# MRC Mental Health Data Pathfinder initiative

In 2018, nine leading universities from across the UK received £10M in grants from the Medical Research Council (MRC) to help researchers harness the power of data science for mental health research. Funding for the grants was secured by the MRC as part of the government's National Productivity Investment Fund (NPIF).

Mental Health Data Pathfinder awards were designed to be used flexibly by the institution to fund preliminary-stage projects to support the development of a mental health data and informatics resource in the longer term. Funds supported a range of projects aiming to address key challenges or demonstrate exemplar studies that might be expected to contribute to a future data platform.

The Pathfinder initiative was aimed at addressing several of the following challenges:
- advanced data discoverability methodologies for research;
- methodologies for data linkage and integration of biological, psychological, clinical, behavioural, lifestyle-related, social, educational or other datasets.
- enabling interoperability between datasets, ethical access and data sharing, and support for data mining;
- harmonisation of measurement of mental disorders and symptoms in cohorts and longitudinal studies and measurement of the outcomes of interventions;
- enhancement of existing cohort and longitudinal studies, e.g. through annotation of mental health-related data, to accelerate progress towards making these data more accessible for research;
- incorporating a broader range of physical and mental health outcome measures in population research and clinical studies, recognising the important link between physical and mental health;
- Producing a feasible design for the next phase of platform development and its long-term sustainability;
- Demonstrating how activities will utilise and link to existing UK informatics infrastructure and resources

The Mental Health Pathfinder awardees are listed below. The grant reference can be used to explore publications arising from the projects using search tools such as Europe PMC.

| University, PI and grant reference | Pathfinder Summary |
|---|---|
| **University of Bristol** <br> *John Macleod* <br> **(MC_PC_17210)** | Cohorts as Platforms for Mental Health research (CaP:MH) – integrating the ALSPAC and Born in Bradford population cohorts to enable pooled analysis across cohorts, and extending data linkage to primary, secondary care and digital data. |
| **University of Cambridge** <br> *Rudolf Cardinal* <br> **(MC_PC_17213)** | Developing new tools for data linkage in mental health research; enhancing routine clinical data with structured psychiatric assessments; and using machine learning to develop early predictors of premature death in people with serious mental illness. |
| **Cardiff University** <br> *Jeremy Hall* <br> **(MC_PC_17212)** | Integrating genetic, clinical and phenotypic data to advance stratification, prediction and treatment in mental health; and enhancing the Welsh Schools Health Research Network to include adolescent mental health measures and linkage to routine health and administrative records. |

| University of Edinburgh *Andrew McIntosh* (**MC_PC_17209**) | Leveraging routinely collected and linked cohort research data to study the causes and consequences of common mental disorders. Creating a structured, secure informatics environment with enhanced mental health phenotyping, and enabling targeted re-contact of cohort participants for future mental health studies. |
|---|---|
| University of Glasgow *Daniel Smith* (**MC_PC_17217**) | Enhancing Scottish mental health cohorts through linkage to other health, education and administrative datasets; and establishing the feasibility of a Scottish Schools Health Research Network with a focus on adolescent mental health. |
| King's College London *Robert Stewart* (**MC_PC_17214**) | Advancing mental health informatics through new approaches to data extraction and enhancement; and working with teachers, adolescents and parents to pilot digital recruitment of school-based e-cohorts and health and education data linkage for mental health research. |
| University of Oxford *Clare Mackay* (**MC_PC_17215**) | Developing an informatics supported platform for experimental medicine, adding value to existing datasets and enriching data collection from patient-reported and remote monitoring measures. |
| Swansea University *Ann John* (**MC_PC_17211**) | MytHICAL- Mental Health Informatics in Children, Adolescents and young adults. The pathfinder builds on the MQ Adolescent Data Platform for mental health research and aims to build the infrastructure to link cohort and routinely collected data relevant to adolescent mental health. |
| University College London *David Osborn* (**MC_PC_17216**) | Advancing methodology and linkages in electronic health records, through using machine learning to add structure to mental health records, and improved linkage of primary and secondary care records, including child and adolescent mental health services. |

Summaries of the Pathfinder projects (as published on Gateway to Research)

**University of Bristol (**MC_PC_17210)

**PI: John Macleod**

**Cohorts as Platforms for Mental Health research (CaP:MH): Integrating the ALSPAC and Born in Bradford population cohorts.**

