

Maximising the value of UK population cohorts

MRC Strategic Review of the Largest UK Population Cohort Studies



MEDICAL RESEARCH COUNCIL (Swindon office)
2nd Floor David Phillips Building
Polaris House
North Star Avenue
Swindon SN2 1FL

MEDICAL RESEARCH COUNCIL (London office)
14th Floor
One Kemble Street
London WC2B 4AN
www.mrc.ac.uk

Published: February 2014

Acknowledgements and Attributions

The MRC Strategic Review was carried out by the MRC Population Health Sciences Group (PHSG). The work was directed by the Cohort Strategic Review Subgroup comprising members of PHSG and the MRC Cross-Board Cohort Advisory Group. We are indebted to the UK cohort funders BHF, ESRC, Cancer Research UK, Breakthrough Breast Cancer and the Wellcome Trust, and in particular members of the cohort study teams, for providing the data for this review. We are grateful for the invaluable contribution of Professor Hazel Inskip and Dr Sarah Crozier who conducted the cohort modelling and projection exercise. Thanks also to members of the research community and stakeholders for their insightful input at the cohort workshop and for providing feedback on the review document. Responsibilities at MRC Head Office for the review were: initial scoping by Dr Jana Voigt; project management by Dr Ghada Zoubiane; data collection and analyses by Jess Phanwises and Kate Sheedy; case studies compiled by Cara Steger; administrative support by Lauren Rooney; and oversight of the work by Dr Janet Valentine. The review was developed under the strategic guidance of Professor Jill Pell.

Membership of the Cohort Strategic Review Subgroup

Professor Jill Pell (Chair)
Professor Richard Hayes
Professor Ian Deary
Professor Hazel Inskip
Professor Frank Kelly
Professor Andrew Steptoe
Professor Laura Rodrigues
Professor Kate Hunt
Professor Frank Kee

Reproduction of images of cohort members with thanks to the Avon Longitudinal Study of Parents and Children (Children of the 90s), The Determinants of Adolescent Social well-being and Health, European Prospective Investigation of Cancer Norfolk, Hertfordshire Cohort Study, Lothian Birth Cohort 1936 and the Southampton Women's Survey.

Contents

List of Figures and Tables	3
Abbreviations	4
Foreword	5
Executive Summary	6
1. Introduction	11
2. Review of Current Portfolio	13
2.1 Methodology	13
2.1.1 Inclusion criteria	13
2.1.2 Data collected	14
2.2 Results from the data gathering	14
2.2.1 Overview of the portfolio	14
2.2.2 Start date of the cohorts	15
2.2.3 Cohort follow-up	16
2.2.4 Age of the cohort participants	17
2.2.5 Current cohort sample size	20
2.2.6 Variables collected	22
2.2.7 Data linkage	26
2.2.8 Biological samples and omics	27
2.2.9 Collaboration across cohorts	29
3. Cohort Portfolio Projection	30
3.1 Methodology	30
3.2 Results	31
3.3 Summary	36
4. Recommendations	37
4.1 Stakeholder engagement	37
4.2 Findings	37
4.2.1 Strengths	37
4.2.2 Gaps and potential limitations in the portfolio	37
4.2.3 Opportunities	39
4.3 Key recommendations	40
4.4 Next steps	44

Annexes

Annex 1	Cohorts included in the Strategic Review	45
Annex 2	Cohort Questionnaire	49
Annex 3	Cohort Data Overview	54
Annex 4	Cohort Data by Age and Number	59
Annex 5	Cohort Data by Consent for Re-contact and Linkage	64
Annex 6	Cohort Data by Anthropometric and Blood Pressure Variables	66
Annex 7	Cohort Data by Physical Health Variables	68
Annex 8	Cohort Data by Mental Health and Cognitive Measures	70
Annex 9	Cohort Data by Lifestyle Variables	72
Annex 10	Cohort Data by Socio-Economic Position	74
Annex 11	Cohort by Data Linkage	76
Annex 12	Cohort Data by Biological Samples	78
Annex 13	Cohort Data by Omics Analysis	80
Annex 14	Cohort Workshop Agenda and Attendees	82

List of Figures and Tables

Figure 1	Cohorts by start date of initial data collection	15
Figure 2	Length of cohort follow-up	16
Figure 3	Age range of cohort participants at recruitment	18
Figure 4	Estimated age range of cohort participants in 2013	19
Figure 5	Estimated current number of cohort participants	20
Figure 6	Estimated current number of participants in the Million Women Study and UK Biobank	21
Figure 7	Proportion of cohorts collecting anthropometric and blood pressure variables	22
Figure 8	Proportion of cohorts collecting physical health variables	23
Figure 9	Proportion of cohorts collecting mental health and cognitive function variables	24
Figure 10	Proportion of cohorts collecting lifestyle variables	25
Figure 11	Proportion of cohorts collecting socio-economic variables	26
Figure 12	Proportion of cohorts with data linkage	27
Figure 13	Proportion of cohorts collecting biological samples	27
Figure 14	Proportion of cohorts with omics analysis	28
Figure 15	Numbers of cohort study participants by age and sex in 2012	32
Figure 16	Numbers of cohort study participants by age and sex in 2012, excluding the Million Women Study and UK Biobank	33
Figure 17	Projected numbers of cohort study participants by age and sex in 2022	34
Figure 18	Projected numbers of cohort study participants by age and sex in 2022, excluding the Million Women Study and UK Biobank	35
Tables		
Table 1	Cohorts receiving core funding from the MRC (partially or fully)	13
Table 2	Cohorts by category	14
Table 3	Cohorts included in the projection	30

Abbreviations

1970 BCS	1970 British Cohort Study
ACONF	Aberdeen Children of the 1950s
ALSPAC	Avon Longitudinal Study of Parents and Children
BBSRC	Biotechnology and Biological Sciences Research Council
BHF	British Heart Foundation
BiB	Born in Bradford
BIS	Department for Business, Innovation and Skills
BRHS	British Regional Heart Study
BWHHS	British Women's Heart & Health Study
CFAS I	Cognitive Function and Ageing Studies I
CFAS II	Cognitive Function and Ageing Studies II
CLOSER	Cohort and Longitudinal Studies Enhancement Resources
CRUK	Cancer Research UK
DASH	Determinants of Adolescent Social well-being and Health
DEXA	Dual Energy X-ray Absorptiometry
ELSA	English Longitudinal Study of Ageing
EPIC Norfolk	European Prospective Investigation of Cancer Norfolk
EPIC Oxford	European Prospective Investigation of Cancer Oxford
ESRC	Economic and Social Research Council
GUS	Growing up in Scotland
HALCyon	Healthy Ageing Across the Life Course
HCS	Hertfordshire Cohort Study
LBC1936	Lothian Birth Cohort 1936
MCS	Millennium Cohort Study
MRC	Medical Research Council
MRI	Magnetic resonance imaging
NCDS/1958 BC	The National Child Development Study / 1958 Birth Cohort
NICOLA	Northern Ireland Cohort for Longitudinal Study of Ageing
NIHR	National Institute of Health Research
NSHD/1946 BC	MRC National Survey of Health and Development Cohort / 1946 Birth Cohort
Rural Uganda GPC	Rural Uganda General Population Cohort
SABRE	Southall and Brent Revisited
SWS	Southampton Women's Survey
TEDS	Twin Early Development Study
Twenty-07	West of Scotland Twenty-07 Study
UKWCS	UK Women's Cohort Study

Foreword

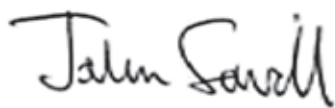
The Medical Research Council has just celebrated 100 years of ground breaking research. One of the most influential MRC funded studies to have a profound impact on global health is the 1950s Richard Doll study of a cohort of GPs which first identified the harmful effects of smoking. Since then, the MRC has followed many population subgroups over time to understand the role of biological, environmental and lifestyle factors shaping human health. Today, maximising the value of these population cohort studies is a priority.

It is timely, therefore, to review the UK's population cohorts, showcasing the rich diversity of studies funded by the MRC and others, from the longest running birth cohort in the world, to the Southampton Women's Survey which famously collected data on mothers before they conceived.

A striking feature to emerge from the MRC Cohort Strategic Review is the number of people in the UK who have participated in cohort studies. Altogether 2.5m people have taken part and currently around 2.2m people – 3.5% of the population – are cohort members. We owe them a debt of gratitude, not only for their time and cooperation but for their belief that participating in research will lead to gains in societal health and wellbeing. Over half a million people are part of UK Biobank and soon the entire cohort will be genotyped. The combination of lifestyle and environmental measures with state of the art biological analyses greatly increases the added value of cohort studies for new scientific advances.

Participants from the UK cohort studies have given consent for their personal data to be linked to NHS records and other data sources such as education and the census. Linking cohort study data to routine health and administrative datasets in safe environments that protect confidentiality increases the scope and scale of research possibilities. Investments by the MRC and partners to create the Farr Institute of Health Informatics Research and UK Health Informatics Research Network will facilitate data linkage to identify the causes of disease, develop personalised treatments and monitor health risks and drug safety. Trustworthy use of personal health data, by participating in studies or through use of anonymised health records, is fundamental to improving patient care and public health and is essential for the UK to remain at the forefront of medical research.

As the MRC enters its second century, the cohort studies are in a prime position to take advantage of high throughput technologies and data linkage to enhance our understanding of protective and risk factors underpinning health, wellbeing and disease.



Professor Sir John Savill
Chief Executive
January 2014

Executive Summary

Introduction

The UK supports an unparalleled collection of large scale population cohort studies which provide a wealth of longitudinal phenotypic, biological and social data for studying health and wellbeing throughout the life course. The ability to link to health and other routine records, collect data and samples from consenting participants and apply cutting edge imaging and omics technologies, places the UK in an optimal position to fully capitalise on these major research assets.

For more than 50 years, the MRC has funded a diverse range of population cohorts that have provided important insights into the determinants of health, wellbeing and disease, and have contributed to public health policy and changes in clinical practice. Maximising the value of longitudinal studies for new scientific discoveries and informing health policy and practice is a key MRC strategic priority.

A comprehensive understanding of the scientific niche of each cohort within the context of the wider UK population cohort landscape is vital to inform funding decisions, strategic discussions and ensure value for money. The aims of this strategic review are to:

- Document the current investment in all major UK population cohorts, including the positioning of large MRC funded cohorts within the UK portfolio
- Model the projected trajectory of the current UK portfolio over the next 10 years
- Highlight scientific and translational opportunities to inform future investments in cohorts and associated studies over the next decade.

Overview of the portfolio

A total of 34 cohorts were included in the review, encompassing 19 cohorts partially or fully funded by the MRC and 15 cohorts funded entirely by others. These cohort studies comprise the majority of large longitudinal population studies in the UK. The combined annual spend supporting the 34 cohort studies is £27.6m, with MRC funding accounting for £9.6m of the total. In addition, two new large population cohorts, one of which is funded in part by the MRC, will commence in 2014.

A key feature of the UK portfolio is the large number of cohorts that have been followed for a long period of time, with half of the cohorts (n=17) having been followed for at least 20 years. One of the oldest cohorts in the portfolio is the MRC National Survey of Health and Development/1946 Birth Cohort, which is the oldest cohort continuously followed from birth in the world. On average, one new large population cohort has been funded every year since 1990.

The age range of the UK cohort portfolio spans the whole life course from birth to over 100 years old, with the Southampton Women's Survey uniquely collecting data on mothers before they conceived. Four of the 34 cohort studies include women only and one study is exclusively male.

The current size of the cohorts ranges from 1.24 million in the Million Women Study to approximately 150 in CFAS I (of the original 18,500 who were over 65 years of age at the time of recruitment). The Million Women Study and UK Biobank together account for three quarters of the total participants in the portfolio. It is estimated that 2.5 million people in the UK have been recruited to large population cohort studies and today there are over 2.2 million people, which is 3.5% of the UK population, who are still taking part.

Variables, linkage and omics

One of the aims of the data gathering exercise was to illustrate at a high level, the types of variables collected by the cohort studies in the UK portfolio.

Virtually all cohort studies collected anthropometric measures. Physical health measures were collected by all but three cohort studies in the portfolio, including 23 studies which all measured respiratory, cardiovascular and musculoskeletal function. Two thirds of the cohort studies collected data on either mental health or cognitive function. A total of 23 studies (16 MRC funded) measured both mental health and cognitive function. All cohort studies collected lifestyle data on alcohol consumption and smoking, and data on physical activity and diet was gathered by all but two studies. Information on education and occupation was recorded by every study, alongside a range of other study-specific socio-economic variables.

All 34 cohort studies have obtained consent from participants for re-contact. A total of 31 cohort studies had linked to different routine datasets across health, administrative and environmental sources. More than half of MRC funded cohorts had linked to primary and/or secondary health records.

Biological samples were collected by all but three cohort studies, with blood being the most common tissue collected. A total of 26 cohort studies had conducted omics studies. Genotyping had been carried out on sub-populations within 23 of the cohorts. Over half of the MRC funded cohort studies (n=10) had used epigenetic studies, mostly on sub-groups of participants. Only nine cohort studies had carried out whole genome sequencing on sub-groups within the cohorts and 10 studies had carried out high throughput metabolomics, commonly using NMR platforms.

Cohort Portfolio Projection

Twenty-eight of the total 34 cohorts in this review were included in an exercise modelling the current and projected profile of the UK cohort portfolio in the next 10 years. Currently, broadly equal numbers of men and women are being studied up to age 20 years, but thereafter the number of women exceeds that of men at all ages. Far greater numbers of women than men are being studied at ages over 50 years, but there are also fewer men between the ages of 20 and 40 years. The Million Women Study largely accounts for the disparity in the numbers of men and women being studied. UK Biobank is also a particularly large study and contributes to the large numbers observed at ages 45-74 years for both men and women.

Extrapolation of data from the 28 cohorts forward over ten years, allowing for attrition, and including estimated numbers for the two newly planned cohorts (the Life Study and NICOLA) showed that in 2022 most of the cohort participants will be within the 55-85 year age range. Unsurprisingly, the number of women studied still greatly exceeds that of men in later life. The Life Study which commences in 2014 will bring in a large number of children of both sexes. It will also collect data on the children's fathers which will boost the low number of men in the 30-50 year age range.

Key recommendations

- **Guidance on use of omics platforms and emerging technologies**

Integration of genomics, epigenetics, metabolomics, imaging and other emerging technologies into cohort studies has the potential to improve our understanding of the aetiology, risk prediction and stratification of disease across different populations. Two thirds of the 34 cohort studies have carried out genotyping and more than half the studies have used epigenetics and/or metabolomics studies. Shared learning on tissue specificity, platform choice, informatics and tools will allow cohort studies to more appropriately adopt evolving technologies, and enable replication of findings and comparisons between studies. There is scope for greater use of imaging to measure associations between early exposures and later phenotypic structural and functional changes. A potential opportunity for the UK is comparative omics or imaging studies across the large well-phenotyped population cohorts.

Recommendation 1

Cohort studies should use standardised or validated sample collection, storage, tools and platforms for evolving technologies, where possible aligning with leaders in the field such as UK Biobank, to enable future cross-cohort comparison.

The MRC, cohort leaders and experts in the field should work together to develop guidance on best practice for high throughput science in large-scale populations, in particular for epigenetic studies.

- **Linkage to routine and research datasets**

Linkage to routine health records, cross-sector administrative and environmental data and research datasets greatly expands the scope of a cohort to carry out clinical, public health and socio-economic research. Recent infrastructure initiatives are increasing secure access to clinical records for research and administrative datasets. This infrastructure has been complemented by major investments from the MRC in partnership with nine UK funders to establish the Farr Institute to build capacity and undertake research using record linkage.

Although almost all of the cohort studies in the UK portfolio are linked to at least one source of routine records, there is scope for more extensive record linkage to enrich the study data and expand opportunities for new discovery science.

Recommendation 2

Broad and enduring consent for linkage to routine data needs to be obtained from cohort participants, wherever possible, for all prospective studies and sweeps.

