

ADR UK response to the Sudlow Review



ADR UK response to the Sudlow Review – Unifying Health Data in the UK

July 2023

1. What are your views on the use of health data in research and healthcare as it stands today?

ADR UK is an ESRC programme transforming the way researchers access the wealth of UK public sector administrative data, to inform policy and improve lives. ADR UK funds projects that link administrative data from across different government departments and devolved administrations to enable cross-cutting research for the public good. The legal basis for much of this work in England is the Digital Economy Act (DEA) which excludes access to [Data] “relating to the provision of health services or adult social care”. Data outside the scope of the DEA can be shared under different legal gateways, and additional approvals may be required. We believe there is great potential to undertake research using linked administrative and health data.

The [Marmot Review](https://www.instituteofhealthequity.org/resources-reports/marmot-review-10-years-on) of health (<https://www.instituteofhealthequity.org/resources-reports/marmot-review-10-years-on>) outlines the links between deprivation and lower life expectancy. By linking health data to administrative data, for example education data, data on household and living conditions from the census, or benefits and income data from HMRC and DWP, researchers are able to track these factors and evaluate social policy interventions on health outcomes.

In the devolved administrations, the ADR UK partnerships for Scotland, Wales and Northern Ireland work within a range of different legal gateways that allow the linkage of administrative data to health data, and this has led to much policy-relevant research. Details of those gateways are available on ADR UK's website:

https://www.adruk.org/fileadmin/uploads/adruk/Documents/The_legal_framework_for_accessing_data_April_2023.pdf.

During the Covid-19 pandemic, the Secretary of State for Health and Social Care issued NHS Digital with a Notice under Regulation 3(4) of the Health Service (Control of Patient Information) Regulations 2002 (COPI), which allowed a range of Covid-related health data for England to be made accessible to researchers via the ONS Secure Research Service. This included data from sources such as NHS Test and Trace data, the Covid Infection Survey linked to Test and Trace data and NHS Vaccination Status to name a few. Additionally, via the Statistics and Registration Service Act, the Public Health Research Database (Linked Census/Mortality/Hospital Episodes Statistics/GP Extraction Service Data) was created to inform policy during the pandemic. It was clear that the pandemic created an opportunity for NHS Digital to reassess their risk appetite for opening up access to large-scale health datasets, and linked health and administrative datasets, but this has not been done at the pace or scale of opening up access to non-health administrative datasets.

2. What are the barriers to uniting health data in the UK?

Health data is excluded from the provisions in the Digital Economy Act that relate to providing access to administrative data for research. This means that for linked datasets that include health and other administrative data, the Digital Economy Act must be used in conjunction with the Health and Social Care Act (HSCA) 2012, which states that the dissemination of data must be for the purposes of the provision of health care or adult social care, or for the promotion of health. A further and significant difference is that access to health data must be covered by individual data

sharing agreements between researchers and NHS Digital, and when these expire all data must be deleted. In practice, this means that the governance processes are not in place to support NHS England making decisions around the creation of curated, linked, de-identified datasets, and then deciding how to open these datasets up for wider research use.

Between 2019 and 2022, ADR UK funded UCL and NHS Digital to create the ECHILD dataset, linking 17 years of National Pupil Database (NPD) data to Hospital Episode Statistics data to create an enduring research asset that would be made available to wider researchers via the ONS Secure Research Service. This dataset is now being used to track the education outcomes of children with mental and physical impairments, to understand how these have been impacted by Covid-19 restrictions.

Work to agree the governance process to provide access to this linked dataset for wider research use was taken forward by a Task and Finish Group set up by NHS Digital in 2021, as part of the original ADR UK funded ECHILD project. The group agreed to adopt a two-stage approach, whereby an interim instance of ECHILD would be established via a DfE Data Sharing Agreement with NHS Digital under S.251 of the HSCA (2012), and an enduring instance will be established via a new NHS Digital Direction under S.254 of the HSCA (2012).

While progress on the first stage is being made, with an agreed process to open up wider access to the ECHILD dataset anticipated to be launched imminently, progress on the second stage has stalled. This is because NHS England is not convinced that the direct healthcare benefits gained from creating this as a standalone enduring asset would pass the threshold required to obtain a Direction from the DHSC Minister to support the work. Additionally, they do not consider they have the remit to make a decision on the creation of a standalone enduring asset that included health data, but which could result in wider benefits unrelated to healthcare (for example, benefits to education). As such, although ADR UK is going to be able to open up wider research access to ECHILD, the process to get this agreed has been extremely costly in terms of both time and effort, compared with similar decisions related to administrative data linkages. Also, we have yet to test how robust the process that resulted in the ECHILD team gaining the necessary approvals to open up access to this data is, with respect to decisions being made in relation to other related health and administrative data linkages.

ADR UK have also funded work to link diabetes audit data to education (<https://www.adruk.org/our-work/browse-all-projects/the-personal-cost-of-health-conditions-in-childhood-349/>) research on which will help to address how disease-specific measures of health conditions might affect school attendance, university participation and educational outcomes. This project is in effect in the queue behind ECHILD to find a sustainable solution to make the linked data available for wider research. We have two other projects which will need to adopt the ECHILD model of sub-licensing if a legal gateway is not established. The Kids' Environment and Health Cohort is linking a number of environmental datasets about small areas across England, on air pollution, energy efficiency of buildings, and proximity to major roads, to de-identified health and education data to allow researchers to examine how local physical and social environments influence children's health and schooling across England (<https://www.adruk.org/our-work/browse-all-projects/kids-environment-and-health-cohort/>). And the Linked local data on children and young people project aims to create a research-ready dataset linking data held on children by local authorities, including education and social care, with

data held by health services to enable researchers to build evidence to support local and national strategies that improve outcomes for children (<https://www.adruk.org/our-work/browse-all-projects/linked-local-data-on-children-and-young-people/>).