We will enhance two of the UK's premier population based birth cohorts to create an intergenerational, integrated resource for mental health research with an unparalleled capacity to consider the importance of the early life course. This resource will be readily discoverable and available through a proven secure informatics platform with the flexibility to incorporate other cohorts as part of a larger resource. The Avon Longitudinal Study of Parents and Children (ALSPAC) and the Born in Bradford (BiB) cohorts together include over 40,000 participants across three generations. Both cohorts are deeply phenotyped through clinic and questionnaire assessments and further enhanced through multi-omic measures and linkage to routine health and social records. We will extend this linkage to a comparable level in both cohorts particularly to primary care patient records ensuring this is higher than in any comparable study based on the multiple strategies we have developed and the strengths of our underlying ethical and technical infrastructure. To this resource we will add additional measures from internet based applications and social media feeds underpinned by our expertise in data science and our strengths in participant engagement. Indicative projects, both individually in each cohort and across the integrated cross-cohort resource, will demonstrate value in enabling mental health discovery. In these projects we will further develop and deploy innovative bioinformatic and biostatistical approaches to facilitate discovery and strengthen causal inference developed in our MRC Integrative Epidemiology Unit to

maximise the value that existing information can add to observational research. We will build on ALSPAC's established, sustainable infrastructure for secure, cost-efficient data sharing based on our own instance of the UK Secure E-research Platform (UKSeRP), extending this to include BiB and providing the foundation of a Mental Health Data Platform to support discovery science and service evaluation at a scale that will underpin our ability to understand and effectively improve mental health in individuals and the population.

**University of Cambridge (**MC_PC_17213)

**PI: Rudolf Cardinal**

**Developing new tools for electronic health record data linkage, structured neuropsychiatric phenotyping, outcome prediction, and democratisation of mental health research.**

With strong NHS partnerships and recent contributions to national mental health (MH) informatics, we shall add novel methods, epidemiology and phenotyping to the MH Platform. We envisage a modular pipeline that de-identifies MH data; supports flexible consent for sharing/contact; and links MH, cognitive, physical, psychosocial and biomarker data.

- Project (P) 1. Our open-source tools de-identify clinical records to create CPFT's Research Database, supporting research and participation. We shall extend them to generate anonymised subsets and link data from consenting patients across MH/community services, acute care, and research organizations, including from existing deeply phenotyped longitudinal cohorts. We emphasize rigorous interface standards and NHS governance over identifiable/pseudonymised data. We shall collaborate on a national natural language processing framework, allowing NHS/research organizations to generate structured data from free text.

- P2. We have created novel open-source neuropsychiatric assessment software. We shall extend it for broad and integrated NHS and research use. This will take automated cognitive testing into routine clinical practice. As a bold but tractable exemplar with research and clinical applications, we shall use it to apply electronic diagnostic algorithms and neuropsychiatric phenotyping, and link these detailed data to clinical records and biomarkers that include immunophenotyping.

- P3. We shall apply P1 tools to a public health crisis: the premature death of those with serious mental illness. We shall link MH, national and acute Trust data and use machine learning to develop early predictors of mortality.

- P4. We shall democratize MH research though broad consultation on generic tiered consent models for data-sharing and participation, by giving the research database direct clinical interfaces, and by enhancing data visualization to help clinicians and service users develop research and the NHS improve local services.

**Cardiff University (**MC_PC_17212)

**PI: Jeremy Hall**

**Integrating genetic, clinical and phenotypic data to advance stratification, prediction and treatment in mental health.**

Mental ill health has a major impact on patients, their carers and the wider community. Psychiatric disorders are estimated to cost the UK economy £70-100 Bn per year. Recently, large scale data sharing has allowed major progress to be made in understanding the genetic basis of many major psychiatric disorders. The MRC Centre for Neuropsychiatric Genetics and Genomics (MRC CNGG) in Cardiff has played a leading role in driving this genomic progress. Key insights that have emerged include the polygenic nature of these conditions, the extensive overlap of risk across current diagnostic categories, and the convergence of risk factors onto specific biological processes. These advances in genetics offer major opportunities for improving stratification and treatment responses in mental disorders, as well as for the development of new biomarkers and biological models. However, to capitalise on this progress there is now a need to extend this large-scale data-driven approach forward by integrating genetic information with clinical, environmental, developmental and biological data at scale in mental health to transform the management of psychiatric disorders. Our overall vision is to build key components of a platform that will enable the extension of the collaborative efforts seen in psychiatric genetics into the clinical and biological platforms required for fundamental progress in psychiatric disorders. Our proposed programme comprises 4 Workstreams to achieve these aims:

- Workstream (WS) 1. Development of a platform for integrating genomic and phenotypic data across large patient cohorts, including with electronic health records.

