

# Multidisciplinary ecosystem to study lifecourse determinants and prevention of early-onset burdensome multimorbidity (MELD-B) – protocol for a research collaboration

Journal of Multimorbidity and Comorbidity

Volume 13: 1–13

© The Author(s) 2023

Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/26335565231204544

[journals.sagepub.com/home/cob](https://journals.sagepub.com/home/cob)

Simon DS Fraser<sup>1,16</sup> , Sebastian Stannard<sup>1</sup> , Emilia Holland<sup>1</sup>, Michael Boniface<sup>2</sup>, Rebecca B Hoyle<sup>3</sup>, Rebecca Wilkinson<sup>4</sup>, Ashley Akbari<sup>5</sup> , Mark Ashworth<sup>6</sup>, Ann Berrington<sup>7</sup>, Roberta Chiovoloni<sup>5</sup>, Jessica Enright<sup>8</sup>, Nick A Francis<sup>1</sup>, Gareth Giles<sup>9</sup>, Martin Gulliford<sup>6</sup>, Sara Macdonald<sup>10</sup>, Frances S Mair<sup>10</sup> , Rhiannon K Owen<sup>5</sup>, Shantini Paranjothy<sup>11,12</sup>, Heather Parsons<sup>13</sup>, Ruben J Sanchez-Garcia<sup>3,14</sup>, Mozhdeh Shiranirad<sup>3</sup>, Zlatko Zlatev<sup>2</sup> and Nisreen Alwan<sup>1,13,15,16</sup>

## Abstract

**Background:** Most people living with multiple long-term condition multimorbidity (MLTC-M) are under 65 (defined as ‘early onset’). Earlier and greater accrual of long-term conditions (LTCs) may be influenced by the timing and nature of exposure to key risk factors, wider determinants or other LTCs at different life stages. We have established a research collaboration titled ‘MELD-B’ to understand how wider determinants, sentinel conditions (the first LTC in the lifecourse) and LTC accrual sequence affect risk of early-onset, burdensome MLTC-M, and to inform prevention interventions.

<sup>1</sup>School of Primary Care, Population Sciences and Medical Education, Faculty of Medicine, Southampton General Hospital, Southampton, UK

<sup>2</sup>School of Electronics and Computer Science, University of Southampton, Southampton, UK

<sup>3</sup>School of Mathematical Sciences, University of Southampton, Southampton, UK

<sup>4</sup>Southampton City Council, Southampton, UK

<sup>5</sup>Population Data Science, Swansea University Medical School, Faculty of Medicine, Health and Life Science, Swansea University, Swansea, UK

<sup>6</sup>School of Life Course and Population Sciences, King’s College London, London, UK

<sup>7</sup>Department of Social Statistics and Demography, University of Southampton, Southampton, UK

<sup>8</sup>Computing Science, University of Glasgow, Glasgow, UK

<sup>9</sup>Public Policy Southampton, University of Southampton, Southampton, UK

<sup>10</sup>School of Health and Wellbeing, General Practice and Primary Care, University of Glasgow, Glasgow, UK

<sup>11</sup>School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, Aberdeen, UK

<sup>12</sup>NHS Grampian Health Board, Aberdeen, UK

<sup>13</sup>Patient and Public Involvement and Engagement, University Hospital Southampton NHS Foundation Trust, Southampton, UK

<sup>14</sup>The Alan Turing Institute, London, UK

<sup>15</sup>NIHR Southampton Biomedical Research Centre, Southampton, UK

<sup>16</sup>NIHR Applied Research Collaboration Wessex, Southampton, UK

## Corresponding author:

Simon DS Fraser, School of Primary Care, Population Science and Medical Education, Faculty of Medicine, University of Southampton, Southampton General Hospital, Tremona Road, Southampton, SO16 6YD, UK.

Email: [s.fraser@soton.ac.uk](mailto:s.fraser@soton.ac.uk)



Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License

(<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without

further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

**Aim:** Our aim is to identify critical periods in the lifecourse for prevention of early-onset, burdensome MLTC-M, identified through the analysis of birth cohorts and electronic health records, including artificial intelligence (AI)-enhanced analyses.

**Design:** We will develop deeper understanding of ‘burdensomeness’ and ‘complexity’ through a qualitative evidence synthesis and a consensus study. Using safe data environments for analyses across large, representative routine healthcare datasets and birth cohorts, we will apply AI methods to identify early-onset, burdensome MLTC-M clusters and sentinel conditions, develop semi-supervised learning to match individuals across datasets, identify determinants of burdensome clusters, and model trajectories of LTC and burden accrual. We will characterise early-life (under 18 years) risk factors for early-onset, burdensome MLTC-M and sentinel conditions. Finally, using AI and causal inference modelling, we will model potential ‘preventable moments’, defined as time periods in the life course where there is an opportunity for intervention on risk factors and early determinants to prevent the development of MLTC-M. Patient and public involvement is integrated throughout.

### Keywords

Life course, multimorbidity, long-term conditions, health, burdensome, complex, artificial intelligence, birth cohorts, routine healthcare datasets, prevention

Received 22 May 2023; accepted: 13 September 2023

## Background

A growing number of people are living with multiple long-term condition multimorbidity (MLTC-M). Factors such as demographic characteristics (e.g. age, ethnicity), lifecourse events (e.g. infections, accidents), behaviours (e.g. smoking, diet) and broader experiences (e.g. the environment people grew up in, their education, work, income) influence the chances of developing MLTC-M. MLTC-M occurs earlier in the lifecourse among people from more socioeconomically and demographically disadvantaged backgrounds.<sup>1</sup> The burden of MLTC-M, and the sequence that people develop conditions, also vary. To meet the significant challenge of preventing early-onset, burdensome and complex MLTC-M there is a need to clarify the meaning of burdensomeness and complexity and, taking a lifecourse approach, to understand the influence of wider determinants (such as social, economic, and environmental factors), the role of sentinel conditions and the sequence of long-term condition (LTC) accrual. We have therefore designed a research collaboration titled ‘MELD-B’ to harness the potential of artificial intelligence (AI) to handle the complexity of analyses required to discover insights from complex national social and routine datasets across the lifecourse.

### *Long term condition accrual and sentinel conditions*

Earlier and greater accrual of long-term conditions (LTCs) may be influenced by the timing and nature of the exposure to key risk factors, wider determinants or other LTCs at different life stages. However, existing studies have limited time frames (5–20 years of follow up mostly during adulthood) and limited analyses of LTC accrual sequence.

