

## Annex 1. Methods and search strategy

We divided the work required for landscaping the world for longitudinal datasets into five staggered stages. All stages overlapped with one another once started and continued until the end of the project.

The breadth of our searches is exemplified in the early stages of our work. The depth of our analyses is illustrated in the later stages.

### Stage 1 – Searching for datasets and identifying longitudinal datasets

The task of searching for and identifying datasets was split between the team based at KCL who focused on datasets in academia and the team based at ODI who focused on datasets outside academia (e.g., industry, government, charities).

#### *Datasets from academia*

The KCL team adopted two different approaches to identifying datasets and repositories of information about longitudinal datasets (e.g., websites, online search platforms and academic journals). Those approaches were described as active search and passive search (**Figure I**). The two approaches were not equally successful, with the active search being the most productive strategy by far (**Figure III**).

- **Active search:** We searched through repositories containing information about longitudinal datasets including (1) reference lists and bibliographies of papers reporting longitudinal findings; (2) academic journals focused on longitudinal studies; (3) university-affiliated microdata catalogues such as the Harvard Dataverse; (4) academic databases such as ProQuest using key terms such as 'longitudinal study', 'cohort study'; and (5) national government websites, non-governmental agencies, or global and regional consortia. Repositories were an especially valuable source of information about datasets. We scanned through 200 repositories which provided information for a total of 7,985 datasets worldwide.
- **Passive search:** We received information about longitudinal datasets from individuals across the world using our outreach and networks in the mental health and epidemiology communities. We disseminated the project and our aims on social media platforms (i.e., Twitter and LinkedIn), via email and through presentations during which we invited people to contact us directly or by filling an online form on the landscaping website ([landscaping-longitudinal-research.com](http://landscaping-longitudinal-research.com) - more details below). ODI and MQ amplified those outreach activities and disseminated the project to their respective professional networks across the world. We received a total of 207 submissions via our website form and emails, mostly from the UK (41%) but also from the USA (14%) and Ireland (12%). Our outreach was global as we also received submissions from Zimbabwe (N = 1), Chile (N = 4) and Indonesia (N = 1). Those submissions generated 82 datasets and 15 repositories, whilst the remaining 109 submissions were duplicates of prior submissions. A further 70 submissions were sent to team members through various methods, for example via Twitter and Microsoft Teams, which generated another 45 datasets. Overall, we received 259 submissions via our passive search.

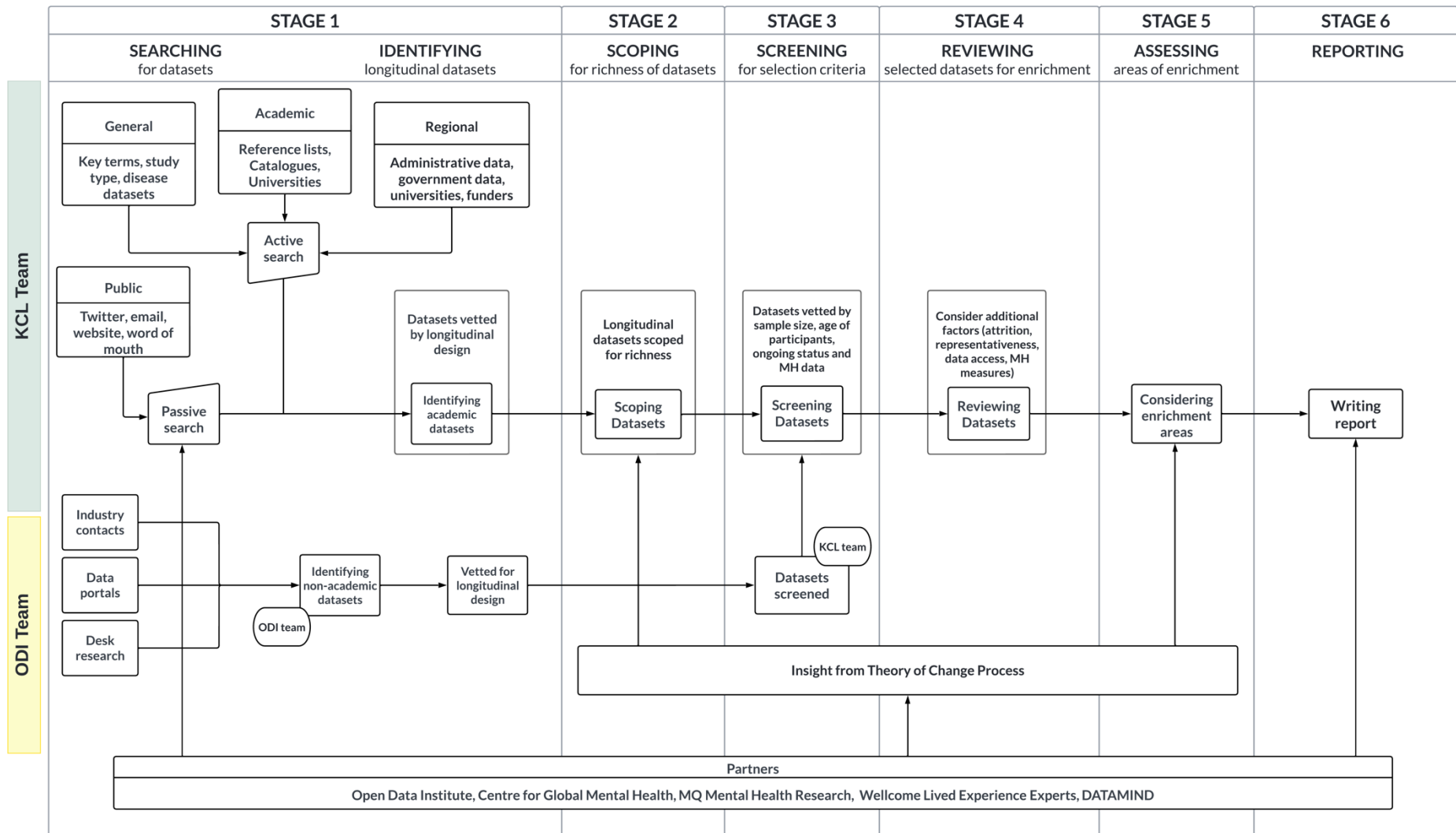
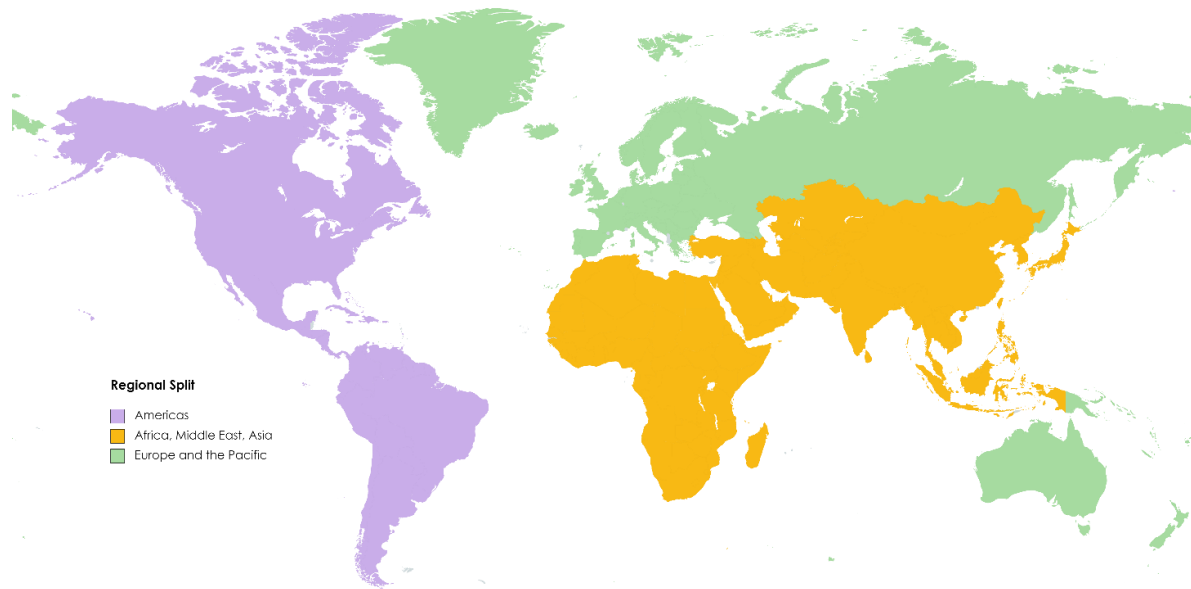


Figure I. Flowchart describing the six stages for landscaping datasets worldwide

We divided the world into three geographical regions for the purpose of this global search (**Figure II**). Based on the location of its participants at inception, each dataset was allocated to one of three regions: 1) North, Central and South America (encompassing 36 countries); 2) Africa, Middle East and Asia (including 104 countries); 3) Europe and the Pacific (covering 57 countries). Three KCL-based research assistants were responsible for screening all datasets for one region each.



**Figure II. Regional breakdown for landscaping datasets worldwide**

*Note: Our use of the term 'the Middle East' is to denote countries geographically located between Africa and Asia. However, we recognise that this term has political connotations that are not relevant to this project.*

These geographical regions are arbitrary but meaningful for the work we conducted. For example, we anticipated having to invest considerable efforts with discovering datasets in specific areas such as Africa and the Middle East where datasets would not be disseminated as widely as those from European countries. We also expected challenges with uncovering datasets from countries where English is not the main spoken language, such as in Asia and South America.

We considered datasets that spanned multiple regions as a separate group (N = 194). Once identified, all datasets were vetted for longitudinal design and moved to the next stage for further examination. We excluded at this stage clinical trials and cross-sectional studies.

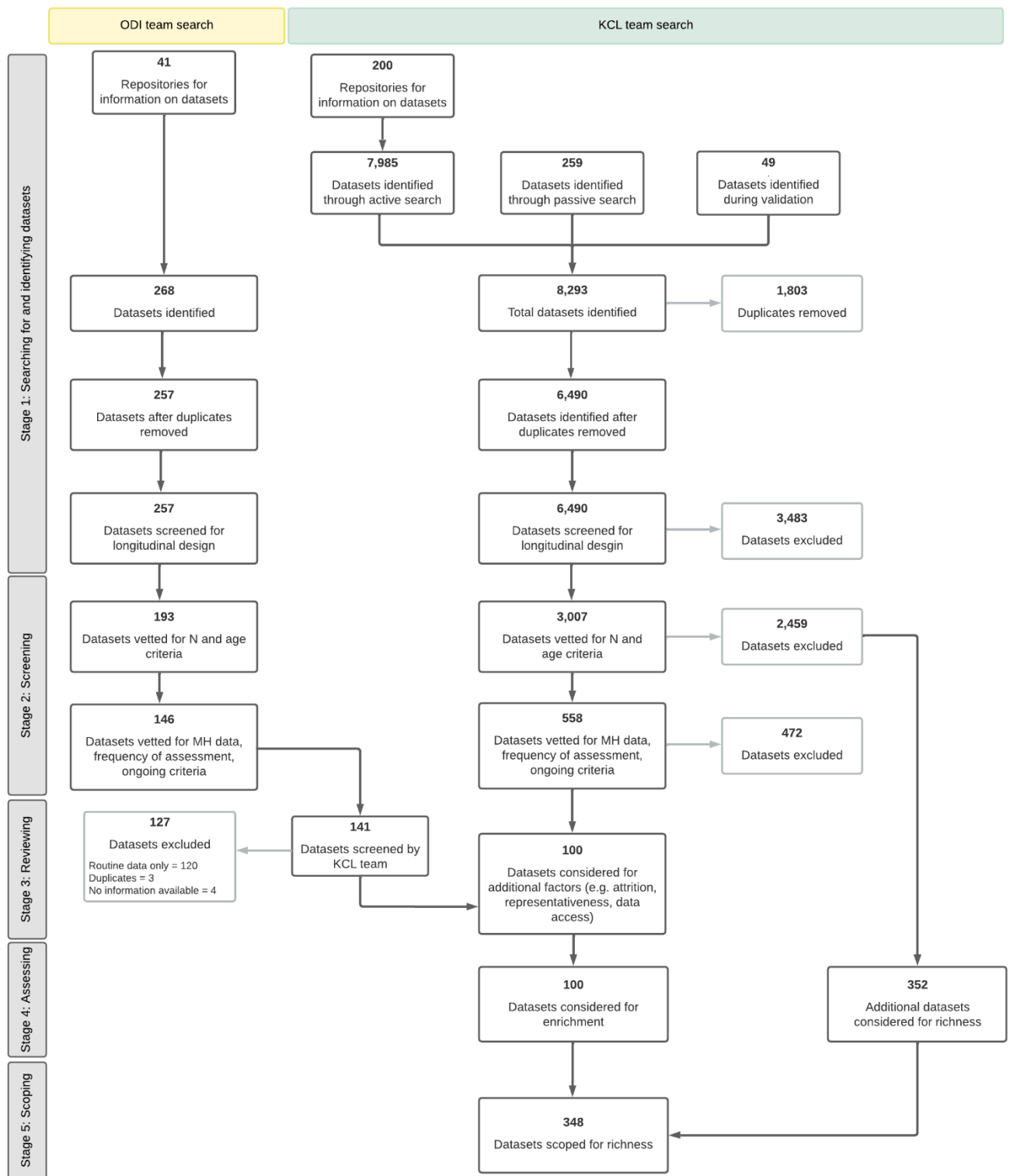


Figure III. PRISMA flowchart reporting the number of datasets processed through the landscaping stages.

### *Stage 1 findings*

The KCL team identified a total of 3,007 longitudinal datasets with repeated assessments on groups of individuals. We considered different study designs, as long as the datasets provided repeated assessments over time on the same group of people. These included household panel surveys and registries of routinely collected data.

Datasets (of any type) were identified in all but a single country around the world; North Korea. Researching North Korea is difficult because the country's political isolation demands no disclosure of healthcare conditions internally or externally, creating a scarcity of data sources (Ro et al., 2019). To combat this, many researchers often investigate North Korean defectors who have settled in South Korea and other parts of the world (Ro et al., 2019). These, however, were considered as datasets from the countries of settlement.

### Stage 2 – Scoping for richness in longitudinal datasets

As part of the process for reviewing longitudinal datasets worldwide, we identified richness in several datasets including datasets that did not meet Wellcome's selection criteria. To account for this, and in addition identify areas of enrichment for longitudinal datasets as requested by Wellcome, we highlighted richness in existing longitudinal datasets.

We captured richness in datasets within 'pockets of value' that were determined by (a) our experience of screening and reviewing more than 3,000 longitudinal datasets from around the world; b) the outcomes of the ToC process capturing suggestions from different stakeholders, and particularly LEEs; and (c) Wellcome's vision for transformative mental health research.

To reduce the influence of individual biases when screening for richness in datasets, we enacted a 3-step process. First, all three KCL-based research assistants were assigned one geographical region to screen for all the pockets of value. At the end of this point, approximately 50% of the datasets in each region were screened out. Second, the research assistants swapped regions and were assigned to a region they had not previously screened in previous stages. The research assistants then screened through the shortlist of datasets in each pocket created by the first step. Research assistants were free to add datasets they thought were missing from each list. At the end of that second step, approximately 15% of the pool of about 3,068 longitudinal datasets remained as examples of the pockets of value. Third, each research assistant 'holistically' screened pocket of values, regardless of the region, to check the remaining datasets were a good fit. This final step reduced the number of datasets to 326, approximately 10.6% of all the longitudinal datasets.

