding factors and achieve exchangeability (e.g., Hernán and Robins 2016). These data and the relevant linkages may be at the individual, household, and/or environmental level.

Infrastructure, funding, and methods for usage of data from non-healthcare sources in health research are much less developed and require urgent attention if we are to move beyond our current, blinkered view. While the Administrative Data Research Centres/Network have made some inroads in this regard, progress has been slower than hoped and many key datasets remain elusive. The barriers are primarily procedural (e.g. uncertainty and delays in information governance processes), and cultural (e.g. less familiarity and confidence in research involvement, and less precedent for data sharing), rather than technical: many of the technical aspects, such as use of trusted research environments, virtual private networks, and pseudonymisation are well-established. Enabling timely and accurate cross-sectoral linkage of data for research and evaluation will lead to benefits for patients, the healthcare system, the economy, and the population as a whole.

Short term gains can be achieved by utilizing and improving linkage to the UK's many well established cohorts and surveys, e.g. UK Biobank, Understanding Society, and the British Cohort studies. Longer term gains could be achieved by greater availability of and linkage to data from local and national government, which are particularly rich in information on the wider determinants of health and in opportunities for evaluating policy and service changes (for instance, through natural experimental approaches) (Craig, Katikireddi et al. 2017).

The use of quantitative data – particularly from administrative sources – must also be informed by stakeholder perspectives, public engagement, and triangulation with other sources of evidence (such as qualitative research), if it is to represent lived experience and avoid misinterpretation (Deeny and Steventon 2015). An example of this is described by the Independent Care Review on routine data on the experiences of children in care in Scotland, which described how 'snapshot' data fail to capture an individual's journey through the system and may not reflect what matters to children (Independent Care Review 2020). Similarly, the mismatch between healthcare access and need – epitomised by the inverse care law – means that relying solely on administrative data from healthcare and social care may introduce important biases, especially with regard to under-estimating need among disadvantaged populations (Watt 2013).

2. Governance and approvals processes require profound redesign if the benefits of health data research are to be realised, and the risks minimised

Information governance (IG) is rightly an important area for any use of data for health research. It is essential that risk assessments for projects address both the potential risks from poor information governance and the benefits to population health that the research is likely to deliver (or, put another way, the risks of *not* undertaking such research).

Our experience suggests that IG processes are a ubiquitous challenge in delivering timely and high-quality research with the potential to make a positive impact on health. Currently, such processes are often lengthy, circuitous, opaque, and inconsistent, in a way that creates duplication and wastage of time and effort and is not proportionate to the nature or degree of risk involved. Moreover, many processes have not been updated to reflect the large-scale shifts that have occurred as a result of the pandemic (e.g. widespread use of remote access to Trusted Research Environments), so do not reflect current practice.

Despite the re-use of data in the public interest being permitted under both common law and GDPR, and established precedent for some healthcare datasets (such as hospitalisations, and cancer registrations), the default approach is often a highly risk-averse one with a starting point of denying access.

This is particularly true for the kind of cross-sectoral, interdisciplinary research that is required to effectively respond to population-scale public health challenges (Academy of Medical Sciences 2016). This has greatly hampered such research and tends to trap funders and researchers into narrow approaches to health research which is individualistic and medicalised, in turn skewing the evidence base available to decision-makers (Rutter, Savona et al. 2017).

In particular, when undertaking more complicated research across multiple geographies (e.g., Health Authorities/Health Boards, Local Authorities, UK countries) and multiple sectors (e.g., health, education, social security) the processes of obtaining all of the necessary permissions and undertaking suitable information governance training that satisfies all data controllers is currently lengthy and burdensome to researchers. This gives rise to the risk that research either cannot take place at all or must use weaker designs, undermining the quality of the research and evaluation evidence produced.

For instance, researchers in our Institute have been working for over 10 years on a cross-sectoral linkage project, hindered by delays in differing approvals processes for healthcare data and data from other government departments, and by changes in the governance landscape which meant previous approvals were no longer seen as valid. As a result, no analysis has yet been able to take place, despite external funding that recognised that the research was of high quality and high potential impact.

Yet huge strides have been made during the Covid-19 pandemic to accelerate access to data in England (less so Scotland) whilst maintaining appropriate scrutiny and safeguards: there are important lessons here for health research beyond the pandemic (Cavallaro, Lugg-Widger et al. 2020, Royal Statistical Society 2021).