The sequence of a) risk factors and b) LTCs may be a key determinant of the nature and burden of LTC clusters.<sup>2–5</sup>

In earlier developmental work we introduced the term ‘sentinel condition’ to describe the first LTC that an individual develops in their lifecourse as part of a subsequent MLTC-M cluster.<sup>6</sup> The timing and nature of sentinel conditions may influence subsequent clusters, which then influence the nature and risk of burden and outcomes. Clinical diagnosis leads to actions such as medication, specialist referral, and self-management advice. Such actions affect future trajectories. The time point of diagnosis and associated actions therefore also become important determinants of future MLTC-M, and this will be explored within the MELD-B collaboration.<sup>2,3</sup>

### *Lifecourse determinants of MLTC-M*

Developmental Origins of Health and Disease (DOHaD) has become an established research field linking aetiology of disease in adulthood with environmental exposures in utero and early life.<sup>7</sup> Preconception and pregnancy are important periods and the concept of ‘fetal programming’ has emerged whereby a stimulus or insult during that period can have permanent effects on structure, physiology and metabolic system of offspring.<sup>8–12</sup> Epigenetics is a biological pathway underlying DOHaD, where permanent effects of transient environmental influences alter epigenetic gene regulation.<sup>12,13</sup> Socioeconomic disadvantage is key in shaping developmental life experiences.<sup>14</sup> Analyses in the Hertfordshire cohort study showed that paternal social class was associated with future multimorbidity.<sup>15</sup> In the Aberdeen Children of the 1950s (ACONF) cohort, lower father’s social class at birth was associated with early-onset

multimorbidity.<sup>16</sup> In the 1970 British Cohort Study (BCS70) those with fathers from unskilled occupational groups (vs. professional) at birth had 43% higher risk of early-onset multimorbidity.<sup>17</sup> Recent policy has also focused on the importance of early life in shaping health and disease. The 2019 Health and Social Care Select Committee report described how “The first 1000 days of life, from conception to age 2, is a critical phase during which the foundations of a child’s development are laid.”<sup>18</sup> Recommendations from research commissioned by the Royal Foundation of the Duke and Duchess of Cambridge included the importance of promoting education and supported wide dissemination of evidence on the primacy of the early years.<sup>19</sup> Despite previous research evidence, there remains a need to explore wider early-life determinants, defined as the period from pre-conception until age 18, on the combinations of LTCs.

### **Burden and complexity**

MLTC-M analyses would be greatly enhanced by a better understanding of burdensomeness and complexity, and what they mean to patients and carers. MLTC-M is commonly defined as having two or more LTCs, but there is a need to move away from LTC counts towards a more sophisticated understanding of MLTC-M, considering the interplay between wider social determinants and disease, the influence of mental and physical conditions, and the importance of disease stage/severity. Conceptualisation of some of the challenges experienced by patients is encapsulated in the ‘cumulative complexity model, which includes the concept of ‘treatment burden, patient workload and ‘capacity’, and was included in the National Institute for Health and Care Excellence multimorbidity guidance.<sup>5,20–23</sup> For people with greater clinical complexity and fewer available resources, total workload might outweigh ‘capacity’ (ability to manage workload conferred by treatment and the demands of everyday life) and risk treatment failure.<sup>24</sup> High treatment burden is associated with poor quality-of-life and lower treatment adherence, potentially leading to worse outcomes and health service inefficiency.<sup>25–29</sup> In a cross sectional study of 835 people with multimorbidity, higher treatment burden was associated with younger age (people aged 55-64 vs. those over 65).<sup>30</sup> In MELD-B we will consider a broad scope of ‘burdensomeness’ and ‘complexity’ that includes treatment burden as well as disease burden and broader psychosocial factors.

### **Aim**

Our aim is to identify critical periods in the lifecourse for prevention of early-onset (under 65), burdensome MLTC-M, identified through the analysis of birth cohorts and

electronic health records. We will use artificial intelligence (AI)-enhanced epidemiological analysis and disseminate our findings to policy makers, establishing pathways to policy and practice impact.

## **Research plan/methods**

### **Study design**

The study uses a mixed methods approach that combines qualitative evidence synthesis and quantitative analysis of birth cohorts and electronic health records. This will be achieved through five complementary work packages, a patient and public advisory board, and an expert advisory group, as shown in [Figure 1](#).

### **Research objectives**

Work Package 1 - ‘Burden and complexity’:

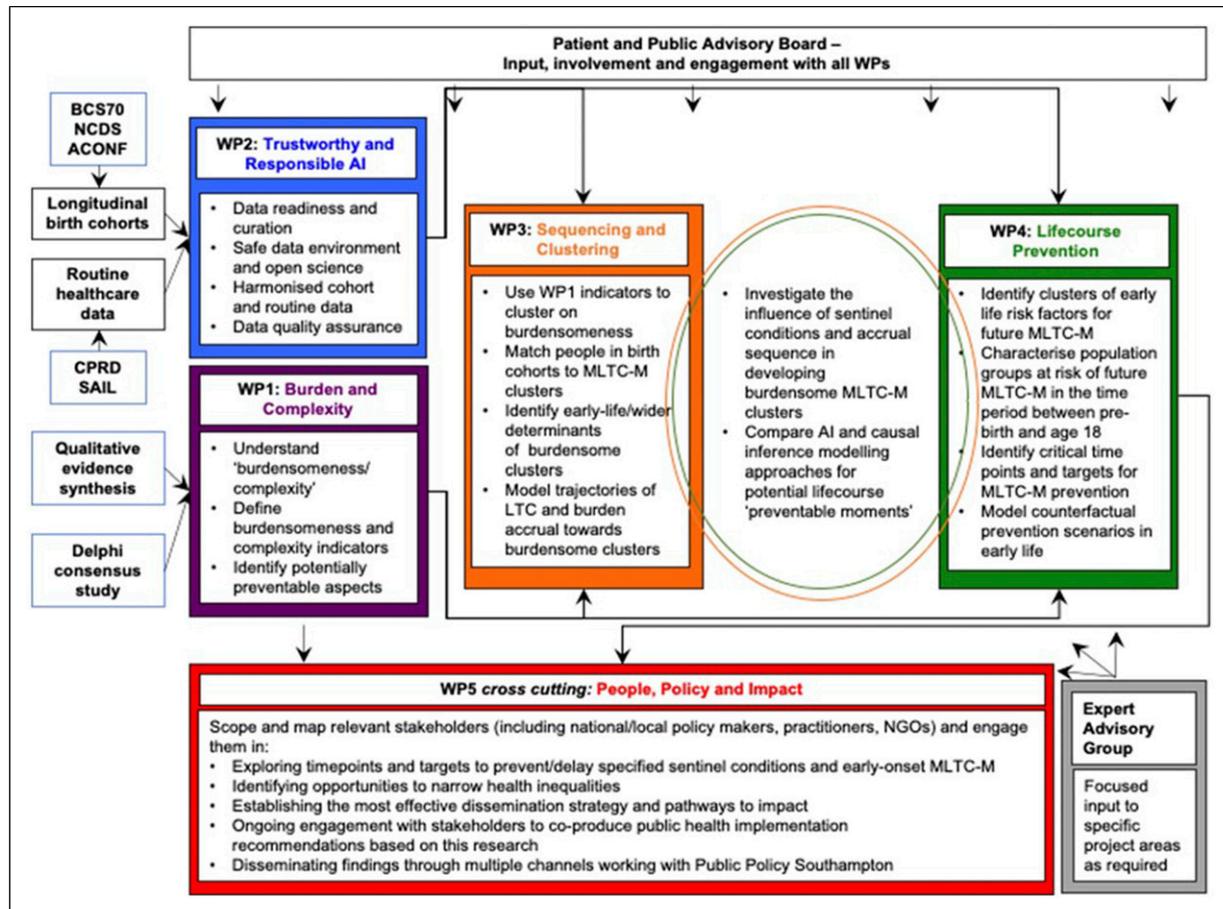
- 1.1 Develop a deeper understanding of what ‘burdensomeness’ and ‘complexity’ mean to people living with early-onset MLTC-M, their carers and healthcare professionals.
- 1.2 Produce a suite of burdensomeness/complexity indicators for routine healthcare data as burdensomeness/complexity domains for use in clustering and clinical practice.