The scientific and interdisciplinary potential of a cohort should be enhanced through linkage to the increasing number of routine health records and administrative datasets available in the UK. Studies can draw on expertise in data linkage within the Farr Institute, UK Health Informatics Research Network and Administrative Data Research Centres.

- **Skills and capacity in analysing complex datasets**

Cohort studies involve increasingly large datasets, particularly if the cohort is linked to health or administrative records, 'omics' or imaging data. Managing, linking and analysing these complex datasets requires a range of interdisciplinary skills. Building sustainable UK research capability in informatics is a major priority for the MRC. Recent and on-going MRC funded initiatives offer a range of training and career development opportunities in analysing complex data.

Recommendation 3

Cohort studies should take advantage of the opportunities for skills development in interrogating large and complex data through collaborations with centres of excellence such as the Farr Institute.

- **Data discovery and sharing**

The ability to discover information about a cohort and have access to data and samples is essential to fully realising the scientific and translational value of these resources. Although MRC funded cohorts are listed in at least one cohort directory, not all cohorts in this review are as readily discoverable.

The MRC has clear data sharing policies¹ which state that sharing data, and where appropriate samples, should be normal practice. Despite the MRC policies and similar policies of other funders, approaches to data access and sharing vary enormously between cohorts and cohort studies are not yet at the expected level of compliance.

¹ MRC policy on sharing of research data from population and patient studies
<http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/datasharing/Policy/PHSPolicy/index.htm>

Recommendation 4

Cohort leads should ensure that their studies are easily discoverable via directories.

Processes are needed to ensure that all MRC funded cohorts comply with MRC data sharing policies. Studies need to be accessible and have transparent governance procedures in place that enable data sharing and where appropriate, access to samples.

- **Cross-cohort analyses**

Cross-cohort collaborations can further enhance scientific opportunities and the translational potential of individual cohorts. Most of the cohorts in this review collect the same core information on common exposures and variables, however, the individual measures and methods used vary enormously. Combining or harmonising phenotypic variables across studies relies on high-quality study meta-data and the use of data standards. Widespread adoption of data standards in UK cohort studies is not yet commonplace, and the quality of meta-data within studies is highly variable.

Recommendation 5

Adoption of core common data standards, sharing knowledge and improving meta-data quality should be encouraged and facilitated by cohort studies, the MRC and other funders.

- **Novel methods for establishing and maintaining cohorts**

Sustained support of large cohort studies with regular data and sample collection, and on-going maintenance of the resource, is very expensive. New digital technologies and remote data capture can provide cost-efficient alternatives for recruitment, retention and data gathering. More studies are needed to evaluate the benefits and limitations of new technologies and data capture. However, investigators should consider adopting less costly remote monitoring on a case-by-case basis.

Recommendation 6

Cost-effective methods, such as the use of new digital technologies and remote data and sample capture, should be adopted where possible and appropriate to reduce the costs of enhancing and maintaining cohort studies.

- **Increasing the research and translational application of cohort studies**

There is potential for even greater application of UK cohort studies to address important public health challenges such as obesity, dementia and alcohol consumption. The cohort studies in the review collect a similar set of core phenotypic information. Data from the cohort studies can potentially be applied to a range of research or policy questions.

Research outcomes from cohort studies are of relevance to a range of stakeholders, including policy makers and practitioners, as well as researchers. Currently, policy makers and other beneficiaries of research are only familiar with a sub-set of UK cohort studies.

Recommendation 7

Effective models of two-way engagement between cohort study teams, policy makers and/or practitioners should be established to increase the impact of cohort study research outputs and potential for translation to inform evidence based policies.

Next steps

This document has been produced to assist the MRC in strategic decisions and policy development. The review will also be relevant to other funders, researchers and policy makers seeking evidence to evaluate policies and interventions.

The recommendations aimed at capacity building and more expansive linkage to routine data sources are already the subject of current initiatives. The MRC will work in partnership with other cohort funders to take forward recommendations on discoverability, data sharing and accessibility, adoption of data standards and improving meta-data quality to ensure a coordinated approach to these areas. Many of the MRC funded and other UK cohort studies are carrying out omic analyses on cohort biosamples. There is scope for the MRC to develop guidance for population studies on best practice, ranging from sample storage to analyses.

This review showcases for the first time the breadth of large population cohorts that exist in the UK. In addition to providing a useful tool to aid funding decisions, the recommendations are intended to strengthen the value of MRC and other UK cohort assets by highlighting areas that will enable new discovery science and improve translation of research outcomes.

1. Introduction

The United Kingdom supports an unparalleled collection of large-scale population cohort studies which provide a wealth of longitudinal phenotypic, biological and social data for studying health and wellbeing throughout the life course. For more than 50 years, the MRC has funded a diverse range of population cohorts that differ in size, age, gender, ethnicity, socio-economic position, geographic location and length of follow-up. The portfolio of MRC supported population cohorts includes the world's longest running birth cohort and the largest longitudinal study of women's health. Some cohort studies are focused on investigating specific exposures or health outcomes. Other studies examine a range of environmental, lifestyle and biological factors that influence population health and wellbeing.

In contrast to numerous other countries, the UK offers a world leading environment for establishing and maintaining cohort studies. Historic and current differences in laws and regulations relating to privacy and data linkage mean that in many European countries large-scale studies involving contact with participants or linkage to national datasets are difficult. Nordic countries have numerous cohorts that are based on data linkage, however, studies with direct participant contact involving measurements and biological sample collection are less common than in the UK. Conversely, the absence of a national health records system in the United States limits longitudinal data collection on cohort participants' health. The ability to link to health and other routine records, collect data and samples from consenting participants and apply cutting edge imaging and omics technologies places the UK in an optimal position to fully capitalise on these major research assets.

MRC funded cohorts have identified many important modifiable risk factors that predispose to disease and disability such as the link between smoking and lung cancer, the influence of early life circumstances on health in later life, and the contribution of socio-economic position to overall health and health inequalities. In addition to providing insights into the determinants of health, wellbeing and disease, outputs from these studies have made significant contributions to public health policy and led to changes in clinical practice. Examples of impact from a selection of MRC cohort studies are described throughout this review.

Large-scale population cohorts are the foundation for understanding the role and dynamic interplay of genetic, lifestyle and environmental influences on human health. However, sustained support for large longitudinal studies is expensive. It is therefore essential that these resources are used in ways to realise their scientific potential and increase societal benefit. Maximising the value of longitudinal studies for new scientific discoveries and informing health policy and practice is a key MRC strategic priority.

Cohorts are funded via response mode at the MRC Boards, are integrated within MRC Units or are supported in strategic partnerships with other funders. In addition to MRC funding, large population cohorts are also supported by the Wellcome Trust, ESRC, CRUK and BHF, as well as other funders. A stocktake of the largest MRC funded population cohorts can only be truly informative if the cohorts are considered within the context of the wider UK population cohort landscape.

A comprehensive understanding of the scientific niche of each cohort within the UK portfolio is vital to inform funding decisions and ensure value for money. Collating information on the largest UK population cohorts into a single review offers an opportunity for the first time to examine the profile and future trajectory of these major investments. In addition, it provides an evidence base for strategic discussions on the strengths, gaps and emerging opportunities for UK cohorts.

The aims of the strategic review are to:

- Document the current investment in all major UK population cohorts, including the positioning of large MRC funded cohorts within the UK portfolio
- Model the projected trajectory of the current UK portfolio over the next 10 years
- Highlight scientific and translational opportunities to inform future investments in cohorts and associated studies over the next decade.

The 34 cohorts included in this review, encompassing 19 MRC funded studies and 15 cohorts funded entirely by others, comprise the majority of large longitudinal population based studies in the UK. Information on the participants, variables and samples collected, and use of data linkage and 'omics' from the 34 population cohort studies are presented. During the past 10 years more than 10 funders, either individually or in funding partnerships, have supported the core infrastructure of the different cohorts. Most of the cohort studies are currently in receipt of funding for continued data collection and maintenance of the core resource, while a small number rely on project funding to maintain the cohort. In addition to the existing 34 cohort studies, two new large population cohort studies, one of which is funded by the MRC in partnership with others, are due to commence in 2014. The combined annual spend by the funders supporting the 34 cohort studies is £27.6m, with the MRC's annual investment accounting for £9.6m of the total.

This review analyses the current cohort portfolio and the projected profile in 10 years' time. It explores how these studies can be used to maximum advantage for further mechanistic, epidemiological and public health research, and to inform policy and practice. The report concludes with a set of recommendations to assist future funding and strategic decisions by the MRC. The recommendations will also be relevant to other cohort funders, researchers and users of the outputs from these valuable studies.

2. Review of Current Portfolio

The MRC funds a diverse range of cohorts from large general population cohorts with more than a million participants to small patient cohorts. Some cohorts are funded entirely by the MRC and some in partnerships with other funders. The intention of this strategic review is to gather data on the largest population cohorts supported wholly or in part by the MRC, as well as cohorts funded by others, to provide a comprehensive overview of the UK population cohort landscape.

2.1 Methodology

2.1.1 Inclusion criteria

The criteria for inclusion in the portfolio were as follows:

- The cohort is a population study and not patient specific;
- The initial sample size at recruitment was $\geq 1,000^2$;
- The cohort is UK based, with the exception of the large MRC funded Rural Ugandan General Population Cohort³;
- Participant follow-up is via any route, including both participant contact and record linkage;
- Cohort studies can be currently collecting data or recently archived.

Nineteen cohorts that received partial or full funding by the MRC to maintain the core cohort resource were identified as fulfilling the above criteria (Table 1).

Table 1. Cohorts receiving core funding from the MRC (partially or fully)

MRC core funded cohorts
11-16 and 16+ Study
Aberdeen Children of the 1950s (ACONF)
Avon Longitudinal Study of Parents and Children (ALSPAC)
Cognitive Function and Ageing Studies I (CFAS I)
Cognitive Function and Ageing Studies II (CFAS II)
Determinants of Adolescent Social well-being and Health (DASH)
European Prospective Investigation of Cancer Norfolk (EPIC Norfolk)
Hertfordshire Cohort Study (HCS)
Lothian Birth Cohort of 1936 (LBC1936)
Million Women Study
MRC National Survey of Health and Development Cohort 1946 Birth Cohort (NSHD/1946BC)
National Child Development Study 1958 Birth Cohort (NCDS/1958BC)
Newcastle 85+
Rural Uganda General Population Cohort (Rural Uganda GPC)
Southampton Women's Survey (SWS)
Twin Early Development Study (TEDS)

² The MRC funded Newcastle 85+ cohort recruited just under 1,000 participants but has been included as it is the only large cohort of the oldest old in the UK.

³ The Rural Ugandan General Population Cohort is the only sizable population cohort funded by the MRC that is outside the UK and therefore has been included in this review.

MRC core funded cohorts

UK Biobank

West of Scotland Twenty-07 Study (Twenty-07)

Whitehall II

In order to examine MRC investments in the context of the wider UK portfolio, the funders of other large population cohort studies, including the Wellcome Trust, CRUK, ESRC and BHF, and in some cases the cohort studies themselves, were approached and kindly provided information on cohorts that fulfil the inclusion criteria.

In total, data on 34 cohorts were collected encompassing 19 MRC and 15 non-MRC funded cohorts (Annex 1 for a list of cohorts included). Collectively, these represent the largest population cohorts in the UK⁴.

Funding has been recently committed for two new cohort studies. As they have not started recruitment yet, they have been excluded from the current portfolio analysis, but have been included in the future cohort portfolio projection. The Life Study, a large national birth cohort study funded by BIS, ESRC and MRC, is due to start recruitment in 2014. The Northern Ireland Cohort for Longitudinal Study of Ageing (NICOLA), funded by Atlantic Philanthropies, the Centre for Ageing Research and Development in Ireland and the ESRC, which aims to recruit 8,500 participants over 50 years of age in Northern Ireland, is due to start recruitment in late 2013.

2.1.2 Data collected

A pilot study collecting information from two MRC cohorts, CFAS II and the NSHD/1946BC, was carried out to inform the data gathering exercise. The finalised questionnaire was then circulated to all 34 cohorts for completion (Annex 2).

Information collected on the 34 cohorts included the cohort profile such as the recruitment date, the number, age and sex of participants and data collection sweeps. Variables collected included anthropometric measurements, physical and cognitive function, lifestyle and socio-economic position. Data on biological samples collected and use of omics and record linkage were also obtained.

2.2 Results from the data gathering

The overall findings presented below are supplemented by more detailed data on the individual cohorts in Annexes 3-13. The results are based on all 34 cohorts unless otherwise specified.

2.2.1 Overview of the portfolio

The UK portfolio can be represented in the categories below⁵:

Table 2. Cohorts by category

Birth Cohorts:	NSHD/1946BC, NCDS/1958BC, 1970BCS, ALSPAC, SWS, MCS
Area-based:	11-16 and 16+ study, the Twenty-07 Study, ACONF, GUS, Rural Uganda GPC
Occupational:	Whitehall II
Household survey:	Understanding society
Twins:	TEDS, Twins UK and Gemini
Ethnic minorities:	DASH, SABRE and BiB
Older cohorts:	CFAS I and II, ELSA, UK Biobank, Boyd Orr, LBC1936, HCS, Newcastle 85+
Diet/specific health outcomes:	EPIC Oxford, EPIC Norfolk, Million Women Study, Breakthrough Generations Study, BHRS, BWHHS, UKWCS

⁴ It is possible there are other large population cohorts in existence in the UK that fit the inclusion criteria that were not identified in this exercise.

⁵ Many of the cohorts could be included in more than one category.

Cohort participants are based in England, Wales and Scotland, with participants in the Millennium Cohort Study and Twins UK recruited from all UK countries including Northern Ireland. The only cohort based outside the UK is the MRC funded Rural Uganda GPC, a cohort of 18,000 participants in South West Uganda monitoring the prevalence and incidence of HIV. This cohort has been included in the review as it is one of the MRC’s large longitudinal population study investments.

The portfolio of cohorts spans the whole life course from birth to over 100 years old, with the SWS uniquely collecting data on mothers before they conceived.

Of the 34 cohorts, four studies include women only and one study, the BRHS, is exclusively male. The BRHS and BWHHS were established to study the incidence of cardiovascular disease in men and women. The all-female studies, Million Women Study, Breakthrough Generations Study and UKWCS examine a range of health outcomes in pre- and post-menopausal women.

All cohorts, with the exception of the 11-16 and 16+ Study, are actively gathering data. Most of the cohorts (n=29) recruited participants for a fixed period and are now closed to recruitment. Only five studies are currently open to on-going recruitment.

A number of the cohorts are inter-generational. Rural Uganda GPC covers the entire population of defined rural communities and new residents are continuously added over time. Understanding Society collects information on families, and the birth cohorts ALSPAC, BiB, SWS and MCS collect information on parents and children.

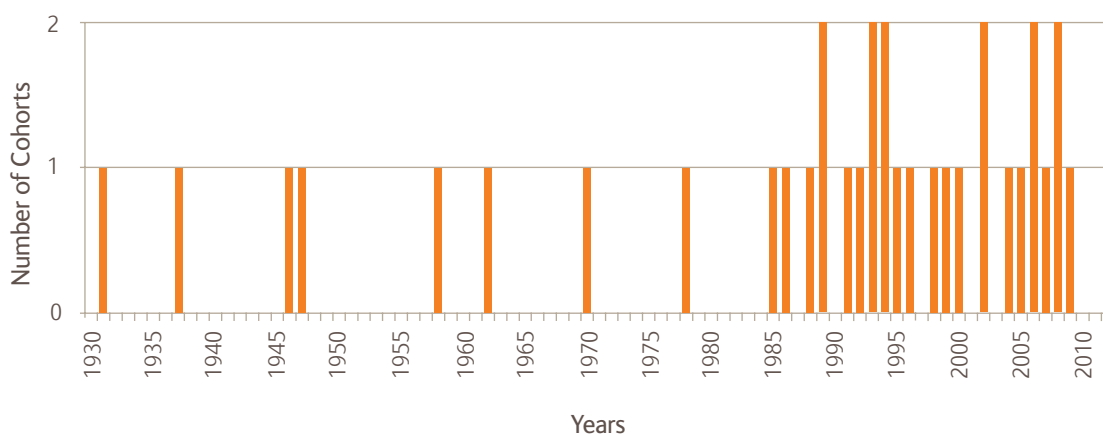
The Southampton Women’s Survey

was initiated in 1998 to explore how mothers’ dietary and lifestyle factors before and during pregnancy influence the health of their offspring. A key finding is that maternal vitamin D concentrations during pregnancy are positively associated with children’s bone health and body composition – a result which is particularly significant given that vitamin D insufficiency is common among pregnant women. Insufficient maternal vitamin D was found to be associated with abnormal fetal bone development and lower bone mineral density in nine-year-old children, as well as with a greater body weight in children at age six. These findings, together with other research, have informed a recommendation by the Department of Health for vitamin D supplementation during pregnancy, and a randomised controlled trial of maternal vitamin D supplementation is currently underway.