To address the issues described above whilst maintaining necessary and proportionate scrutiny and safeguards, we propose:

- Closer integration between approvals processes that focus on research importance and methodological appropriateness (e.g. peer review for funding, publication) and those that focus on appropriate research conduct and minimising harms to the public and participants (e.g. review for ethics and information governance) – with each element of the process having a clearly defined scope and remit, and the order of decision-making dovetailing to ensure necessary scrutiny and efficiency of review. Each of these elements are mutually relevant – for instance, the quality and impact of the research affects the risk-benefit decisions made by IG panels – so greater integration can ensure well-informed and efficient decision-making.
- This streamlined governance process should ideally have jurisdiction across each devolved nation, with delegated powers from across health geographies and local authorities, to facilitate use and linkage of data at the population-scale. While such bodies must be closely involved in governance, there also must be independent oversight to avoid conflicts of interest, as there are likely to be instances where data controllers may wish to avoid sharing data where research findings may contradict or undermine current practice or policy. It should have a single clear point of access for researchers, supported by specialist staff (as described in section 7), with a defined pathway and milestones so that expectations and accountability are clear.
- Governance panels should include data controllers, researchers, service users, members of the public, and other research users (e.g. decision-makers in local, devolved, and national government) and have a clear remit for making data work efficiently for the overall benefit of the population
- Clear ownership of and accountability for decision-making, avoiding 'attrition by indecision'
- Structured and transparent risk assessment and criteria for decision-making which are made available to the public and potential research users, and which set expectations as to what safeguards are likely to be required for different types of data and uses
    - For instance, Scottish Government's Data Access Panel (now subsumed into a different panel) previously published a standard decision-making matrix which was used to inform risk-benefit assessments and therefore decisions about whether and how projects should go ahead: this was helpful for researchers in setting expectations and understanding criteria by which applications would be judged.
- Appropriate emphasis on each of the 'Five Safes' (data; projects; people; settings; and outputs), rather than, for instance, focusing on purely technical safeguards at the expense of training and practical understanding and vigilance by data users
- A recognition that it is never possible to eliminate risk completely and that bureaucracy can paradoxically increase risk by making it harder to identify core governance procedures and by creating false security that undermines common-sense vigilance
- A solutions-focused approach which recognises that most information governance challenges are soluble with appropriate organisational, personnel, and technical safeguards. Once the project has been deemed of sufficient methodological quality and potential impact, the default approach should be to identify ways in which the research can take place within the necessary IG parameters.
- To maximise utilisation of existing datasets in new research and ensure that proposed projects are feasible, data controllers and curators should facilitate rapid, relatively low-threshold access to low-risk de-identified extracts or synthetic data that can be used for developing proposals and undertaking sample size calculations – prior to dedicated funding being obtained.

Any changes to the practice and policy for data access and sharing for health research should be based on best evidence and evaluated prospectively for their impacts and potential unintended consequences.

3. A strategic approach to data acquisition and curation should be adopted, matched to priority research areas

At present, the availability and ease of access of datasets relevant to health – especially administrative data from beyond healthcare – is somewhat piecemeal and ad hoc, determined more by precedent, historical accident, and divergent approaches to data retention, than by a strategic approach or consistent governance position.

Instead, the acquisition and curation of datasets should be aligned to national research priorities, with a diverse range of primary and secondary data resources made available on an ongoing basis to approved researchers with support from dedicated data management teams. As described in Section 2, these should be accessible via single point of entry with an integrated governance process combining multiple approvals across geographies and datasets.

This requires sustained support via long-term central funding which recognises their value as key assets to our research culture and potential for innovation, supports skill development in open science, and integrates public and service user involvement at each step of the process. One approach worth exploring would be a partnership model between local/devolved/national governments, NHS, and universities (as exist for Clinical Research Facilities).

Joint working within and across UK nations

The value of such approaches could be greatly enhanced by better joint working within and across the UK nations, to address the current fragmented landscape for health data. This has been highlighted as part of the Royal Statistical Society's lessons from the pandemic (Royal Statistical Society 2021) and has also been our experience – for instance, within Scotland where there are both local and Scotland-wide Safe Havens which differ in data availability and approach.