Work Package 2 - ‘Trustworthy and Responsible AI’:

- 2.1 Provide the safe data environment and readiness for AI analyses across routine healthcare data (Secure Anonymised Information Linkage (SAIL) Databank and Clinical Practice Research Datalink (CPRD)) and birth cohort data (National Child Development Study (NCDS), Aberdeen Children of the 1950s (ACONF), 1970 British Cohort Study (BCS70)).
- 2.2 Harmonise specified LTCs across birth cohorts and routine data.

Work Package 3 ‘Sequencing and Clustering’:

- 3.1 Use the burdensomeness/complexity indicators and apply AI methods to identify novel burdensome early-onset MLTC-M clusters in routine data.
- 3.2 Develop and apply semi-supervised learning to match individuals in birth cohorts into the routine data MLTC-M clusters and identify early-life and later determinants of the burdensome clusters using the matched datasets.
- 3.3 Describe and model trajectories of LTC and burden accrual towards burdensome clusters.



**Figure 1.** Study work package structure.

Work Package 4 ‘Lifecourse prevention’:

- 4.1 Identify and characterise clusters of early life exposures (risk factors for early-onset, burdensome MLTC-M and sentinel conditions) and characterise population groups at risk of future MLTC-M in early-life (prebirth-18 years).
- 4.2 Identify critical time points and key lifecourse targets for MLTC-M prevention and model counterfactual prevention scenarios acting on combined risk factors at the specified timepoints (prebirth-18 years).

Cross WP3 and WP4 objectives:

- 4.3 Investigate the influence of sentinel conditions and sequence of accrual of wider determinants, conditions and burdensome factors in the development of early-onset, burdensome MLTC-M clusters.
- 4.4 Compare AI and causal inference modelling for potential early-life (0-18) ‘preventable moments’

trajectory, defined as time periods in the life course where there is an opportunity for intervention on risk factors and early determinants to prevent the development of MLTC-M, and exploring alternative trajectories based on models of policies/strategies/interventions and outcomes.

Work Package 5 ‘People, Policy, and Impact’:

- 5.1 Identify key stakeholders and engage them in: 1) exploring timepoints and targets to prevent/delay specified sentinel conditions and early-onset, burdensome MLTC-M, 2) opportunities to narrow health inequalities and, 3) optimal dissemination strategies and pathways to impact.
- 5.2 In partnership with the patient and public advisory board, engage with stakeholders to co-produce public health implementation recommendations based on our research, and produce a policy and practice engagement strategy to disseminate findings through multiple channels.

## Dataset summary

The majority of work within this Research Collaboration involves the use of two large, pseudonymised linked routine healthcare datasets and three anonymised linked birth cohort datasets. Work package 1 also involves the recruitment of individuals to take part in a Delphi study; the details of this are given in the relevant work package sections below.

### Longitudinal birth cohort studies datasets

The *Aberdeen Children of the 1950s (ACONF)* includes children born in Aberdeen, Scotland between 1950 and 1956, in total there are 12,150 cohort members, and participants were traced in their forties (2002) and linked to hospital and mental health admissions, maternity records, cancer registers, and death records. The *National Child Development Study (NCDS)* has followed all children born in England, Scotland and Wales in one week in 1958. In total 17,415 cohort members have been followed up over 12 sweeps – birth, 7, 11, 16, 23, 33, 42, 44, 46 50 and 55. The *1970 British Cohort Study (BCS70)* has followed all children born in England, Scotland, Wales and Northern Ireland in one week in 1970. In total 17,196 cohort members have been followed up over 10 sweeps – birth, 5, 10, 16, 26, 30, 34, 38, 42, 46. Both the NCDS and BCS70 have collected information on socioeconomics, family background, cognitive development, educational, employment, partnerships, fertility, health and health-related behaviour and can be linked to hospital episode statistics.

### Routine health datasets

The *Secure Anonymised Information Linkage (SAIL) Data-bank* contains a range of anonymised linked individual-level, population-scale data sources for the population of Wales, including approximately 3.1 million people. Linked SAIL data sources contain health, administrative and anonymised demographic/geographic information for the population of Wales, collected between 2000 and 2022. The *Clinical Practice Research Database (CPRD)* includes both CPRD GOLD and CPRD Aurum providing data on over 60 million people between 2010 and 2022 and over 45 million people available for data linkage in England. Data is collected from GP practices and primary care data fully coded electronic health records, and includes pharmacy and pathology records (Vision® or EMIS®).

## Methods, design and analysis for specific work packages

### Work package 1: ‘Burden and complexity’

The aim of this work package is to develop a deeper understanding of what ‘burdensomeness’ and ‘complexity’

mean to people living with early-onset MLTC-M, their carers and healthcare professionals, and to use this to inform data curation (WP2), AI analyses (WP3) and future preventative strategies (WP4/5). We will develop a list of burdensomeness/complexity attributes that might be identified and characterised in routine healthcare data as burdensomeness/complexity indicators. We will refine these indicators with people who live with (or care for people with) MLTC-M, healthcare professionals and MLTC-M experts and identify burden that might be prevented/reduced.

**Qualitative evidence synthesis.** We will conduct a qualitative evidence synthesis (QES) to collate findings of relevant qualitative studies on the experience of living with MLTC-M.<sup>31</sup> We will summarise the scope/QES question using the ‘PerSPecTIF’ framework<sup>32</sup> and register the protocol on ‘PROSPERO’.<sup>33</sup> Literature searching will follow Cochrane’s Handbook for Systematic Reviews of Interventions, ‘Qualitative Evidence’ guidance,<sup>34</sup> and we will search Database including: MEDLINE (EBSCO), PsycINFO (EBSCO), CINAHL (EBSCO), Cochrane Library, EMBASE (EBSCOhost) as well as grey literature and backward/forward manual searching. Date restriction (post 2000) will be applied to ensure relevance, and inclusion criteria will include qualitative studies, studies exploring lived experience/management of multimorbidity and mixed methods studies with a qualitative element. Exclusion criteria include single-condition studies with comorbidities and studies in children (older adults are not excluded but age distribution of participants will be considered and reported). However, we anticipate the inclusion and exclusion criteria will be developed iteratively over the course of the study. Two researchers will screen independently, and a third will adjudicate conflicts.

Data extraction will be completed in two stages with the support of public contributors: Stage 1: contextual details (including population, context, methodology, recruitment, data collection, analysis), and Stage 2: individual study findings. Data synthesis methods will be finalised when the included studies are established.<sup>35</sup> We will follow ‘RETREAT’ guidance to choose methodology.<sup>36</sup> ‘Best Fit Framework Synthesis’ may be most appropriate as it adopts a predominantly deductive approach with a pre-specified framework while allowing inductive elements.<sup>37,38</sup> The cumulative complexity model will be the framework as it is patient-centred.<sup>22</sup> Thematic Synthesis or meta-ethnography will also be considered and we will learn from a review of qualitative systematic reviews.<sup>39,40</sup> We will evaluate our QES using NVIVO (data management), CORE-Q (quality appraisal),<sup>41</sup> ENTREQ (reporting)<sup>42</sup> and GRADE CERQual (degree of confidence in our findings).<sup>43</sup>

**Consensus study.** Using findings from the QES, we will build consensus among multimorbidity experts and patients/carers of people with MLTC-M on:

- Which aspects of burdensomeness/complexity can be identified and characterised in routine healthcare data?
- Which aspects of burdensomeness/complexity might be reduced or prevented?
- Which indicators are most important to patients and carers?