### Stage 3 – Screening longitudinal datasets according to selection criteria

The screening stage was an objective process, based on criteria proposed by Wellcome. To be considered further, longitudinal datasets had to meet all four criteria:

- Consist of at least 8,000 participants at inception.
- Include, or have potential to include, data on participants at some point between the ages of 14 and 30, as this is the age when mental health conditions often are first identified.
- Has collected some data annually, or at least every 3 years, and/or has the ability to contact participants to invite more intense data collection.
- Have collected data in the last three years, have plans to collect new data or are in contact with participants.

Throughout the project, we conducted an extensive validation exercise to ensure our search left no stone unturned. Our validation strategy, carried out by the KCL team, was three-pronged.

First, we considered datasets that appeared more than once on our spreadsheets as a way of validating our methods. Duplicated datasets indicated that our search pool was getting smaller as we continuously encountered the same datasets. In total, we identified 1,801 duplicate datasets during the first stage of our global search.

Second, we conducted desk validation using focused searching. We looked for unfamiliar datasets through cohort profile repositories in journals such as PLoS ONE and BMJ Open. We utilised key terms including 'cohort', 'longitudinal' and 'household panel' on search engines such as Bing, Google Datasets and Scopus. We also used country, region and disease-focused consortia to search more narrowly in countries where we had not previously identified any datasets.

Third, we contacted longitudinal research experts and asked them to screen our lists of datasets from specific regions for any missing datasets. Those 24 experts represented all regions of the world and are members of various institutions. We divided our list of longitudinal datasets into 14 sub-regions: Africa, Asia, Australia, Brazil, Canada, Central and South America, Ireland, the Middle East, New Zealand, Scandinavia, the rest of Europe and the USA. We sent lists to these experts and asked them to check whether we missed any datasets from their regions. We also asked the experts to highlight any datasets they knew of that were upcoming or in the first wave of data collection. From this validation process, 49 additional datasets were further identified for consideration.

#### Stage 4 – Reviewing selected longitudinal datasets for enrichment

We reviewed in more depth all longitudinal datasets from academia or other sectors that met the criteria listed above to assess their value and potential areas of enrichment (N = 100). We gathered and reviewed detailed information using resources available online. At this stage, we adopted a holistic approach to evaluate all remaining datasets. This approach was more subjective than the previous two stages. Information provided for each dataset was considered by the PI of this project, along with the KCL-based research assistant who identified and reviewed it.

Information we considered at this stage was related to:

- Sample characteristics, attrition and recruitment methods.
- Data collection and sources of data in the dataset.
- Details of mental health data in the dataset, including which conditions, reporters and instruments.
- Other types of data collected, such as demographics; mood, emotion & personality; activity; life experiences; relationships; and biological, environmental, cognitive and neuroimaging data.
- Data discoverability and accessibility.
- Involvement of LE experts.

#### Stage 5 – Assessing areas of enrichment for longitudinal datasets

To inform this key stage, and to gather the views of a wide range of stakeholders, the team at MQ organised a ToC development process to consider different perspectives about the barriers faced by stakeholders and priorities of various fields working around mental health research and intervention. Participants of the ToC process were not made aware of the shortlisted datasets. This activity and its outputs are described below.

We reviewed all datasets considering the priorities established with Wellcome and feedback received from the ToC process. We considered the datasets both independently and as a group to evaluate their individual strengths and how they complement one another when examined together. To do this, we used detailed information about each selected dataset, including the granularity of mental health data, potential for testing causal associations and outcome measures. We discussed areas of enrichment specific to each dataset and for the overall group of selected datasets.

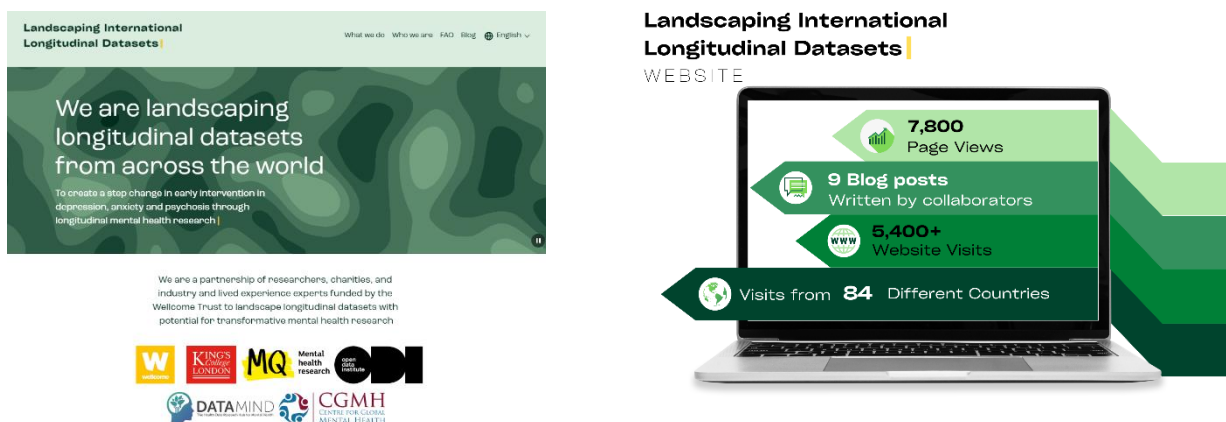
## Stage 6 – Reporting findings

We summarised the findings of the landscaping project, considerations for potential enrichment and highlights in richness of datasets in this report for Wellcome.

### Engagement with community and LEE involvement

Throughout the project, we engaged with worldwide research community and the public using various platforms (Figure IV).

*The Landscaping website:* We created a website dedicated to the project to support engagement with a broad range of stakeholders ([landscaping-longitudinal-research.com](http://landscaping-longitudinal-research.com)). This platform allowed the public to meet the partnership, view the longitudinal datasets identified and participate in the search process by submitting potentially relevant datasets via an online form.



**Figure IV. Summary statistics for engagement via the Landscaping website.**

The website was a successful way to engage with people outside the partnership and to share information with them. However, it was not an efficient way of getting information from the public about datasets.

*Blog posts:* Members of the partnership contributed blog posts to disseminate the project and our observations during the early landscaping stages. Altogether, we wrote a total of nine blogs, uploaded them on the project website and shared them on Twitter and LinkedIn. These blogs were read, further shared and commented on by the public.

*Social media posts:* We regularly posted about the project on Twitter and LinkedIn. This was especially important in the early stages of the project when we needed to engage with the community outside the partnership. We contacted via email people who showed interest in the project by liking, sharing or commenting on our posts (Figure V).



## Landscaping International Longitudinal Datasets



**Figure V. Summary statistics for engagement via Twitter**

*Online meetings:* We engaged with various stakeholders to answer questions and encourage them to contribute information about datasets they may work with. These meetings were often in response to requests from people outside the partnership.

*LEE groups:* We worked with LEEs throughout the project. A representative of Wellcome’s LEE group joined meetings and also wrote a blog post. LEEs from DATAMIND, MQ LEE international contacts and a representative from Wellcome LEE group joined the ToC process.

### Transparency as integral to the process of landscaping

Transparency of the landscaping process was important throughout the project. We ensured we were working in an open, honest and straightforward way:

*Within the partnership:* The research assistants based at KCL shared office space allowing unrestricted flow of exchanges and creative solutions within the team. The ODI-based research assistant occasionally joined in December 2022 and January 2023 to facilitate the harmonization of the findings for datasets from outside academia. We all worked on documents stored on a shared drive to enable all members of the KCL-based team to view and contribute to live updates.

The KCL-based team met twice weekly to discuss progress, challenges and next steps. Key points and important decisions were recorded in documents saved on the shared drive. The PI of the project met at least fortnightly with the ODI-based team to discuss progress, concerns

and next steps, and with the MQ-based team during the planning of the ToC process. The Wellcome LEE also joined these meetings about the ToC process.

We extensively documented each phase of the landscaping process including where repositories and datasets were identified from, by whom, and when. By virtue of this system, it is possible to track the origin of each piece of information, from when it first enters the spreadsheet to how, when, and why final decisions were taken. All partners had viewing access to the spreadsheets and could follow in real time updates.

Partners carbon copied the project's email address ([landscaping-wellcome@kcl.ac.uk](mailto:landscaping-wellcome@kcl.ac.uk)) when they contacted people outside the partnership about the Landscaping project. This facilitated transparency in communication within the partnership, consistency with sharing information with external collaborators, and easy follow-up when required.

*With Wellcome:* The PI of the project met fortnightly with Wellcome staff to discuss progress, concerns, gaps and next steps. Wellcome staff had viewing access to all documents and could follow updates. The PI presented an update of the findings to Wellcome staff in October 2022, December 2022 and February 2023 to share the team's progress. The PI provided written deliverables at regular intervals - according to a pre-established schedule - to allow Wellcome to ascertain the project's proper management and progress.

*With the public:* To ensure fairness while searching for, identifying, and screening datasets during the landscaping process, transparency with the public was of utmost importance. To achieve this, we crafted the following three-pronged approach:

- via social media: We regularly posted updates about the project, so our work was known by a wide community of people (see engagement section above). This included twice-weekly updates and calls for submissions on Twitter, weekly updates on LinkedIn, and the dissemination of 9 blogs about the projects. These updates on social media also referenced partners and Wellcome so they could amplify the messages to their respective networks. Moreover, some posts were translated into different languages to increase outreach and accessibility for those who do not speak English;
- via our website: The project website allowed the public to meet the partnership, view the longitudinal datasets identified so far and participate in the process by completing an online form. The bi-monthly blogs were uploaded on the website for easy access to the public. The website was translated into five of the most spoken research languages to improve the accessibility of the information about our project (i.e., French, Spanish, Portuguese, Chinese and Arabic);
- via the website dashboard: A dashboard of the identified longitudinal datasets allowed the public to access the landscaping data in a user-friendly way and view the process as it took place. We updated the list of datasets on a weekly basis and so the public could view our progress and submit details about datasets we did not identify yet;
- via online meetings: Members of the partnership met with members of the public – most frequently online and when required – to explain the project and its aims.

## **Annex 2. ODI full report, including list of organisations contacted**

### **ODI table of contents**

- Description of the ODI’s approach to identifying non-academic datasets
  - 1. Using ODI and KCL’s network of industry contacts to identify resources that could provide relevant datasets
    - Explanation of gaps from this approach
    - Lessons learned about the availability and types of non-academic datasets identified through this approach
  - 2. Using ODI expertise to identify additional portals, and search them effectively
    - Explanation of gaps from this approach
    - Lessons learned about the availability and types of non-academic datasets identified through this approach
  - 3. Use ODI’s expertise in data research to identify sources of data that may not be captured by 1 and 2, or that can build upon initial findings in 1 and 2.
    - Explanation of gaps from this approach
    - Lessons learned about the availability and types of non-academic datasets identified through this approach
- Recommendations regarding non-academic datasets

### **ODI executive summary**

The ODI was tasked with performing a landscape review of non-academic mental health datasets in order to support this study’s goals of identifying ten longitudinal mental health datasets to enrich and, ultimately, create a sea change in mental health research.

We consulted 41 sources containing more than 1 million datasets, identified 268 initial datasets that fit our initial criteria, and 127 that meet the criteria for this study and which have been included in discussions of which datasets to enrich.

Further, and importantly, our work has served to identify gaps in the non-academic landscape, ways of filling those gaps through further work and areas for further research. In doing so, our research has not only contributed to the goal of enriching the selected mental

health datasets, but can also serve as a basis for enriching the landscape of mental health research as a whole - thereby ultimately helping achieve the long-term goal of a sea change in mental health research.

In this section we outline: the methodologies we used to conduct our landscape review; our results; the gaps we identified through each approach; lessons to inform future research; and recommendations for future work in this area.

The ODI engaged in three complementary streams of engagement and research in order to better understand the availability of non-academic datasets, and begin to identify those that demonstrated promise for this landscaping exercise. The first was through using the ODI's and KCL's industry contacts to a) identify relevant datasets and b) discuss general attitudes and barriers to sharing non-academic data. The second was through an automated approach to investigating open data portals, to a) identify relevant datasets and b) assess the feasibility of using automated approaches to searching open data portals as a means of identifying relevant datasets. The third was via desk research, building on initial findings from the first two strategies to hone in on the most promising types of non-academic datasets.

The key takeaways from the three approaches are as follows:

1. Industry contacts were helpful in identifying potentially relevant datasets, but most of the datasets identified through this approach were part of academic studies and were therefore stewarded by the relevant academic organisation rather than the industrial organisation. Generally speaking, data stewarded by the pharmaceutical industry is unlikely to be appropriate for this exercise given the exclusion criteria. Most studies are too small, and have a randomised control (rather than observational) design. The most promising source of non-academic data identified through this means was healthcare administrative data, which allows for longitudinal study of individuals throughout their healthcare journey.
2. Open data portals, particularly those without a domain focus relevant to this exercise, are generally not effective tools for finding promising datasets relevant to mental health. However, we demonstrated that it is possible to circumvent the poor search functionality by using portals' API functionality and text analysis of metadata to query records in bulk using multiple search terms. We identified only a handful of relevant datasets with this approach, and, from the point of identifying a relevant search result, they required considerable additional desk research to identify the underpinning dataset. This is attributable to a number of factors, chief among which are poor metadata and the 'data' primarily being aggregated to form indicators and statistics. However, with a more relevant portal, this approach could be fruitful.

3. There are a large number of datasets based on healthcare administrative data which meet the initial inclusion criteria, but that are unlikely to fit the final criteria that have been defined in this exercise, as they are not enrichable in the same way as academic studies, and do not include validated measures of mental health beyond those used in clinic.

This led us to the following conclusions and recommendations:

Despite not meeting the criteria of the present study, administrative healthcare datasets have significant potential in furthering our understanding of mental health. However, these datasets require extensive understanding of context and most likely considerable data linkage before they can be used to their fullest potential. In addition, it is difficult to find these datasets, and the information describing them (metadata) are extremely limited.

A potential approach, which aligns with previous Wellcome Trust-funded initiatives through the Public Health Research Data Forum (e.g., reports on increasing findability and linkage of public health data), is to create critical infrastructure to improve the findability and accessibility of these data, for example, the creation of a central repository of metadata describing these assets in detail. In addition, funding initiatives to promote the use of these data in alignment with Wellcome's priorities could encourage activity in this space.

### **The ODI's approach to identifying non-academic datasets**

The ODI engaged in three complementary streams of engagement and research in order to better understand the availability of non-academic datasets, and begin to identify those that demonstrated promise for this landscaping exercise. The first was through using the ODI's and KCL's industry contacts to a) identify relevant datasets and b) discuss general attitudes and barriers to sharing non-academic data. The second was through an automated approach to investigating open data portals, to a) identify relevant datasets and b) assess the feasibility of using automated approaches to searching open data portals as a means of identifying relevant datasets. The third was via desk research, building on initial findings from the first two strategies to hone in on the most promising types of non-academic data.

#### ***1. Using ODI and KCL's network of contacts to identify resources***

One major challenge when conducting landscape reviews outside of academia is the lack of systems for tracking research. Within academia, wide-scale and widely-adopted systems exist that allow researchers to conduct targeted searches across the majority of academic journals to identify relevant articles, datasets and citations and to relatively quickly build a fairly comprehensive view of a given field of inquiry. But outside of academia, when trying to

conduct similar research, those systems rarely exist - and even when they do, they cover a much smaller section of the terrain, with much less consistency and standardisation.

This means that landscape reviews of non-academic sources cannot rely on the tried and tested methods used in academia. Instead, they must be conducted through more time-consuming methods such as emailing network contact and cold-calling potential holders of datasets.