Ensuring consistent and comparable data within and across UK nations, and facilitating access processes, is essential if studies are to achieve sufficient internal validity (for instance, through the use of appropriate controls and study sizes that provide adequate statistical power) and external validity. This is particularly important in the context of devolution, where policy divergence creates opportunities for evaluating population-wide impacts of changes in the social and structural determinants of health – for instance through natural experimental methods (Katikireddi, Smith et al. 2016).

Criteria used in approvals processes need to move away from a focus on 'patients' and on direct impact within specific geographical jurisdictions to reflect that potential benefits of research may be incremental ones accrued across whole populations and that evidence generated in one jurisdiction may have positive impacts elsewhere. One example is that researchers in Scotland evaluating major policy changes have found it difficult to obtain data on hospitalisations from England as a counterfactual control setting, due to the legislative requirement to demonstrate benefit to health(care) and social care in England. The whole of the UK stands to benefit from rigorous evaluations of policies implemented in any of the UK

countries. An unduly lengthy and laborious process of obtaining permission to access comparator data risks undermining the quality of evaluations and evidence available for decision-making.

Data provenance and context

Critical to this curation process will be an understanding of data provenance and context. This is especially important as the use of secondary data from beyond the healthcare system increases, as we believe it must. Data relevant to health research span the gamut from high levels of rigour, geographical consistency, and researcher oversight (e.g. Census, and other ONS data); through smaller-scale cohort studies and surveys, and administrative data with some degree of quality control (e.g. hospitalisation records); to messy and poorly characterised administrative data with limited or no researcher oversight of collection.

Below are some examples from our research of how an understanding of context, selection into datasets, and data flows are critical to correct analysis and interpretation of health-relevant data:

- hospital episode statistics require an understanding of how someone comes to be admitted to hospital, including thresholds and referral pathways, and how these may vary over time, between geographical areas, or for different patient groups
- Covid-19 testing data from January 2020 onwards require an understanding of the availability of testing and the criteria for access at each point in time, with the population able to enter that dataset varying from symptomatic people in hospital with severe illness; symptomatic people in the community; asymptomatic health and social care staff, or care home residents; to asymptomatic people in the community. This requires clear documentation of the timing and sequencing of different testing regimes in order to understand potential bias threats to the research.
- local authority data on people accessing statutory homelessness services and their linkage to health data require an understanding not only of the criteria and thresholds for such services but how individuals experiencing homelessness do and do not access them, and therefore who is missing from the data

This is as much an epidemiological challenge as a technical one, requiring:

- close attention to the underlying processes by which records are generated and where data may be missing or misleading
- close liaison with staff, service users, and often policy-makers from the relevant sectors
- detailed, standardised, and easily available meta-data and documentation
- dedicated staff with the time, knowledge, and skillset to document and advise on dataset provenance, strength and limitations (see Section 7 on Career Structure)

4. Technical factors, including analytic platforms, trusted research environments, and data flows, are important but not the major barrier

Our experience has been that current technical infrastructure works relatively well, in comparison to other aspects of the system: while these standards should continue to be maintained and upgraded, the key blockers to good research lie elsewhere (as described in Sections 1-3). For instance, the ongoing challenges in Scotland in accessing primary care data

for research are primarily political and organisational, rather than technical – since the infrastructure and safeguards already exist.

Our experiences with Trusted Research Environments (TRE) such as regional and national Safe Havens have largely been positive, but their value could be optimised by:

- Greater resource investment to ensure timeliness of data extraction and linkage; disclosure control; and technical support where issues arise
- Integration of approvals and datasets across Safe Haven networks and national geographies. For instance:
    - local Safe Havens in Scotland may only have access to local hospitalisation data which can lead to biases relating to out-migration and care provision across regional boundaries
    - Ensuring that UK-wide resources can be linked throughout the UK – for instance, there are major gaps in the linkage of records from Scottish participants in UK Biobank and CPRD to other health datasets (such as hospitalisations)
    - researchers approved to work in one Safe Haven currently need to start from scratch when seeking to obtain approved researcher status for another Safe Haven, which creates additional delays and inefficiency
- More widespread use of synthetic datasets to enable preparatory work to be undertaken outside the TRE, to maximise efficient use of time and resources
- Measures to enhance compatibility with open science, such as secure links to code repositories
- Continuation of arrangements for remote access (e.g. from home) even once Covid-19 restrictions are eased, with appropriate security provisions. This maximises efficiency by minimising travel time to access centralised physical TREs and equity of access, as previously some researchers would have found it difficult to work in the limited number of physical TREs due to geographical location, caring responsibilities, disability etc.
- Ensuring that TRE have sufficient capacity to run complex analyses rapidly