The Delphi technique is widely used in LTC research.<sup>44-46</sup> Ours will comprise three rounds. In round one, participants will independently rank a series of questions about potential indicators related to ‘burdensomeness’ and ‘complexity’ using a four-point Likert scale. Health professionals/MLTC-M experts will indicate whether they believe the indicator can be characterised in routine data, and patients and carers will be asked to rank its importance to them. We will use a similar process to seek consensus on potentially preventable burdensomeness/complexity aspects from patient, carer and expert perspectives. Each domain will include free-text response sections, including opportunity to suggest other indicators that might be important to collect in the future. Panel recruitment will involve non-probability purposive sampling of about twelve healthcare professionals/MLTC-M experts and twelve adults who live with MLTC-M (or represent people who do) to achieve sufficient numbers and a range of participants by age, sex, ethnicity, profession, and geographical location.<sup>47</sup> Participants will be primarily UK-based (as indicators will be applied in UK data). Participants will be required to respond across all rounds, but may withdraw at any time, and written, informed consent will be obtained from participants. Indicators and preventable aspects from the QES will be used to develop survey rounds.<sup>46</sup> We will use study team, PPI Advisory Board and Expert Advisory group input to ensure clarity and ‘sense check’ questions, ensuring they cover aspects important to patients/carers. We will pilot the survey with iterative feedback to improve structure and readability. In Rounds 2 and 3, content will be iterated to incorporate qualitative comments from preceding stages, fed back in quantitative

form. Consensus will be defined as >75% (the median level of agreement threshold from a systematic review of Delphi studies).<sup>48</sup> of participants agreeing or disagreeing about:

- 1) Importance to patients,
- 2) Identifiability in routine data and,
- 3) Potential for prevention.

### Work package 2: ‘Trustworthy and responsible AI’

The aim of this work package is to design, build and operate a trustworthy and responsible environment for research into AI pipelines exploring early onset, burdensome MLTC-M using our routine and birth cohort datasets. By adopting and complying fully with principles defined by the DHSC AI Code of Conduct “A guide to good practice for digital and data-driven health technologies”<sup>49</sup> and the “Five Safes” framework,<sup>50</sup> We will ensure AI activities deliver benefits to society that are ethical, valuable, fair, safe, legal and conducted reflecting principles of open science including data/code transparency, interoperability, and standardisation. Equality, Diversity and Inclusion will be assessed relating to Public Sector Equality Duty under Section 149(1) and (3) of the Equality Act (2010).

**Safe people, settings and data.** We will address the need for “Safe People” by providing training to the research team in the guiding principles of the AI Code of Conduct and access to eLearning materials for clinical good practice, data security awareness and working with Trusted Research Environments (TREs). We will operate an information governance framework to manage data assets and associated legal and ethical risks, obtaining necessary approvals for storage, processing and linking. Having established “Safe data” through assessment of risk of reidentification and data owner constraints, curation pipelines will be provisioned in Swansea, Southampton, Aberdeen, UK Data Service (See Table 1).<sup>51,52</sup> These nationally distributed TREs will provide “Safe Settings” to ensure data-related activities are undertaken securely and safely in accordance with data protection law and data licensing constraints. We will then explore processes supporting 1) cross-institutional collaboration for MELD-B researchers, 2) Machine Learning Operations (MLOps, experiment management, provenance

**Table 1.** Trusted research environment and datasets.

TRE	Datasets
SAIL Databank, Swansea	SAIL linked data sources
NHS DSPT Compliant Secure Research Environment, Southampton	CPRD
Grampian Data Safe Haven, Aberdeen	ACONF record-linked to routinely available secondary care data
UK Data Service	BSC70 and NCDS record-linked to hospital episode statistics

tracking, containerisation, etc.) in each setting for reliable and repeatable data engineering and model development, and 3) federated learning, testing and validation of models.

**Data readiness and curation.** We will implement data engineering and curation processes to assess, prepare and harmonise data for analysis. MELD-B brings together routine and birth cohort data requiring development of reusable data engineering functions to prepare data such as metadata annotations, harmonisation of variables related to burdensomeness state space and generation of optimal data structures for analysis (e.g., timeseries events, graphs). We will develop specifications and tools to assess data readiness levels considering emerging data requirements from novel MELD-B algorithms for clustering, sequencing and counterfactual analyses. We will address challenges of accuracy, semantic consistency, missing data, bias and power, and implement processes for lifecourse burdensome state reconstruction by:

- Semantically aligning data with SNOMED CT medical terminology through existing mappings (e.g., ICD10) and generate new mappings for prioritised variables in birth cohort data by processing metadata using semantic alignment and natural language processing.
- Profiling data to generate summary metadata including occurring patterns, frequencies and distributions, and distinct or missing values in a column, data types of attributes.
- Combining generated metadata with contextual information captured by WP1 structured in formal ontology to provide a contextualised metadata summary for reporting, audit and harmonisation across the AIM programme.

We will define a data quality model for analysis of lifecourse burdensomeness allowing systematic measurement and assessment against metrics for completeness, uniqueness, timeliness, validity, accuracy, and consistency. We will address bias and fairness as critical elements of model fairness and ensure outcomes do not discriminate against groups, and to clearly explain limitations in models. We will ensure that principles of fairness and responsible AI practices are adopted by everyone involved in data selection and algorithmic decisions. Bias will be addressed qualitatively by engaging stakeholders, guided by the expert advisory group. We will also address bias by applying a formal model of data pre-processing operations (feature selection, feature engineering, imputation and listwise deletion, re-sampling, outlier removal, smoothing/normalisation and encoding) to record operator's effect on the data, logging operations during processing through code-instrumentation.

**Safe outputs and open science.** We will ensure FAIR (Findable, Accessible, Interoperable and Reusable) data

stewardship, model curation and research integrity. AI pipelines will be developed using containers (Docker, Singularity) and described using container orchestration in accordance with MLOps best practice. Models will be developed in interactive environments such as Jupyter, tracked using an experiment manager (e.g., MLFlow) and built into software libraries for integration into workflows (e.g., TensorFlow). Artefacts such as software libraries, images, and notebooks will be made available through open-source licenses (subject to code review and safe output) allowing results to be replicated by others and research outputs repeated/compared by others. We expect to curate and make available metadata and data assets subject to governance and license constraints.

**LTC inclusion.** Clinical Co-Is will review code lists and agree LTC definition using existing clinical code lists including, but not limited to, 40 LTCs from a paper by Hanlon et al exploring associations between MLTC-M and adverse health outcomes in UK Biobank and the SAIL Databank, and 59 from a Delphi study by Ho et al that reached consensus on LTCs that should 'always' or 'usually' be included in MLTC-M analyses (noting that there is some overlap with LTCs in Hanlon et al).<sup>53,54</sup> We will explore whether additional LTC codes from the CALIBER platform, Cambridge, and SAIL are needed as our definition of burdensome emerges.<sup>55-57</sup> Clinicians will match birth cohort health variables to Read/SNOMED codes in routine data. We will then harmonise LTC definitions across BCS70/NCDS/ACONF/SAIL and CPRD where possible.