For this landscape review, we approached a diverse array of ODI contacts. We contacted 400 individuals across 209 organisations (see list below). We included those who might steward (collect, maintain and share) relevant data directly, such as healthcare providers, government bodies, healthcare insurers, and pharmaceutical companies. We also contacted a range of organisations who we believed may use relevant datasets in their work (e.g., as an evidence base). These included highly relevant organisations such as the Health Foundation, but also a range of charities, businesses and social enterprises that we believed to work in an adjacent space. We described the study criteria in our emails, and asked our contacts to help us identify relevant studies.

### *Results*

- We received 33 replies from 26 organisations, leading to 15 data sources for us to investigate.
- In addition, non-academic leads were provided by the KCL team's contacts, who provided 16 sources for us to investigate.
- We also reached out to 3 relevant contacts with which the immediate team had working relationships, from NHS Digital and healthcare insurance companies.
- Finally, after researching additional potential sources, we reached out to 10 organisations with which we had no existing relationship, including mental health charities, learned societies, health tracking applications, private mental healthcare providers and patient groups. This did not lead to any sources being identified.
- In total, 31 sources were identified by external contacts, generating 97 candidate datasets.

### *Gaps in our approach*

Contacts that responded to our initial emails were generally those with which we the ODI has close working relationships. This resulted in the following gaps:

- No private providers of mental healthcare responded to us. This is a potential source of data which could help significantly with furthering understanding of mental health, although the data is highly sensitive. Any therapies funded by the NHS should however be recorded in NHS data, and so mental healthcare provided through the NHS is likely to be recorded in healthcare administrative data (discussed further below).
- Only one pharmaceutical company responded, and not one considered to be a leader in mental health. We therefore cannot draw broad conclusions about the pharmaceutical industry, but have supplemented this with our own knowledge and desk research.
- No app developers responded to us directly, although some candidate datasets originating from apps were identified by our contacts, such as the Zoe Health Study. Data from mood-tracking apps, and data generated by wearables, could have significant potential in helping to understand mental health and, as such, is a gap that ought to be addressed with further research. It is worth noting that the Alan Turing Institute has a “[Data science for mental health](#)” interest group that appears to be interested in the potential of data generated by smartphones, for example.

### *Lessons learned about availability of ‘non-academic’ data*

Many of our non-academic contacts, rather than directing us towards non-academic resources, directed us towards academic studies. A number of the datasets identified by non-academic contacts had already been captured by KCL, validating their approach. Some were new and of potential relevance. In many cases, companies did steward potentially relevant data but were unable to share it. We explore the reasons for this below, and outline gaps with respect to our understanding of non-academic data.

### *Barriers to data sharing*

We reached out to a number of different types of organisations within the health sector including pharmaceutical companies, private healthcare insurers and health charities. A large health insurance company expressed these typical concerns about sharing health data. These

barriers may impact enrichment activities and the wider development of the landscape of data in the health sector.

- Commercial concerns
  - Additional costs of sharing data
  - Commercial advantage (gained through having exclusive access to data)
- Data sensitivity
  - Individual data protection
  - Privacy principle of 'least access'.
- Restrictive services contracts
- Concerns about data quality
  - Completeness
  - Accuracy
- Black box processing
  - Opaque AI generated insight
- Data governance

(For further insights from the sector, see 'Interviews to examine attitudes to data sharing in the health sector')

### *Data stewarded by pharmaceutical companies*

Randomised control trials (RCTs) were excluded from the process, leaving only pharmaceutical-funded studies without an RCT design. We considered that the majority of data stewarded by pharmaceutical companies were likely to be derived from RCTs, as robustly demonstrating the efficacy of new drugs is aligned with pharmaceutical companies' business goals.

Contacts from Roche highlighted Vivli as an organisation who share participant-level data from completed clinical trials. A number of pharmaceutical companies contribute data from clinical trials to Vivli, which allows filtering based on study design and size. Only one study on Vivli was sufficiently large to meet the criteria and not a RCT.

Patient-reported outcomes, which factor in more holistic wellbeing measures of drug efficacy, are on the rise, which is promising for the study of factors impacting mental health more generally. However, the lengths of most pharmaceutical-sponsored studies are insufficient to meet our study criteria.

We conclude that data stewarded by the pharmaceutical industry is unlikely to provide a significant source for our purposes. While we cannot rule out that there are large studies focusing on mental health generating data relevant to this study, we view it as very likely



that these studies will be RCTs, and as unlikely that they will span a sufficient period of time. 'Real-world' pharmaceutical data are likely to be more relevant; in effect, this is disease registry data.

Many of the datasets identified by non-academic contacts were large studies such as disease registries, or datasets based on healthcare administrative data. These are generally considered to fall under the banner of 'big data' - data generated in huge volumes that require advanced analytics to make sense of.

#### *Data stewarded by insurers/healthcare industry/government - healthcare administrative/registry data*

In many countries across the world, citizens are required to have healthcare insurance. Healthcare insurance companies keep detailed records of their customer's claims and their corresponding illnesses. Because healthcare insurance is often mandatory in these countries, the coverage of these datasets is very high. Increasingly, governments require this administrative healthcare data to be made available for research purposes. A similar process exists in the UK through the NHS, which makes certain datasets available through NHS Digital. The NHS in effect is part health insurer, part government body.

The availability of this kind of data across the world is variable, and dependent on the quality and digitalisation of healthcare systems. The purpose of these data being collected is generally to improve patient and system outcomes, keep track of costs, and to guide policy. However, it has the potential to follow millions of individuals over the course of their lives, so it may have significant value to improving the understanding of mental health.

We discuss 'big data' in section 3, below.

#### *Gaps in our understanding*

It is notable that a number of non-academic contacts pointed us towards academic studies. This could be because they simply did not know of non-academic datasets that met our inclusion criteria, or that such datasets do not exist. However, it could be that by specifying the inclusion criteria, that were developed primarily with academic studies in mind, we were unintentionally prompting our contacts to identify academic datasets.

## *2. Using ODI expertise to identify additional portals, and search them effectively.*

A number of the resources identified by our contacts were data portals. These were then supplemented by internal ODI knowledge to identify further platforms which might host relevant data.

Platforms were reviewed manually where possible, in order to ensure all potentially relevant datasets were captured. However, a number of portals had datasets numbering in the thousands, or even millions.

The ODI piloted a programmatic approach using portals' API functionality and text analysis of metadata to query records in bulk using multiple search terms. This allowed us to investigate data portals in as much depth as possible with the limited time available to a study of this type.

### *Results*

We explored 3 portals: data.gov.uk, NHS England open data portal, and data.europa.eu.

- Data.gov.uk lists around 55,000 resources. Initial filters identified around 4000 datasets, and by iterating and testing the effectiveness of different search terms, this was reduced to 3 relevant datasets, and 2 made it through.
- Data.europa.eu holds over 1.5 million datasets, from 36 countries. Using the lessons learned from the data.gov.uk portal, and an iterative approach, we reduced the number of datasets for manual review to 187. Of these 19 were identified to be possibly relevant, and 2 made it through to the screening process.
- NHS England's Data Catalogue had 841 datasets from various NHS bodies. After filtering and review, this resulted in 4 relevant datasets, of which all 4 made it through to the screening process.

In total, 26 datasets were initially identified by using an automated strategy with data portals. Eight of these were relevant and longitudinal, and four made it through the screening stage.

### *Gaps in our approach*

There are a significant number of portals (597) that were identified internally within the ODI that were not explored as part of this study, owing to feasibility. Data.europa.eu however brings in metadata from a wide range of data portals across Europe. While we were unable to investigate all data portals, we have generated an approach that could be used for future attempts to identify relevant datasets via portals, and have been able to draw some broad albeit tentative conclusions about the use of data portals in general.

### *Lessons learned about portals as a source of relevant datasets:*

Two of the portals that we searched were not specific to any domain, meaning that there was a lot of noise returned by using seemingly relevant search terms. For example, ‘depression’ and ‘survey’ are used, with different meanings, in different domains. Generally, search terms that may have returned useful results via academic searches (e.g., longitudinal) returned no results in data portals. Alternatives like “time series” were attempted, but returned a large number of irrelevant results. These portals had relatively poor search functionality, making it difficult to identify relevant results - hence the necessity of the API approach.

Generally, open data portals provided only aggregated datasets/statistics. Offering no ability to focus on individuals, or to identify with any specificity when events occur, meant that these aggregated statistics were excluded from the study. In some cases, further desk research identified the sources underpinning these aggregated statistics, which had potential for longitudinal study of individuals. This was the case for the 2 datasets that were identified via data.europa.eu, for example. These two datasets were based on healthcare administrative data.

From this, we concluded that open data portals are of most value when they are focused on a specific domain. They did identify a number of potentially valuable sources of data, but they are of most use when you have clearly defined search criteria, and as such are likely to be most effectively used to identify supplementary datasets after other angles have been exhausted.

The most relevant NHS datasets identified through this approach (and more) were also identified through our discussions with our contacts at NHS Digital. This mirrors findings from previous ODI research that portals are most effective when used in combination with a ‘librarian’, with intimate knowledge of the data.<sup>1</sup> In the absence of advanced portal search

---

<sup>1</sup> ODI (2022) Portals report - in print. Exact reference to be provided in subsequent draft

functionality, a relational approach to identifying data maybe a more efficient means of locating relevant data, given the present status of the (mental) healthcare data ecosystem.

***3. Use ODI's expertise in data research to identify sources of data that may not be captured by 1 and 2, or that can build upon initial findings in 1 and 2.***

Building on the relevant datasets identified in streams 1 and 2, we researched further the datasets that might exist outside of academia and the pharmaceutical industry. This was achieved through identifying key terms that frequently appeared while finding details about resources identified in streams 1 and 2, and subsequently using these as search terms.

*Results*

This culminated in the identification of a body of literature on 'big data' for mental healthcare. Here, big data refers to data generated in huge volume that require advanced analytics to make sense of. In the context of health data, relevant datasets generally fall into three categories<sup>2</sup>:

- Disease registries - these are generally large studies that aim to collect data on people with specific chronic diseases, and can play a role in understanding longer-term effects of new drugs when they are initially brought to market
- Administrative healthcare or claims data (including those produced by health insurers) - or 'electronic medical/health records' (EMR/EHR) - are discussed above, which follow individuals' interactions with healthcare systems
- Prescription data
  - These datasets were eventually excluded because they followed prescriptions, rather than individuals. However, the majority passed to the screening stage, owing to their longitudinal nature.

In addition to this, we identified psychiatric case registers as a highly relevant source of data. These sit somewhere between administrative healthcare data and registry data. They are traditionally collected in specialist state-provided mental health secondary care institutions.

---

<sup>2</sup> Heads of Medicines Agencies, European Medicines Agency (2018) 'Observational data (Real World Data) Subgroup report' <https://hnrws.cn/uploads/file/20220317/17/30b5707c475c41c84881d40d164306d9.pdf>

Similar to medical records, case registers were previously physical documents, but are increasingly digitised.

This approach resulted in 6 sources, and 145 candidate datasets, making it the most productive approach.

### *Gaps in our approach*

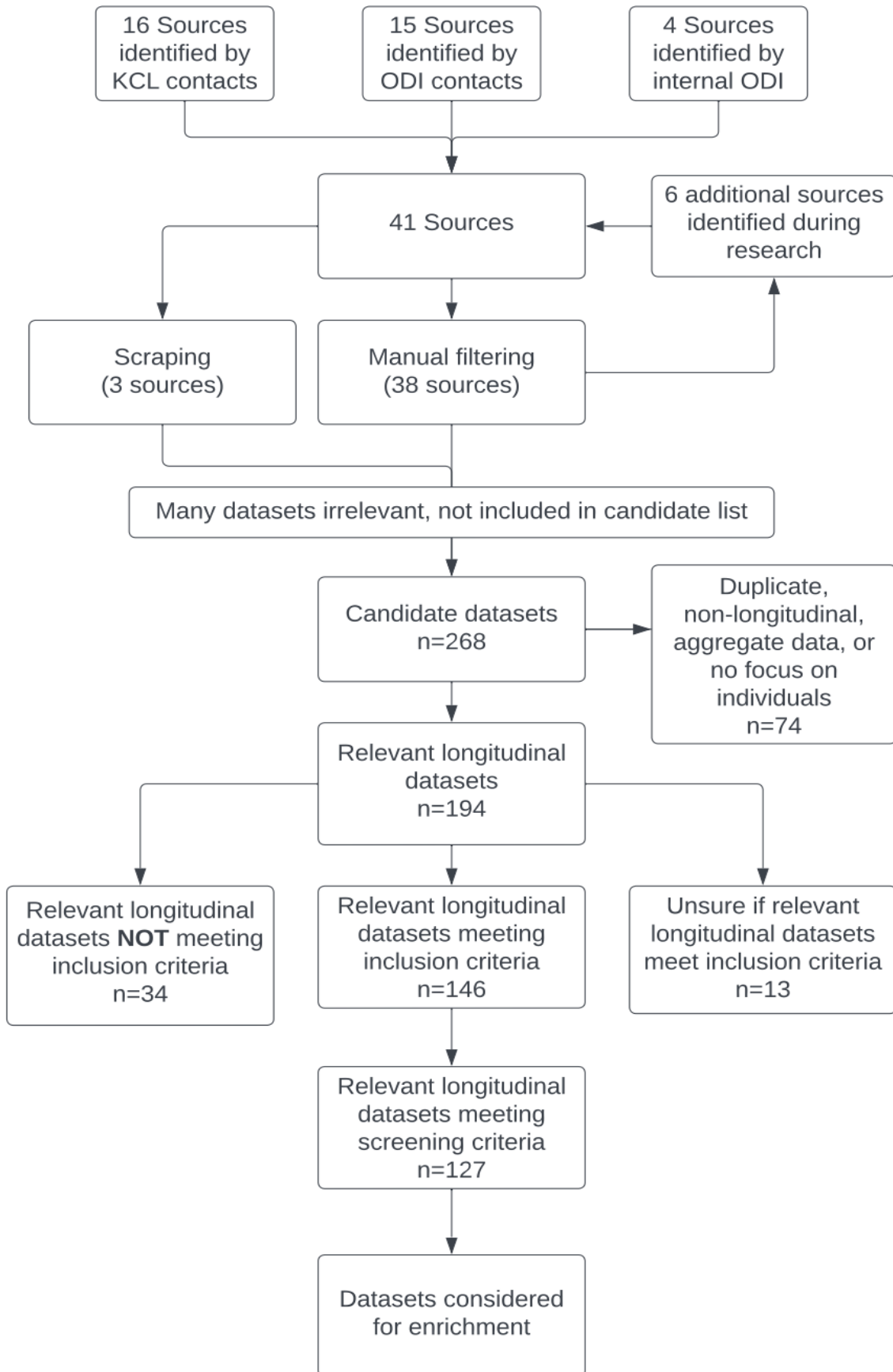
We were unable to perform a systematic literature search on ‘big data’ in (mental) healthcare, although we were able to identify a number of studies that had attempted to collate relevant resources - though only one was specific to mental health. None appear to have taken place in the last 4 years. There may well be a number of papers published recently on the use of big data in mental health, and a systematic literature review would be able to identify them.

In addition, at present we have been unable to screen all of these datasets owing to the limited time available within a study of this kind.