2.2.2 Start date of the cohorts

Figure 1 shows the distribution of cohorts by the start date of data collection. The MRC NSHD/1946BC, which has regularly collected data on participants throughout the life course, has been funded by the MRC since 1962. The cohorts HCS, Boyd Orr and LBC1936 are based on data collected in initial studies in 1931, 1937 and 1947, respectively. Decades after the original data were collected, participants were recruited into cohort studies with funding for prospective data collection. The most recent cohort in the portfolio is Understanding Society, which was funded in 2009. On average, one new large population cohort has been funded every year since 1990. Two new studies, NICOLA and the Life Study, are due to start recruiting in 2014.

Figure 1. Cohorts by start date of initial data collection



2.2.3 Cohort follow-up

A key feature of the UK portfolio is the large number of cohorts that have been followed for a long period of time, either via participant contact or through data linkage. The length of follow-up of the 34 cohorts in the portfolio is shown in Figure 2.

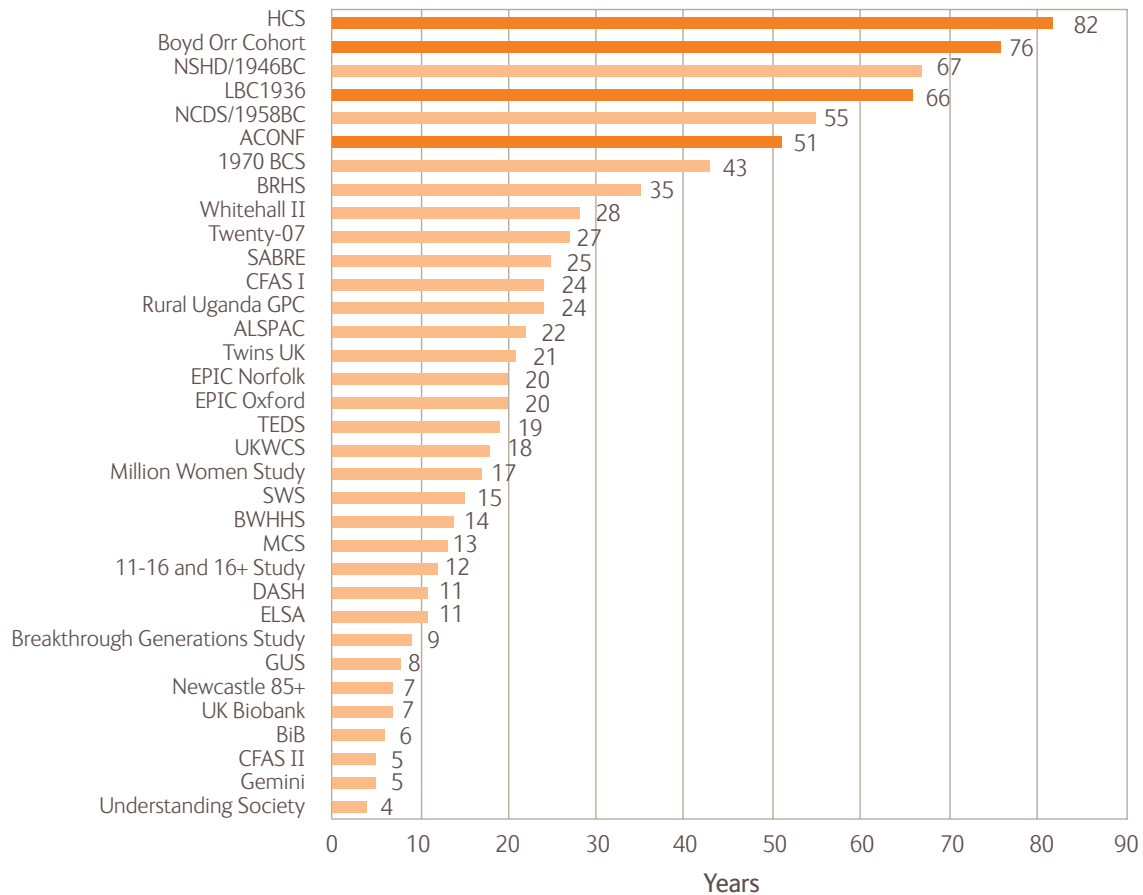
Half of the cohorts (n=17) have been followed for at least 20 years. The three longest continuously running studies are the birth cohorts, NSHD/1946BC, NCDS/1958BC and 1970BC. The NSHD/1946BC has been followed for more than 66 years and is the oldest cohort continuously followed from birth in the world.

The cohorts ACONF, Boyd Orr, LBC1936 and HCS (shown in orange in Figure 2) were established years after earlier studies which collected data on participants. ACONF builds on the Aberdeen Child Development Survey which collected data on individuals in 1962. Boyd Orr is based on the long-term follow-up of a dietary study of children in pre-war Britain. HCS was established from data collected on babies born in Hertfordshire between 1931 and 1939, and LBC1936 follows up children who took part in the Scottish Mental Survey 1947 at the age of 11 years. For the purpose of this document, these four cohorts are referred to as historical cohorts.

All cohort studies have obtained consent from participants for re-contact (Annex 5). The ability to recall participants within a cohort is important not only for further cohort sweeps, but also to enable sub-populations to take part in more detailed phenotyping and genotyping studies. The 11-16 and 16+ Study has been archived and is no longer collecting new data. The length of follow-up shown is therefore based on the last sweep which was conducted in 2002-04.

Figure 2. Length of cohort follow-up

Data based on the time from initial data collection to 2013. Historical cohorts are highlighted in dark orange.



2.2.4 Age of the cohort participants

Investigators provided information on the age of participants at recruitment and their current age. The four cohorts that collected data on parents as well as children have been divided into two separate age groups (Figures 3 and 4).

Figure 3 illustrates that the age of cohort studies within the portfolio at recruitment ranged from birth to old age. In the case of SWS, data was obtained on mothers prior to conception offering a unique longitudinal profile of this birth cohort across their life course. The recruitment profile includes a cluster of birth cohorts, two teenage cohorts, 15 cohorts from late teens to middle age and six which included participants solely over 60 years of age.

The estimated current age range of cohort participants is provided in Figure 4. For the five cohorts that are continuing to recruit, the age range reported is similar to that in Figure 3. In cohorts closed to recruitment the increased age of active participants corresponds to the length of follow-up. Details of the age and sex distribution of the cohorts can be seen in Section 3, Figures 15 and 16.

The MRC National Survey of Health and Development (NSHD) has continuously followed participants from their birth in 1946 to the present day, making it the oldest birth cohort study in the world. A key finding from the Study has been that early life – specifically, child development and home background – plays a major role in many aspects of adult health such as blood pressure, obesity, respiratory health, mental health, reproductive ageing, physical and cognitive capability, and survival. The NSHD has informed UK health care, education and social policy for more than 60 years, and its findings have been published in eight books and more than 600 papers. As the participants enter old age, the next stage of the study will provide important insights into the ageing process.

Figure 3. Age range of cohort participants at recruitment

Data are illustrated in years. Some birth cohorts comprise two age ranges that relate to the child participants and their parents. Dark orange bars relate to the age of participants in historical cohorts during the initial study, on which the cohort is based.

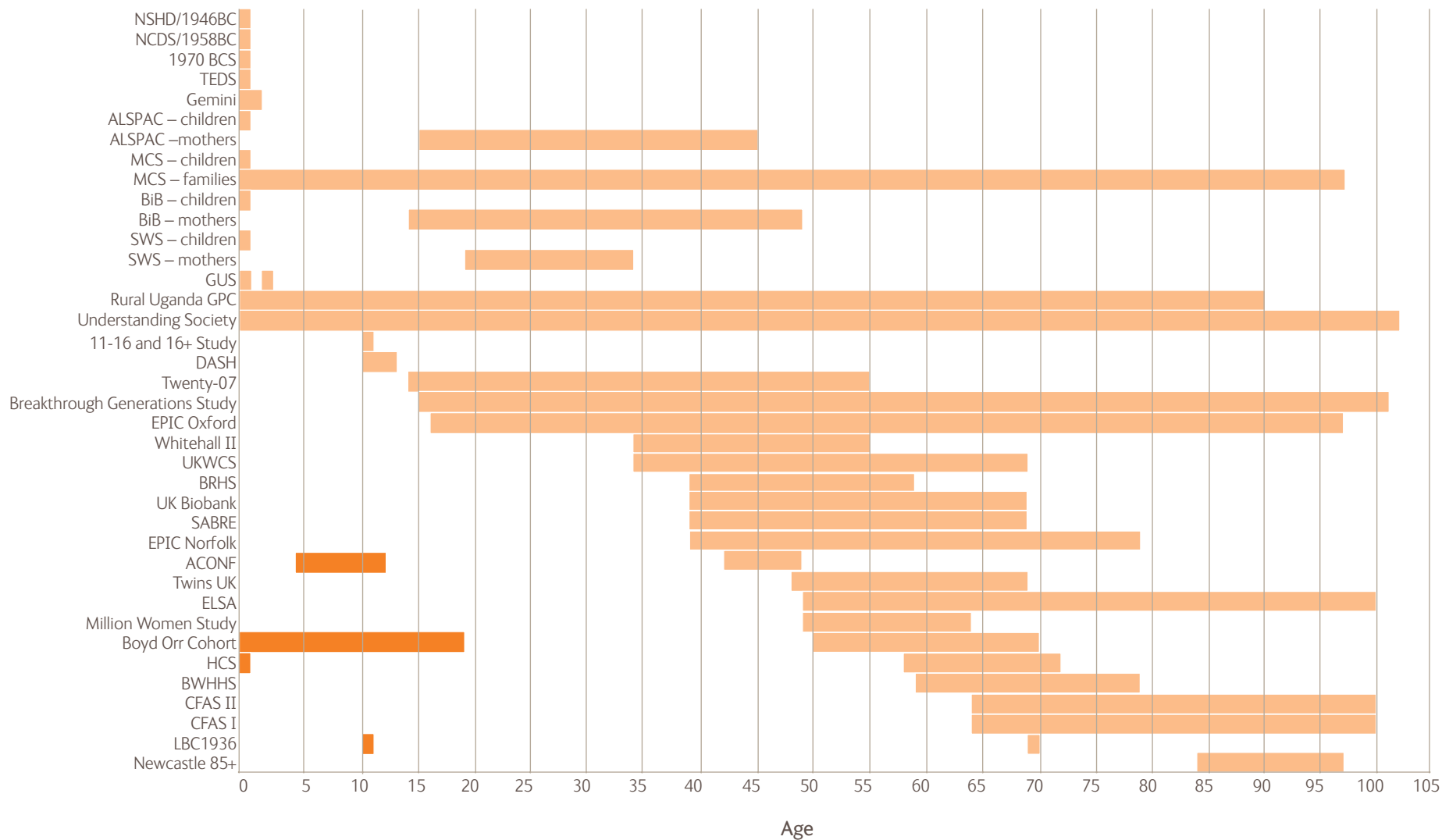
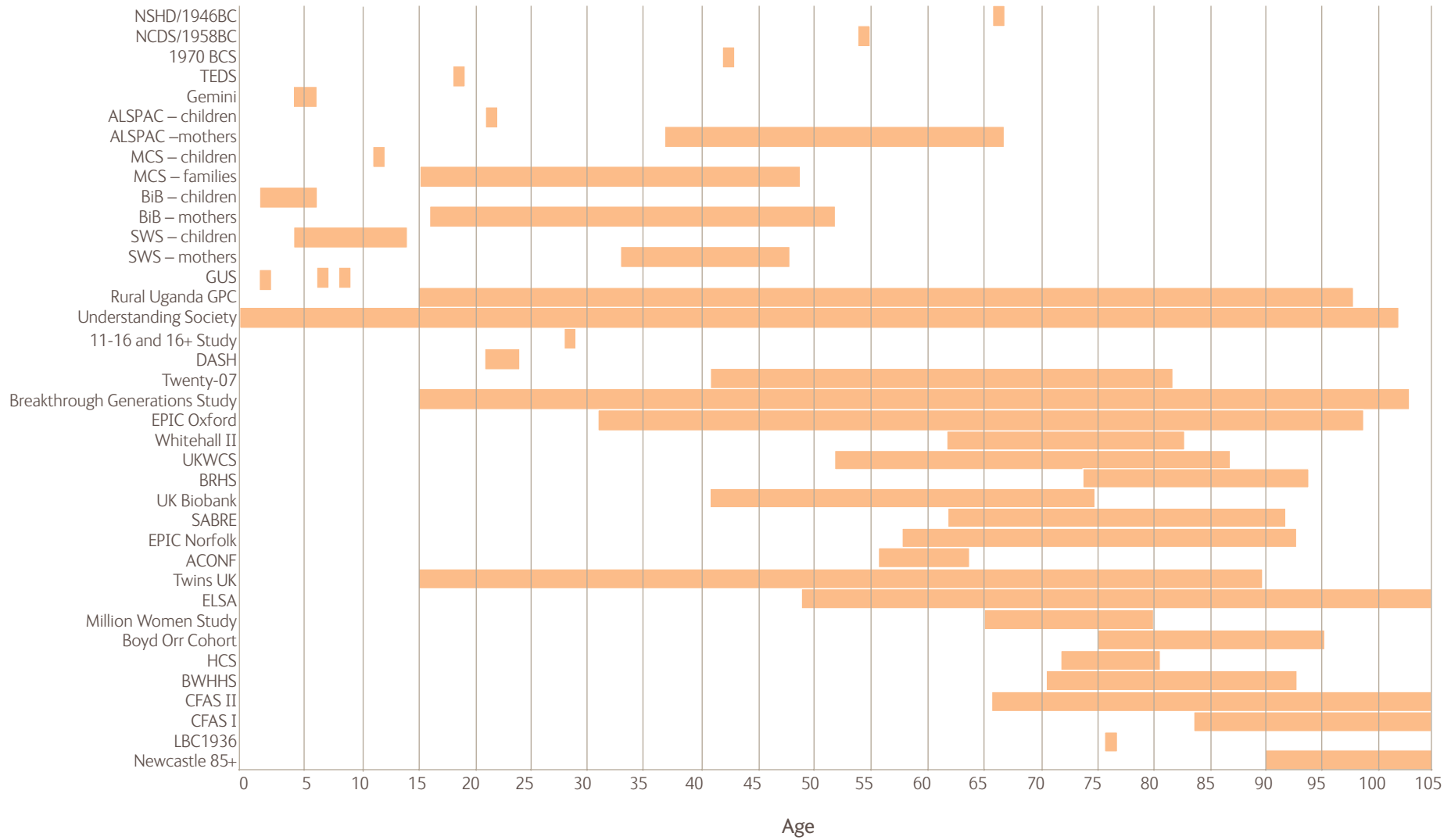


Figure 4. Estimated age range of cohort participants in 2013

Data are illustrated in years. Some birth cohorts may comprise two age ranges that relate to the child participants and their parents.