With regard to data flows, we would strongly advocate for the use of ongoing secure relational data banks (such as SAIL in Wales), rather than models based on a ‘use and destroy’ approach (as currently applies to most data linkage research in Scotland). The former have a number of benefits which substantially enhance the value and efficiency of research:

- Efficiency – considerable time spent on extracting, linking, and cleaning the datasets at the outset of a project; for this investment to be wasted, as occurs in the ‘use and destroy’ model, is a waste of public funds and researcher time
- Reproducibility – ‘use and destroy’ models are poorly compatible with the growing emphasis on reproducibility, in turn risking undermining the internal and external validity of research findings
- Innovation – new methodologies can be applied to existing datasets to refine and compare answers to critical research questions
- Proportionality of governance – re-use of such datasets by approved researchers in secure environments without additional data transfer or linkage steps is relatively low risk but at present often requires starting lengthy governance processes from scratch
- Timeliness – timely access to service use data from healthcare, social care, and other public services should be available to allow for continuous outcome monitoring,

enabling rapid identification and action on problems as they arise, as well as facilitating responsive evaluation.

5. The public must play a central role in guiding the use of data for health research, at every stage of the process

Given that the most common legal basis for the use of data for health research is 'public benefit', the process must involve members of the public and users of services at every step of the process, with a particular focus on governance and dissemination. There is a growing body of evidence in this field about effective ways of engaging and sharing power (e.g.,Aitken, Tully et al. 2019).

Evidence suggests that the majority of the public – including among marginalised groups such as care-experienced children and people experiencing homelessness – not only approve of but *expect* the use of routine data for research with societal benefit (Aitken, de St. Jorre et al. 2016, Luchenski, Clint et al. 2017, Aitken, Tully et al. 2019, Independent Care Review 2020). Data sharing between public bodies and the private sector is more contentious and is generally an area of significant concern, due to lower levels of public trust and concern about the profit motive (Aitken, de St. Jorre et al. 2016). There is currently a potential window of opportunity created by increased public interest in data during the pandemic, which should be harnessed as a route for greater engagement and dialogue.

We note an implicit assumption within the terms of reference that public sector data should be being made more readily available to the private sector. We recognise the potential value of involvement of third sector and private organisations in this area – for instance, through commissioning of specific technical skills – but would highlight that evidence shows this is consistently the area of greatest public concern, despite high levels of acceptability for data linkage within the public sector (Aitken, de St. Jorre et al. 2016). We therefore urge caution in this regard, and extensive public involvement before any such plans are progressed. As previous incidents have demonstrated, public trust is (justifiably) hard to earn but easy to destroy – there is a risk that poorly thought-out data sharing initiatives with the private sector threaten other forms of research for which there is currently a stronger public mandate.

Consideration should also be given to appropriate processes and safeguards to enable private sector data to be used and shared for public benefit – too often, this discussion is framed in terms of a one-way flow from the public to private sector, when the latter holds a wealth of data relevant to health. For instance, we have previously demonstrated the value of commercial data from Experian – a marketing firm – in enhancing existing routine measures of socioeconomic position (Wami, Dundas et al. 2019). Examples relevant to our research include retail data (on health-promoting and health-harming commodities such as foodstuffs, tobacco, and alcohol) as well as data from wearable devices and social media.

6. Greater value must be placed on open science in approvals processes and reward structures for research

Open science is important to allow for adequate scrutiny of methods, provenance of data, code and model checking to ensure the results of any analysis are robust. A high value should be placed on the sharing of data provenance (ie detailing the meta-data/datasets and variables – not necessarily the actual data); code to create derived variables; and code to run

analysis. Expectations and incentives set by funding bodies and journals are crucial in shaping research culture around issues such as open science and will therefore be critical to this endeavour.