### *Lessons learned about ‘big data’ for mental health*

- None of these datasets are open by default, as they contain sensitive information. They generally appear to have application processes through which researchers submit proposals, although we were unable to confirm this for all datasets.
- Particularly with psychiatric case registers, these are often derived from secondary healthcare rather than primary, meaning that it is only the most severely affected individuals that are being recorded. This is an important factor to consider before drawing any conclusions from these datasets. Some case registers also include data derived from primary care, or provide the possibility for linkage with primary care datasets.
- Another significant barrier to the use of these data is its poor findability. Coming to the end of a four-month search process, we are still identifying new candidate ‘big’ datasets, often through publications that attempt to collate resources. There have been a number of such attempts, but few of them have been exhaustive, and, to our knowledge, only one has been specific to mental health. None appear to have taken place in the last 4 years.

- A small body of literature already exists that discusses the strengths and limitations of 'big data' for understanding (mental) health conditions, for example:
  - Schofield P (2017), "Big data in mental health research – do the ns justify the means? Using large data-sets of electronic health records for mental health research" <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5451644/#R4>
  - Pacurariu A et al. (2018), "Electronic healthcare databases in Europe: descriptive analysis of characteristics and potential for use in medicines regulation" <https://bmjopen.bmj.com/content/bmjopen/8/9/e023090.full.pdf>
  - Morgan VA, Jablenskey AV (2010), From inventory to benchmark: quality of psychiatric case registers in research <https://www-cambridge-org.iclibezp1.cc.ic.ac.uk/core/journals/the-british-journal-of-psychiatry/article/from-inventory-to-benchmark-quality-of-psychiatric-case-registers-in-research/81D13552E293952C6B4EA14FB82C57B0>
  - Candore G et al. (2019), Can We Rely on Results From IQVIA Medical Research Data UK Converted to the Observational Medical Outcome Partnership Common Data Model? <https://ascpt.onlinelibrary.wiley.com/doi/pdf/10.1002/cpt.1785>



<b>Name of Organisation</b>	<b>Link</b>	<b>Replies</b>
Abbvie	<a href="https://www.abbvie.co.uk/">https://www.abbvie.co.uk/</a>	x
ACCES Employment	<a href="https://acesemployment.ca/">https://acesemployment.ca/</a>	x
Accountabilitylab	<a href="https://accountabilitylab.org/">https://accountabilitylab.org/</a>	x
Active Norfolk	<a href="https://www.activenorfolk.org">https://www.activenorfolk.org</a>	x
Active Points	<a href="https://activepoints.co.uk/">https://activepoints.co.uk/</a>	x
Active Prospects	<a href="https://activeprospects.org.uk">https://activeprospects.org.uk</a>	x
ADB	<a href="https://www.adb.org/">https://www.adb.org/</a>	x
Advanced Medical Technology Association (AdvaMed)	<a href="https://www.advamed.org/">https://www.advamed.org/</a>	x
Alberta Health Services	<a href="https://www.albertahealthservices.ca/">https://www.albertahealthservices.ca/</a>	x
American Jewish Joint Distribution Committee	<a href="https://www.jdc.org/">https://www.jdc.org/</a>	x
AP-HP	<a href="https://www.aphp.fr/">https://www.aphp.fr/</a>	x
Arena Pharmaceuticals, Inc.	<a href="https://www.arenapharmaceuticals.co.uk">https://www.arenapharmaceuticals.co.uk</a>	x
Ashoka	<a href="https://www.ashoka.org/en-gb">https://www.ashoka.org/en-gb</a>	x
ASHUA	<a href="https://ashua.es/">https://ashua.es/</a>	x
Associação Paulista de Medicina	<a href="https://associacaopaulistamedicina.org.br">https://associacaopaulistamedicina.org.br</a>	x
AZ Delta	<a href="https://www.azdelta.be/nl">https://www.azdelta.be/nl</a>	x
Bay Of Plenty District Health Board	<a href="https://www.bopdhb.health.nz/">https://www.bopdhb.health.nz/</a>	x
Bayer	<a href="https://www.bayer.com/en/">https://www.bayer.com/en/</a>	x
BC Co-operative Association	<a href="https://bcc.a.coop/">https://bcc.a.coop/</a>	x
Beyond Blue Ltd	<a href="https://www.beyondblue.tech/">https://www.beyondblue.tech/</a>	x
Bhcommunityworks	<a href="https://www.bhcommunityworks.org.uk/">https://www.bhcommunityworks.org.uk/</a>	x
Bill & Melinda Gates Foundation	<a href="https://www.gatesfoundation.org/">https://www.gatesfoundation.org/</a>	✓
BioIndustry Association	<a href="https://www.bioindustry.org/">https://www.bioindustry.org/</a>	x
Birmingham Social Enterprise City Steering Group	<a href="https://www.i-se.co.uk/about-us/">https://www.i-se.co.uk/about-us/</a>	✓
Bliss	<a href="https://www.bliss.org.uk/">https://www.bliss.org.uk/</a>	x
BMA	<a href="https://www.bma.org.uk/">https://www.bma.org.uk/</a>	x



BoP Hub	<a href="https://www.bophub.org/">https://www.bophub.org/</a>	x
British Council	<a href="https://www.britishcouncil.org/">https://www.britishcouncil.org/</a>	x
Brotherhood of St Laurence	<a href="https://www.bsl.org.au/">https://www.bsl.org.au/</a>	x
Cabarrus Health Alliance	<a href="https://www.cabarrushealth.org/">https://www.cabarrushealth.org/</a>	x
CAG Chennai	<a href="https://www.cag.org.in/">https://www.cag.org.in/</a>	x
Cardiobase	<a href="https://cardiobase.com/">https://cardiobase.com/</a>	x
Care Quality Commission	<a href="https://www.cqc.org.uk/">https://www.cqc.org.uk/</a>	x
Center for Civilians in Conflict	<a href="https://civiliansinconflict.org/">https://civiliansinconflict.org/</a>	x
Center for Global Development	<a href="https://www.cgdev.org/">https://www.cgdev.org/</a>	x
Center for International Blood & Marrow Transplant Research (CIBMTR)	<a href="http://www.cibmtr.org">www.cibmtr.org</a>	x
Centro Clinico NeMO	<a href="https://www.tricals.org/en/locations/centro-clinico-nemo/">https://www.tricals.org/en/locations/centro-clinico-nemo/</a>	x
Changing Lives	<a href="https://www.changing-lives.org.uk/">https://www.changing-lives.org.uk/</a>	✓
Charite	<a href="https://www.charite.de/en/">https://www.charite.de/en/</a>	✓
Chatham House	<a href="https://www.chathamhouse.org/">https://www.chathamhouse.org/</a>	x
Chelsea and Westminster Hospital	<a href="https://www.chelwest.nhs.uk/">https://www.chelwest.nhs.uk/</a>	x
Children's Hospital of Philadelphia	<a href="https://www.chop.edu/">https://www.chop.edu/</a>	✓
CHU de Nantes	<a href="https://www.chu-nantes.fr/">https://www.chu-nantes.fr/</a>	x
Citizens UK	<a href="https://www.citizensuk.org/">https://www.citizensuk.org/</a>	x
ClimateWorks Foundation	<a href="https://www.climateworks.org/">https://www.climateworks.org/</a>	x
Clinical Research Development Ireland	<a href="http://crdi.ie/">http://crdi.ie/</a>	x
Community Action Suffolk	<a href="https://www.communityactionsuffolk.org.uk">https://www.communityactionsuffolk.org.uk</a>	x
Community Development Halton	<a href="https://cdhalton.ca/">https://cdhalton.ca/</a>	✓
Conway Medical Center	<a href="https://www.conwaymedicalcentre.nhs.uk/">https://www.conwaymedicalcentre.nhs.uk/</a>	x
Cre8 care	<a href="https://www.cre8care.org/">https://www.cre8care.org/</a>	x

Crohn's and Colitis UK	<a href="https://crohnsandcolitis.org.uk/">https://crohnsandcolitis.org.uk/</a>	x
CSIS	<a href="https://www.csis.org/">https://www.csis.org/</a>	x
Dacorum Borough Council	<a href="https://www.dacorum.gov.uk/">https://www.dacorum.gov.uk/</a>	x
Dartford Borough Council	<a href="https://www.dartford.gov.uk/">https://www.dartford.gov.uk/</a>	x
DesAcc	<a href="https://www.desacc.com/">https://www.desacc.com/</a>	x
Deutsche Gesellschaft für Internationale Zusammenarbeit GmbH (GIZ)	<a href="https://www.giz.de/de/html/index.html">https://www.giz.de/de/html/index.html</a>	✓
DI	<a href="https://www.uke.de/">https://www.uke.de/</a>	x
Diaceutics	<a href="https://www.diaceutics.com/">https://www.diaceutics.com/</a>	x
DICA	<a href="https://www.dica.gov.mm/">https://www.dica.gov.mm/</a>	x
D-Tree International Inc	<a href="https://www.d-tree.org/">https://www.d-tree.org/</a>	x
Eagle Genomics	<a href="https://www.eaglegenomics.com/">https://www.eaglegenomics.com/</a>	x
Eastern Health	<a href="https://www.easternhealth.org.au/">https://www.easternhealth.org.au/</a>	x
EFPIA	<a href="https://www.efpia.eu/">https://www.efpia.eu/</a>	x
Elizabeth Glaser Pediatric AIDS Foundation	<a href="https://pedaids.org/">https://pedaids.org/</a>	✓
EMD UK	<a href="https://emduk.org/">https://emduk.org/</a>	x
Emory University	<a href="https://www.emory.edu/home/index.html">https://www.emory.edu/home/index.html</a>	x
Erasmus MC	<a href="https://www.erasmusmc.nl/nl-nl/">https://www.erasmusmc.nl/nl-nl/</a>	x
Esmee Fairbairn Foundation	<a href="https://esmeefairbairn.org.uk/">https://esmeefairbairn.org.uk/</a>	x
European Commission	<a href="https://commission.europa.eu/index_en">https://commission.europa.eu/index_en</a>	x
European Union	<a href="https://european-union.europa.eu/index_en">https://european-union.europa.eu/index_en</a>	✓
F. Hoffmann-La Roche Ltd.	<a href="https://www.roche.com/">https://www.roche.com/</a>	✓
Fairtrade	<a href="https://www.fairtrade.org.uk/">https://www.fairtrade.org.uk/</a>	x
Fairtrade International	<a href="https://www.fairtrade.net/">https://www.fairtrade.net/</a>	x
Fakultní nemocnice v Motole	<a href="https://www.fnmotol.cz/">https://www.fnmotol.cz/</a>	x
Fiocruz	<a href="https://portal.fiocruz.br/en">https://portal.fiocruz.br/en</a>	x
Folkehelseinstituttet	<a href="https://www.fhi.no/en/">https://www.fhi.no/en/</a>	x
Fondation René Touraine	<a href="https://www.fondation-r-touraine.org/?lang=en">https://www.fondation-r-touraine.org/?lang=en</a>	x

Ford Foundation	<a href="https://www.fordfoundation.org/">https://www.fordfoundation.org/</a>	x
Future Care Capital	<a href="https://futurecarecapital.org.uk/">https://futurecarecapital.org.uk/</a>	✓
Gemelli Employee Benefits	<a href="https://www.gemelliemployeebenefits.co.uk/">https://www.gemelliemployeebenefits.co.uk/</a>	x
Genentech	<a href="https://www.gene.com/">https://www.gene.com/</a>	x
Genomics England	<a href="https://www.genomicsengland.co.uk/">https://www.genomicsengland.co.uk/</a>	✓
GFA Consulting Group	<a href="https://www.gfa-group.de/">https://www.gfa-group.de/</a>	x
Golding Homes	<a href="https://www.goldinghomes.org.uk/">https://www.goldinghomes.org.uk/</a>	x
Guy's & St Thomas' Charity	<a href="https://gsttcharity.org.uk/">https://gsttcharity.org.uk/</a>	x
Guy's and St Thomas' NHS Foundation Trust	<a href="https://www.guysandstthomas.nhs.uk/">https://www.guysandstthomas.nhs.uk/</a>	✓
Health and Social Care Alliance Scotland	<a href="https://www.alliance-scotland.org.uk/">https://www.alliance-scotland.org.uk/</a>	x
Health Education England	<a href="https://www.hee.nhs.uk/">https://www.hee.nhs.uk/</a>	x
Health Fabric	<a href="https://www.healthfabric.co.uk/">https://www.healthfabric.co.uk/</a>	x
Health IQ	<a href="https://www.healthiq.com/">https://www.healthiq.com/</a>	x
Health Research Authority	<a href="https://www.hra.nhs.uk/">https://www.hra.nhs.uk/</a>	x
Henry Schein	<a href="https://www.henryschein.co.uk/">https://www.henryschein.co.uk/</a>	x
Hoag	<a href="https://www.hoag.org/">https://www.hoag.org/</a>	x
Humanity United	<a href="https://humanityunited.org/">https://humanityunited.org/</a>	x
Institute of Baltic Studies	<a href="https://www.ibs.ee/en/">https://www.ibs.ee/en/</a>	✓
Institute of Myology	<a href="https://www.institut-myologie.org/">https://www.institut-myologie.org/</a>	x
Institute of Public Health of Serbia	<a href="https://www.batut.org.rs/index.php?lang=2">https://www.batut.org.rs/index.php?lang=2</a>	x
International SOS	<a href="https://www.internationalsos.com/">https://www.internationalsos.com/</a>	x
Johnson & Johnson Services, Inc.	<a href="https://jnj.co.uk/">https://jnj.co.uk/</a>	x
Kaiser Permanente	<a href="https://healthy.kaiserpermanente.org/front-door">https://healthy.kaiserpermanente.org/front-door</a>	x
Kids Help Phone	<a href="https://kidshelpphone.ca/">https://kidshelpphone.ca/</a>	x
Leicester-Shire & Rutland Sport	<a href="https://www.connectsport.co.uk/organisation/leicester-shire-rutland-sport">https://www.connectsport.co.uk/organisation/leicester-shire-rutland-sport</a>	✓
Leiden University Medical Center	<a href="https://www.universiteitleiden.nl/en/medicine-lumc">https://www.universiteitleiden.nl/en/medicine-lumc</a>	x
LGiU	<a href="https://lgiu.org/">https://lgiu.org/</a>	x
LifeLabs	<a href="https://www.lifelabs.com/">https://www.lifelabs.com/</a>	x
Local Trust	<a href="https://localtrust.org.uk/">https://localtrust.org.uk/</a>	x