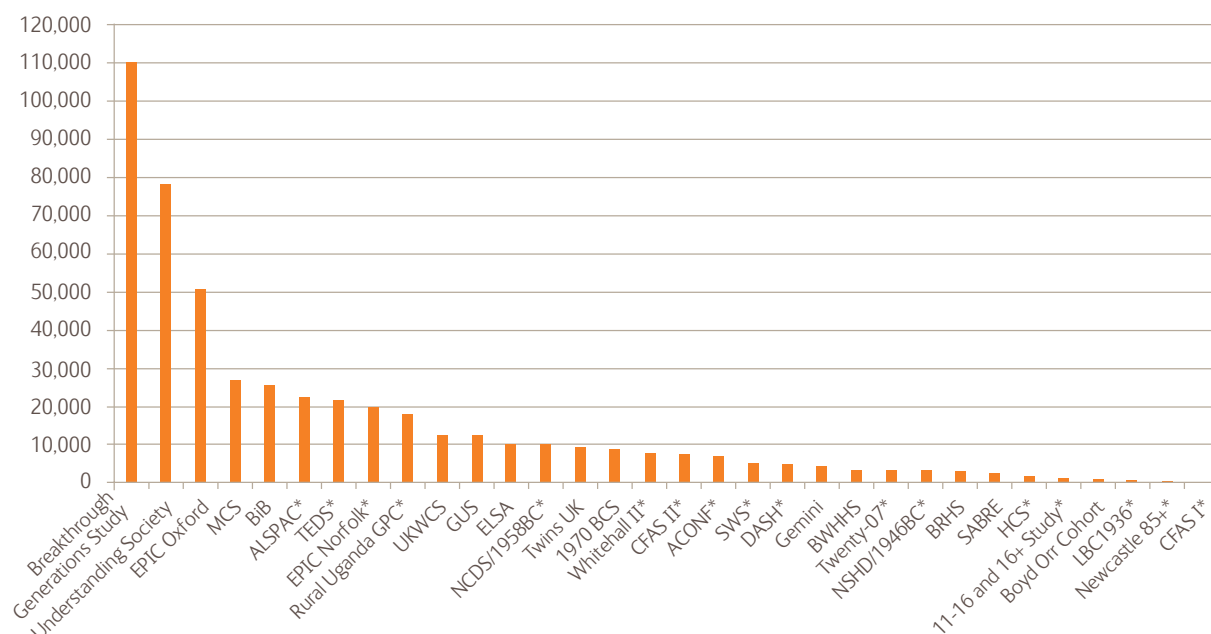
2.2.5 Current cohort sample size

The data in Figures 5 and 6 depict the current sample size of each of the cohorts in the portfolio based on estimates by the investigators. The current size of the cohorts ranges from 1.24 million in the Million Women Study to approximately 150 in CFAS I (of the original 18,500 who were over 65 years of age at the time of recruitment). The Million Women Study and UK Biobank together account for three quarters of the total participants in the portfolio. A further eight cohorts currently have over 20,000 participants. Eighteen cohorts have current participant numbers between 19,000 and 2,500 and only six cohorts have less than 2,000 active participants. Unsurprisingly, five of the six smaller cohorts comprise participants in old age with mortality, illness and disability being the most frequent cause of attrition. Two of these studies are historical cohorts. Based on this analysis, it is estimated that 2.5 million people in the UK have been recruited to large population cohort studies and today there are over 2.2 million people, 3.5% of the UK population, who are still taking part.

Figure 5. Estimated current number of cohort participants

Excluding the Million Women Study and UK Biobank

*Cohorts are MRC funded



The Avon Longitudinal Study of Parents and Children (ALSPAC) is a birth cohort study investigating genetic and environmental influences on health and development across the life course. Prior to ALSPAC, the initiation of the UK 'Back to Sleep' campaign in the early 1990s to reduce the risk of Sudden Infant Death Syndrome was met with scepticism from some scientists who were concerned about the risk of slowed motor development. ALSPAC results provided evidence to support this campaign by showing that putting infants to sleep on their backs was not associated with an increased health risk. This outcome also supported the development of a similar campaign in the USA.

Figure 6. Estimated current number of participants in the Million Women Study and UK Biobank
*Cohorts are MRC funded



UK Biobank was set up in 2006 to determine how genes, lifestyle and environment interact to cause a wide range of diseases. By collecting information and samples on half a million adults aged 40-69, UK Biobank has become a major national resource which is hoped will improve our understanding of disease prevention, diagnosis and treatment. Participants will be followed over the next thirty years, mostly by linkage to NHS health records and death and cancer registries. In 2012 UK Biobank was made available for use by UK and international researchers to assist in discovering the causes and treatments of disease.

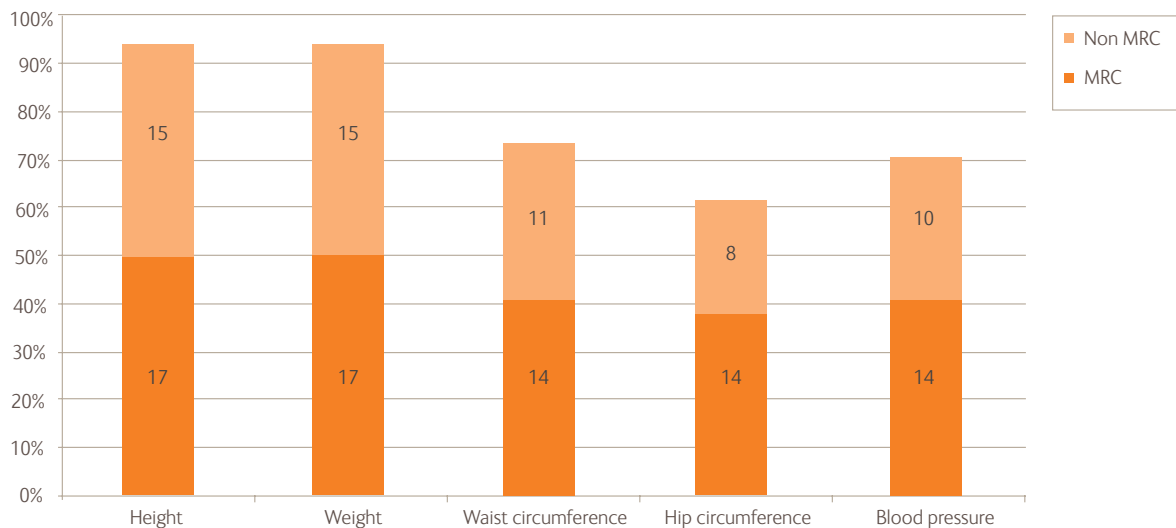
2.2.6 Variables collected

The purpose of the data gathering exercise was to illustrate at a high level, the types of variables collected by the cohort studies in the portfolio. The tables in Annexes 6-10 provide a breakdown of anthropometric, physical health, mental health and cognition, lifestyle and socio-economic measures collected for each cohort. Figures 7-11 below illustrate data pooled from the cohorts for each of the variables. The data collected on the variables by individual studies varied considerably as a result of different methods of collection and a range of instruments used. Individual studies also collected many other variables, particularly physical health measures that are not documented in this review.

- **Anthropometric and blood pressure measurements**

All cohort studies, with the exception of CFAS I and II, collected anthropometric measures (Figure 7 and Annex 6). Height and weight were collected by all these 32 studies. Blood pressure was measured by the majority of the MRC cohort studies. Other measures that were collected by some studies, but not shown in the figure, included body fat and arm circumference.

Figure 7. Proportion of cohorts collecting anthropometric and blood pressure variables



The Determinants of Adolescent Social well-being and Health (DASH) study was set up to explore the long-term influence of social conditions on the health and well-being of ethnic minority adolescents across London. DASH results revealed better mental health among adolescents from ethnic minorities compared to White British adolescents. African boys had significantly higher blood pressure at age 16 compared to White boys, and Black Caribbean and Nigerian/Ghanaian boys and girls had a higher BMI compared to their White peers. DASH findings have been used by schools and local communities in workshops, revised lesson plans and newsletters to parents.

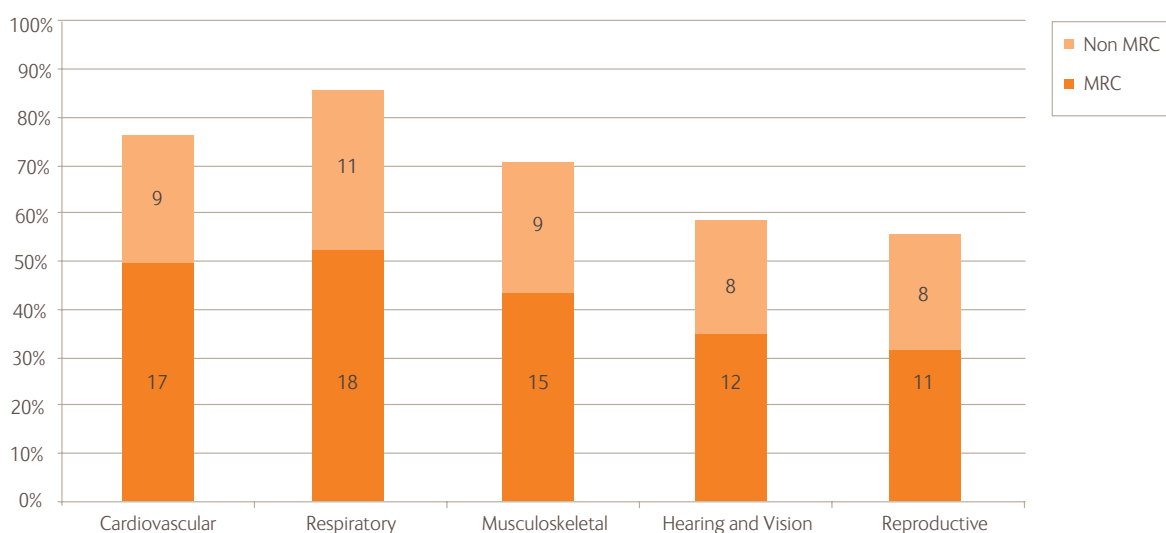
- **Physical health variables**

Physical health measures were collected by all but three cohort studies (Figure 8 and Annex 7). As expected, the majority of the MRC funded studies collected physical measures. The only MRC funded cohort with no physical data was TEDS, which was set up as a twins birth cohort exploring the influences of genes and environment on cognition and behaviour.

Respiratory, cardiovascular and musculoskeletal measures were all collected by 23 cohort studies, and nine studies collected all five measures illustrated in the figure.

Other variables reported but not shown in the figure included information on clinical conditions, imaging and dental health. Only four cohort studies reported collecting information related to infection. MCS measured exposure to common childhood infections through saliva collection, Boyd Orr reported infections in children from the historical data collection, NSHD/1946BC collected information from mothers about children's infections and the Rural Uganda GPC cohort study recorded prevalence and incidence of HIV infection.

Figure 8. Proportion of cohorts collecting physical health variables



It is estimated that 33 million people worldwide are infected with HIV, of whom two-thirds live in sub-Saharan Africa. **The Rural Uganda General Population Cohort** was set up in 1989 to examine trends in HIV prevalence and incidence, and their determinants, in rural Uganda. The cohort has revealed trends of the epidemic in the area before and after the advent of anti-retroviral therapy, and identified changes in sexual behaviour and other risk factors which have shaped the epidemic. These findings have been used for planning national HIV/AIDS programmes in Uganda, and contributed to the international global AIDS epidemic reports by UNAIDS.

- **Mental health and cognitive measurements**

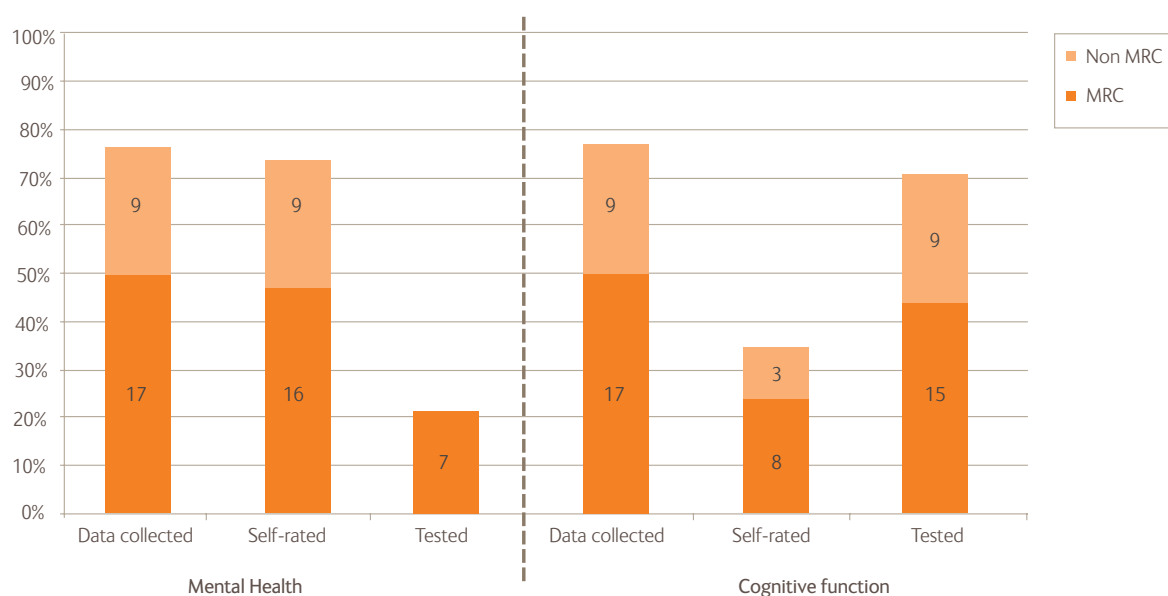
As shown in Figure 9 and Annex 8, two thirds of the cohort studies measured either mental health or cognitive function, with 23 studies (16 MRC funded) measuring both. The only MRC cohort with neither mental health nor cognitive measures was the Rural Uganda GPC.

The majority of mental health assessment was through self-report. Seven cohort studies used clinically tested measures, through professional diagnosis or via health records, and these were all MRC funded cohorts.

Conversely, assessment of cognitive function in most cohorts was by testing. Only two MRC funded cohort studies did not use cognitive tests, but used self-rated measures of cognitive function instead.

The MRC funded cohorts CFAS I and II, LBC1936 and TEDS were established with a specific scientific focus on mental health and/or cognitive function at various stages of the life course. All four of these cohort studies have tested cognitive function, and have tested or were in the process of clinically assessing, the mental health of the participants.

Figure 9. Proportion of cohorts collecting mental health and cognitive function variables



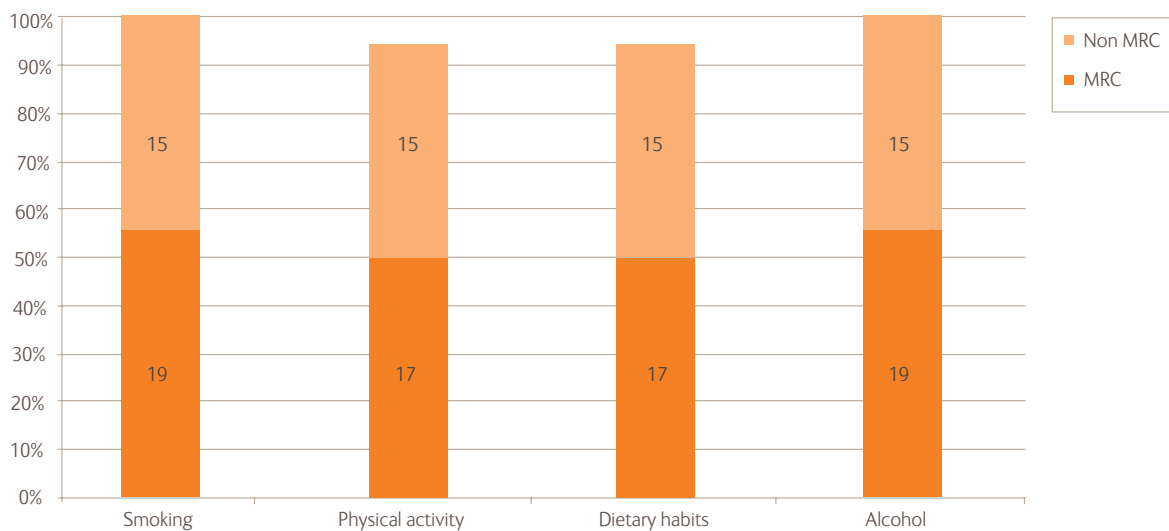
The MRC Cognitive Function and Ageing Study (CFAS I and II) are two population-based studies, investigating health and cognitive ageing in adults aged 65 years and older across the UK. Accurate assessment of dementia prevalence enables planning of appropriate health service. However prior to these studies, the number of people in the UK with dementia was unclear. Results from CFAS I and CFAS II have revealed a reduction in dementia prevalence over the past twenty years. In 2011 there were 214,000 fewer cases of dementia than predicted, which translates to a 1.8 per cent lower overall prevalence than expected.

