

## HDR UK Innovation Gateway – DATAMIND collection (Projects)

### Road Builder Innovation 1 - Discoverable Schools

URL: <https://popdatasci.swan.ac.uk/centres-of-excellence/datamind/>

Data science approaches are key to addressing the major problem of poor MH in children and young people (CYP). We will build on Pathfinder successes to develop a joined-up data research infrastructure. Areas of particular strength include multiple existing cohorts and recent developments in schools-based MH research, i.e. the Schools Health Research Network (SHRN), Wales and the Schools Health and Wellbeing Improvement Research Network (SHINE), Scotland. Since 2018 (MRC Glasgow Pathfinder) SHINE has grown rapidly, with 275 schools (135,000 pupils) from all Local Authorities. SHRN, established 2013, includes every secondary school in Wales (n=211), 120,000 pupils. Our MRC Cardiff Pathfinder award: introduced standardised internationally used measures e.g. SDQ to SHRN; implemented consent procedures for linking the prospective individual school-wide MH and social data to EHRs; then linked with data in the SAIL Databank<sup>8</sup> via the ADP (<https://adolescentmentalhealth.uk>; John) through the MRC Swansea Pathfinder; feasibility tested collection of DNA in school in a subsample of pupils.

SHRN and SHINE schools have collected a wide range of MH data, associated measures (e.g. sleep, social media use, relationships) and linked MH EHR diagnoses from pupils and school level data from school leadership teams. They support the development of school networks in SW England and Manchester (Abel, Whelan). There is an opportunity to harmonise and consolidate UK data on CYP MH, making these data more widely visible and available. The networks have established policy partnerships with education, health, social care and COVID Advisory Groups to promote data access and use to support translational pathways for impact. The networks also aim to promote data literacy within school communities and engage CYP in the analysis and interpretation of data. SHINE is currently working on a small-scale pilot project to develop an interactive MH data dashboard prototype for use in schools. Working with our established youth advisory groups ALPHA and TRIUMPH, we will develop and test the dashboard for its utility, accessibility and relevance to inform future development.

#### **Outputs:**

Development and piloting of a school dashboard for improved data discoverability, access and use by schools communities, working with teachers on a newly formed 'Teachers Panel' with Core Activity 1; catalogue existing network MH measures and stakeholder engagement to identify a common set of subjective MH measures for UK school networks building on diagnosed disorders work (with Core Activity 2); make SHRN and SHINE data discoverable through the Hub and Gateway, working with existing processes to access data; identify the necessary systems, structures and resources for a UK wide scalable and sustainable school network dashboard Success indicators will include: i) successful establishment of a Teacher panel, ii) identification of a common set of subjective mental health measures, iii) school data discoverability to stakeholders through a dashboard and researchers through the Gateway.

## Road Builder Innovation 2 - Discoverable excluded and under-served groups

URL: <https://popdatasci.swan.ac.uk/centres-of-excellence/datamind/>

Administrative data have the benefits of efficiency and population coverage, providing robust effect estimates for smaller vulnerable and underserved populations that are poorly represented in surveys and trials research. They also offer the opportunity for the creation of e-cohorts at scale, and comparison across jurisdictions with the potential for natural experiments to evaluate policy differences, e.g. arising from devolution, and assess the intersection with deprivation. However, researchers must be cognisant of the underlying administrative processes and aware of any attendant caveats. Unfortunately, knowledge and proficiency here lags considerably behind that associated with routine health services data and is often compounded by a dearth of the requisite metadata. There is a need to consolidate this increasing expertise across the UK. Some of this is being advanced through component parts of ADR UK e.g. in children known to social services, homeless populations, ethnic minorities but there is a need to collate examples of how administrative data can enable real world insights into factors impacting on the MH of underserved populations. Examples include identifying MH needs of Traveller communities in routine data (John, SAIL, MQ funded; Prisoners, SAIL, ADR funded). This work will link to Road Builder Innovation 5 audit tool.