Lohmann & Rauscher GmbH & Co. KG	<a href="https://www.lohmann-rauscher.com/en/">https://www.lohmann-rauscher.com/en/</a>	x
LUX MED	<a href="https://www.luxmed.pl/">https://www.luxmed.pl/</a>	x
MacArthur Foundation	<a href="https://www.macfound.org/">https://www.macfound.org/</a>	x
Médecins Sans Frontières	<a href="https://www.msf.org/">https://www.msf.org/</a>	✓
Medizinische Universität Wien	<a href="https://www.meduniwien.ac.at/web/">https://www.meduniwien.ac.at/web/</a>	x
Medtronic	<a href="https://www.medtronic.com/uk-en/index.html">https://www.medtronic.com/uk-en/index.html</a>	x
Mellon	<a href="https://www.mellon.com/">https://www.mellon.com/</a>	x
Mencap	<a href="https://www.mencap.org.uk/">https://www.mencap.org.uk/</a>	x
Merck	<a href="https://www.merckgroup.com/en">https://www.merckgroup.com/en</a>	x
Merck Sharp & Dohme (MSD)	<a href="https://www.msd-uk.com/">https://www.msd-uk.com/</a>	x
Migrations & Développement	<a href="https://www.migdev.org/">https://www.migdev.org/</a>	x
Ministry of Health	<a href="https://moh.gov.my/">https://moh.gov.my/</a>	x
Murray PHN	<a href="https://www.murrayphn.org.au/">https://www.murrayphn.org.au/</a>	x
National Health and Medical Research Council	<a href="https://www.nhmrc.gov.au/">https://www.nhmrc.gov.au/</a>	x
National Network to End Domestic Violence	<a href="https://nnev.org/">https://nnev.org/</a>	x
NHS	<a href="https://www.nhs.uk/">https://www.nhs.uk/</a>	x
NHS Confederation	<a href="https://www.nhsconfed.org/">https://www.nhsconfed.org/</a>	x
NHS England	<a href="https://www.england.nhs.uk/">https://www.england.nhs.uk/</a>	✓
NHS Scotland	<a href="https://www.nhslothian.scot/Pages/default.aspx">https://www.nhslothian.scot/Pages/default.aspx</a>	x
NHS Wales	<a href="http://wales.nhs.uk/">http://wales.nhs.uk/</a>	x
NICVA	<a href="https://www.nicva.org/">https://www.nicva.org/</a>	x
North Bristol NHS Trust	<a href="https://www.nbt.nhs.uk/">https://www.nbt.nhs.uk/</a>	x
Novartis	<a href="https://www.novartis.com/">https://www.novartis.com/</a>	x
NYU Langone Medical Center	<a href="https://nyulangone.org/">https://nyulangone.org/</a>	x
Ochsner Health	<a href="https://www.ochsner.org/">https://www.ochsner.org/</a>	✓
Olympic Medical Center	<a href="https://www.olympicmedical.org/">https://www.olympicmedical.org/</a>	x
OneMedical Group	<a href="https://onemedicalgroup.co.uk/">https://onemedicalgroup.co.uk/</a>	x
Ontario Nonprofit Network	<a href="https://theonnc.ca/">https://theonnc.ca/</a>	x

Open Society Foundations	<a href="https://www.opensocietyfoundations.org/">https://www.opensocietyfoundations.org/</a>	x
Optos	<a href="https://www.optos.com/">https://www.optos.com/</a>	x
Optum	<a href="https://www.optum.com/">https://www.optum.com/</a>	x
Oxford University Hospitals NHS Foundation Trust	<a href="https://www.ouh.nhs.uk/">https://www.ouh.nhs.uk/</a>	x
Oxfordshire Community Foundation	<a href="https://oxfordshire.org/">https://oxfordshire.org/</a>	x
PALGA Foundation	<a href="https://www.palga.nl/en/over-ons-1/about-stichting-palga/">https://www.palga.nl/en/over-ons-1/about-stichting-palga/</a>	✓
Pathways Inc	<a href="https://www.pathwaysforyou.org/">https://www.pathwaysforyou.org/</a>	x
Pfizer	<a href="https://www.pfizer.co.uk/">https://www.pfizer.co.uk/</a>	x
Pfizer Inc.	<a href="https://www.pfizer.com/">https://www.pfizer.com/</a>	x
Phoenix Children's Hospital	<a href="https://www.phoenixchildrens.org/">https://www.phoenixchildrens.org/</a>	x
Planned Parenthood Mar Monte	<a href="https://www.plannedparenthood.org/planned-parenthood-mar-monte">https://www.plannedparenthood.org/planned-parenthood-mar-monte</a>	x
Polycystic Kidney Disease (PKD) Charity	<a href="https://pkdcharity.org.uk/">https://pkdcharity.org.uk/</a>	x
Positive View	<a href="https://positiveview.org.uk/">https://positiveview.org.uk/</a>	✓
Quanticate	<a href="https://www.quanticate.com/">https://www.quanticate.com/</a>	x
Radboud UMC	<a href="https://www.radboudumc.nl/en/patient-care">https://www.radboudumc.nl/en/patient-care</a>	x
Rainforest Alliance Inc	<a href="https://www.rainforest-alliance.org/">https://www.rainforest-alliance.org/</a>	x
RAND Corporation	<a href="https://www.rand.org/">https://www.rand.org/</a>	x
Region Kalmar län	<a href="https://www.regionkalmar.se/">https://www.regionkalmar.se/</a>	x
ResMed	<a href="https://www.resmed.co.uk/">https://www.resmed.co.uk/</a>	x
RSPB	<a href="https://www.rspb.org.uk/">https://www.rspb.org.uk/</a>	x
Sage Bionetworks	<a href="https://sagebionetworks.org/">https://sagebionetworks.org/</a>	✓
Sainsbury Family Charitable Trusts	<a href="https://www.sfct.org.uk/">https://www.sfct.org.uk/</a>	✓
Salford Royal Foundation Trust	<a href="https://www.northerncaresalliance.nhs.uk/">https://www.northerncaresalliance.nhs.uk/</a>	x
Samaritans	<a href="https://www.samaritans.org/">https://www.samaritans.org/</a>	x
Sanitas	<a href="https://www.sanitas.es/">https://www.sanitas.es/</a>	x
Santa Casa da Misericórdia de Lisboa	<a href="https://www.scml.pt/">https://www.scml.pt/</a>	x
Saudi Patient Safety Center	<a href="https://www.spsc.gov.sa/English/Pages/Home.aspx">https://www.spsc.gov.sa/English/Pages/Home.aspx</a>	x

Save the Children Federation	<a href="https://www.savethechildren.org/">https://www.savethechildren.org/</a>	x
Shelter	<a href="https://www.shelter.org.uk/">https://www.shelter.org.uk/</a>	x
SIOP Europe	<a href="https://siope.eu/">https://siope.eu/</a>	x
Smith & Nephew	<a href="https://www.smith-nephew.com/en-gb">https://www.smith-nephew.com/en-gb</a>	x
Somos Más	<a href="https://somosmas.org/">https://somosmas.org/</a>	x
South London and Maudsley NHS Foundation Trust	<a href="https://slam.nhs.uk/">https://slam.nhs.uk/</a>	✓
South Yorkshire Housing Association	<a href="https://www.syha.co.uk/">https://www.syha.co.uk/</a>	x
Spire Healthcare Group plc	<a href="https://www.spirehealthcare.com/">https://www.spirehealthcare.com/</a>	x
Sunbeam Family Services	<a href="https://sunbeamfamilyservices.org/">https://sunbeamfamilyservices.org/</a>	x
Surrey Heath Borough Council	<a href="https://www.surreyheath.gov.uk/">https://www.surreyheath.gov.uk/</a>	x
Sylva Ry	<a href="https://www.sylva.fi/">https://www.sylva.fi/</a>	x
SyncroPhi	<a href="http://www.syncrophi.com/">http://www.syncrophi.com/</a>	x
Takeda Pharmaceutical Company Limited	<a href="https://www.takeda.com/en-gb/">https://www.takeda.com/en-gb/</a>	x
The Chicago Community Trust	<a href="https://www.cct.org/">https://www.cct.org/</a>	x
The Children's Society	<a href="https://www.childrenssociety.org.uk/">https://www.childrenssociety.org.uk/</a>	x
The David and Lucile Packard Foundation	<a href="https://www.packard.org/">https://www.packard.org/</a>	x
The Health Foundation	<a href="https://www.health.org.uk/">https://www.health.org.uk/</a>	✓
The Henry Smith Charity	<a href="https://www.henrysmithcharity.org.uk/">https://www.henrysmithcharity.org.uk/</a>	x
The King's Fund	<a href="https://www.kingsfund.org.uk/">https://www.kingsfund.org.uk/</a>	x
The McGovern Foundation	<a href="https://mcgovernfoundation.com/">https://mcgovernfoundation.com/</a>	x
The Rockefeller Foundation	<a href="https://www.rockefellerfoundation.org/">https://www.rockefellerfoundation.org/</a>	✓
The Social Market Foundation	<a href="https://www.smf.co.uk/">https://www.smf.co.uk/</a>	x
The State of Queensland (Queensland Health)	<a href="https://www.health.qld.gov.au/">https://www.health.qld.gov.au/</a>	✓
The Trussell Trust	<a href="https://www.trusselltrust.org/">https://www.trusselltrust.org/</a>	x

The University of Vermont Health Network	<a href="https://www.uvmhealth.org/">https://www.uvmhealth.org/</a>	x
Tirol Kliniken	<a href="https://www.tirol-kliniken.at/page.cfm?vpath=index">https://www.tirol-kliniken.at/page.cfm?vpath=index</a>	x
UCB S.A.	<a href="https://www.ucb.com/">https://www.ucb.com/</a>	x
UK BioIndustry Association	<a href="https://www.bioindustry.org/">https://www.bioindustry.org/</a>	x
UMC Utrecht	<a href="https://www.umcutrecht.nl/nl">https://www.umcutrecht.nl/nl</a>	x
UMCG	<a href="https://www.umcg.nl/">https://www.umcg.nl/</a>	x
UNICEF	<a href="https://www.unicef.org/">https://www.unicef.org/</a>	✓
Uniklinik RWTH Aachen	<a href="https://www.ukaachen.de/">https://www.ukaachen.de/</a>	x
United Nations Foundation	<a href="https://unfoundation.org/">https://unfoundation.org/</a>	x
Universitätsklinikum Essen	<a href="https://www.uk-essen.de/">https://www.uk-essen.de/</a>	x
Universitätsklinikum Hamburg Eppendorf	<a href="https://www.uke.de/">https://www.uke.de/</a>	x
University College London Hospitals NHS Foundation Trust	<a href="https://www.uclh.nhs.uk/">https://www.uclh.nhs.uk/</a>	x
University Hospitals Birmingham NHS Foundation Trust	<a href="https://www.uhb.nhs.uk/">https://www.uhb.nhs.uk/</a>	x
University Medical Centre Ljubljana	<a href="https://www.kclj.si/index.php?dir=/about_us">https://www.kclj.si/index.php?dir=/about_us</a>	x
University of Tsukuba	<a href="https://www.tsukuba.ac.jp/en/">https://www.tsukuba.ac.jp/en/</a>	x
UZ Gent	<a href="https://www.uzgent.be/">https://www.uzgent.be/</a>	x
UZ Leuven	<a href="https://www.uzleuven.be/en">https://www.uzleuven.be/en</a>	x
Vancouver Island Health Authority	<a href="https://www.islandhealth.ca/">https://www.islandhealth.ca/</a>	x
Vereniging Ouders, Kinderen & Kanker	<a href="https://zorgen.nl/verenigingen/vereniging-ouders-kinderen-en-kanker-vokk/">https://zorgen.nl/verenigingen/vereniging-ouders-kinderen-en-kanker-vokk/</a>	x
Vertex Pharmaceuticals	<a href="https://www.vrtx.com/">https://www.vrtx.com/</a>	x
Webbhälsa AB	<a href="https://www.kry.se/en/about/">https://www.kry.se/en/about/</a>	x
Whiston Hospital	<a href="https://www.sthk.nhs.uk/whiston-hospital">https://www.sthk.nhs.uk/whiston-hospital</a>	x
Women's Aid Organisation	<a href="https://www.womensaid.org.uk/">https://www.womensaid.org.uk/</a>	x
World Bank	<a href="https://www.worldbank.org/en/home">https://www.worldbank.org/en/home</a>	x

### *Suitability of 'big data' for this landscaping exercise*

The majority of the non-academic datasets that we identified and investigated fell under the 'big data' category. While these datasets met a number of important criteria for candidate datasets, they are likely to fall short at the final selection process.

Firstly, there is limited possibility for 'enriching' these datasets in the manner envisioned by this project, as the collection of additional categories of data is highly unlikely. These datasets are collected routinely, in a 'business as usual' manner, and so asking every healthcare professional in the UK (for example) to begin collecting data according to validated mental health measures is a very significant undertaking and would represent a significant shift in national healthcare policy.

Secondly, the purpose of collecting these datasets is not to understand the drivers of (mental) health and illness, but to improve outcomes - either from the perspective of a government, healthcare system, or business. These datasets can be used to track individuals longitudinally, but this context is important to bear in mind. For example, they are unlikely to collect data on drivers of mental health, such as socioeconomic status, and disaggregation is limited by factors that are captured in clinic (meaning disaggregation by race, for example, may not be easily achievable).<sup>3</sup> The key point here is that these datasets are unlikely to meet the criteria for a high-quality academic study, nor are they likely to be amenable to richer data collection.

However, that is not to say that they are not useful, and that they cannot be enriched through different means. Many of the datasets that we identified can be linked to other datasets that *do provide* contextual information such as socioeconomic status, which serves a similar purpose to enrichment. This is particularly true in countries that assign individuals personal identification numbers that allow for linkage across a range of different longitudinal datasets, such as the Scandinavian countries. Other approaches to linkage are possible, such as through the use of names, but are less reliable than the use of identifiers. Some countries, such as Wales (SAIL) and Australia, have institutions whose primary purpose is to facilitate the linkage of data for research purposes.

### *Recommendations regarding non-academic datasets - enriching the data ecosystem*

Making best use of big data for understanding mental health is likely to be far from simple. First, as we have found, it is difficult to locate. Once it has been found, in order to derive

---

<sup>3</sup> Big data in mental health research – do the *ns* justify the means? Using large data-sets of electronic health records for mental health research <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5451644/#R4>



maximum value, these datasets require linkage to other datasets, understanding of the context in which data were collected, and the ability to thoroughly clean data.

This highlights another significant opportunity for enrichment: the enrichment of the mental health data *ecosystem*.

By data ecosystem, we mean “the people, communities, and organisations that are stewarding data, creating things from it, deciding what to do based on it, influencing any of those activities, or are affected by any of those activities”.

Investing in this ecosystem, through providing critical infrastructure to make relevant data findable, is a potential avenue for Wellcome to consider.

This could be through:

- Engaging and convening stakeholders inside and outside academia to highlight the challenges of finding and sharing data and exploring and co-creating solutions
- Establishing or adopting existing standards and FAIR principles to make data findable and reusable within academia and without
- Advocating for open science and open research outside of academia in order to encourage a culture of sharing, collaboration and reuse to limit the number of datasets kept closed
- Explore options for motivating organisations to share data, whether through incentives (financial, social), requirements (policies, regulations) and novel data sharing approaches such as data portals and the establishment of data institutions

For instance, our research suggests that mental health data outside of academia is not easily findable, and so a mental health-specific portal could have significant value for mental health research. This could be combined with challenge prizes to incentivise the use of these data. We recommend focusing on metadata in the short term, as metadata is crucial to findability. Improving findability is the first step to making data FAIR (findable, accessible, interoperable and reusable).

This could be achieved through a central metadata repository that contains critical information on the datasets, including how they have been used before. While some of these criteria will mirror those used in this exercise, for example, size, terms of access and representativeness, defining the quality of these datasets will require additional criteria. For example:

- Reliability and usability: for example, have diagnoses been validated, have standardised collection processes been followed?
- Applicability to the general population: for example, does data come from primary care, secondary care, or both?
- Degree of disaggregation possible
- Degree and ease of linkability to other datasets
- Environmental factors impacting outcomes (e.g., disparities in access to healthcare)
- Possibility for free-text extraction and analysis

This would tie in well with previous Wellcome Trust-funded initiatives through the Public Health Research Data Forum (e.g., reports on increasing [findability](#)<sup>4</sup> and [linkage](#)<sup>5</sup> of public health data). Similar approaches to creating repositories describing study cohorts have already been attempted in [Alzheimer's](#), and [dementia](#).

Our attempts to find data through data portals were hampered by the lack of consensus terminology used to describe relevant datasets. A consistent approach to creating and finding metadata using specified vocabularies is therefore also likely to improve the findability of data immensely.