### **Work package 3: 'Sequencing and clustering'**

The aim of this work package is to cluster individuals within the space of burdensomeness indicators and analyse determinants of clusters and sequence of acquisition of burdensome features for individuals in those clusters. Burdensome clusters will initially be identified in CPRD/SAIL that include information on later life. We will then develop methodology to connect birth cohort data with these clusters, permitting inference across the complementary birth cohort and routine datasets to identify early-life determinants of burden. Finally, we will identify and analyse the sequence of sentinel conditions and subsequent accrual of burden in harmonised birth cohort and routine data.

**Clustering.** We will apply AI technologies for clustering and cluster interpretation iteratively, with a human expert in the loop, on our routine GP datasets to elicit the structure of the burdensomeness space and the prevalent trajectories of evolution of MLTC-M within it. We will investigate how transience of health conditions and components of burden affect cluster stability, persistence and membership over

time to inform our dynamical modelling of trajectories to burdensome MLTC-M. Methods to achieve this may include agglomerative, hierarchical, k-prototypes and graphical clustering, and XGBoost-SHAP cluster interpretation.

*Inference across datasets.* The birth cohort datasets contain much richer information on early life than the routine healthcare datasets; however birth cohort data are not linked to routine healthcare data and we do not know whether an individual represented in a birth cohort dataset appears in a routine healthcare dataset or vice versa. Therefore, to explore early-life predictors of burdensome MLTC-M we will identify individuals in our birth cohort datasets with patterns of clinical and social variables similar to those predicting burdensome cluster membership in the routine datasets. This will allow us to identify individuals in the birth cohort datasets who are likely to belong (now or in the future) to the burdensome clusters and use explainable AI methods and causal DAG-based models to determine early-life predictors of the burdensome clusters. We will use semi-supervised learning to classify individuals from our birth cohort datasets into the clusters identified in our routine datasets, starting from overlapping clinical and social variables recorded in the two types of datasets. The specific variables used to infer across datasets will be developed and refined throughout the project.

*Lifecourse trajectories and dynamical modelling.* We will explore the sequence of exposure to lifecourse risk factors, occurrence of the sentinel LTC and sequence of accrual of other LTCs and burdensome features in birth cohort and routine datasets, and characterise those that lead to early MLTC-M in the burdensome clusters. This may indicate how the order of acquisition of conditions influences the development of MLTC-M and suggest points of intervention. We will build dynamical models of the acquisition of LTCs, elements of burden and risk factors leading to burdensome MLTC-M and compare them with data on the dates of acquisition of these features by individuals in the routine and birth cohort datasets. We will investigate the risk of developing a given condition or burdensome element, conditional on the sequence and timing of prior acquisition of LTCs and aspects of burden over the lifecourse. By exploring counterfactual scenarios, such as a change in sequencing or timing of conditions or risk factors, we aim to identify key timepoints at which to address specific risk factors in the prevention of onset of the specified burdensome MLTC-M clusters. Methods to achieve this may include statistical multistate modelling and approximate Bayesian computation in combination with individual-based modelling.

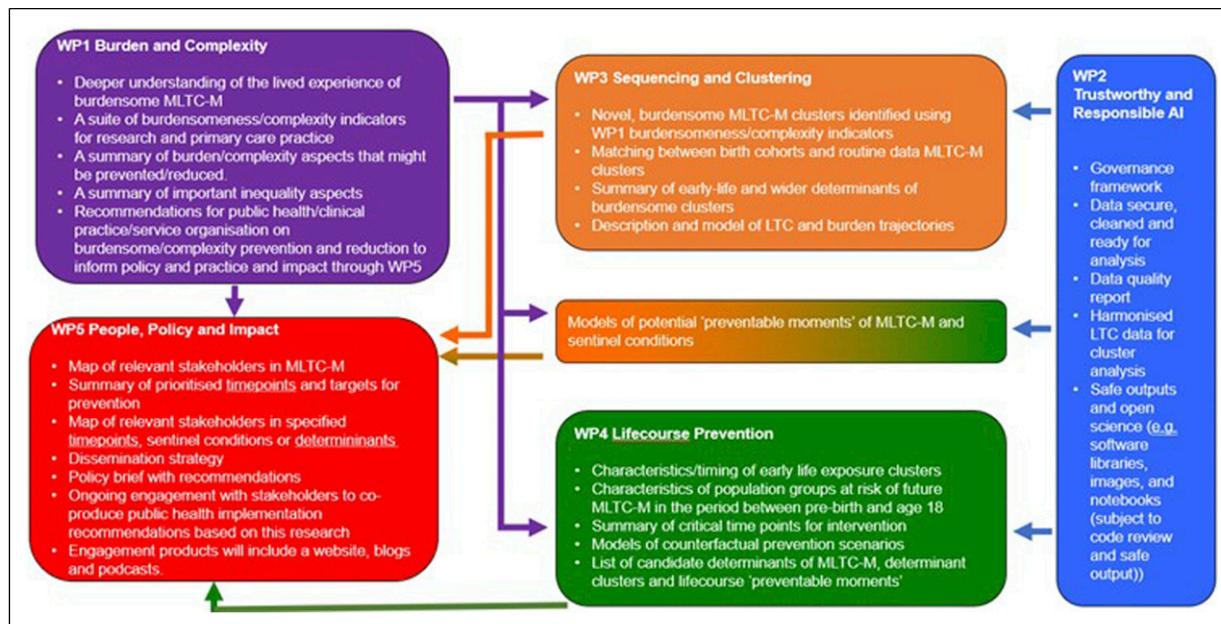
#### *Work package 4: 'Lifecourse prevention'*

The aim of this work package is to identify the important early-life (from preconception, pregnancy, and birth to childhood and adolescent) characteristics of population groups at risk of future early-onset multimorbidity and use this characterisation to model the nature and timing of targeted public health prevention scenarios of early-onset, burdensome/complex MLTC-M through examining counterfactual scenarios of lower risk. We will do this through:

*Characterisation of early health, social, economic and environmental risk factors.* We will explore the early health, behavioural, social, economic and environmental characteristics of population groups and their risk of sentinel conditions and early-onset, burdensome MLTC-M. We will do this through describing clusters of early life exposures (potential determinants of early-onset, burdensome MLTC-M and sentinel conditions) and characterising their components, pattern, and time trend. We will start with a long list of candidate determinants from the three birth cohorts and SAIL Databank, and develop criteria for selection guided by a conceptual model, degree of association with target outcomes, public and patient input, scoping the relevant literature and policy directions.

*Causal inference.* Utilising a causal inference approach to visualise apriori knowledge and assumptions about what confounds and mediates relationships of interest we will explore adopting a directed acyclic graph (DAG)-based approach informed by the generated early life clusters based on the project data, prior evidence, biological plausibility and patient and public involvement. The outcomes will be informed by findings from work packages 1 and 3.