- WS 2. Development of a school-based adolescent cohort for identifying the antecedents of early-onset psychiatric disorder.

- WS 3. Developing a platform for advancing electrophysiological imaging biomarkers in mental health disorders.

- WS 4. Establishing methodologies for data sharing and standardisation in neuronal cellular model systems.

**University of Edinburgh (**MC_PC_17209)

**PI: Andrew McIntosh**

**Leveraging routinely collected and linked research data to study the causes and consequences of common mental disorders.**

Our vision is to link consented research studies and genomic resources with clinical and administrative data to create a structured secure infrastructure for studies into the causes and consequences of psychiatric disorders. Our ambition to implement this infrastructure across the whole of the UK.
We seek to accomplish our initial aims in Edinburgh, where there is world-leading health informatics, rich data resources covering the whole lifespan and a multidisciplinary environment with access to domain-specific knowledge and expertise. Our application builds upon key MRC and charitable investments, such as the Farr Network, HDR-UK, DPUK and important recent developments in Edinburgh such as the Bayes Institute and the 'City Deal' (£1Bn funding).
A participant-unique NHS identifier is in universal use throughout Scotland, where there is existing

and rapidly developing NHS and academic medical informatics capacity on which to build resources and expertise specifically in mental health. There are rich data resources, including prospective cohorts covering the whole lifespan, population-wide datasets, studies covering specific patient and at-risk groups and developing biorepositories. This combination of infrastructure and resources makes Edinburgh's potential for mental health data science without parallel in the UK.

An Edinburgh-based MRC Data Pathfinder award would facilitate a step-change in multidisciplinary mental health research, including collaborators and disciplines from throughout the UK.

**University of Glasgow (**MC_PC_17217)

**PI: Daniel Smith**

**Enhancing mental health cohorts through linkage to health, education and administrative datasets.**

There are substantial opportunities in Scotland to develop health informatics and routine data linkage to investigate the determinants and outcomes of mental health disorders. The University of Glasgow has an internationally-recognised record for informatics research, an established local and national infrastructure for data linkage, and strong partnerships with the NHS, third sector organisations and both local and national government. Our goal is to demonstrate the utility and versatility of routine data by developing our work further within both clinical and population-level cohorts. We will use two main approaches: linkage of primary data sources to secondary data; and development of mental health eCohorts. Four exemplar projects within this framework are proposed:

- Developing the West of Scotland Psychosis Clinical Information System as a resource for future epidemiological and pharmacogenomic research, with a focus on the interface between physical and mental health;

- Establishing a Scotland-wide Schools Health Research Network to facilitate large-scale adolescent mental health research and innovation in schools;

- Extending the Scottish Learning Disabilities Observatory's neurodevelopmental eCohort linkages of the annual Pupil Census and Scotland's Census 2011 - across a range of cross-sectoral education, health and social datasets - to study a broader range of physical, mental and educational outcomes; and

- Extending several current Scotland-wide mental health data linkage projects, including assessing the potential for repurposing antihypertensives as treatments for mood and cognitive disorders and exploiting the unique 'SHARE' population register for pharmacogenomics work in bipolar disorder.

**King's College London (**MC_PC_17214)

**PI: Robert Stewart**

**Developing algorithms to extract mental health data from electronic health records; and establishing the feasibility of working with schools and education data for mental health research.**

King's College London Institute of Psychiatry, Psychology and Neuroscience is the largest academic centre for mental health research in Europe, and has built up key strengths and leadership in clinical informatics using the rich data contained within mental health electronic records, in psychosis outcome prediction, and in longitudinal studies of mental health in childhood and adolescence. Building on our extensive experience and resources in these areas, we will deliver projects to provide the necessary groundwork for a world-leading national mental health research platform.

- Project 1 will take a range of algorithms already used at KCL and develop them for implementation in any electronic mental health record, generating protocols to enhance and render available research- and clinically-relevant data from secondary care.

- Project 2 will facilitate data linkages in a future platform by setting up a proof-of-principle service to 'network' individually negotiated healthcare linkages (with National Pupil Database education records in this case study), enabling the transition from single- to multi-site research.

- Project 3 will design and pilot a transferable model for monitoring treatment response in first episode psychosis, nested within routine clinical care, by harmonising, standardising and integrating clinical, biological and imaging data.

- Project 4 will develop and pilot a portable digital platform for recruiting and retaining school-aged mental health cohorts.