The European Medicines Agency and the Heads of Medicines Agencies have a 'Joint big data working group', which includes the promotion of standards for making 'big data' more usable, which could dovetail with this work.

Our initial research suggests that King's College London is also the locus of expertise in 'big data' for mental health: Professor Robert Stewart, Dr Katrina Davis and Dr Peter Schofield have published on the use of healthcare administrative data to investigate mental health, for example. We recommend that KCL, in addition to global experts, should be consulted before undertaking any further work in this space.

It is worth noting that 'big data' should in no way replace academic studies, but can form part of a complementary programme to further our understanding of mental health.

---

<sup>4</sup> Public Health Research Data Forum (2014), "Enhancing Discoverability of Public Health and Epidemiology Research Data: Summary"

<sup>5</sup> Public Health Research Data Forum (2015), "Enabling Data Linkage to Maximise the Value of Public Health Research Data: Summary"

We also identify a number of non-sector specific recommendations with regards to how the mental healthcare ecosystem can be enriched to encourage the ethical sharing of mental health data in ‘Interviews to examine attitudes to data sharing in the health sector’.

***Interviews to examine attitudes to data sharing in the health sector.***

A number of organisations were contacted to arrange structured discussions in order to understand their attitudes to data sharing and their data sources. The initial idea was to identify data sets for inclusion, however it became clear that there are attitudes to data sharing in the non-academic sector that are barriers to more data being available for research. This insight has been summarised here, though not directly attributed. The organisations included represent the key types of organisation in the sector, including pharmaceutical, health insurance, health providers and a health charity. NHS Digital has a role acting as a data steward for health data in the UK.

Date	Organisation	Type	Role
06.01.23	UCL,NHS,Cambs BRMC	Research	Independent consultant <sup>6</sup>
26.10.22	Akrivia	Private data steward	CEO and co-founder
11.11.22	MSF	Health charity	Deputy Medical Director
02.12.22	MSF	Health charity	Data Specialist
18.11.22	Roche	Pharmaceutical	Global Data Policy Strategy
14.11.22	NHS SLAM	Health provider	Quality Improvement Analytics
10.11.22	Bupa APAC	Health insurer	Data Office Operations Manager
26.10.22	NHS Digital	Health data steward	Director of Data Curation and Open Data

---

<sup>6</sup> Author of Reference 3

The ODI's mission is to encourage data sharing for the creation of value. It is no surprise that many of the reasons that organisations provide for not sharing data are similar to those from across a range of industry sectors. Each sector has different priorities for data sharing and in the types of data that when shared can have the biggest impact.

### *Interview structure*

- 30 minutes
- Introductions and project aims
  - This project is to look at non-academic data sources for longitudinal data that can be used to support mental health research.
  - Here is the link to the website
  - <https://www.landscaping-longitudinal-research.com/>
- Data sets generated, collected or used
- Attitudes to data sharing
- Available data
- Any other points

### *Interview findings*

The following is a summary of the views from the interviewees about the barriers to sharing healthcare data and suggestions of how to make more data available and more usable.

The organisations are on a spectrum of data sharing. There is a range from 'nothing is shared', where this is unlikely to change in the near future, to organisations with data sharing portals, and where the individual datasets can be requested, to organisations curating and stewarding data for reuse. In the latter category are NHS Digital, providing reusable healthcare data and statistics and Akrivia, collecting and providing data on a commercial basis and at no cost to the NHS.

## *Barriers to sharing healthcare data*

- Commercial concerns

A particular organisation sees data and the insight generated as an asset with commercial value and as a differentiator from market competition. This organisation has a data strategy that aims to maximise this benefit and does not see value in sharing data.

Sharing data can incur additional costs. If the sharing organisation does not see any commercial benefit then this makes it additionally difficult to justify.

- Data sensitivity

- Individual data protection
- Privacy principle of least access

Collected health data relates to individual patients. The location of that patient and their medical history are sensitive data too. When collecting personal data consent is required. Consent is often related to the initial purpose of the data collection and does not allow for reuse. In one organisation there is a principle of 'least access' as a way to improve the privacy of the data. This organisation sees maintaining privacy as a higher priority than the need to share data.

- Restrictive services contracts

One organisation collects and supports data on behalf of third parties as part of a service. These data is protected by the service contract and cannot be shared.

- Concerns about data quality.

There are concerns that sharing data will expose poor quality data. It is felt that this may reflect badly on the organisation. In some cases it may become a liability, where information can be shown to be inaccurate.

- Black box processing

In one example given, an AI chatbot was being tested to generate insight and improve customer interaction. The test has been stopped because it wasn't clear how the chatbot was learning and what data it was learning from. In this situation an organisation will be reluctant to share data. There is a reciprocal concern about using data generated in this way of which any 'receiving' organisation needs to be aware.

- Data governance.<sup>7</sup>

In one example the organisation's data team has developed plans to share data and has published a data sharing principle. The team has not yet been granted permission by the internal data owners. This is one example of the impact of data governance.

### *How to make data more findable and reusable*

- An intermediary as a data steward

Intermediary data stewards will collect and share data in ways that address some of the concerns and barriers listed about. Depending on the situation they could process sensitive data, including aggregation and analysis to produce insights, in ways not open to other organisations.

- Common keys for data linkage

In order to aggregate data and analyse against different factors there needs to be a common key that links to the individual data record and puts data records within context. This has to be achieved within the constraints of consent.

- Anonymisation

This process ensures that any individual is protected and unidentifiable when data is being processed. A number of organisations are examining masking techniques and software tools that would allow improved processing while protecting the individuals.

- Associated consent for reuse

Data subjects may be willing to allow reuse of their data. The data subject may require certain conditions and may wish to limit reuse to particular cases. By adding the consent record to the actual data record it would be clear what the data can and can't be used for.

- Better metadata and use of standards for reuse

It was suggested that applying standards and improving metadata - according to sector norms - would provide a big improvement to the applicability of data for reuse.

One internal data sharing example was provided where a list of the data to be shared and the standards for reporting that data have provided input to a very informative and visually appealing dashboard. Working backwards from the particular use cases

---

<sup>7</sup> Work produced by the ODI in support of a WHO health data summit in 2021

<https://www.theodi.org/project/data-landscape-and-health-data-governance-for-the-world-health-organization-who/>

captured displayed by the dashboard was seen as a focused way of developing internal standards.

- Tagging data sets through Machine Learning  
Identified 'relevant' datasets can be tagged and used as a 'learning' reference to help identify additional relevant datasets.

### *Communications and Social Media Support*

#### *Social Media*

The ODI has supported project communications through the social media platforms LinkedIn and Twitter. The ODI has retweeted and reshared original KCL posts as they have been identified. In addition the ODI has created original social media content for the project.

Social: 26709 impressions, 434 engagements

#### *Blog posting on Website and WID*

The ODI created a blog describing the aims and the approach for the project, [Transforming mental health research through longitudinal data sets](#). This is displayed on the ODI website and was featured as an article in our weekly newsletter The Week in Data (TWID).

Statistics (23.01.2023)

Web Page views: 172 (factoring in that the new webpage has only been up for 2ish weeks)

The Week in Data: featured in a TWID that had 6163 opens, with 56 clicks on the project page link

## Annex 3. Datasets included in each pocket of value

### 1. Richness in mental health measures

#### Strong measurement of depression and anxiety (N = 55)

1. [23andMe - The Affective disorders, Environment, and Cognitive Trait \(AFFECT\) study](#)
2. [Adolescent Brain Cognitive Development \(ABCD\) Study](#)
3. [Akrivia](#)
4. [All Our Families Study](#)
5. [Amsterdam Born Children and their Development](#)
6. [Assessment from Preschool to Puberty— Longitudinal Epidemiological \(APPLE\) study in Hirosaki, Japan](#)
7. [Australian Temperament Project](#)
8. [Avon Longitudinal Study of Parents and Children \(ALSPAC\)](#)
9. [Bachpan Cohort](#)
10. [Born in Bradford](#)
11. [Born in Bradford: Better Start](#)
12. [Born in Guangzhou Cohort Study \(BIGCS\)](#)
13. [California Families Project \(CFP\)](#)
14. [Concord Health and Ageing in Men Project - CHAMP](#)
15. [Copenhagen Child Cohort Study \(CCC2000\)](#)
16. [DaTravV datasets](#)
17. [Dunedin Study](#)
18. [Dutch 'Tracking Adolescents' Individual Lives' Survey'; TRAILS](#)
19. [Environmental \(E-Risk\) Risk Longitudinal Twin Study](#)
20. [Findata datasets](#)
21. [French E3N Cohort Study](#)
22. [Generation R](#)
23. [Generation Scotland \(incl STRADL\)](#)
24. [Growing Up in Scotland: Birth Cohort 1](#)
25. [Health and Employment After Fifty \(HEAF\) Cohort](#)
26. [iPSYcare](#)
27. [Lifelines](#)
28. [Mental Health Services Data Set \(MHSDS\)](#)
29. [Millennium Cohort Study \(MCS\)](#)
30. [Mom2B](#)
31. [National Child Development Study \(NCDS\)](#)
32. [National Survey of Health and Development \(NSHD\)](#)
33. [NEMESIS-II](#)
34. [Netherlands Study of Depression and Anxiety \(NESDA\)](#)
35. [Newcastle 85+ Study](#)
36. [Next Steps \(fka LSYPE 1\)](#)
37. [NNDC \(National Network of Depression Centers\) Cohort](#)
38. [Norwegian Mother and Child Cohort Study \(MoBa\)](#)
39. [Our Future \(LSYPE 2\)](#)
40. [Pelotas \(Brazil\) Birth Cohort Study - 2015](#)
41. [Personality and Total Health Through Life](#)
42. [Quebec Longitudinal Study of Child Development](#)
43. [Resilience, Ethnicity & Adolescent Mental Health \(REACH\)](#)
44. [Stratify](#)
45. [Study of Women's Health Across the Nation \(SWAN\)](#)
46. [Taiwan Birth Cohort Study \(TBCS\)](#)
47. [The Irish Longitudinal Study on Ageing \(TILDA\)](#)
48. [Trondheim Early Secure Study \(TESS\)](#)
49. [Twins Early Development Study \(TEDS\)](#)
50. [UK Biobank](#)
51. [UKHLS: Understanding Society & BHPS](#)
52. [Western Australian schizophrenia high-risk e-Cohort](#)
53. [Whitehall II](#)
54. [Wirral Child Health and Development Study](#)
55. [Young Lives Cohort](#)



### Valuable resources for psychosis research (N = 44)

1. [7T GluCEST](#)
2. [Accelerating Medicines Partnership Program - Schizophrenia \(AMP-SCZ\)](#)
3. [Adolescent Brain Cognitive Development \(ABCD\) Study](#)
4. [Aetiology and Ethnicity in Schizophrenia & Other Psychoses \(AESOP-10\)](#)
5. [Akrivia](#)
6. [Australian Temperament Project](#)
7. [Butajira Course and Outcome Study for Severe Mental Illness](#)
8. [Copenhagen Child Cohort Study \(CCC2000\)](#)
9. [Danish High Risk and Resilience Study](#)
10. [DaTravV datasets](#)
11. [Enhancing Neuroimaging Genetics through Meta-Analysis \(ENIGMA\)](#)
12. [Estonian Biobank of the Estonian Genome Center](#)
13. [European network of national schizophrenia networks studying gene-environment interactions \(EU-GEI\)](#)
14. [Feasibility and Acceptability of a Smartphone App to Assess Early Warning Signs of Psychosis Relapse](#)
15. [Fepsy early detection of psychosis](#)
16. [Findata datasets](#)
17. [First Episode of Psychosis cohort](#)
18. [GAMMA-F Study](#)
19. [Generation Scotland \(incl STRADL\)](#)
20. [GROUP \(Genetic Risk and Outcome of Psychosis\)](#)
21. [iBerry study: a longitudinal cohort study of adolescents at high risk of psychopathology](#)
22. [International Research Programme on Psychosis in Diverse Settings \(INTREPID\)](#)
23. [iPSYcare](#)
24. [Kungsholmen Project](#)
25. [LifeLines](#)
26. [Mauritius Child Health project](#)
27. [McLean OnTrack \(psychosis\)](#)
28. [Mental Health Services Data Set \(MHSDS\)](#)
29. [National quality register for psychosis care \(PsykosR\)](#)
30. [Netherlands Mental Health Survey and Incidence Study \(NEMESIS-II\)](#)
31. [Netherlands Mental Health Survey and Incidence Study \(NEMESIS\)](#)
32. [New England Family Study](#)
33. [North American Prodrome Longitudinal Study \(NAPLS\)](#)
34. [Our Future Health](#)
35. [ProNET](#)
36. [PsyCymru](#)
37. [SCOPE: Studying the sociocultural context of psychosis in Ethiopia to improve outcomes](#)
38. [Stratify](#)
39. [Trajectories and Predictors in the Clinical High Risk for Psychosis Population: Prediction Scientific Global Consortium \(PRESCIENT\)](#)
40. [UK BioBank](#)
41. [UKHLS: Understanding Society & BHPS](#)
42. [UP'S: A Cohort Study on Recovery in Psychotic Disorder Patients](#)
43. [Western Australian schizophrenia high-risk e-Cohort](#)
44. [YHCD \(Youth Health Care Divisions\)](#)

### Other mental health conditions (N = 30)

1. [AcuteLines](#)
2. [ADVANCE Armed Services Trauma Rehabilitation Outcome Study](#)
3. [Afghanistan Mental Health and Childhood Adversities](#)
4. [Airwave Health Monitoring Study](#)
5. [Australian Gulf War Veterans' Health Study Cohort](#)
6. [Biological pathways of risk and resilience in Syrian refugee children \(BIOPATH\)](#)
7. [Boricua Youth Study](#)
8. [Brisbane Longitudinal Twin Study](#)
9. [Building a New Life in Australia \(BNLA\)](#)
10. [Edinburgh Study of Youth Transitions and Crime](#)
11. [FAMILY Cohort](#)

12. [Flint Adolescents Study \(FAS\)](#)
13. [Impact of the Treatment of PTSD on Comorbid Insomnia and Pain](#)
14. [King's Cohort \(UK\)](#)
15. [LOGIC \(Longitudinal Outcomes of Gender Identity in Children\)](#)
16. [Longitudinal Changes in Posttraumatic Stress Disorder After Resettlement Among Yazidi Female Refugees Exposed to Violence](#)
17. [Michigan State University Twin Registry \(MSUTR\)](#)
18. [Mid Atlantic Twin Registry](#)
19. [Millennium Cohort Study \(USA military\)](#)
20. [Million Veteran Program, USA](#)
21. [Mom2B Cohort](#)
22. [National Research Council Twin Registry \(NAS-NRC Twin Registry\)](#)
23. [Prechter Longitudinal Study of Bipolar Disorder](#)
24. [Project Ice Storm](#)
25. [Seeding Success data resources](#)
26. [Stress in Pregnancy \(SIP\) Study](#)
27. [Taiwan Birth Cohort Study \(TBCS\)](#)
28. [The Project Positive Adolescent Training through Holistic Social \(PATHS\) Programs](#)
29. [WELL Living Laboratory in China \(WELL-China\)](#)
30. [Wenchuan Earthquake Adolescent Health Cohort Study](#)