*Identification of critical time-points for public health interventions.* We will examine the nature and critical time-points for potential public health interventions of early-onset multimorbidity and the elements within clusters of risk factors that are most important to act on for early prevention. Using counterfactual scenarios of prevention, we will conduct comparative analysis to specify what is most effective in reducing early-onset multimorbidity risk; action on individual risk factors (classical public health approach) or simultaneous action on a combination of risk factors at specific time-points between pre-birth and 18 years. We will explore the use of methods such as adjusted population attributable fractions or G-methods or which deal with time varying exposures to estimate the effect of hypothetical interventions at critical points before adulthood (pre-birth to 18 years) on the prevention of multimorbidity and burdensomeness.<sup>58</sup> Such methods estimate potential outcomes under less restrictive conditions than standard regression methods and there is less risk of



**Figure 2.** MELD-B research collaboration outputs by work package.

biased estimates due to over-adjustment.<sup>59</sup> These also follow a DAG approach to represent causal assumptions and compare time-varying joint and dynamic interventions.

### Work Package 5: 'People, policy, and impact'

The aim of this work package is to connect emerging findings with relevant stakeholders in order to identify appropriate intervention opportunities, effective means of dissemination and effective policy outputs. Through learning from each work package, and engagement and co-production with key stakeholders, we aim to identify prevention opportunities across the life course and establish pathways to policy and practice impact. We will establish a Policy and Practice Engagement Strategy and undertake:

1. Stakeholder analysis to identify key MLTC-M stakeholders and identify key stakeholders relevant to emerging findings.
2. Policy workshops to review burdensomeness/complexity and prioritising policy/practice outputs.
3. On-going stakeholder engagement, in partnership with the PPI Advisory Board.

These activities we will help us to prioritise populations, timepoints and determinants to target for the prevention/delay of specified sentinel conditions and early-onset, burdensome MLTC-M, and to identify opportunities to narrow health inequalities. We will also determine the most effective ways to disseminate findings and ensure pathways to impact. Finally, we will co-produce public health and

practice implementation recommendations based on this research and determined by its findings.

**PPI advisory board.** Our PPI lead will, along with a PPI Officer, will lead the PPIE structure in MELD-B, with a team of PPIE contributors forming a Patient and Public Advisory Board. We will seek diversity among PPIE contributors, considering age, sex, multimorbidity experience, ethnicity, and background, and contributors will be drawn from across our collaboration sites. The PPI Advisory Board will meet regularly, and meetings will be chaired by our PPI lead, and co-chaired by a PPI member. PPI members will be paid according to the NIHR payment guidance.<sup>60</sup> We will create a safe, supportive environment for PPI contributors to bring their knowledge and experience. In addition, we will work with PPI groups in Glasgow, which is an ideal context to explore multimorbidity and socioeconomic deprivation. We will also consult with the public and communities beyond the PPI Board and Glasgow groups as guided by our developing findings and objectives, linking into diverse communities and networks.

### Outputs

The outputs we aim to generate from the MELD-B Research Collaboration are summarised in [Figure 2](#).

### Ethics approvals and dissemination

The study will be conducted in accordance with the UK Policy Framework for Health and Social Care Research.

Ethics approval has been obtained from the University of Southampton Faculty of Medicine Ethics committee (ERGO II Reference 66810).

The regulatory authorities will be notified of any subsequent non-substantial amendments deemed necessary for the study. The appropriate approvals will be sought for any substantial amendments, and the appropriate approvals and processes will be followed for linked datasets (NCDS/BCS70/ACONF).

The dissemination of our findings will be achieved through academic (conferences, academic papers, harmonising datasets for wider user and working with other award holders) and non-academic (social media, website, blog and a podcast) channels. Central to our dissemination plans is to ensure that our findings have maximum impact, particularly to non-academic audiences. We will undertake broad engagement activities with our Patient and Public Advisory Board and other key stakeholders to identify potential for impact from outputs. Our extensive connections in local and national Public Health will help facilitate this.

## Discussion/conclusion

### Challenges

The MELD-B research collaboration will face several practical and operational challenges, we recognise that new challenges will emerge and change as the project progresses, and it is not possible to anticipate every eventuality. However, challenges include gaining access to five separate datasets, harmonisation and volume- and quality- control of variables across these datasets. Methodological challenges include ensuring we utilise the most appropriate methods to use as these are likely to vary depending on the questions being addressed through the life of the project as new ideas emerge. It is also important we ensure that our Patient and Public Advisory Board are embedded in the project from its conception and that PPI colleagues are supported to provide meaningful contributions across all work packages. The development and understanding of new definitions and terminology across disciplines of public health, epidemiology, data science and computer science and engaging relevant stakeholders before we have findings will also be challenging. Additionally, an important aspect we face is the scale of the project – we are likely to have a team of over 40 members, across multiple locations and institutions in the UK, with different skills sets and disciplines who are working together to achieve multiple objectives in a short space of time. Therefore, managing the dynamics associated with such a large, complex project is a recognised issue.

### Potential impact

Traditional models of prevention (primary, secondary, tertiary) are limited in application to multimorbidity, particularly when considering the accrual of conditions over the lifecourse. The Academy of Medical Sciences emphasised prevention of early-onset MLTC-M clusters and targeting individual risk and its research priorities included identifying ‘the burden caused by common clusters of conditions’.<sup>61</sup> It highlighted conditions in common clusters with much less evidence on prevention, such as depression.<sup>61,62</sup> Public Health England has highlighted the importance of a lifecourse approach and identifying critical points for intervention: ‘Unlike a disease-oriented approach, which focuses on interventions for a single condition often at a single life stage, a lifecourse approach considers the critical stages, transitions, and settings where large differences can be made in promoting or restoring health and wellbeing.’<sup>63</sup> Moreover, public health interventions focusing on wider determinants are highly cost-effective, averaging a fivefold return on investment.<sup>64</sup> From a public health perspective, characterising the sequence of accrual of conditions and their determinants across the whole lifecourse would give the opportunity to identify critical timepoints for population-level prevention efforts.

In summary, the MELD-B Research Collaboration aims to fill several key gaps in the research evidence in MLTC-M and thereby influence policy and practice. It will achieve this by developing a deeper understanding of the lived experience of ‘burdensomeness’ and ‘complexity’ of multimorbidity, identifying new clusters of burdensome MLTC-M and their key early-life risk factors, mapping trajectories across the lifecourse towards burdensome clusters in those under 65, and modelling prevention scenarios to inform policy.

### Acknowledgements

We would like to acknowledge all other members of the MELD-B Consortium: Paul Smart, Nic Fair, Lingzi Wang, Kate Mathews, Aiysha Qureshi, Heather Parsons, Kelly Cheung, Jim McMahon, Jack Welch, Rosie Martin, Lynn Laidlaw, Rita Rajababoo and Saroj Parekh.