- **Lifestyle variables**

Lifestyle information was collected by all cohort studies, mostly through self-reporting. The four most common variables were smoking, physical activity, diet and alcohol, with data on alcohol consumption and smoking universally collected (Figure 10 and Annex 9). All but two MRC funded cohorts collected data on either physical activity or diet.

A wide range of other lifestyle measures not shown in the figure was gathered on the cohorts such as sleep activity, sexual behaviour and the use of illegal drugs.

Figure 10. Proportion of cohorts collecting lifestyle variables



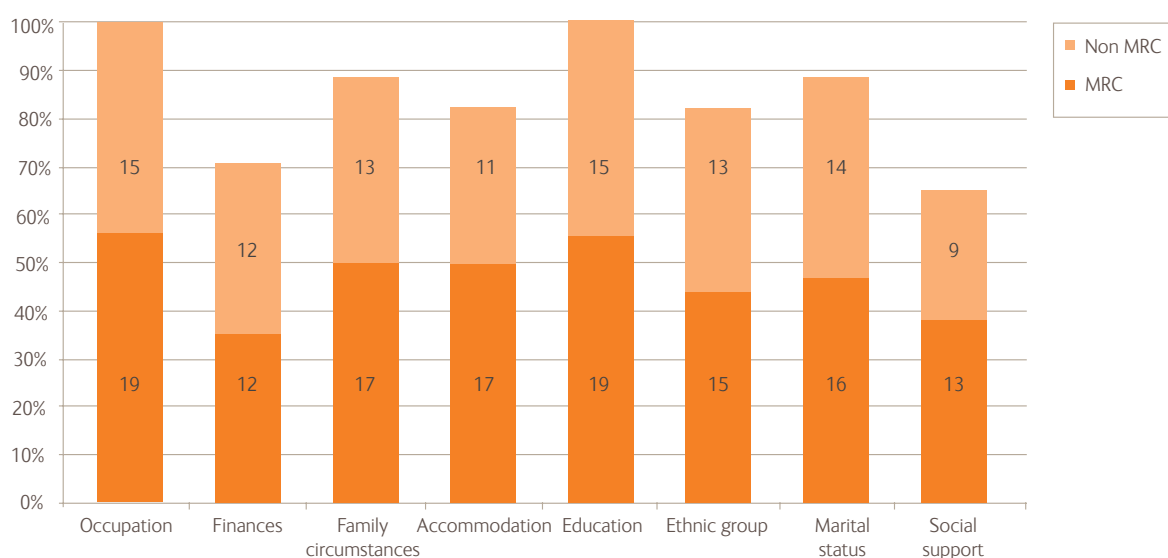
The 11 to 16 and 16+ study is a school-based survey initiated to explore patterns in health and health behaviours among adolescents from the Glasgow area. The relationship between alcohol use and antisocial behaviour is complex. Understanding this relationship among adolescents is particularly significant given that levels of alcohol consumption and conduct disorder are rising in this group. The 11 to 16 and 16+ study showed that the use and misuse of alcohol impacts slightly on antisocial behaviour in the shorter term. However it is antisocial behaviour that is the main predictor of alcohol use, misuse and alcohol-related trouble in young adolescents.

• Socio-economic variables

Information on socio-economic position was collected by the 34 cohort studies, with data on occupation and education common to all studies (Figures 11 and Annex 10). In addition, 22 studies including 12 MRC funded cohort studies gathered information on all four of the other variables: family circumstances, accommodation, ethnic group and marital status. Fourteen studies collected data on all the variables, of which five studies are MRC funded.

The majority of socio-economic data were obtained through self-reports. Five cohorts (three MRC funded) were linked to education records and five cohorts (two MRC funded) were linked to Her Majesty's Revenue and Customs (HMRC) taxation records and Department of Work and Pensions (DWP) records.

Figure 11. Proportion of cohorts collecting socio-economic variables



The West of Scotland Twenty-07 study followed three cohorts from the Glasgow area, aged 20 years apart, from 1987 to 2007 to investigate how people's social environment affects different aspects of their health. A key finding was that adults from more disadvantaged circumstances have poorer health across a range of measures and that the longer a person spends in such circumstances, the more harmful it is for their health. Certain health behaviours, such as smoking and unhealthy eating, were confirmed in the study to be more common among those in disadvantaged circumstances.

2.2.7 Data linkage

Participant consent to link to routine data sources had been obtained by all but two cohort studies, Gemini and 11-16 and 16+ Study (which is now archived), as shown in Annex 5.

All studies with consent, with the exception of the Rural Uganda GPC, linked to a range of different data sources across health care as well as administrative data, the census, births and deaths, and environmental data such as air and noise pollution (Figure 12 and Annex 11). These data show that record linkage, in particular to health, birth and death records, is now a common method of data collection by cohort studies.

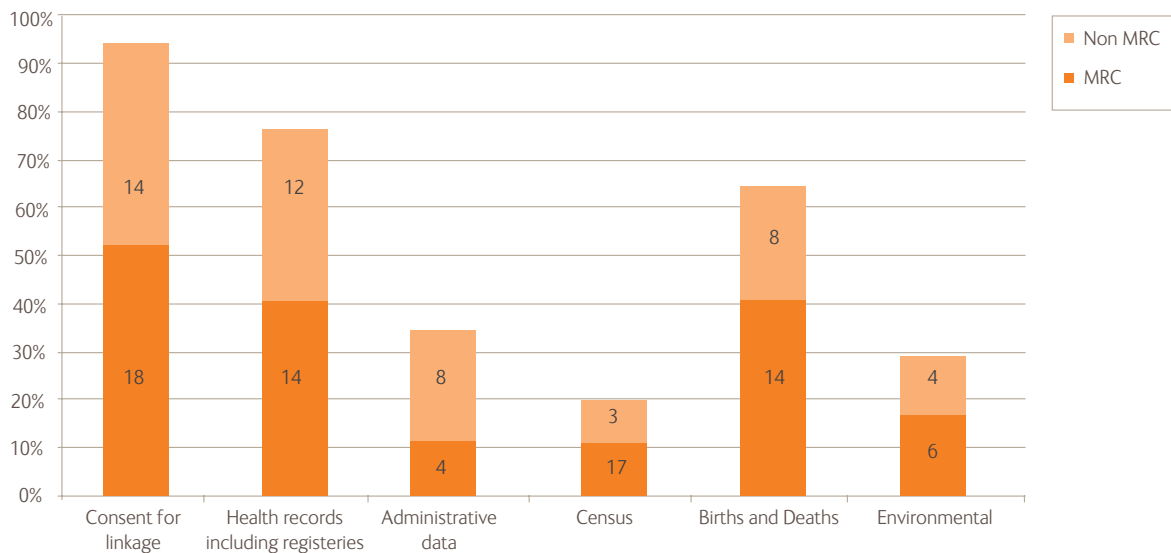
More than half of MRC funded cohorts linked to primary and/or secondary health records. Some of the more recent birth cohorts funded by the MRC (ALSPAC, SWS and TEDS) were also linked to education data.

Ten cohort studies, including six funded by the MRC, reported linkage with environmental records. Of the six MRC funded cohorts, three were linked to air pollution and three to geographical datasets.

The Millennium Cohort was the only study that linked to all five categories of routine data in Figure 12, including national education data in England, Scotland, Wales and Northern Ireland.

The two largest studies, UK Biobank and the Million Women Study, used data linkage as the primary mechanism for future participant follow-up.

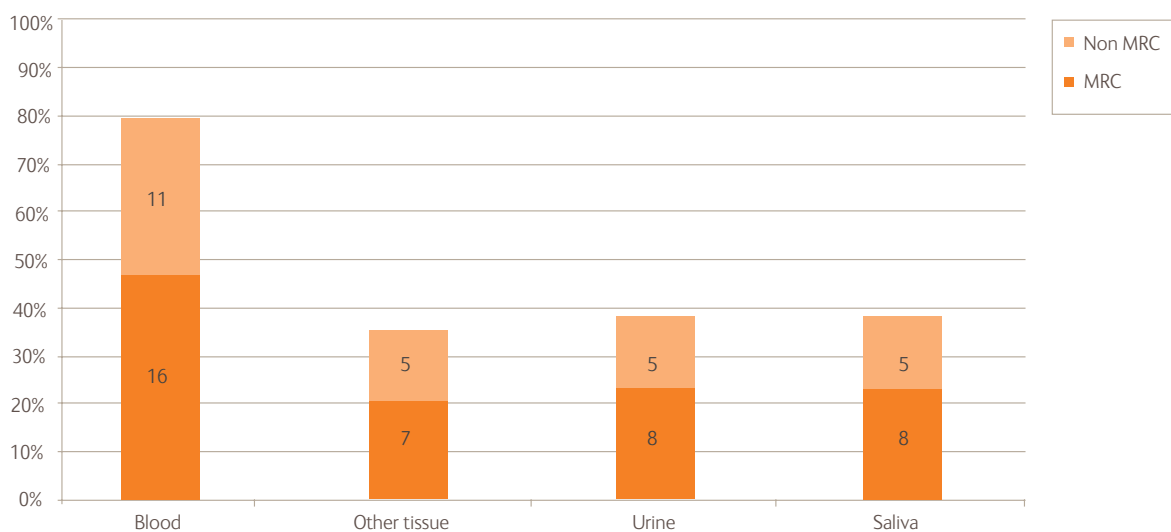
Figure 12. Proportion of cohorts with data linkage



2.2.8 Biological samples and omics

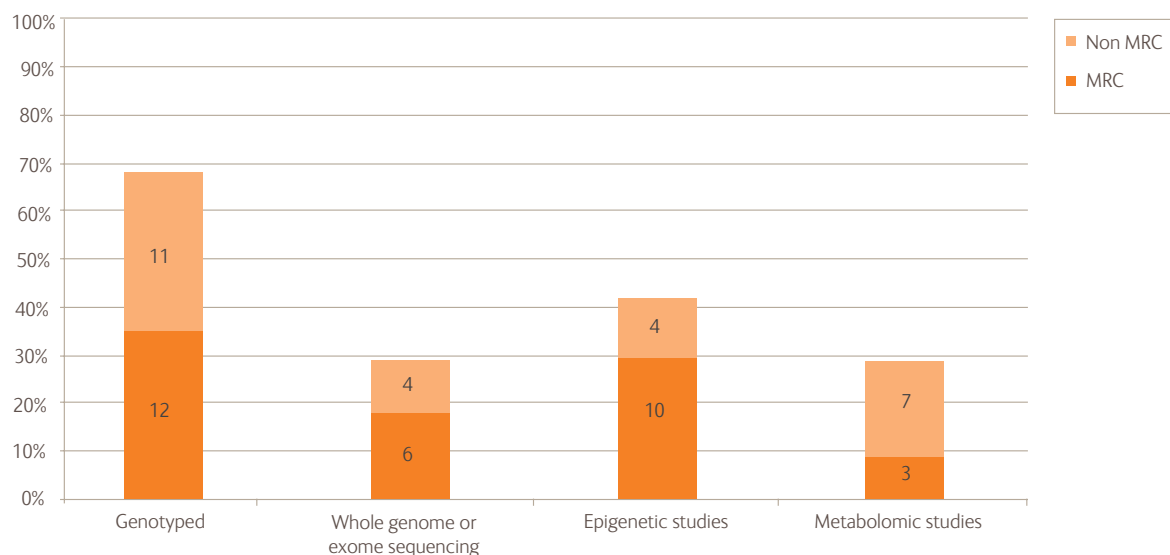
Biological samples were collected by all but three cohort studies. Blood was the most common tissue collected, particularly among the MRC funded cohorts, as illustrated in Figure 13. Other tissues collected included buccal cells, placenta, post-mortem brains, muscle, hair and teeth. All studies that collected urine also collected blood. Of those cohort studies that did not take blood samples, the two twin cohorts Gemini and TEDS collected buccal samples, CFAS II collected post-mortem brains and the Millennium Cohort collected saliva and baby milk teeth.

Figure 13. Proportion of cohorts collecting biological samples



A total of 26 cohort studies have conducted omics studies (Figure 14). Genotyping was the most common and was carried out on sub-populations of participants in 23 of the cohorts. The total number of participants in UK cohorts who have been genotyped will increase significantly once UK Biobank has completed its planned genotyping of all 500,000 participants.

Figure 14. Proportion of cohorts with omics analysis



The three cohort studies that did not genotype participants but reported other omics studies were CFAS II, which conducted epigenetics on post-mortem brains, and HCS and SABRE, which both carried out epigenetic and metabolomic studies.

Epigenetic studies can be costly and to date have been largely exploratory in large-scale population studies. Over half of the MRC funded cohort studies have used epigenetic studies, mostly on sub-groups of 80-1,000 cohort participants. Tissues used in these studies included blood, umbilical cord, buccal cells, brain and muscle.

Whole genome sequencing is expensive. Unsurprisingly, only nine cohort studies have used this technique, largely on sub-groups within the cohort.

High throughput metabolomics is an evolving technology. To date 10 studies have trialled this methodology. Based on the information provided, NMR was the most common platform used.

2.2.9 Collaboration across cohorts

Three quarters of cohort studies reported that they were members of a national or international cohort consortium. Many cohorts are part of more than one of these consortia. The consortia reported by the investigators are epidemiological, investigating phenotypic and genotypic variables across populations, or are focused on identifying genetic variants associated with particular disease traits.

Several cohort studies belong to ageing cohort consortia such as HALCyon⁶, IALSA⁷ and the longitudinal study of ageing consortium⁸, an international grouping of cohorts, including ELSA, which collects common variables based on the US Health and Retirement Survey. HALCyon includes five of the large population cohorts within this review: HCS, Boyd Orr, ELSA, NSHD/1946BC and NCDS/1958BC.

Examples of other UK cohort consortia include the UCL-London-School-Edinburgh-Bristol (UCLEB) consortium and the EAGLE⁹ Consortium. UCLEB is a population-based prospective collaboration investigating cardiovascular genomics, including the BRHS, BWHHS, NCDS/1958BC and NSDH/1946BC cohorts. EAGLE is studying the genetic basis of phenotypes in antenatal and early life and childhood and includes the NCDS, ALSPAC and TEDS cohorts.

The Healthy Ageing across the Life Course (HALCyon)

study is a collaboration of nine UK cohort studies to examine the social, psychological and biological factors that contribute to healthy ageing. By bringing together these studies, HALCyon can compare and replicate findings across different ageing cohorts. HALCyon has shown how physical capability levels – or the ability to undertake the physical tasks of everyday living – predict survival and subsequent morbidity, differ by gender, decline with age, are influenced by childhood socioeconomic circumstances, and vary by body size and neighbourhood characteristics. Other areas of focus for HALCyon include cognitive capability, psychological and social well-being, nutrition and diet, and biomarkers of ageing such as telomere length and HPA axis.