### **Mental health across the life course (N = 82)**

1. [3D Cohort Study \(Design, Develop, Discover\) \(3D\)](#)
2. [Airwave Health Monitoring Study](#)
3. [All Our Families Study](#)
4. [Amsterdam Born Children and their Development \(ABCD\) Study](#)
5. [Australian Longitudinal Study of Women's Health](#)
6. [Avon Longitudinal Study of Parents and Children \(ALSPAC\)](#)
7. [Biobank Japan](#)
8. [Birth to Twenty Plus \(Birth to Thirty\)](#)
9. [Born in Guangzhou Cohort Study \(BIGCS\)](#)
10. [Boston University Twin Project \(BUTP\)](#)
11. [Brisbane Longitudinal Twin Study](#)
12. [British Cohort Study 1970 \(BCS70\)](#)
13. [British Columbia Generations Project \(BCGP\)](#)
14. [California Families Project \(CFP\)](#)
15. [Canadian DOHaD Cohort Registry](#)
16. [CHILD Cohort Study](#)
17. [Child Health and Development Study \(CHDS\)](#)
18. [Continuation of Dating It Safe: A Longitudinal Study on Teen Dating Violence](#)
19. [Costa Rican Longevity and Healthy Aging Study](#)
20. [Course and Outcome of Bipolar Youth \(COBY\)](#)
21. [Dunedin Study](#)
22. [Dutch 'TRacking Adolescents' Individual Lives' Survey', TRAILS](#)
23. [English Longitudinal Study of Ageing \(ELSA\)](#)
24. [Environmental Risk Longitudinal Twin Study \(E-Risk\)](#)
25. [FinnTwin12](#)
26. [Great Smoky Mountain Study \(GSMS\)](#)
27. [Growing Up in Australia \(the Longitudinal Study of Australian Children\)](#)
28. [Growing Up in New Zealand](#)
29. [Growing Up in Quebec](#)
30. [Growing Up Today Study](#)
31. [Harvard Pilgrim Health Care Institute Project Viva](#)
32. [Harvard Study of Adult Development](#)
33. [Hispanic Community Health Study/Study of Latinos \(HCHS/SOL\)](#)
34. [Jamaican 1986 Birth Cohort Study](#)
35. [Kids, Families and Places](#)
36. [lidA Cohort Study](#)
37. [Longitudinal Study of Generations](#)
38. [Longitudinal Surveys of Australian Youth](#)
39. [Mater-University Study of Pregnancy \(MUSP\)](#)
40. [Mauritius Child Health project](#)
41. [Mayo Clinic Biobank](#)
42. [Michigan Twins Study \(MTwins\)](#)
43. [Mid Atlantic Twin Registry](#)

44. Millenium Cohort Study (USA veterans)
45. Million Veteran Program, USA
46. Montreal Heart Institute (MHI) Biobank
47. National Child Development Study (NCDS)
48. National Longitudinal Study of Adolescent Health (ADD Health)
49. National Longitudinal Survey of Mature and Young Women (NLSW)
50. National Longitudinal Survey of Older and Young Men (NLSM)
51. National Longitudinal Survey of Youth (NLSY79)
52. National Longitudinal Survey of Youth (NLSY97)
53. National Longitudinal Survey of Youth 1979 Child and Young Adult (NLSCYA)
54. National Research Council Twin Registry (NAS-NRC Twin Registry)
55. Netherlands Study of Depression and Anxiety (NESDA)
56. New England Family Study
57. New South Wales Child Development Study (NSW-CDS)
58. Next Steps (formerly known as LSYPE 1)
59. Norwegian Mother and Child Cohort Study (MoBa)
60. Nurses' Health Study II
61. Nurses Health Study III
62. Nurses' Health Study (NHS) I
63. Our Future (LSYPE 2)
64. Pelotas (Brazil) Birth Cohort Study - 1982
65. Pelotas (Brazil) Birth Cohort Study - 1993
66. Personality and Total Health Through Life
67. Pittsburgh Girls Study
68. Project Talent Twin and Sibling Study
69. RADAR old cohort
70. RADAR young cohort
71. Raine Study
72. Reykjavik Study
73. Study of Health in Pomerania
74. Taiwan Birth Cohort Study (TBCS)
75. Ten to Men (the Australian Longitudinal Study on Male Health)
76. Twins Early Development Study (TEDS)
77. TwinsUK
78. UKHLS: Understanding Society & BHPS
79. Victorian Adolescent Health Cohort Study
80. Wirral Child Health and Development Study
81. Young Lives Cohort (Combined)
82. Zurich Longitudinal Studies

## **2. Value in targeted populations**

### **High-risk groups (N = 27)**

1. Accelerating Medicines Partnership Program - Schizophrenia (AMP-SCZ)
2. Adolescents at High Risk for Familial Bipolar Disorder
3. ADVANCE Armed Services Trauma Rehabilitation Outcome Study
4. Alcohol and Other Drug Treatment Services National Minimum Data Set (AODTS NMDS)
5. Australian Gulf War Veterans' Health Study Cohort
6. Bachpan Cohort
7. Biological pathways of risk and resilience in Syrian refugee children (BIOPATH)
8. Boston Hospital Workers Health Study (BHWHS)
9. Brazilian High Risk Cohort Study
10. British Autism Study of Infant Siblings (BASIS)
11. Child Protection National Minimum Data Set (CP NMDS)
12. C-MaMiE: Child outcomes in relation to Maternal Mental disorders in Ethiopia
13. Danish High Risk and Resilience Study
14. Desenvolvimentu Isin-d'iak Labarik no Inan (DILI) birth cohort study
15. Ethiopia Mental Health Study
16. Etiological and Prospective Family Study in Southwestern Pennsylvania
17. Health and Wellbeing of UK Armed Forces Personnel: A Cohort Study
18. iBerry study: a longitudinal cohort study of adolescents at high risk of psychopathology

19. Longitudinal Changes in Posttraumatic Stress Disorder After Resettlement Among Yazidi Female Refugees Exposed to Violence
20. Longitudinal Studies on Child Abuse and Neglect (LONGSCAN)
21. Michigan Longitudinal Study (UMICH)
22. Millennium Cohort Study (USA military)
23. National Youth in Transition Database (NYTD)
24. NHS-Check
25. Trajectories and Predictors in the Clinical High Risk for Psychosis Population: Prediction Scientific Global Consortium (PRESCIENT)
26. Western Australian schizophrenia high-risk e-Cohort
27. Wirral Child Health and Development Study

### **Under-represented groups (N = 23)**

1. Adolescent Brain Cognitive Development (ABCD) Study
2. Aetiology and Ethnicity in Schizophrenia & Other Psychoses (AESOP-10)
3. Boricua Youth Study
4. California Families Project (CFP)
5. Cambridge Centre for Attention, Learning and Memory
6. Ethiopia Mental Health Study
7. Family and Community Health Study FACHS (FACHS)
8. First Episode of Psychosis cohort
9. Fragile Families & Child Wellbeing Study
10. Generations: A Study of the Life and Health of LGB People in a Changing Society
11. Healthy Living in an Urban Setting (HELIUS)
12. Healthy Young Men's (HYM) study
13. Hispanic Community Health Study/Study of Latinos (HCHS/SOL)
14. IMAGINE ID
15. Impact of the intrauterine and early childhood environments on neurocognitive and metabolic development in African American youth: focus on the gut-brain axis
16. Life and Living in Advanced Age: a Cohort Study in New Zealand (LiLACS NZ)
17. LOGIC (Longitudinal Outcomes of Gender Identity in Children)
18. Millennium Cohort Study (MCS)
19. Peru Migrant Study
20. Resilience, Ethnicity & Adolescent Mental Health (REACH)
21. Seeding Success data resources
22. Texas Department of Criminal Justice
23. Trans Youth of Color Study

## Populations from under-represented locations (N = 15)

### **Africa, the Middle East & Asia**

1. Alliance for Maternal and Newborn Health Improvement (AMANHI) biobanking study
2. Afghanistan Mental Health and Childhood Adversities
3. Desenvolvimentu Isin-d'iak Labarik no Inan (DILI) birth cohort study
4. LIFE course study in CARdiovascular disease Epidemiology (LIFECARE)
5. Malawi Longitudinal Study of Families and Health (MLSFH)
6. Mauritius Child Health project
7. Performance Monitoring for Action (PMA) project
8. West Africa Birth Cohort Study

### **Americas**

9. Growth and Obesity Chilean Cohort Study (ECO)
10. Jamaican 1986 Birth Cohort Study
11. Jamaican 2011 Birth Cohort Study
12. Peru Migrant Study

### **Europe & the Pacific**

13. Growing Up in Hungary
14. Reykjavik Study

### **Cross-regional**

15. Young Lives Cohort

## Ageing populations (N = 32)

1. 10 K: a large-scale prospective longitudinal study in Israel
2. 45 and Up Study
3. A Population Based Prospective Cohort Study to Unravel the Causes of Stroke and Cognitive Decline: A Cross-Cultural Perspective
4. Amirkola Health and Ageing Project (AHAP)
5. Australian Longitudinal Study of Women Health
6. Brazilian longitudinal study of aging [ELSA-Brazil]
7. British Cohort Study 1970 (BCS70)
8. Canadian Longitudinal Study on Aging (CLSA)
9. China longitudinal ageing study
10. Cognitive Health in Ageing Register: Investigational, Observational, and Trial studies in dementia research (CHARIOT): Prospective Readiness cOhort Study (PRO) Sub Study
11. Concord Health and Ageing in Men Project - CHAMP
12. Dunedin Study
13. Edad con Salud
14. English Longitudinal Study of Ageing (ELSA)
15. ESPRIT Project
16. Health and Ageing in Africa: A Longitudinal Study of an INDEPTH Community in South Africa (HAALSI)
17. Health and Employment After Fifty Study (HEAF)
18. Health and Retirement Study (HRS)
19. Health, Aging and Retirement Transitions in Sweden - HEARTS
20. Irish Longitudinal Study on Ageing (TILDA)
21. LoCARPoN
22. Mexican Health and Aging Study
23. National Child Development Study (NCDS)
24. National Social Life, Health, and Ageing Project (NSHAP)
25. National Survey of Health and Development (NSHD)
26. NICOLA
27. Panel on Health and Ageing of Singaporean Elderly (PHASE)
28. Personality and Total Health Through Life
29. PROTECT Study
30. Survey of Health, Ageing and Retirement in Europe
31. The 90+ Study
32. Whitehall II

### **Value in the long run (N = 18)**

1. [Accelerating Medicines Partnership Program - Schizophrenia \(AMP-SCZ\)](#)
2. [Adolescent Health Study](#)
3. [All of Us Research Program](#)
4. [Alliance for Maternal and Newborn Health Improvement \(AMANHI\) biobanking study](#)
5. [Assessment from Preschool to Puberty— Longitudinal Epidemiological \(APPLE\) study in Hirosaki, Japan](#)
6. [British Autism Study of Infant Siblings \(BASIS\)](#)
7. [Children of the 2020s](#)
8. [Generation R Next](#)
9. [Growing Up in Hungary](#)
10. [GUIDE \(Growing Up In Digital Europe: EuroCohort\)](#)
11. [Healthy Brains Network](#)
12. [Mutaba'ah—Mother and Child Health Study](#)
13. [Ontario Birth Study](#)
14. [Our Future Health](#)
15. [Qatar Birth Cohort Study](#)
16. [SCOPE: Studying the sociocultural context of psychosis in Ethiopia to improve outcomes](#)
17. [Shanghai Children's Health, Education and Lifestyle Evaluation, Preschool \(SCHEDULE-P\) study](#)
18. [Tohoku Medical Megabank Project Birth and Three-Generation Cohort Study \(TMM-BirThree Cohort Study\)](#)

### **3. Diversity of data**

#### **Factors contributing to mental health conditions (N = 25)**

1. [3D Cohort Study \(Design, Develop, Discover\) \(3D\)](#)
2. [Adolescent Brain Cognitive Development \(ABCD\) Study](#)
3. [All Our Families Study](#)
4. [Amsterdam Born Children and their Development \(ABCD\) Study](#)
5. [Avon Longitudinal Study of Parents and Children \(ALSPAC\)](#)
6. [Birth to Twenty Plus \(Birth to thirty\)](#)
7. [Born in Bradford](#)
8. [Child and Adolescent Twin Study in Sweden \(CATSS\)](#)
9. [Christchurch Health and Development Study](#)
10. [Danish National Birth Cohort](#)
11. [Dunedin Study](#)
12. [Environmental Risk \(E-Risk\) Longitudinal Twin Study](#)
13. [Generation R](#)
14. [Great Smoky Mountain Study \(GSMS\)](#)
15. [Growing Up in Australia \(the Longitudinal Study of Australian Children\)](#)
16. [Growing Up in Hungary](#)
17. [Italian Twin Registry](#)
18. [ORIGINS Project](#)
19. [Raine Study](#)
20. [Swedish Twin Registry](#)
21. [Triple B Pregnancy Cohort Study](#)
22. [Trondheim Early Secure Study \(TESS\)](#)
23. [Twins Early Development Study \(TEDS\)](#)
24. [Victorian Adolescent Health Cohort Study](#)
25. [Western Australian schizophrenia high-risk e-Cohort](#)

### Neuroimaging data (N = 17)

1. [Adolescent Brain Cognitive Development \(ABCD\) Study](#)
2. [Brain-MRI Project](#)
3. [Danish Developmental studies - HUBU \("Hjernens Udvikling hos Børn og Unge": Brain maturation in children and adolescents\)](#)
4. [Effects of ECT in treatment of depression: study protocol for a prospective neuroradiological study of acute and longitudinal effects on brain structure and function \(The GEMRIC Bergen, Norway center\)](#)
5. [Electroconvulsive therapy-induced brain plasticity determines therapeutic outcome in mood disorders \(The GEMRIC centers in Lausanne, Switzerland and Berlin, Germany\)](#)
6. [Enhancing Neuroimaging Genetics through Meta-Analysis \(ENIGMA\)](#)
7. [FinnBrain Birth Cohort Study](#)
8. [Generation R](#)
9. [Generation R Next](#)
10. [Healthy Brain Network](#)
11. [IMAGEN Study](#)
12. [Jerusalem fMRI](#)
13. [Nathan Kline Institute – Rockland Sample \(NKI-RS\)](#)
14. [National Consortium on Alcohol and Neurodevelopment in Adolescence](#)
15. [Reproducible Imaging-Based Brain Growth Charts for Psychiatry](#)
16. [Tohoku Medical Megabank Project Birth and Three-Generation Cohort Study \(TMM BirThree Cohort Study\)](#)
17. [UK Biobank](#)

### Wearables and phone apps (N = 22)

1. [Adolescent Brain Cognitive Development \(ABCD\) Study](#)
2. [Amsterdam Born Children and their Development \(ABCD\) Study](#)
3. [Children of the 2020s](#)
4. [Cognitive and Mood Assessment Data in Major Depressive Disorder Using Digital Wearable Technology](#)
5. [Feasibility and Acceptability of a Smartphone App to Assess Early Warning Signs of Psychosis Relapse](#)
6. [Future Proofing Study](#)
7. [GAMMA-F Study](#)
8. [GEMRIC Linköping, Sweden center](#)
9. [GEMRIC Utrecht, the Netherlands center](#)
10. [Generation R](#)
11. [Healthy Brain Network](#)
12. [Mello Study](#)
13. [Mom2B](#)
14. [Mood Disorder Cohort Research Consortium \(MDCRC\)](#)
15. [Net4Health](#)
16. [Next-Generation Psychiatric Assessment: Using Smartphone Sensors to Monitor Behavior and Mental Health](#)
17. [NIMH Intramural Longitudinal Study of the Endocrine and Neurobiological Events Accompanying Puberty](#)
18. [Personality and Total Health Through Life](#)
19. [Regional brain volumes, diffusivity, and metabolite changes after electroconvulsive therapy for severe depression \(The GEMRIC Copenhagen, Denmark center\)](#)
20. [Remote Assessment of Disease and Relapse in Major Depressive Disorder \(RADAR-MDD\): A Multi-Centre Prospective Cohort Study Protocol](#)
21. [Stony Brook Temperament Study](#)
22. [Stratify](#)