3. What are the solutions for unlocking these barriers and realising the potential of health data in the UK?

We believe that work should continue on securing a legal basis for creating enduring research assets that link health and administrative data in England. The approach in Wales has shown the value in making linkable health and administrative data available to researchers, where The Welsh Government has provided for the processing of health data through the Digital Government (Welsh Bodies) Regulations 2018. Examples of what can be achieved are available at ADR UK's website <https://www.adruk.org/news-publications/publications-reports>. A particularly valuable insight which demonstrates the value of the Welsh model is illustrated in this publication <https://saildatabank.com/groundbreaking-research-examines-how-the-sequence-of-disease-acquisition-affects-life-expectancy/>.

Members of the Scottish Centre for Administrative Data Research, part of ADR Scotland, have been using health data for several years to develop innovative and policy relevant research in a number of areas of core governmental importance.

During the public health emergency in March 2020, the Scottish Government set up a Covid19 Data Taskforce to support operational decision-making and evidence-based policy, led by ADR Scotland Co-Director. This work continued by Scottish Government's Covid-19 Data and Intelligence Network (www.gov.scot/groups/data-and-intelligence-network) which enabled collaborative research and analysis for Covid-19 decision-making and understanding using data. The Covid-19 data holding created an environment for key datasets to be linked safely and quickly. By May 2021, Scotland had a detailed data catalogue with 29 datasets allowing data linkage projects to happen quickly, with 72 approved research projects being supported by ADR Scotland within the first 12 months of the pandemic.

ADR Scotland researchers pivoted their research to respond to policy-critical areas namely around community based Covid-19 mortality (www.scadr.ac.uk/our-research/health-and-social-care/community-based-covid-19-mortality), policing the pandemic (www.scadr.ac.uk/our-research/community-safety-equality-and-wellbeing/covid-19-policing-pandemic) and a commissioned project by Scottish Government on risk factors for Covid-19 (www.scadr.ac.uk/our-research/health-and-social-care/investigating-socioeconomic-household-and-environmental-risk). Our researchers also developed a new household linkage variable – the CURL tool (www.scadr.ac.uk/news-and-events/news-innovative-new-residential-linkage-tool-launched) – which helped enhance understanding of household transmission during the pandemic. All these projects and research, highlight the benefit of having quick access to several datasets and also the value of administrative data research linking to large-scale health datasets to answer pressing policy questions and in the longer term, enabling us to explore why post pandemic there is still a large increase in numbers of people dying at home (www.scadr.ac.uk/our-research/health-and-social-care/deaths-home-during-covid-19-pandemic-scotland).

In addition to the above research, data from ambulance call outs, accident and emergency visits, and hospital admissions have been used to examine public health policing issues, such as patterns of repeat violent victimization (www.scadr.ac.uk/our-research/community-safety-equality-and-wellbeing/violence-and-vulnerability) and the role of alcohol availability and deprivation in increasing rates of violence (including weapon-related injury). Moreover, individual-level data from Police Scotland and Public Health Scotland have recently been linked to examine patterns of non-compliance with the Covid-19 Public Health Regulations in an effort to inform the UK and Scottish Covid-19 Inquiries around the impact of police enforcement strategies during lockdown (www.scadr.ac.uk/sites/default/files/Fourth%20FPN%20Data%20report%20FINAL%20-%20Aug%202022.pdf).

4. What types of data should be prioritised now to be made available to support research and healthcare?

Across the ADR UK programme, we are committed to advancing the utilisation of linked administrative data sources to unlock wider academic research and enable new discoveries that have the potential to lead to better informed policy decisions and more effective public services. Cancer Research UK recently launched their Cancer Data Driven Detection (CD3) programme, with the ambition of revolutionising our understanding of cancer risk using a huge variety of data to advance early detection of the disease. ADR UK strongly agree with the concept of CD3 and believe there is an opportunity to develop methodologies and facilitate research utilising linked health and administrative sources to improve the prediction of cancer risk and detection. These are major public health challenges, for which – until now – there has been a scarcity of relevant, linked data available to researchers. Our partners in ADR Wales (who manage the SAIL Databank) now hold extensive health, administrative and environmental data linked for the population of Wales and are assisting in forwarding CR UK's CD3 mission, to test what is now feasible in this area of research. However, inevitably the insights that can come from the SAIL Databank related to the population of Wales (around 1.3 million people) will be limited in the context of rarer cancers and risk factors, compared to what would be possible if comparable data were available for researchers to access for the population of England (around 56.5 million people):

ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates).

The academic consultation workshops carried out by Cancer Research UK in early 2022, which members of ADR UK and our economic and social research community attended and participated in, highlighted the clear potential for an initiative like CD3 to capitalise on the skillsets and expertise of a broad range of scientific fields by connecting around a single healthcare challenge. Such an initiative would, I am confident, elicit a wealth of policy-relevant research insights, not only in terms of cancer research specifically, but also from the perspective of supporting research into inequalities, diversity and inclusion, through the linkage and utilisation of administrative data to cancer registration (and other health) data.