This could be achieved by:

- Governance bodies making code-sharing a condition of use of the data at the point of approval
- Journals making code-sharing a condition of publication
- Incentivising open science and good data management practice through:
    - Individual-level academic and NHS recognition and reward schemes (e.g. criteria for promotion)
    - Institutional-level recognition schemes, akin to the Athena Swan awards, which become pre-conditions of research funding and badges of prestige.

Most researchers have had little formal training in writing and documenting reproducible code and may lack confidence doing so. Provision of training, guidelines and templates for producing reproducible code will ensure that researchers have more confidence in sharing code and any code that is shared will be more interpretable. Greater use of pair coding – widely used in software development – may be valuable as a means of reducing errors; increasing replicability; and supporting rapid skills development.

Improving job security will also reduce barriers to sharing code: at present, such activities are often deferred to the end of projects when the analysts responsible for creating the code may no longer be in post, and access to the necessary files – or tacit knowledge – becomes more complex. There is a time commitment associated with sharing code and data, which should be costed into funding proposals and incorporated into project planning from the outset.

We note that some aspects of open science may be more challenging to implement where there is private sector involvement, given concerns about intellectual property and commercial sensitivity – for instance, where a company has developed a specific algorithm which they wish to be proprietary. Any mechanism for data sharing with for-profit entities must (a) transparently address potential conflicts between profit motives and the open science agenda and (b) ensure that costs for involvement reflect potential returns for the company.

7. Staffing models should recognise and reward the specialist, skilled work of data management and access

The process of curating, managing, and negotiating access to datasets is a skilled and specialist one, requiring dedicated roles. At present, many aspects of this process are undertaken by individual (often junior) researchers on an ad-hoc basis, which is inefficient, duplicates effort, and results in a lack of institutional memory.

Instead, the infrastructure for secondary data utilisation and linkage should include dedicated staff resource for supporting researchers to scope; plan; obtain permissions; negotiate access; and undertake research with core datasets. These individuals should be valued as a critical enabler of research, with clear career pathways and appropriate recognition (for instance, being named on research papers). Our experience working with such co-ordinators – for instance, at the Urban Big Data Centre at the University of Glasgow – has convinced us of their value.

8. Our experiences highlight both successes and challenges

Successes

- The Urban Big Data Centre at the University of Glasgow has previously operated a streamlined, single-point-of-entry process for access to data, which integrates advice on dataset choice and availability; application for all necessary approvals; and brokerage between data controllers and researchers to obtain permissions, sign agreements, and transfer data
- Our experiences working with the SAIL (Secure Anonymised Information Linkage) databank in Wales have been positive. SAIL provides access to a range of health-relevant, population-scale datasets via an ongoing secure relational linkage, rather than ad-hoc linkages undertaken for individual projects and subsequently discarded. It offers a timely and relatively low-cost service, with high levels of security, and includes both primary data (from trials, surveys, and cohort studies) with secondary data (from administrative sources within and beyond healthcare). It is an exemplar model from which the rest of the UK has a great deal to learn.
- Another smaller-scale example of a system which benefits from extensive co-ordinator support, proportionate governance processes, timely approvals and provision of data, and close public involvement is the CRIS system from the South London and Maudsley NHS Foundation Trust (SLaM) – albeit with a primary focus on mental health
- We have also worked extensively with UK Biobank and would commend their open access, governance and approvals processes, with data made available securely to approved researchers.
- Our researchers have benefited from training and consultation services on governance procedures and legislation from MRC Regulatory Support Centre and University of Glasgow Data Protection Office – especially at times of change, such as the introduction of GDPR
- We find that there is enormous enthusiasm for better use of data from across local, devolved, and national government; the third sector; the NHS; and social care – which has only been enhanced by the Covid-19 pandemic. We must ensure we harness this enthusiasm effectively by timely, efficient, and proportionate governance processes; high-quality research; and close partnership working with professional stakeholders and the public.
- Where the co-ordinator role described in section 7 has been available, it has been beneficial – however, under-resourcing means that such roles have often been spread too thinly, resulting in the researchers ending up taking on the bulk of the work in navigating approvals and negotiating access themselves.