### Declaration of conflicts of interest

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: RKO is a member of the National Institute for Health and Care Excellence (NICE) Technology Appraisal Committee, member of the NICE Decision Support Unit (DSU), and associate member of the NICE Technical Support Unit (TSU). She has served as a paid consultant to the pharmaceutical industry, providing unrelated methodological advice. She reports teaching fees

from the Association of British Pharmaceutical Industry (ABPI) and the University of Bristol.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Institute for Health Research (NIHR) under its Programme Artificial Intelligence for Multiple and Long-Term Conditions (NIHR203988). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

### ORCID iDs

Sebastian Stannard  <https://orcid.org/0000-0002-6139-1020>

Ashley Akbari  <https://orcid.org/0000-0003-0814-0801>

Frances S Mair  <https://orcid.org/0000-0001-9780-1135>

### References

- Barnett K, Mercer SW, Norbury M, et al. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *Lancet* 2012; 380(9836):37–43.
- Ashworth M, Durbaba S, Whitney D, et al. Journey to multimorbidity: longitudinal analysis exploring cardiovascular risk factors and sociodemographic determinants in an urban setting. *BMJ Open* 2019; 9(12):e031649.
- Ruel G, Lévesque JF, Stocks N, et al. Understanding the evolution of multimorbidity: evidences from the North West Adelaide Health Longitudinal Study (NWAHS). *PLoS One* 2014; 9(5):e96291.
- Xu X, Mishra GD, Dobson AJ, et al. Progression of diabetes, heart disease, and stroke multimorbidity in middle-aged women: A 20-year cohort study. *PLoS Med* 2018; 15(3): e1002516.
- May CR, Eton DT, Boehmer K, et al. Rethinking the patient: using Burden of Treatment Theory to understand the changing dynamics of illness. *BMC Health Serv Res* 2014; 14: 281.
- Stannard S, Holland E, Crozier SR, et al. Early-onset burdensome multimorbidity: an exploratory analysis of sentinel conditions, condition accrual sequence and duration of three long-term conditions using the 1970 British Cohort Study *BMJ Open* 2022; 12:e059587. doi: [10.1136/bmjopen-2021-059587](https://doi.org/10.1136/bmjopen-2021-059587).
- Gluckman PD, Buklijas T and Hanson MA. The developmental origins of health and disease (DOHaD) concept: Past, present, and future. In Rosenfeld CS (ed) *The Epigenome and Developmental Origins of Health and Disease*. Academic Press 2016 (pp. 1–15).
- Fleming TP, Watkins AJ, Velazquez MA, et al. Origins of lifetime health around the time of conception: causes and consequences. *Lancet* 2018; 391(10132):1842–1852.
- Hanson MA and Gluckman PD. Developmental origins of health and disease: moving from biological concepts to interventions and policy. *Int J Gynaecol Obstet* 2011;115 Suppl 1:S3–S5.
- Wilding S, Ziauddeen N, Smith D, et al. Are environmental area characteristics at birth associated with overweight and obesity in school-aged children? Findings from the SLOPE (Studying Lifecourse Obesity PrEdictors) population-based cohort in the south of England. *BMC Med* 2020;18(1):43.
- Barouki R, Gluckman PD, Grandjean P, et al. Developmental origins of non-communicable disease: implications for research and public health. *Environmental Health* 2012;11(1): 42.
- Holliday R. Epigenetics: a historical overview. *Epigenetics* 2006;1(2):76–80.
- Waterland RA and Michels KB. Epigenetic epidemiology of the developmental origins hypothesis. *Annu Rev Nutr* 2007; 27:363–388.
- Harper S, Lynch J and Smith GD. Social determinants and the decline of cardiovascular diseases: understanding the links. *Annu Rev Public Health* 2011; 32:39–69.
- Humphreys J, Jameson K, Cooper C, et al. Early-life predictors of future multi-morbidity: results from the Hertfordshire Cohort. *Age Ageing* 2018; 47(3):474–478.
- Johnston MC, Black C, Mercer SW, et al. Impact of educational attainment on the association between social class at birth and multimorbidity in middle age in the Aberdeen Children of the 1950s cohort study. *BMJ Open* 2019; 9(1): e024048.
- Gondek D, Bann D, Brown M, et al. Prevalence and early-life determinants of mid-life multimorbidity: evidence from the 1970 British birth cohort. *BMC Public Health* 2021; 21(1): 1319. doi: [10.1186/s12889-021-11291-w](https://doi.org/10.1186/s12889-021-11291-w).
- 2019 Health and Social Care Select Committee report ‘First 1000 days of life’. <https://publications.parliament.uk/pa/cm201719/cmselect/cmhealth/1496/1496.pdf> (2019, accessed 04 May 2023).
- Ipsos MORI ‘State of the Nation: Understanding Public Attitudes to the Early Years report’. [https://mk0royalfoundatcnh10.kinstacdn.com/wp-content/uploads/2020/11/Ipsos-MORI-SON\\_report\\_FINAL\\_V2.4.pdf](https://mk0royalfoundatcnh10.kinstacdn.com/wp-content/uploads/2020/11/Ipsos-MORI-SON_report_FINAL_V2.4.pdf) (2020, accessed 05 May 2023).
- Mair FS and May CR. Thinking about the burden of treatment. *BMJ* 2014; 349. doi: [10.1136/bmj.g6680](https://doi.org/10.1136/bmj.g6680).
- Sav A, King MA, Whitty JA, et al. Burden of treatment for chronic illness: a concept analysis and review of the literature. *Health Expect* 2015;18(3):312–324.
- Shippee ND, Shah ND, May CR, et al. Cumulative complexity: a functional, patient-centered model of patient complexity can improve research and practice. *J Clin Epidemiol* 2012; 65(10):1041–1051.
- National Institute for Health and Care Excellence (NICE). *Multimorbidity: clinical assessment and management (NG56)* <https://www.nice.org.uk/guidance/ng56> (2016, accessed 14 January 2023).