Amongst the many international cohort consortia reported by the cohort studies is the EPIC consortium¹⁰, which includes EPIC Norfolk and EPIC Oxford. The purpose of the EPIC consortium is to study diet and health in over half a million (520,000) people in ten European countries: Denmark, France, Germany, Greece, Italy, The Netherlands, Norway, Spain, Sweden and the United Kingdom.

Building on the success of HALCyon, the recently funded CLOSER consortium¹¹ is seeking to develop methodology and share good practice to enable integration and standardisation of variables for meta-analyses and comparative cohort studies. Eight cohort studies within this review are members of the CLOSER consortium: NSDH/1946BC, NCDS/1958BC, 1970 BCS, HCS, ALSPAC, MCS, SWS and Understanding Society.

EPIC Norfolk is part of the European Prospective Investigation of Cancer (EPIC), which examines the links between diet and cancer among adults in Europe. The study found that men and women who did not smoke, were physically active, had a moderate alcohol intake and consumed five or more servings of fruit and vegetables a day lived an average of 14 years longer than people without any of these behaviours. This demonstrates the marked effect that modest lifestyle changes can have on longevity. These findings contributed to the Department of Health “Small Change Big Difference” public health initiative, which aimed to show people how to improve their health by making small changes to their lifestyles.

6 HALCyon <http://www.halcyon.ac.uk/>

7 IALSA <https://www.ilifespan.org/?q=IALSA>

8 <http://www.ifs.org.uk/ELSA/links>

9 EAGLE <http://www.copsac.com/content/eagle-consortium>

10 EPIC <http://epic.iarc.fr/>

11 CLOSER <http://www.closerprogramme.co.uk/>

3. Cohort Portfolio Projection

In addition to reviewing the current portfolio of population cohort studies in the UK, the strategic review also models the projected trajectory of the cohort portfolio over the next 10 years.

3.1 Methodology

Twenty-eight of the total 34 cohorts in this review (Table 3), for which data were available at the time of the projection, were included in the modelling exercise. The following information was requested:

- The age and sex breakdown of the cohort in five year age bands at the time of recruitment and at the two most recent follow-up waves.
- The rate of attrition in five year age groups in recent years, or between the most recent waves, through loss of contact, death, or drop out for any reason. Sex-specific information on attrition was requested if the attrition differed between men and women.
- The mortality rates of the cohort by age and sex over the past five years or between the most recent waves. This was of particular importance for older cohorts as the mortality rate is low at younger ages.

Table 3. Cohorts included in the projection

11-16 & 16+ Study	HCS
ACONF	MCS
ALSPAC	Million Women Study
1970 BC	NCDS/1958 BC
BiB	Newcastle 85+
Boyd Orr	NSHD/1946 BC
Breakthrough Generations Study	SWS
BRHS	TEDS
CFAS II	Twins UK
DASH	UK Biobank
ELSA	UKWCS
EPIC Norfolk	Understanding Society
EPIC Oxford	Twenty-07
GUS	Whitehall II

Where necessary, the above requested data were supplemented with information from published papers and study websites.

Reported attrition was used to estimate numbers at the end of 2012, where the most recent wave had taken place in earlier years. Where no information on attrition was provided, estimates were made by extrapolation from previous waves or, if necessary, assuming a standard attrition, estimated subjectively, bearing in mind the type of cohort, the nature of the follow-up and the age of the participants.

It was noted that some studies have passive follow-up through death notifications as well as, or instead of, actual contact with participants. The distinction was not always made in the figures provided, but in general passive follow-up figures were used when available.

The numbers of participants by age (in five year bands) and sex in 2012 were then derived for each cohort, and summated across the cohorts to provide the totals for each age and sex band.

For the projection to 2022, the approach was as follows:

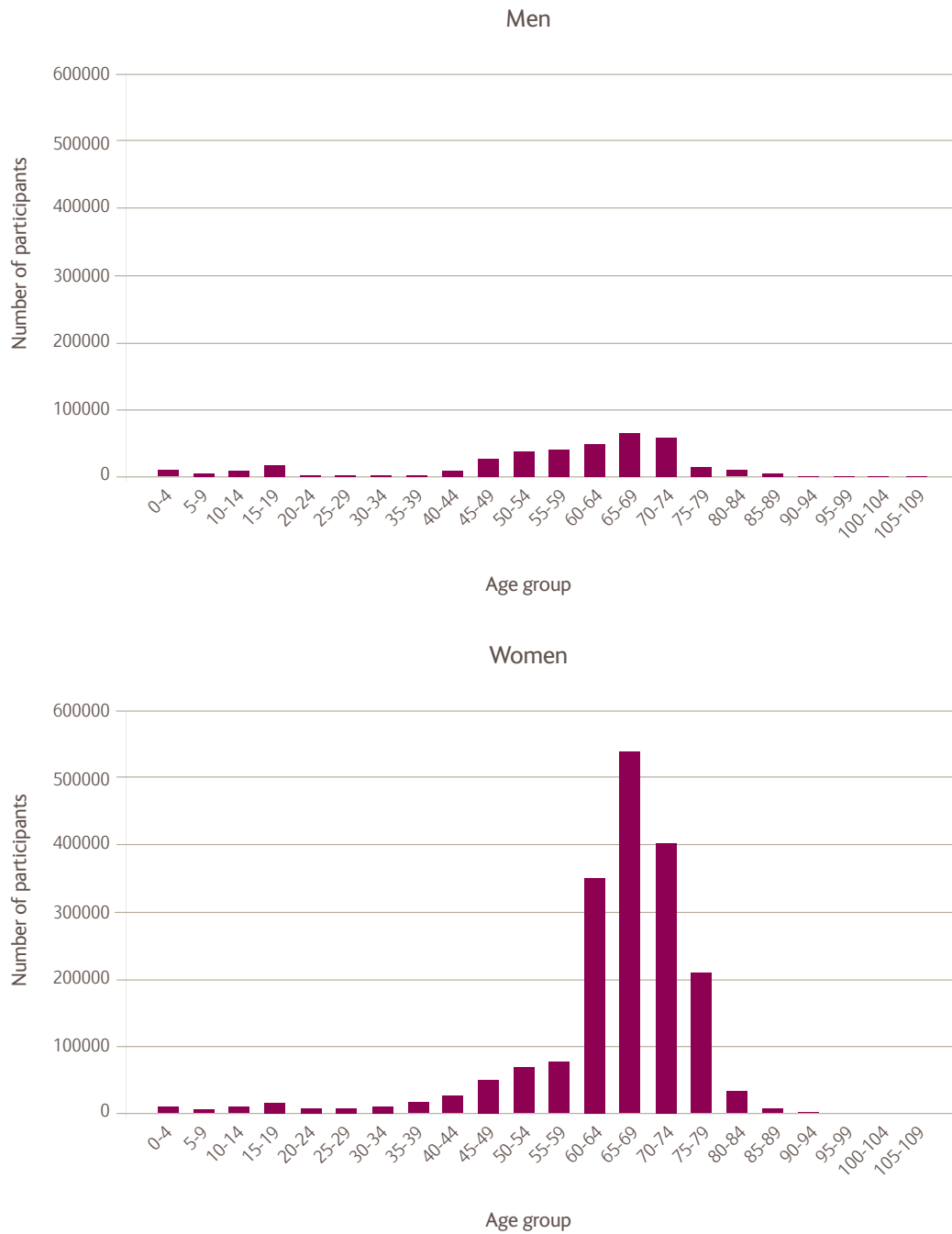
- Within each cohort, the study numbers were moved across by 10 years (two age bands).
- Broad assumptions were made about attrition to 2022, which occurs through death and dropout. For 'passive' follow-up, attrition is almost entirely from deaths, as dropout is minimal. Each cohort was considered separately, and estimates made subjectively, based on age- and sex-specific population mortality rates and knowledge of attrition in the cohorts.
- Estimated numbers for the recently funded Life Study and NICOLA were included, both of which are about to start, increasing the total number of cohorts to 30.
- Totals across cohorts were then derived by sex and age.

The first Whitehall Study in 1967 found that male British civil servants in the lowest employment grades were much more likely to die prematurely than those in the highest grades. Coronary heart disease (CHD) accounted for a large part of the mortality difference. Men in the lowest grades were found to have three times the mortality rate from CHD. **The Whitehall II Study** was set up in 1985 to determine the factors that underlie this social gradient in death and disease, and to include women in the cohort. Whitehall II findings led to the publication of the Marmot Review, 'Fair Society, Healthy Lives', in 2010 which outlined the most effective strategies for reducing health inequalities in England.

3.2 Results

The numbers of male and female participants across five yearly age groups, as of 2012, in the 28 cohorts considered are summarised in Figure 15. Broadly equal numbers of men and women are being studied up to age 20 years, but thereafter the number of women exceeds that of men at all ages. Far greater numbers of women than men are being studied at ages over 50 years, but there are also fewer men between the ages of 20 and 40 years (around 2,000-3,000 fewer in each 5 year age band). The two largest cohort studies dominate the participant numbers. The Million Women Study (1.36 million women recruited) largely accounts for the disparity in the numbers of men and women being studied. UK Biobank is also a particularly large study (500,000 recruited) and contributes to the large numbers observed at ages 45-74 years for both men and women.

Figure 15. Numbers of cohort study participants by age and sex in 2012



The Million Women Study investigates how reproductive and lifestyle factors influence health in a cohort of over one million women. A key finding of the study is that current users of hormone replacement therapy (HRT) have a higher risk of breast cancer, with risk increasing with long-term use. Combined oestrogen-progestagen HRT had a greater risk than oestrogen-only or tibolone HRT. This finding led the Committee on Safety of Medicines to release a safety update on HRT in 2003, in addition to advice on the prescribing and use of HRT. It is thought that decreased HRT use as a result of changes in prescribing have prevented approximately 10,000 cases of breast cancers in the UK.

The numbers by age and sex, excluding the Million Women Survey and UK Biobank, are shown in Figure 16. There is still an excess of women but the numbers are more evenly spread across the age groups. The dearth of young adult males across the cohorts is particularly obvious.

Figure 16. Numbers of cohort study participants by age and sex in 2012, excluding the Million Women Study and UK Biobank

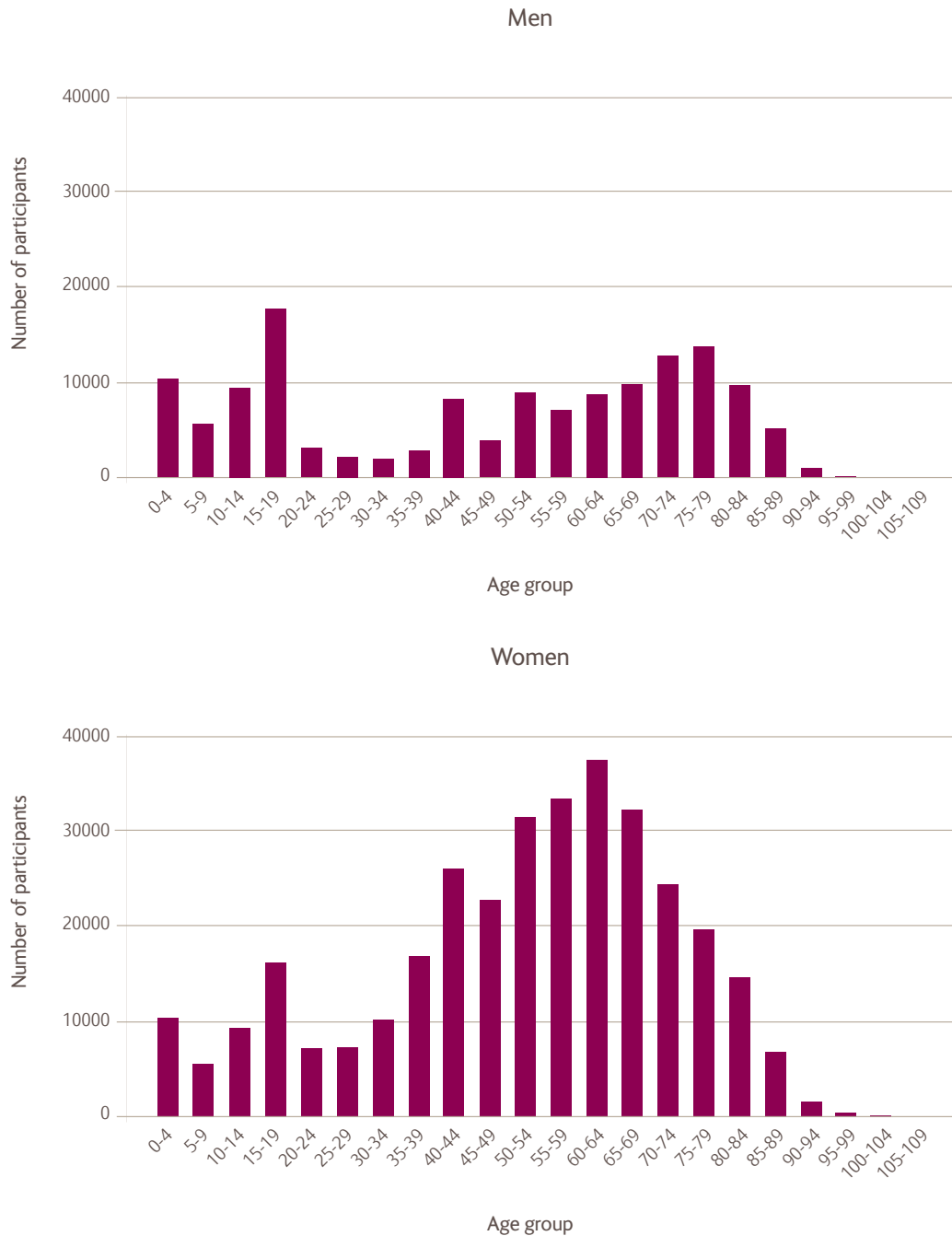
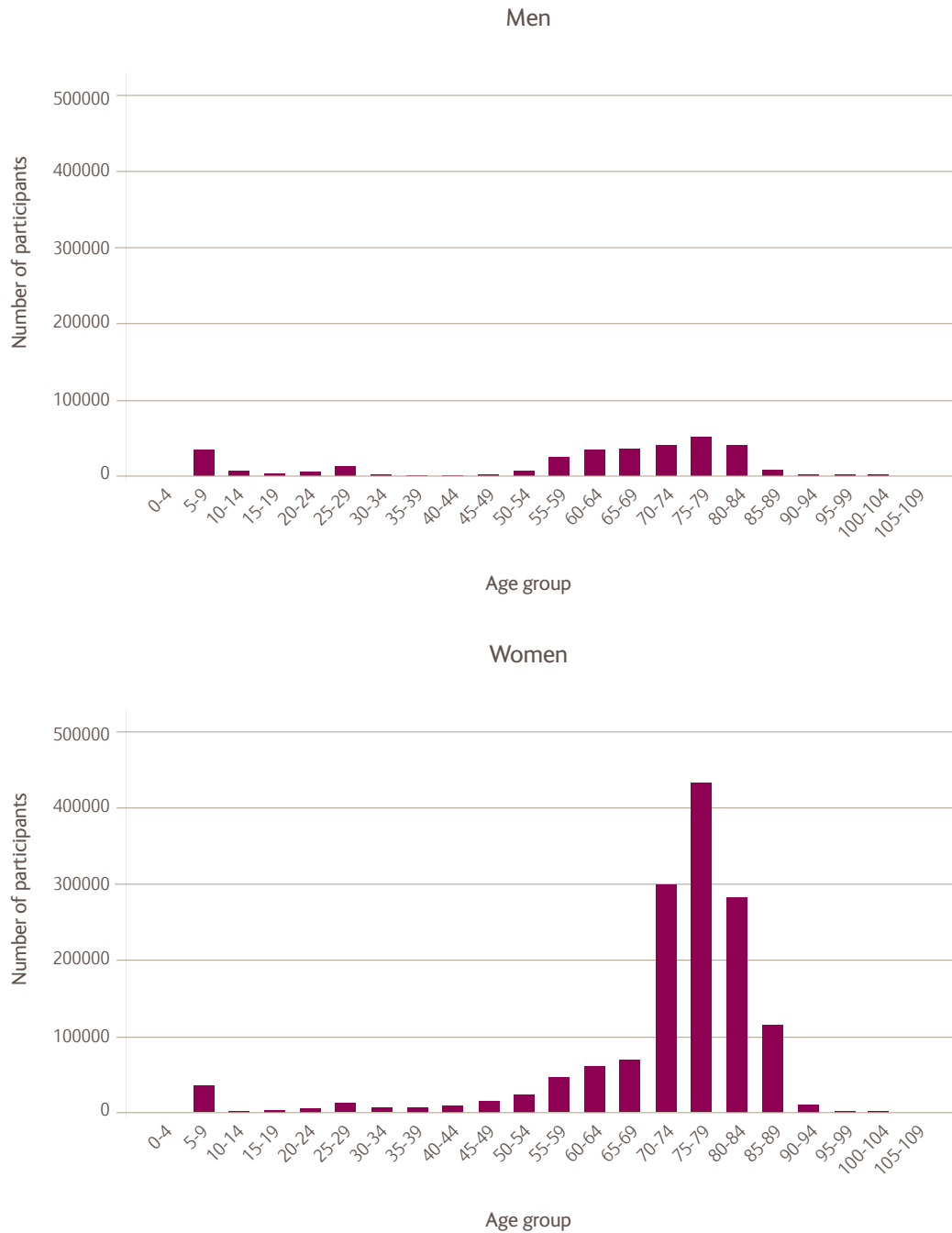