Challenges

- Researchers cannot usually apply for data until funding is obtained: given that the approvals process is almost always lengthy and involved, this is challenging given the short-term nature of most funding and the expectation on the part of funders that projects begin immediately after funding is approved
- Researchers cannot usually apply for data until funding is obtained: given that the approvals process is almost always lengthy and involved, this is challenging given the

short-term nature of most funding and the expectation on the part of funders that projects begin immediately after funding is approved

- Delays and difficulties in approvals; contracting; and data extraction and linkage are endemic, and have the following negative impacts:
    - Barriers to methodological innovation and empirical discovery because projects using secondary data are perceived as high-risk by funders and researchers; this is particularly true for data from outside the healthcare system
    - Persistence of major gaps in the evidence base and perpetuation of health inequalities because of particular difficulties achieving approvals for projects which involve population groups or health conditions perceived as higher risk (such as children, mental health, people involved in justice system) or data outside the healthcare system
    - Potential for negative impacts on methodological quality, due to difficulties in accessing data that would enhance rigour and transferability: one example of this from our experience is in the use of geographical controls from across the UK as part of natural experiments of policy change
    - Missed opportunities to evaluate time-limited interventions or exposures, missed 'policy windows' in which research can achieve influence and impact, and the risk of research becoming irrelevant by the time it is completed
    - Negative career impacts on those choosing to work with administrative data, due to the opportunity cost of time spent navigating lengthy, involved, and often opaque governance processes, which often:
        - Threatens timely completion of projects, often resulting in scaling-back or abandonment
        - Reduces people's research outputs, which in turn affects their career prospects given current incentives structure in academia
        - Does not meaningfully develop an individual's skills or experience in planning, conducting, disseminating, or leading research
        - Dissuades others from pursuing research using similar methods (particularly more innovative and interdisciplinary approaches)
- Lack of integration across the multiple health geographies within the UK reduces the value of existing resources: for instance, UK Biobank data is not fully linked to Scottish hospitalisation data. Similarly, while CPRD covers the entire UK, linkage to other health records within Scotland is not currently available
- Access to primary care data is an ongoing challenge despite primary care representing the majority of healthcare contacts in the UK; this must therefore be addressed as a priority. This has been a particular challenge in Scotland, even during the pandemic, despite evident lawful basis 'in the public interest'. There have been instances where research participants have consented for their primary care records to be accessed – for instance, as part of UK Biobank – yet there have been attempts to overrule this at the individual practice level. Similar standards to those in place for secondary healthcare data should apply to primary healthcare data: that is, that lawful uses of data in the public interest, where expectations are managed through best endeavours at transparency (e.g. information provision and the right to object), and safeguards are in place to minimise disclosure risk (e.g. use of Trusted Research Environments, pseudonymisation, and data minimisation), can and should be facilitated at a national level. Incentivisation for data sharing may be appropriate here.

References

Academy of Medical Sciences (2016). Improving the health of the public by 2040: optimising the research environment for a fairer, healthier future. https://acmedsci.ac.uk/policy/policy-projects/health-of-the-public-in-2040

Aitken, M., J. de St. Jorre, C. Pagliari, R. Jepson and S. Cunningham-Burley (2016). "Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies." BMC Medical Ethics 17(1): 73.

Aitken, M., M. P. Tully, C. Porteous, S. Denegri, S. Cunningham-Burley, N. Banner, C. Black, M. Burgess, L. Cross and J. van Delden (2019). "Consensus statement on public involvement and engagement with data-intensive health research." International Journal of Population Data Science 4(1).

Aldridge, R. W., D. Lewer, S. V. Katikireddi, R. Mathur, N. Pathak, R. Burns, E. B. Fragaszy, A. M. Johnson, D. Devakumar, I. Abubakar and A. Hayward (2020). "Black, Asian and Minority Ethnic groups in England are at increased risk of death from COVID-19: indirect standardisation of NHS mortality data." Wellcome Open Research 5(88).

Brown, D., M. Allik, R. Dundas and A. H. Leyland (2019). "All-cause and cause-specific mortality in Scotland 1981–2011 by age, sex and deprivation: a population-based study." European Journal of Public Health 29(4): 647-655.

Cavallaro, F., F. Lugg-Widger, R. Cannings-John and K. Harron (2020). "Reducing barriers to data access for research in the public interest—lessons from covid-19." BMJ Opinion https://blogs.bmj.com/bmj/2020/07/06/reducing-barriers-to-data-access-for-research-in-the-public-interest-lessons-from-covid-19/

Craig, P., S. V. Katikireddi, A. Leyland and F. Popham (2017). "Natural Experiments: An Overview of Methods, Approaches, and Contributions to Public Health Intervention Research." Annual Review of Public Health 38(1): 39-56.