The projects will be implemented by a co-applicant team that mixes senior, long-established academics at KCL with recently attracted experts and a cadre of future leaders, thus contributing both experience and potential to the national platform. We additionally draw on longstanding national collaborations and partnerships in data science applied to mental health research.


**University of Oxford (**MC_PC_17215)

**PI: Clare Mackay**

**Developing an informatics supported platform for experimental medicine.**

Making data findable and interoperable, adding value to existing datasets and enriching data collection from person-centric and patient reported measures represents the most valuable opportunity to accelerate understanding and improved treatment of mental health conditions. Building on our experience of leading national and international programmes enabling access, interoperability and reuse of data for mental health - including the platform enabling use of Electronic Medical Records for mental health across England (CRIS), the platform for data aggregation for dementia (DPUK) and the pan-European platform to enable data interoperability from cohort and real-world data (EMIF) we propose here a programme of work including underpinning themes and four indicative projects. These projects will gain from the linkage we propose to establish between routine care data and primary, secondary, schools and cohort data

and will use devices to collect data remotely ranging from patient reported outcome measures to cognition, sleep and behaviour. To these enormous volumes of data, including the full mental health medical records from 10k people we have already linked to UK Biobank and the generation of 10k proteomes linked to mental health records, plus the streaming data from connected devices, we will bring advanced analytics from the Mathematics Institute and Natural Language Processing from computer scientists using neural networks. All our work will be enveloped in a governance and ethics stream with patient and public involvement. We leverage more than £40m in existing grant funding plus new cash and in-kind contribution from industry partners of over £12m. The outcome will be an informatics supported platform for experimental medicine, increased understanding and development of novel therapeutics for depression and a bold drive to determine the relationship between intervention and outcome in mental wellbeing in young people. These themes and proposed projects will work with our proposed partners to establish the preparatory work for a platform for Mental Health Data capitalising on the unique opportunities presented by the NHS, by strengths in mental health research and by outstanding digital sciences in the UK and in Oxford.

**Swansea University (**MC_PC_17211)

**PI: Ann John**

**MytHICAL- Mental Health Informatics in Children, Adolescents and young adults. How do my feelings become numbers?**

Swansea University (SU) has established an international reputation as a leader in health informatics, reflecting significant RCUK and Government investment (SAIL Databank, Farr, ADRC). Our ambition is to become a UK hub for data-driven, internationally-relevant research to improve children and young people's mental health (MH). SU is particularly strong in population health research, data sharing platforms (e.g. UKSeRP), innovative governance and a collaborative ethos that underpins our many multidisciplinary research activities. Our key strengths in MH research lies in our advanced data science and informatics infrastructure and cross-disciplinary expertise.
Data science at SU has contributed much to MH research by linking diverse datasets for in-depth exploration of aetiology and the role of bio-psycho-social determinants of MH, bringing together cross-disciplinary research and data at scale, and solving problems inherent in standardisation and validation. This Award will build on our experience hosting the MQ Adolescent Data Platform, creating electronic and hybrid e-cohorts, phenotype validation, pharmaco epidemiology, and genetic and conceptual work. It will give widespread access to new Development and Well-Being Assessment and childhood maltreatment data, educational data linked to disorder-specific cohorts, health, environmental and social media data, and new partnerships with paediatricians, geographers, computer scientists and psychiatrists.
We aim to extend our expertise through studies with focussed validation and preparation of novel data and high quality transferable metadata, making future MH informatics studies more deeply phenotyped and more efficiently executed. In short, we will place children and young people within settings that may influence their MH using electronic data to better understand the causes of the causes. We will achieve this through up to eight studies to address the key challenges of mental health informatics.

**University College London (**MC_PC_17216)

**PI: David Osborn**

**Advancing methodology and linkages in electronic health records for mental health research.**

UCL's cross-disciplinary mental health research is world leading, with current mental health grant activity in excess of £100M. Our Mental Health Data Pathfinder proposal brings together research leaders from across our mental health and neuroscience Divisions and Institutes, and integrates our unparalleled infrastructure and expertise in Population Health Sciences and Health Informatics. The application builds on our track record in neurosciences, epidemiology and mental health database research. Our project is ambitious and deliverable, representing a range of research disciplines that will meet several pathfinder data challenges, and will deliver real impact on data quality and methods in mental health research.

The UCL mental health data pathfinder aims to enhance UCL's Electronic Mental Health Record capacity through new linkage, new data discovery methods, involving large databases of adults and/or children in primary and secondary care. This will determine the effectiveness, precision and safety of mental health interventions.