## Biological and genetic data (N = 67)

1. [23andMe - The Affective disorders, Environment, and Cognitive Trait \(AFFECT\) study](#)
2. [Accelerating Medicines Partnership Program - Schizophrenia \(AMP-SCZ\)](#)
3. [AcuteLines](#)
4. [Adolescent Brain Cognitive Development \(ABCD\) Study](#)
5. [Alberta's Tomorrow Project](#)
6. [All of Us Research Program](#)
7. [All Our Families Study](#)
8. [Alliance for Maternal and Newborn Health Improvement \(AMANHI\) biobanking study](#)
9. [Auria Biobank](#)
10. [Australian Temperament Project](#)
11. [Avon Longitudinal Study of Parents and Children \(ALSPAC\)](#)
12. [Baby ELEFANT- Environmental and LifeStyle Factors in metabolic health throughout life-course Trajectories](#)
13. [Barts BioResources](#)
14. [Biobank of Eastern Finland](#)
15. [Borealis Biobank of Northern Finland](#)
16. [Born in Bradford](#)
17. [Born in Guangzhou Cohort Study \(BIGCS\)](#)
18. [Boston University Twin Project \(BUTP\)](#)
19. [Canadian Healthy Infant Longitudinal Development \(CHILD\) Study](#)
20. [CARTaGENE \(CaG\)](#)
21. [Central Finland Biobank](#)
22. [Children's Hospital of Philadelphia \(CHOP\) Biorepository](#)
23. [Copenhagen Aging and Midlife Biobank \(CAMB\)](#)
24. [Dutch 'Tracking Adolescents' Individual Lives' Survey'; TRAILS](#)
25. [Enhancing Neuroimaging Genetics through Meta-Analysis \(ENIGMA\)](#)
26. [Environmental Risk Longitudinal Twin Study \(E-Risk\)](#)
27. [Estonian Biobank of the Estonian Genome Center](#)
28. [Finnish Clinical Biobank Tampere](#)
29. [Generation R](#)
30. [Generation Scotland](#)
31. [Genes and Health \(ELGH\)](#)
32. [Genetic and Environmental Predictors of Combat-Related PTSD](#)
33. [Great Smoky Mountain Study \(GSMS\)](#)
34. [GROUP \(Genetic Risk and Outcome of Psychosis\)](#)
35. [Growing Up in Singapore Towards healthy Outcomes \(GUSTO\) birth cohort study](#)
36. [Growing Up Today Study](#)
37. [Health Outcomes and Measures of the Environment \(HOME\) study](#)
38. [Helsinki Biobank](#)
39. [Kaiser Permanente, including KP Research Program on Genes, Environment and Health \(RGEH\) - in the KP Research Bank](#)
40. [Kids, Families and Places](#)
41. [LifeLines](#)
42. [Mayo Clinic Biobank](#)
43. [Michigan Twins Study \(MTwins\)](#)
44. [Mid Atlantic Twin Registry](#)
45. [Millennium Cohort Study \(MCS\)](#)
46. [Minnesota Twin Family Study \(MTFS\)](#)
47. [Montreal Heart Institute \(MHI\) Biobank](#)
48. [National Child Development Study \(NCDS\)](#)
49. [National Longitudinal Study of Adolescent Health \(ADD Health\)](#)
50. [Northern Ireland Cohort for the Longitudinal Study of Ageing \(NICOLA\)](#)
51. [Norwegian Mother and Child Cohort Study \(MoBa\)](#)
52. [Norwegian Twin Panel](#)
53. [Nurses' Health Study II](#)
54. [Nurses' Health Study III](#)
55. [Ontario Birth Study](#)
56. [Ontario Health Study \(OHS\)](#)
57. [PsyCymru](#)
58. [Qatar Biobank population based cohort study](#)
59. [Sequence Bio - NL Genome Project \(pilot study\)](#)



60. Signature Bank - Banks of Biological, Psychosocial and Clinical Data on People Suffering from Mental Health Problems
61. Taiwan Biobank
62. THL Biobank

63. Tohoku Medical Megabank Community Cohort Study
64. Twins Early Development Study (TEDS)
65. UK Biobank
66. UMCG Biobank
67. Wisconsin Twin Research

#### **Routinely collected data (N = 36)**

1. 45 and Up Study
2. Administrative Data Research (ADR) UK - Northern Ireland
3. Agenzia Regionale di Sanità della Toscana (ARS)
4. Akrivia
5. All of Us Research Program
6. Atlantic Partnership for Tomorrow's Health Study, or Atlantic PATH
7. Born in Guangzhou Cohort Study (BIGCS)
8. British Columbia Generations Project (BCGP)
9. Centre for Data and Knowledge Integration for Health (CIDACS)
10. Clinical Record Interactive Search System (CRIS)
11. DANLIFE: Danish LifeCourse Study
12. Growing Up in Australia (the Longitudinal Study of Australian Children)
13. Health and Employment After Fifty (HEAF) Cohort
14. Kaiser Permanente, including KP Research Program on Genes, Environment and Health (RGEH) - in the KP Research Bank
15. lidA Cohort Study
16. Lifelines
17. Lifelines NEXT
18. Manitoba Multigenerational Cohort
19. Mental Health Services Data Set (MHSDS)
20. National Income Dynamics Study (NIDS)
21. Next Steps (formerly known as LSYPE 1)
22. Norwegian Life Course, Ageing and Generation Study (NorLAG)
23. Norwegian Mother and Child Cohort Study (MoBa)
24. Our Future (LSYPE 2)
25. Our Future Health
26. PsyCymru
27. Puberty Cohort in the Danish National Birth Cohort (DNBC)
28. Seeding Success data resources
29. Signature Bank - Banks of Biological, Psychosocial and Clinical Data on People Suffering from Mental Health Problems
30. Swedish Longitudinal Occupational Survey of Health (SLOSH)
31. Texas Department of Criminal Justice
32. Tohoku Medical Megabank Project Birth and Three-Generation Cohort Study (TMM BirThree Cohort Study)
33. TwinsUK
34. UK Biobank
35. Western Australian schizophrenia high-risk e-Cohort
36. Young Netherlands Twin Register (YNTR)

#### **4. Mental health embedded in wider context**

##### **Mental health within social context (N = 39)**

1. [Americans' Changing Lives](#)
2. [Belfast Youth Development Study](#)
3. [Born in Bradford: Better Start](#)
4. [Building a New Life in Australia \(BNLA\)](#)
5. [California Families Project \(CFP\)](#)
6. [Centre for Data and Knowledge Integration for Health \(CIDACS\)](#)
7. [China Family Panel Studies \(CFPS\)](#)
8. [China Health and Nutrition Survey](#)
9. [Chinese Longitudinal Healthy Longevity and Happy Family Study \(CLHLS-HF\)](#)
10. [Chitwan Valley Family Study](#)
11. [Family and Community Health Study \(FACHS\)](#)
12. [Gazel Cohort](#)
13. [German Ageing Survey \(DEAS\)](#)
14. [Growing Up in Australia \(the Longitudinal Study of Australian Children\)](#)
15. [Growing Up in Hungary](#)
16. [Health and Ageing in Africa: A Longitudinal Study of an INDEPTH Community in South Africa \(HAALSI\)](#)
17. [Indonesian Family Life Survey \(IFLS\)](#)
18. [Integrated Data Infrastructure](#)
19. [Journeys Home: a Longitudinal Study of Factors Affecting Housing Stability](#)
20. [lidA Cohort Study](#)
21. [Longitudinal Surveys of Australian Youth](#)
22. [Malawi Longitudinal Study of Families and Health \(MLSFH\)](#)
23. [Mexican Family Life Survey - Mexican Health and Aging Study \(MHAS\)](#)
24. [Millennium Cohort Study \(MCS\)](#)
25. [National Income Dynamics Study \(NIDS\)](#)
26. [Netherlands Autism Register](#)
27. [Norwegian Life Course, Ageing and Generation Study \(NorLAG\)](#)
28. [Our Future \(LSYPE 2\)](#)
29. [Panel Study of Income Dynamics \(PSID\)](#)
30. [Performance Monitoring for Action \(PMA\) project](#)
31. [Social Protection Survey \[Encuesta de Previsión Social\]](#)
32. [Survey of Income and Program Participation Series](#)
33. [Swedish Longitudinal Occupational Survey of Health \(SLOSH\)](#)
34. [Taiwan Birth Cohort Study \(TBCS\)](#)
35. [Ten to Men \(the Australian Longitudinal Study on Male Health\)](#)
36. [UKHLS: Understanding Society & BHPS](#)
37. [Understanding the lives of adolescents and young adults \(UDAYA\)](#)
38. [West of Scotland Twenty-07](#)
39. [Young Lives Cohort \(COMBINED\)](#)

##### **The impact of the COVID-19 pandemic (N = 28)**

1. [Adolescent Brain Cognitive Development \(ABCD\) Study](#)
2. [All of Us Research Program](#)
3. [Bachpan cohort](#)
4. [Born in Bradford](#)
5. [British Cohort Study 1970 \(BCS70\)](#)
6. [Chilean Maternal & Infant Cohort Study II \(CHIMINCS-II\)](#)
7. [COVICAT - CONTENT cohort](#)
8. [COVID-19 Well-Being Study](#)
9. [Danish National Birth Cohort \(DNBC\) - COVID-19 collection](#)
10. [Edad con Salud](#)
11. [Generation Scotland \(incl STRADL\)](#)
12. [HealthWise Wales](#)
13. [IMAGINE ID](#)
14. [Impact of Covid-19 on the health of physicians, nurses and other healthcare workers: an interprovincial cohort study](#)
15. [The Irish Longitudinal Study on Ageing \(TILDA\)](#)
16. [LifeLines](#)
17. [Loddon Mallee Healthcare Worker Covid Study](#)
18. [Mental Health Million Project](#)
19. [Millennium Cohort Study \(MCS\)](#)

- |  |  |
|--|--|
| 20. <a href="#">National Child Development Study (NCDS)</a>          | 25. <a href="#">Rotterdam Study</a>  |
| 21. <a href="#">National Survey of Health and Development (NSHD)</a> | 26. <a href="#">Santé, pratiques, relations et inégalités sociales en population générale pendant la crise COVID-19 (SAPRIS)</a> |
| 22. <a href="#">Next Steps (formerly known as LSYPE 1)</a>           | 27. <a href="#">UKHLS: Understanding Society &amp; BHPS</a>  |
| 23. <a href="#">NHS-Check</a>  | 28. <a href="#">Young Lives Cohort</a>   |
| 24. <a href="#">Pelotas (Brazil) Birth Cohort Study - 2015</a>       |  |

### Connectivity between datasets (N = 22)

- |  |   |
|--|---|
| 1. <a href="#">Accelerating Medicines Partnership Program - Schizophrenia (AMP-SCZ)</a>                      | 10. <a href="#">Enhancing NeuroImaging Genetics through Meta-Analysis (ENIGMA)</a>            |
| 2. <a href="#">Ageing Trajectories of Health – Longitudinal Opportunities and Synergies (ATHLOS) project</a> | 11. <a href="#">Global ECT-MRI Research Collaboration (GEMRIC)</a>                            |
| 3. <a href="#">All of Us Research Program</a>  | 12. <a href="#">Integrated Data Infrastructure</a>  |
| 4. <a href="#">Asian Pharmacoepidemiology Network (AsPEN)</a>  | 13. <a href="#">Net4Health</a>  |
| 5. <a href="#">CanPath, including Alberta’s Tomorrow Project</a>   | 14. <a href="#">Next Steps and Our Future</a>   |
| 6. <a href="#">Collaborating Consortium of Cohorts Producing NIDA Opportunities (C3PNO)</a>                  | 15. <a href="#">NICOLA</a>  |
| 7. <a href="#">Dutch ‘TRacking Adolescents’ Individual Lives’ Survey’, TRAILS</a>                            | 16. <a href="#">Norwegian Twin Registry</a>   |
| 8. <a href="#">Dynamic Analyses to Optimize Ageing (DYNOPTA) Project</a>                                     | 17. <a href="#">Project ELEFANT</a>   |
| 9. <a href="#">E3N and E3N-Generations Cohorts</a>   | 18. <a href="#">PsyCymru</a>  |
|  | 19. <a href="#">Research Advancement through Cohort Cataloguing and Harmonization (ReACH)</a> |
|  | 20. <a href="#">Survey of Health, Ageing and Retirement in Europe</a>                         |
|  | 21. <a href="#">UK Biobank</a>  |
|  | 22. <a href="#">UMCG Biobank</a>  |

### Natural disasters and geopolitical factors (N = 4)

- |  |   |
|--|---|
| 1. <a href="#">Afghanistan Mental Health and Childhood Adversities</a>                             | 3. <a href="#">Stress in Pregnancy (SIP) Study</a>                    |
| 2. <a href="#">Biological pathways of risk and resilience in Syrian refugee children (BIOPATH)</a> | 4. <a href="#">Wenchuan Earthquake Adolescent Health Cohort Study</a> |

### Interventions embedded in longitudinal datasets (N = 21)

- |   |   |
|---|---|
| 1. <a href="#">Accelerating Medicines Partnership Program - Schizophrenia (AMP-SCZ)</a>   | 6. <a href="#">Comparing Effectiveness of Treating Depression With &amp; Without Comorbidity to Improve Fetal Health</a>  |
| 2. <a href="#">Akrivia</a>  | 7. <a href="#">Early- and late-onset depression in late life: A prospective study on clinical and structural brain characteristics and response to Electroconvulsive Therapy (The GEMRIC Amsterdam, the Netherlands center)</a> |
| 3. <a href="#">All of Us Research Program</a>   |   |
| 4. <a href="#">Andhra Pradesh Children and Parents Study</a>  |   |
| 5. <a href="#">Bilateral ECT induces bilateral increases in regional cortical thickness (The GEMRIC Nijmegen, the Netherlands center)</a> |   |

8. Early Developmental Stages of Psychopathology Study
9. Effects of ECT in treatment of depression: study protocol for a prospective neuroradiological study of acute and longitudinal effects on brain structure and function (The GEMRIC Bergen, Norway center)
10. Electroconvulsive therapy-induced brain plasticity determines therapeutic outcome in mood disorders (The GEMRIC centers in Lausanne, Switzerland and Berlin, Germany)
11. Ethiopia Mental Health Study
12. GEMRIC Linköping, Sweden center
13. GEMRIC Utrecht, the Netherlands center
14. Grey matter volume increase following electroconvulsive therapy in patients with late life depression: a longitudinal MRI study (The GEMRIC Leuven, Belgium center)
15. Increase in hippocampal volume after electroconvulsive therapy in patients with depression: a volumetric magnetic resonance imaging study (The GEMRIC Lund, Sweden center)
16. McLean OnTrack (psychosis)
17. Measuring Health and Well-being of Young People in the Transfer Project
18. Regional brain volumes, diffusivity, and metabolite changes after electroconvulsive therapy for severe depression (The GEMRIC Copenhagen, Denmark center)
19. R-LINK: optimizing response to lithium treatment through personalized and multimodal evaluation of individuals with Bipolar I disorder
20. Yorkshire Health Study
21. Zambia Family (ZAMFAM) Project