Deeny, S. R. and A. Steventon (2015). "Making sense of the shadows: priorities for creating a learning healthcare system based on routinely collected data." BMJ Qual Saf 24(8): 505-515.

Douglas, M., S. V. Katikireddi, M. Taulbut, M. McKee and G. McCartney (2020). "Mitigating the wider health effects of covid-19 pandemic response." BMJ 369: m1557.

Hernán, M. A. and J. M. Robins (2016). "Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available." Am J Epidemiol 183(8): 758-764.

Independent Care Review (2020). The Promise. https://www.carereview.scot/conclusions/independent-care-review-reports/

Katikireddi, S. V., K. E. Smith, D. Stuckler and M. McKee (2016). "Devolution of power, revolution in public health?" Journal of Public Health 39(2): 241-247.

Luchenski, S., S. Clint, R. Aldridge, A. Hayward, N. Maguire, A. Story and N. Hewett (2017). "Involving People with Lived Experience of Homelessness in Electronic Health Records Research." International Journal for Population Data Science 1.

Marmot, M. G., J. Allen, T. Boyce, P. Goldblatt and J. Morrison (2020). Health equity in England: The Marmot Review 10 years on. London, Institute of Health Equity.

Marmot, M. G., J. Allen, P. Goldblatt, E. Herd and J. Morrison (2020). Build Back Fairer: The COVID-19 Marmot Review. The Pandemic, Socioeconomic and Health Inequalities in England. London, Institute of Health Equity.

McCartney, G., A. Leyland, D. Walsh and D. Ruth (2021). "Scaling COVID-19 against inequalities: should the policy response consistently match the mortality challenge?" Journal of Epidemiology and Community Health 75(4): 315.

Mutambudzi, M., C. Niedwiedz, E. B. Macdonald, A. Leyland, F. Mair, J. Anderson, C. Celis-Morales, J. Cleland, J. Forbes, J. Gill, C. Hastie, F. Ho, B. Jani, D. F. Mackay, B. Nicholl, C. O'Donnell, N. Sattar, P. Welsh, J. P. Pell, S. V. Katikireddi and E. Demou (2020). "Occupation and risk of severe COVID-19: prospective cohort study of 120 075 UK Biobank participants." Occup Environ Med.

Niedzwiedz, C. L., C. A. O'Donnell, B. D. Jani, E. Demou, F. K. Ho, C. Celis-Morales, B. I. Nicholl, F. S. Mair, P. Welsh, N. Sattar, J. P. Pell and S. V. Katikireddi (2020). "Ethnic and socioeconomic differences in SARS-CoV-2 infection: prospective cohort study using UK Biobank." BMC Medicine 18(1): 160.

Royal Statistical Society (2021). Statistics, Data and Covid: Ten statistical lessons the government can learn from the past year. London, Royal Statistical Society.

Rutter, H., N. Savona, K. Glonti, J. Bibby, S. Cummins, D. T. Finegood, F. Greaves, L. Harper, P. Hawe, L. Moore, M. Petticrew, E. Rehfuess, A. Shiell, J. Thomas and M. White (2017). "The need for a complex systems model of evidence for public health." The Lancet 390(10112): 2602-2604.

Walsh, D., G. McCartney, J. Minton, J. Parkinson, D. Shipton and B. Whyte (2020). "Changing mortality trends in countries and cities of the UK: a population-based trend analysis." BMJ Open 10(11): e038135.

Wami, W. M., R. Dundas, O. R. Molaodi, M. Tranter, A. H. Leyland and S. V. Katikireddi (2019). "Assessing the potential utility of commercial 'big data' for health research: Enhancing small-area deprivation measures with Experian™ Mosaic groups." Health & Place 57: 238-246.

Watt, G. (2013). "What can the NHS do to prevent and reduce health inequalities?" British Journal of General Practice 63(614): 494.

When was the response submitted?

17 March 2021

Find out more about our research in this area

https://www.gla.ac.uk/researchinstitutes/healthwellbeing/research/

Who to contact about this response

Emily Tweed - Emily.Tweed@glasgow.ac.uk