### **Outputs:**

Create a resource of examples of how administrative data can be cross-linked to identify such groups including the use of techniques such as geo-spatial mapping by postcode. This in turn can be used characterise their needs and interaction with public services and investigate outcomes. Success Indicators: We will create case studies exemplifying the discoverability of looked after children, prisoners, people from ethnic minorities, Traveller Communities and Refugees and Asylum Seekers populations using linkage across different datasets and jurisdictions.

### Road Builder Innovation 3. Linking physical and mental health in severe mental illness

URL: <https://popdatasci.swan.ac.uk/centres-of-excellence/datamind/>

We will juxtapose two challenge areas, physical and mental health (MH), and severe mental illness (SMI), building on MRC MH. Pathfinders from UCL and KCL. The premature mortality gap in people with SMI (such as schizophrenia) is a major health inequality which we have exposed using UK routine clinical data. UK research in this field has been translated into policy nationally and internationally, but more needs to be done. While this is a key area for DHSC and PHE, current HDRUK Hubs have limited MH consideration. A number are keen to work with us to address this (Breathe, LoS; Gut Reaction; BHF). This Innovation will facilitate a step change to determine the socio-demographic, biological, pharmacological, and health care pathways implicated in the excess of deaths and multimorbidity in those with SMI, often an excluded population. Our data curation and cataloguing will provide a UK wide platform for research across the spectrum of MH disorders, through the life-course.

We will deliver a discoverable portal of UK data sources available to determine the longitudinal mechanisms underpinning mortality and multimorbidity in SMI, as an exemplar for transdiagnostic research at the interface of physical and MH, available through the Gateway. We will describe the key variables available to describe and explain SMI multimorbidity within the UK's routine and bespoke datasets. This will include geographical and temporal availability and the content/timing of existing UK linkages. The initial focus on SMI provides a proof of concept for other MH disorders (e.g. PTSD, EDs). These are under-represented in birth cohort and biobank resources. Although SMI multimorbidity is a research and policy priority, UK data sources offer far broader opportunities for mechanistic and interventional research, particularly the physical health interface with common MH problems. This includes highly prevalent conditions across the affective and anxiety spectrum interfacing with cardiovascular, musculoskeletal, neurological and gastrointestinal domains. Our outputs therefore additionally focus on the UK's unrivalled cohort and biobank resources in terms of multimorbidity research opportunities. This work will align with Road Builder Innovation 5.

**Outputs:** characterisation of physical health exposures and outcomes (including EHR linkages) implicated in the excess of deaths and multimorbidity among people with SMI. Success indicators: use of the portal via the Gateway by other Hubs and researchers.

#### **Road Builder Innovation 4: Widening availability of MH care text analytics capabilities**

**URL:** <https://popdatasci.swan.ac.uk/centres-of-excellence/datamind/>

Realising the potential of EHRs to support research depends on NLP algorithms, owing to the predominance of text fields as an information source in MH services. Translating achievements from single- to multi-site studies requires a platform and service through which prototype algorithms can be enabled whilst not compromising their IP or the security of the healthcare data being processed.

The same challenge faces algorithm development in the commercial sector where direct EHR data access may be challenging and IP and commercial confidentiality concerns preclude handling of algorithms by healthcare/academic partners. This blockage to algorithm development and implementation significantly impedes MH research using EHR data. The Kings MRC Pathfinder programme successfully delivered its NLP functionality to other academic MH Trusts (e.g. Oxford, UCL), and has built an Azure hosted prototype platform, the MH Text Analytics Cloud for secure hosting of algorithms for providers and NHS-compliant transfer and processing of EHR text. This Road Builder will support the development and implementation of a service to support this resource.

#### **Outputs:**

Implementation of the MH-TAC platform and service, enabling processing of NHS EHR text with NLP and other algorithms, and achieving sustainability within 18 months.