24. Bohlen K, Scoville E, Shippee ND, et al. Overwhelmed patients: a videographic analysis of how patients with type 2 diabetes and clinicians articulate and address treatment burden during clinical encounters. *Diabetes Care* 2012; 35(1):47–49.
25. Fiese BH, Wamboldt FS and Anbar RD. Family asthma management routines: connections to medical adherence and quality of life. *J Pediatr* 2005; 146(2):171–176.
26. Pifferi M, Bush A, Di Cicco M, et al. Health-related quality of life and unmet needs in patients with primary ciliary dyskinesia. *Eur Respir J* 2010; 35(4):787–794.
27. Vijan S, Hayward RA, Ronis DL, et al. Brief report: the burden of diabetes therapy: implications for the design of effective patient-centered treatment regimens. *J Gen Intern Med* 2005; 20(5):479–482.
28. Ho PM, Rumsfeld JS, Masoudi FA, et al. Effect of medication nonadherence on hospitalisation and mortality among patients with diabetes mellitus. *Arch Intern Med* 2006; 166(17):1836–1841.
29. Rasmussen JN, Chong A and Alter DA. Relationship between adherence to evidence-based pharmacotherapy and long-term mortality after acute myocardial infarction. *JAMA* 2007; 297(2):177–186.
30. Morris JE, Roderick PJ, Harris S, et al. *Treatment burden for patients with multimorbidity: cross-sectional study with exploration of a single-item measure*. BJGP; 2020. <https://bjgp.org/content/early/2020/12/13/BJGP.2020.0883>
31. Carroll C. Qualitative evidence synthesis to improve implementation of clinical guidelines. *BMJ* 2017; 356, j80.
32. Booth A, Noyes J, Flemming K, et al. Formulating questions to address the acceptability and feasibility of complex interventions in qualitative evidence synthesis. *BMJ Global Health* 2019; 4(Suppl. 1), e001107.
33. PROSPERO, *International prospective register of systematic reviews*. Centre for Reviews and Dissemination, University of York. <https://www.crd.york.ac.uk/prospero/> (2021, 01 August 2021).
34. Noyes J, Booth A, Cargo M, et al. Chapter 21: Qualitative evidence. In: Higgins JPT, Thomas J, Chandler J, et al (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.3 (updated February 2022). Cochrane, 2022.
35. Noyes J, Booth A, Cargo M, et al. Cochrane qualitative and implementation methods group guidance series—paper 1: Introduction. *J Clin Epidemiol* 2018; 97, 35–38.
36. Booth A, Noyes J, Flemming K, et al. Structured methodology review identified seven (RETREAT) criteria for selecting qualitative evidence synthesis approaches. *J Clin Epidemiol*. 2018;99:41–52. doi: [10.1016/j.jclinepi.2018.03.003](https://doi.org/10.1016/j.jclinepi.2018.03.003)
37. Booth A, Noyes J, Flemming K, et al. *Guidance on choosing qualitative evidence synthesis methods for use in health technology assessments of complex interventions*. Integrate-HTA. <http://www.integrate-hta.eu/downloads/> (2016, accessed 15/06/2022).
38. Carroll C, Booth A, Leaviss J, et al. “Best fit” framework synthesis: refining the method. *BMC Med Res Methodol* 2013;13:37. doi: [10.1186/1471-2288-13-37](https://doi.org/10.1186/1471-2288-13-37).
39. Thomas J and Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Med Res Methodol* 2008; 8:45. doi: [10.1186/1471-2288-8-45](https://doi.org/10.1186/1471-2288-8-45).
40. May CR, Cummings A, Myall M, et al. Experiences of long-term life-limiting conditions among patients and carers: what can we learn from a meta-review of systematic reviews of qualitative studies of chronic heart failure, chronic obstructive pulmonary disease and chronic kidney disease? *BMJ Open* 2016; 6(10):e011694.
41. Tong A, Sainsbury P and Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32- item checklist for interviews and focus groups. *Int J Qual Health Care* 2007; 19:349–357.
42. Tong A, Flemming K, McInnes E, et al. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. *BMC Med Res Methodol* 2012;12:181. doi: [10.1186/1471-2288-12-181](https://doi.org/10.1186/1471-2288-12-181).
43. Lewin S, Bohren M, Rashidian A, et al. Applying GRADE-CERQual to qualitative evidence synthesis findings—paper 2: How to make an overall CERQual assessment of confidence and create a summary of qualitative findings table. *Implement Sci* 2018;13(10). doi: [10.1186/s13012-017-0689-2](https://doi.org/10.1186/s13012-017-0689-2).
44. Jones J and Hunter D. Consensus methods for medical and health services research. *BMJ*. 1995;311:376–380.
45. Rowe G and Wright G. The Delphi technique as a forecasting tool: issues and analysis. *Int J Forecast* 1999; 15:353–375.
46. Hasson F, Keeney S and McKenna H. Research guidelines for the Delphi survey technique. *J Adv Nurs* 2000; 32(4):1008–1015.
47. Murphy MK, Black NA, Lamping DL, et al. Consensus development methods, and their use in clinical guideline development. *Health Technol Assess* 1998;2:1–88.
48. Diamond IR, Grant RC, Feldman BM, et al. Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies. *J Clin Epidemiol* 2014; 67:401–409
49. Department of Health & Social Care, *a guide to good practice for digital and data-driven health technologies*. <https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology> (2021, accessed 23 September 2021).
50. UK Health Data Research Alliance. (UKHDRA). *Trusted Research Environments (TRE): A strategy to build public trust and meet changing health data science needs*. Green Paper v2.0. [https://ukhealthdata.org/wp-content/uploads/2020/07/200723-Alliance-Board\\_Paper-E\\_TRE-Green-Paper.pdf](https://ukhealthdata.org/wp-content/uploads/2020/07/200723-Alliance-Board_Paper-E_TRE-Green-Paper.pdf) (2020, accessed 23 September 2021).

51. Herrett E, Gallagher AM, Bhaskaran K, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* 2015; 44(3):827–836.
52. Ford DV, Jones KH, Verplancke JP, et al. The SAIL Databank: building a national architecture for e-health research and evaluation. *BMC Health Serv Res* 2009; 9: 157. doi: [10.1186/1472-6963-9-157](https://doi.org/10.1186/1472-6963-9-157). PMID: 19732426; PMCID: PMC2744675.
53. Ho ISS, Azcoaga-Lorenzo A, Akbari A, et al. Measuring multimorbidity in research: Delphi consensus study. *BMJ Medicine* 2022; 1:e000247. doi: [10.1136/bmjmed-2022-000247](https://doi.org/10.1136/bmjmed-2022-000247).
54. Hanlon P, Jani BD, Nicholl B, et al. Associations between multimorbidity and adverse health outcomes in UK Biobank and the SAIL Databank: A comparison of longitudinal cohort studies. *PLoS Med* 2022;19(3):e1003931. doi: [10.1371/journal.pmed.1003931](https://doi.org/10.1371/journal.pmed.1003931).
55. Lyons J, Akbari A, Agrawal U, et al. Protocol for the development of the Wales Multimorbidity e-Cohort (WMC): data sources and methods to construct a population-based research platform to investigate multimorbidity. *BMJ Open* 2021;11(1):e047101. doi: [10.1136/bmjopen-2020-047101](https://doi.org/10.1136/bmjopen-2020-047101).
56. Kuan V, Denaxas S, Gonzalez-Izquierdo A, et al. A chronological map of 308 physical and mental health conditions from 4 million individuals in the English National Health Service. *Lancet Digit Health* 2019;1(2):e63–e77. doi: [10.1016/S2589-7500\(19\)30012-3](https://doi.org/10.1016/S2589-7500(19)30012-3).
57. Cambridge CPRD code list. [https://www.phpc.cam.ac.uk/pcu/research/research-groups/crmh/cprd\\_cam/codelists/v11/](https://www.phpc.cam.ac.uk/pcu/research/research-groups/crmh/cprd_cam/codelists/v11/). (2018, accessed on 20 January 2023).
58. Naimi AI, Cole SR and Kennedy EH. An introduction to g methods. *Int J Epidemiol*. 2017;46(2):756–762.
59. Mansournia MA, Etminan M, Danaei G, et al. Handling time varying confounding in observational research. *BMJ* 2017; 359:j4587.
60. National Institute for Health and Care Research (NIHR): *Payment guidance for researchers and professionals*. <https://www.nihr.ac.uk/documents/payment-guidance-for-researchers-and-professionals/27392#payment> (2022, accessed 22 August 2023).
61. The Academy of Medical Sciences. *Multimorbidity: a priority for global health research*. <https://acmedsci.ac.uk/file-download/82222577> (2018, accessed 24 January 2023).
62. Luppino FS, de Wit LM, Bouvy PF, et al. Overweight, Obesity, and Depression: A Systematic Review and Meta-Analysis of Longitudinal Studies. *Arch Gen Psych* 2010; 67(3), 220–229.
63. Public Health England, health matters: Prevention - a life course approach. <https://www.gov.uk/government/publications/health-matters-life-course-approach-to-prevention/health-matters-prevention-a-life-course-approach> (2019, accessed 05 December 2022).
64. Masters R, Anwar E, Collins B, et al. Return on investment of public health interventions: a systematic review. *J Epidemiol Community Health* 2017; 71:827–834.