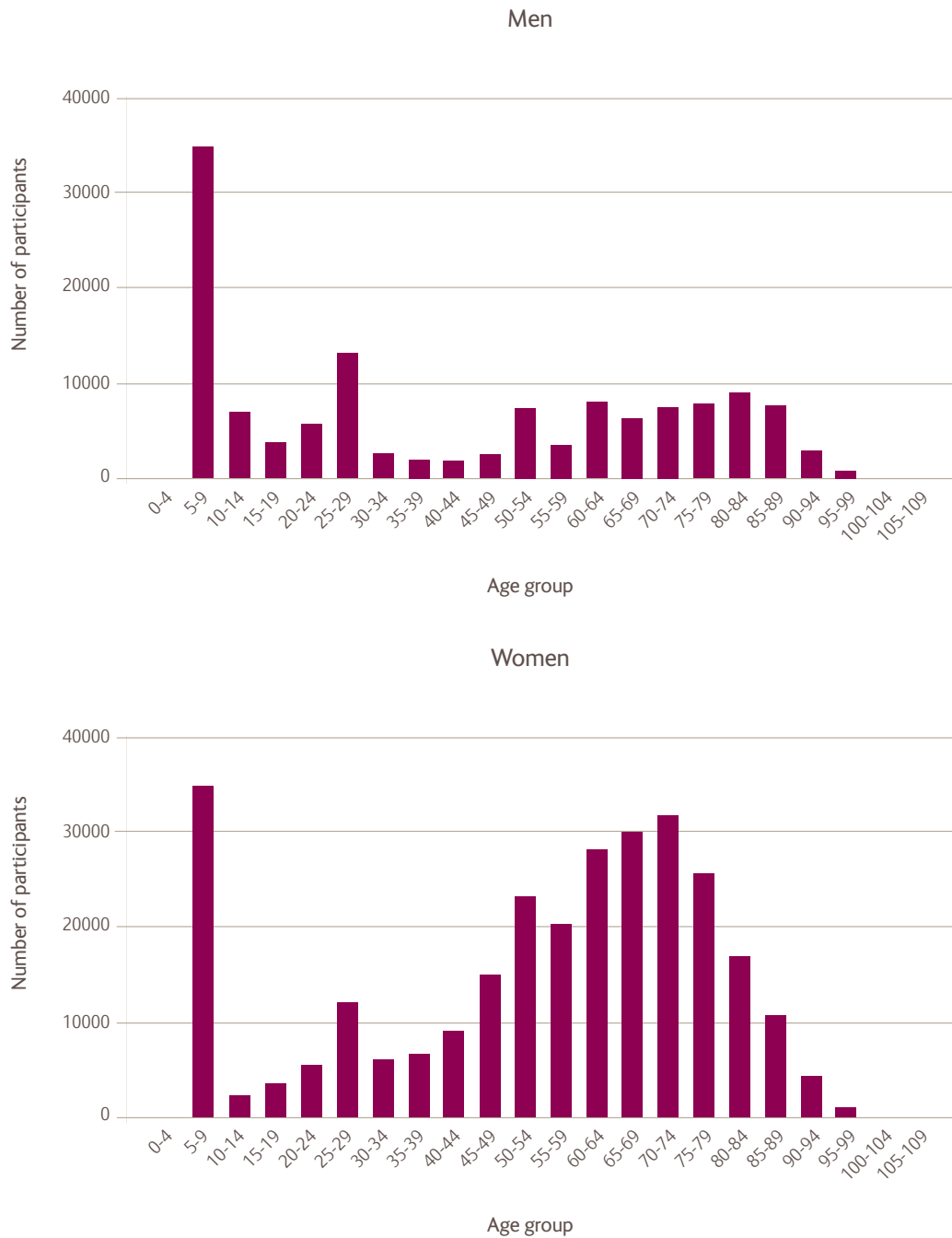
Figure 17 illustrates the extrapolation of data from all cohorts forward to 2022, allowing for attrition, and including the two new planned cohorts (the Life Study and NICOLA). The numbers of women being studied still greatly exceed that of men in later life, and the Million Women Study and UK Biobank are still the major contributors to those being studied at older ages. The numbers of men and women are broadly similar up to 30 years of age, but the new Life Study will contribute large numbers of children of both sexes. By 2022, there will be a dearth of men being studied at ages 30 to 49 years, reflecting the small numbers currently aged 20 to 39 years.

Figure 17. Projected numbers of cohort study participants by age and sex in 2022



Finally, the projected numbers for 2022, excluding the Million Women Study and UK Biobank are presented in Figure 18. The new Life Study dominates the numbers as recruitment goes as planned. In 2022, there will still be a much greater number of women than men under study, particularly from age 30 onwards, even without consideration of the Million Women Study.

Figure 18. Projected numbers of cohort study participants by age and sex in 2022, excluding the Million Women Study and UK Biobank



3.3 Summary

The projection of the trajectory of the cohorts over the next 10 years showed that in 2022 most of the cohort participants will be within the 55-85 year age range. Unsurprisingly, the number of women studied still greatly exceeds that of men in later life. The Million Women Study and UK Biobank dominate the numbers. Much of their follow-up is passive through record linkage, although sub-groups are being contacted for specific studies. The Life Study which commences in 2014 will bring in a large number of children of both sexes. It will also collect data on the children's fathers which will boost the number of men in the 30-50 year age range.

Six cohorts that are part of the overall review are not included in the cohort projection as the data were not available or not collected in advance of this projection exercise. Participants in four of these cohorts, BWHHS, CFAS I, LBC1936 and SABRE, although collectively not large in number, would all be over 70 in 2023. As an all-female cohort, the BWHHS would add more older women, further increasing the gender in-balance in the portfolio. Rural Uganda GPC is a cohort of 18,000 men and women aged 16-100 years. Recruitment to the cohort is on-going so it will be a source of younger men in the future, but not enough to significantly alter the overall portfolio shape in the next decade. The Gemini participants, who are girls and boys currently of 4-5 years of age, will increase the numbers of teenagers in the portfolio in 2022. The Longitudinal Study of Young People in England (LSYPE), funded by the Department of Education, recruited 16,000 13-14 year olds in England in 2004. The study has very recently received funding from the ESRC. Although this cohort has not been included in the review or in the projection exercise, cohort members will be in their early thirties in 10 years' time, adding to the number of younger adults in the overall 2022 cohort portfolio.

The Hertfordshire Cohort Study (HCS) recruited participants from Hertfordshire who were born between 1931 and 1939 to explore the effects of genes and early environmental exposures on health and ageing later in life. A key focus of HCS is on osteoporosis, which is most commonly associated with ageing and seen largely in post-menopausal women. Study results demonstrated that healthy weight gain in infancy improves bone strength later in life. Findings from HCS on various risk factors for osteoporosis contributed to the development of the WHO Fracture Risk Assessment Tool (FRAX®), which calculates the 10-year probability of fracture to the hip, and to the spine, forearm or shoulder.

4. Recommendations

4.1 Stakeholder engagement

Many of the UK population cohorts are internationally renowned and used by investigators globally. Other cohorts in this review are less well known and offer scientific potential that could be further exploited in new areas or collaborations. Past discussions on maximising the value of UK cohorts have led to initiatives to support cohort discovery and collaborations¹², and many UK cohort studies, including those funded by the MRC, have benefited from such developments.

The MRC sought advice from a breadth of stakeholders, including policy makers, industry and academics from a wide range of disciplines, on how to build on previous initiatives and further enhance the scientific and translational potential of the UK population cohort portfolio.

This process was greatly assisted by having access, for the first time, to information on the whole portfolio of the 34 largest cohorts. Delegates were invited to a workshop to review the collated information, discuss the strengths and gaps of the current portfolio, and identify opportunities.

Discussions were focused around four themes:

- Beyond epidemiology: How cohort biological samples could be used in mechanistic studies, in particular in high throughput science.
- Cross-cohort collaboration: How cohorts can work together to address major scientific challenges.
- Impact on health policy: How to effectively translate outcomes from the cohort studies to inform policy.
- Portfolio balance: Identifying portfolio gaps and scientific questions that cannot be currently addressed by the existing cohorts.

The workshop agenda and list of attendees are in Annex 14.

The portfolio analysis and outcomes from debate at the workshop were discussed by the MRC Population Health Sciences Group, the four MRC Research Boards and the Cohort Strategy Advisory Group. Observations from this expert input are presented below and form the basis of the following recommendations to maximise the value of these significant investments.

4.2 Findings

4.2.1 Strengths

The UK has an unrivalled set of large population cohorts covering the life course from preconception and pregnancy to old age, and across different socio-economic and ethnic groups. More than half of the studies have been supported by the MRC over a long time frame. These studies have impacted on health by providing insights into the aetiology and natural history of disease and contributing to the discovery of new biomarkers. The use of serial measurement has enabled monitoring risk trajectories over time and cohort studies have been successfully used to evaluate therapeutic and policy interventions in real world sub-populations.

4.2.2 Gaps and potential limitations in the portfolio

• Profile of the portfolio

Older adults dominate the portfolio, and men aged 20–40 years are particularly under-represented. The 20–40 age group are at higher risk from detrimental health behaviours such as alcohol and drug misuse. Young adult men can be recruited into studies, but are often difficult to retain. Increased use of social media and digital technologies for data gathering may be attractive to this age group and help to address the lack of engagement. The potential benefits of these remote data capture technologies on recruitment and retention need to be balanced against possible adverse effects on representativeness or generalisability. These issues require further exploration.

- **Variables collected**

Very few cohorts collect information on exposure to infectious diseases and vaccination, or collect and store biosamples in a way that is suitable for immunological studies. Banked blood specimens from representative population samples offer scope for serological studies to estimate prevalence and incidence of different infections by age, sex and immunity to vaccination. Measuring antibody titres, cytokine levels and immunophenotyping requires blood samples to be collected and stored using standardised techniques and this needs to be considered at the outset of sample collection. Expert advice should be obtained when planning biosample collections to ensure collected samples can be used to test a wide variety of biomarkers and phenotypes. Linkage to primary health records will provide some information on vaccination and infection.

A number of the cohorts have been linked to routine environmental exposure data. However, often this is at an area-based level, which is subject to limitations. Environmental exposures such as pollutants and weather can be important precursors to ill health. Currently, few cohort studies measure individual environmental exposure levels. There is scope for greater use of new technologies for monitoring personal exposures to better understand the influence of the physical environment on health and disease.

- **Data access**

Discovery of cohorts and access to their data and samples is fundamental to fully realising the scientific and translational value of these resources. Of the 34 cohort studies in this review, 24 (including all MRC funded cohorts) are part of the MRC Data Gateway cohort directory and seven other cohorts are discoverable on other directories. Information about the remaining three studies can only be found on their individual websites. Despite MRC data sharing policies for population and patient studies¹³, and similar policies of several other funders, approaches to data access and sharing vary enormously between cohorts. Cohort access arrangements that are in the public domain range from request forms available on the study website to no visible contact details or policies for data sharing.

Long-term preservation of cohort data and tissues can be problematic, particularly when a study moves into the archived phase where it is no longer collecting any new data. JISC and research organisations have begun to address this by working together to develop institutional repositories to store large datasets. Repositories such as the UK Data Service also offer data preservation, discovery and access functions.

- **Engagement with policy and practice**

Research findings from the cohort studies are of interest to policy makers and practitioners as well as to researchers, as evidenced by cross-Government department co-funding of three cohorts in this review. However, only a small number of cohorts in the UK portfolio are known to Government. Opportunities for the wealth of data in the cohort studies to inform policy are being missed. Methods to improve communication such as providing accessible narrative summaries of achievements and regular meetings with relevant policy makers and researchers will assist the potential for research translation. Increased awareness of policy priorities by cohort studies might help to shape the research agenda to increase the impact of these valuable resources.

- **Closer links with industry**

Currently, few of the cohort studies in this review collaborate with industry. Population cohorts offer industry the opportunity to validate biomarkers or risk exposures in real world, well-phenotyped populations. The benefits of using serial samples from longitudinal studies need to be weighed up against issues such as depletable samples, frequency of collection and sample size. Despite these potential limitations, a number of pharmaceutical companies are already working with UK population cohorts. Greater knowledge of the UK cohort portfolio may facilitate collaboration with industry and this may be an avenue more cohort studies may wish to explore in the future.

4.2.3 Opportunities

- **Omics, imaging and emerging technologies**

Emerging techniques such as epigenetics, metabolomics and microbiomics, together with continually evolving imaging and genomics technologies, enable deeper genotyping and phenotyping of cohort studies. Genetic studies complement observational studies as they are less subject to confounding, are increasingly affordable and can be done on small amounts of stored samples. Furthermore, platforms for genotyping are relatively robust, allowing comparison of genetic variants and validation of potential biomarkers across different cohort studies. Tools and techniques for other emerging technologies such as epigenetics and large-scale metabolomics studies are rapidly evolving and are not yet well standardised for large population studies, which can restrict comparisons between cohorts.

With repeated exposure measures over multiple time points, cohorts ought to be ideally placed for studies of epigenetic modification of the genome mediated by environmental factors. Issues of tissue specificity, heterogeneous cell types within samples, differences in technological platforms, bias introduced by sampling, storage and analyses methods and individual variations in epigenetic patterning can make reproducibility and functional relevance difficult to interpret in large-scale population studies. Despite these challenges the cohort studies in this review, including many of the MRC funded studies, are carrying out epigenetic studies. The majority of these studies are based on blood samples and are using samples from a single time point. However, it is through such exploratory studies in well-characterised longitudinal populations that knowledge will improve, in particular assessing the value of makers from accessible tissues.

Cohorts, with their wealth of longitudinal data on exposures, offer great potential for discovery and validation of biomarkers at scale. Currently, ten cohort studies have carried out metabolomics studies investigating areas such as genetic and environmental contribution to metabolite nutritional patterning, ageing metabolomics profiles and metabolite predictors for incident diabetes across different ethnic groups. The MRC and NIHR have recently funded the Phenome Centre¹⁴, which is a legacy from the London 2012 Olympics anti-doping facilities. The Phenome Centre is a resource for large-scale sample metabolic screening and biomarker identification that can be utilised for population scale studies.

Imaging, such as echocardiography, DEXA and MRI, is a valuable tool in aetiological and prognostic studies and can be used to link early cohort exposures to structural and functional changes and health outcomes over time. Some of the cohort studies in the portfolio, including many MRC funded cohorts, are already using imaging. There is, however, scope for more extensive use of imaging modalities to add value to cohort resources and to enhance our understanding of structural changes associated with disease progression.