Success indicators: i) multiple agencies, both public sector and commercial, hosting algorithms on this facility with acceptable and standardised documentation; ii) multiple MH-TAC users and a facilitated user community; iii) a successful mixed academic/commercial funding model for sustainability; iv) successful integration and interfaces with related HDR UK initiatives (e.g. Text Analytics Resource) and TRE infrastructure.

## Road Builder Innovation 5: Digitally Enhanced Trials

URL: <https://popdatasci.swan.ac.uk/centres-of-excellence/datamind/>

The Mental Health Research Framework<sup>9</sup> specifically identified routine capture of MH data in physical health studies was a critical area to address. The NIHR CRN collects data on over 850,000 participants across its clinical portfolio every year in England and Wales. The vast majority of participants are from studies concerned with physical ill health. A significant gap in these data is information about participants' MH and wellbeing. It links well with NIHR and MRC calls to develop and deliver research in areas with highest disease burden. The Digitally Enhanced Trials group (University of Manchester) is developing a core MH dataset for existing and new NIHR CRN portfolio studies in physical health. We shall enable the routine capture of MH data across the CRN portfolio, making discoverable a valuable new research resource to understand the links between mental and physical health, enhancing our understanding of co- and multimorbidity. The core MH dataset has been defined with extensive PPI and input from key mental and physical health stakeholders.

To understand the representativeness of clinical trial data, the DET group is developing a digital tool to promote equity in research that provides routine equity auditing of MH trial data submitted to the CRN that can be applied across future trial datasets. We shall exploit our links with the CRN (Abel, NIHR National Mental Health Lead) to audit the representativeness of recruited participants across UK studies in MH routinely for sex, age and ethnicity. We shall compare recruited participants against expected samples for local, regional and national populations and MH condition. The equity audit will enable identification of specific underserved groups within the NIHR MH portfolio as a whole and within individual MH clinical trials helping to improve planning for future recruitment and identify regional variation in disease burden and outcomes.

### **Outputs:**

Deliver a core mental health dataset collection tool fit-for-purpose to capture key MH data from individuals participating in physical health clinical trials. It will comply with FAIR principles and FHIR standards and be publicly available as Open Source software; deliver an equity in research digital tool that analyses and displays the representativeness of MH clinical trial datasets. Success indicators: delivery and use of core MH dataset and audit tool, discoverable within the Gateway making this Road Builder into core activity of the Hub.

## Road Builder Innovation 6: Drug discoverability

URL: <https://popdatasci.swan.ac.uk/centres-of-excellence/datamind/>

The lack of new pharmaceutical agents for MH disorders continues to perpetuate the global disability associated with MH. There has been a relative lack of potential drug targets, and few studies that would enable existing drugs to be targeted to new indications or to the individuals most likely to benefit. This has discouraged major investment in MH, perpetuating the current impasse in drug development. While there are challenges, this is a time of unparalleled opportunity for re-engaging Industry. Genetic discoveries in MH have accelerated over the last five years, with over 500 significant loci implicating new and existing drug targets, alongside developments in other 'omics technologies. There has also been a movement to greater pre-competitive collaboration between multiple pharmaceutical companies, mirroring shifts in academia towards multi-site working, and a greater public appreciation of the need for drug development brought into focus by the pandemic.

The Hub will capitalise on these new opportunities and deal with existing impediments to closer collaboration with pharma, the major drivers of new and more effective drug development:

- We will identify the needs of Industry by engaging both individually and collectively with multiple companies, such as through the Medicines Discovery Catapult/ Psychiatry Consortium (LoS).
- We will work with Core Activity 1 to identify why data sharing with Industry has less public confidence than with the NHS and Academia, and specifically engage with individuals who show low trust to identify a plan of action to nurture public trust.
- The Hub will explore a platform for anonymised data sharing with Pharmaceutical and other Industries, that will be incorporated into our business development and sustainability plans (Core Activity 3). It is clear from our discussions that there is a strong appetite to access routinely collected datasets and genomic data for drug development and precision medicine approaches.
- The Hub will provide a roadmap for greater engagement with Industry to unblock drug development through socially licenced engagement and the provision of suitable TRES.