- **Data linkage**

The NHS is the largest single health care provider globally that holds longitudinal health data on the whole population. Linking cohorts to health records and other cross-sector administrative data opens up a multitude of clinical, public health and social science research possibilities. In addition, data linkage offers a cost-effective means of prospective follow-up.

Although almost all the cohort studies in this review are linking to routine records, linkage is often to a single health data source. Currently, few studies are linking to both primary and secondary health care records or to the range of other routine health and administrative datasets that are available to researchers. There is great potential for more extensive linkage to NHS records, which will both directly benefit the study through richer data collection and expand the opportunity for new interdisciplinary collaboration and discovery science across diseases and conditions. In the future, studies will benefit from recent investments to increase access to routine health data for research and to build research capacity in linking large and complex datasets¹⁵.

14 MRC-NIHR National Phenome Centre <http://www1.imperial.ac.uk/phenomecentre/>

15 Clinical Practice Research Datalink <http://www.cprd.com>, Health Informatics Research Centres <http://www.mrc.ac.uk/Ourresearch/ResearchInitiatives/E-HealthInformaticsResearch/index.htm> and Farr Health Informatics Research Institute <http://www.farrinstitute.org/>

- **Cross-cohort collaborations**

Cohort consortia and cross-cohort collaborations can provide new research opportunities such as the study of rare phenotypes, as well as increase sample sizes to more accurately predict risk and validate findings. Comparisons between cohorts can enable intergenerational and period effects to be investigated and allow evaluation of policies and other natural experiments that differ over time or by location.

The funders support a number of initiatives aimed at harmonising and standardising meta-data to pool longitudinal studies in future research¹⁶. While harmonisation and standardisation of core measures are important, this must be balanced with the need for scientific innovation and the use of new tools and instruments that exploit the unique features of a cohort study. The majority of UK population cohorts are involved in national and international cohort consortia, however, the issue of widespread use of standards by cohorts remains an on-going challenge. The MRC will continue to strongly encourage adoption of standards in data management and prospective data collection to enable greater cross-cohort comparisons, leading to wider research applications beyond the individual studies.

- **Digital technologies and remote data capture**

Traditional methods of cohort recruitment, maintenance, and data and sample collection are costly and can have high attrition levels. New digital technologies, including social media and the internet, can offer cost-effective alternatives for recruitment, retention and data gathering. Remote data collection such as internet-based cognitive testing and sending blood spots or accelerometers through the post can be used in place of face-to-face contact, where appropriate. Further work is needed to assess whether the effect of these different methodologies on data accuracy, selection and retention bias, alters the ability to test hypotheses.

4.3 Key recommendations

- **Guidance on use of omics platforms and emerging technologies**

Integration of genomics, epigenetics, metabolomics, imaging and other emerging technologies into cohort studies has the potential to greatly improve our understanding of the aetiology, prediction and stratification of disease across different populations. High throughput science is a rapidly evolving area and not all technologies are as well-validated as genomics for widespread use in large-scale population cohort studies.

Despite these challenges, two thirds of the cohort studies in this review have carried out genotyping on participants and more than half the studies have used epigenetics and/or metabolomics studies. Although large-scale population epigenetics and metabolomics studies are at a comparatively early stage, application of these techniques on cohort studies will potentially advance our understanding of individual variations and tissue specificity. Use of common platforms and standardised methods for sample collection, storage and analyses across cohort studies will enable more reliable comparisons between studies and replication of findings. Shared learning on tissue specificity, platform choice, informatics and tools and technologies will allow cohort studies to more appropriately adopt emerging and evolving technologies. Greater application of imaging to measure sub-clinical disease and associations between early exposures and later phenotypic structural and functional changes will also enrich individual studies.

A potential opportunity for the UK is to carry out cross-cohort 'omics' or imaging comparative studies for the discovery and validation of causal traits and biomarkers across large well-phenotyped longitudinal cohorts.

Recommendation 1

Cohort studies should use standardised or validated sample collection, storage, tools and platforms for evolving technologies, where possible aligning with leaders in the field such as UK Biobank, to enable future cross-cohort comparison.

The MRC, cohort leaders and experts in the field should work together to develop guidance on best practice for high throughput science in large-scale populations, in particular for epigenetic studies.

- **Linkage to routine and research datasets**

Linkage to routine health records, cross-sector administrative and environmental data and research datasets greatly expands the scope of a cohort to carry out clinical, public health and socio-economic research. Data linkage also offers a cost-effective method of follow-up in addition to, or instead of, future participant contact.

Recent infrastructure initiatives such as the Clinical Practice Research Datalink¹⁷ and Safe Havens have increased secure access to clinical records for research. This infrastructure has been complemented by major investments from the MRC, in partnership with nine other UK funders, to establish Health Informatics Research Centres and the Farr Institute to undertake research using record linkage¹⁸. Access to data from government departments should improve through the Administrative Data Research Centres (ADRCs) which will be based in each of the UK countries¹⁹. A shared objective of all these activities is to capitalise on the wealth of information in the NHS and non-health sectors by developing methodologies, standards and best practice to combine and interrogate large datasets.

The MRC is committed to supporting capacity building, infrastructure, methods development and knowledge sharing in the use and linking of large datasets, and continues to make significant investments in these areas. Although most cohort studies in the UK portfolio are linked to at least one source of routine records, there is scope for more extensive record linkage to these studies. Cohort studies should take advantage of the increased availability of health and administrative records for research and the growing expertise, networking and research infrastructure in the UK for linking large and complex datasets.

Recommendation 2

Broad and enduring consent for linkage to routine data needs to be obtained from cohort participants, wherever possible, for all prospective studies and sweeps.

The scientific and interdisciplinary potential of a cohort should be enhanced through linkage to the increasing number of routine health records and administrative datasets available in the UK. Studies can draw on expertise in data linkage within the Farr Institute, UK Health Informatics Research Network.

- **Skills and capacity in analysing complex datasets**

Cohort studies involve increasingly large datasets with each new round of data collection, particularly if the cohort is linked to health or administrative records, 'omics' or imaging data. Managing, linking and analysing these complex datasets requires a range of interdisciplinary skills.

Building capacity in informatics is a major priority for the MRC. As well as offering strategic skills fellowships in these areas, the MRC supports skills development in complex data within MRC funded cohorts, centres and units. In the past 12 months the MRC, in partnership with other funders, has made significant investments in the Health Informatics Research Centres and Network, Farr Institute of Health Informatics Research²⁰ and Medical Bioinformatics, which offer a range of training and career development opportunities in analysing complex data. Collaborations between cohort study teams and these centres of excellence will enhance capacity building and the sharing of knowledge and skills in analysing large datasets.

Given the continuing expansion in size, complexity and availability of research data, the MRC is committed to building a sustainable research capability in the UK in critical skills to exploit big data.

Recommendation 3

Cohort studies should take advantage of the opportunities for skills development in interrogating large and complex data through collaborations with centres of excellence such as the Farr Institute.

17 Clinical Practice Research Datalink <http://www.cprd.com>

18 Health Informatics Research Centres <http://www.mrc.ac.uk/Ourresearch/ResearchInitiatives/E-HealthInformaticsResearch/index.htm> and Farr Institute of Health Informatics Research <http://www.farrinstitute.org/>

19 Administrative Data Research Centres <http://www.esrc.ac.uk/funding-and-guidance/funding-opportunities/26526/administrative-data-research-centres-2013.aspx>

20 Farr Health Informatics Research Institute <http://www.farrinstitute.org/>

- **Data discovery and sharing**

The ability to discover information about a cohort and have access to data and samples is essential to fully realising the scientific and translational value of these resources. There are multiple cohort directories in existence and the vast majority of large UK cohorts, including all MRC funded cohorts, are listed in at least one. However, not all cohorts in this review are readily discoverable. Bringing together the portfolio of large UK cohorts into a single directory will raise awareness and increase the utility and value for money of these valuable UK resources.

The MRC has clear data sharing policies for population and patient studies²¹ which state that sharing data, and where appropriate samples, should be normal practice. Applicants are required to submit data management plans as part of their proposals, which outline arrangements for data sharing and governance processes. These also include long-term plans for curation and sustainability of the resource. The cohort study's access and sharing policies should comply with the MRC policy and be transparent and publically available. Although study policies are in place, some MRC funded cohort studies are not yet at the expected level of compliance. Monitoring procedures need to be put in place to enable all cohort studies to reach the expected standards of access, sharing and governance within a designated timeframe.

Recommendation 4

Cohort leads should ensure that their studies are easily discoverable via directories.

Processes are needed to ensure that all MRC funded cohorts comply with MRC data sharing policies. Studies need to be accessible and have transparent governance procedures in place that enable data sharing and where appropriate, access to samples.

- **Cross-cohort analyses**

Individual cohort studies have made major contributions to our understanding of factors that influence health over the life course. Cross-cohort collaborations can further enhance scientific opportunities and the translational potential of individual cohorts. Comparing or combining cohort populations increases the statistical power, enables replication of findings from individual studies, validation of biomarkers and studies of intergenerational and period effects.

Although pooling genotypic information across population studies to identify genetic variance and causal traits is common practice, combining phenotypic variables across studies is more challenging. Most of the cohorts in this review collect the same core information on common exposures and variables, however, the individual measures and methods used vary enormously. Combining phenotypic data relies on high-quality study meta-data and the ability to harmonise these different data. Use of common meta-data standards such as DDI3.2 facilitates data harmonisation and cohort meta-analyses.

Widespread adoption of data standards in MRC and other UK cohort studies is not yet commonplace and the quality of meta-data within these studies is highly variable. The use of core common standards can be balanced with methodological innovation and study-specific needs. Establishing fora for sharing knowledge on the use of standards, improving meta-data quality and lessons from new methodologies will benefit individual studies as well as enable new research through meta-studies.

Many UK cohort studies are involved in consortia and cross-cohort collaborations that are developing innovative methods for combining and comparing complex phenotypic data, such as the HALCYon collaboration of eight UK ageing cohorts²². The BIS/ESRC/MRC funded CLOSER initiative²³ is working to harmonise meta-data across a number of UK cohort studies to answer questions on body size and socio-economic position. There is scope for the MRC and other funders to encourage the adoption of common standards for prospective data collection to facilitate future cross-cohort research applications.

21 MRC policy on sharing of research data from population and patient studies
<http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/datasharing/Policy/PHSPolicy/index.htm>

22 HALCYon <http://www.halcyon.ac.uk/>

23 CLOSER <http://www.closerprogramme.co.uk/>

Recommendation 5

Adoption of core common data standards, sharing knowledge and improving meta-data quality should be encouraged and facilitated by cohort studies, the MRC and other funders.

- **Novel methods for establishing and maintaining cohorts**

Sustained support of large cohort studies with regular data and sample collection and on-going maintenance of the resource is very expensive. Frequency of new data sweeps and the extent of variables and samples collected are a balance between optimal study design and practical considerations such as participant burden and costs. New digital technologies provide a cost-efficient means of remote data capture and identifying, recruiting and retaining participants. The use of remote methods can increase value for money by enabling more frequent data gathering and overcoming geographical constraints.

Despite cost savings, there can be disadvantages of remote techniques. Concerns about recruitment and retention, selection bias and limitations on types of data and samples that can be collected remotely need to be taken into account. More studies are needed to evaluate the benefits and limitations of new technologies and data capture. To increase value for money, investigators should consider the merits of adopting less costly remote monitoring, such as the use of the internet, social media or sending samples by post on a case-by-case basis.

Recommendation 6

Cost-effective methods such as the use of new digital technologies, and remote data and sample capture, should be adopted where possible and appropriate to reduce the costs of enhancing and maintaining cohort studies.

- **Increasing the research and translational application of cohort studies**

The UK cohort studies have made major contributions to population health and wellbeing. There is potential for even greater application of these valuable resources to address important public health challenges such as obesity, dementia, alcohol consumption and healthy ageing. Each cohort within the portfolio has been established with a specific scientific purpose; for instance, investigating the incidence of cancer or cardiovascular disease in healthy populations. However, the cohort studies collect a similar set of core phenotypic information that can potentially be applied to a range of research or policy questions. There should be greater use of existing UK cohort studies to address major public health challenges, particularly the novel use of cohorts in studies outside their usual scientific focus.

Demonstrating impact from funded research is an increasing priority for funders and investigators alike. Research outcomes from cohort studies are of relevance to a range of stakeholders, including policy makers and practitioners, as well as researchers. Currently, policy makers and other beneficiaries of research are only familiar with a sub-set of UK cohort studies, many of which are not MRC funded. Scientific papers and traditional academic routes are not the best method for conveying relevant messages to busy policy makers. Methods to improve communication will assist translation of study findings. Cohort study teams need to play a more proactive role in engaging potential research users, such as publishing research outcomes in an accessible format for stakeholders or holding regular meetings to discuss policy needs and research outputs.

Recommendation 7

Effective models of two-way engagement between cohort study teams, policy makers and/or practitioners should be established to increase the impact of cohort study research outputs and potential for translation to inform evidence based policies.

4.4 Next steps

The UK cohort portfolio has a good coverage of population sub-groups, although the numbers of young adults, especially young men, are comparatively low. The breadth of populations encompassed and the range of repeat measures mean that the existing portfolio of cohort studies could be used to address a wide range of scientific questions. Any future investment in new population cohorts will need to be well-justified, demonstrating that there is a clear scientific gap that cannot be adequately addressed by the current portfolio.

Cohort funding decisions have traditionally been taken in isolation, on a case-by-case basis. Given the significant on-going costs of establishing and maintaining a cohort, it is important to take the overall UK cohort portfolio into account when making individual funding decisions. An expert Cross-Research Board Advisory Group has been established to develop a consistent approach to managing the MRC's extensive investment in cohort studies. The Group review all applications for new population cohort studies and renewals of large population cohorts. Proposals are assessed not only for quality, but also for their unique scientific niche and value for money within the context of other funded cohort studies. The comprehensive overview of all the large UK population cohort studies within this review will assist the MRC, and other funding bodies, in future funding decisions.

The recommendations emerging from the review enforce the current MRC population and patient data sharing policy and highlight areas that will enable wider research applications and improve translation of research outcomes. Some recommendations such as capacity building, and to some extent more expansive linkage of cohorts to routine data sources, are already the subject of current initiatives. Other recommendations, including engaging with policy makers and the greater adoption of data standards, improving meta-data quality and adherence to MRC policy, will require steps to be put in place to encourage good practice or compliance in a timely manner. For these recommendations to impact on the whole UK cohort portfolio, the MRC will need to work in partnership with other cohort funders to develop a coordinated approach to ensuring all cohort studies are discoverable, accessible and adopt common standards where possible.

Many of the MRC funded and other UK cohort studies have already carried out, or are intending to carry out, omic analyses on cohort biosamples. High throughput science on large-scale population studies is a rapidly evolving field. There is scope for the MRC to develop guidance for population studies on best practice, ranging from sample storage to analyses. There may also be value in carrying out an in-depth review of cohort biosamples to ascertain which studies are amenable to omics analyses and whether there are benefits in developing a common approach across suitable studies.

This review showcases for the first time the breadth of large population cohorts that exist in the UK. In addition to providing a useful tool to aid funding decisions, the recommendations are intended to strengthen the value of MRC and other UK cohort assets by highlighting areas that will enable new discovery science and improve translation of research outcomes. The document has been produced to assist the MRC in funding decisions, policy development and strategic discussions. It is envisaged that the landscape review and recommendations will also be relevant to other funders, cohort study teams and collaborators, and will be of interest to policy makers seeking evidence to develop or evaluate policies and interventions.

Annexes

Annex 1. Cohorts included in the Strategic Review

The following cohorts funded by the MRC and other funders have been included in the data gathering exercise.

Cohort name	Principal Investigator(s)	Organisation	Current Core Facility Funders
11-16 and 16+ Study	Dr Helen Sweeting	MRC Social and Public Health Sciences Unit http://www.sphsu.mrc.ac.uk	Medical Research Council
1970 British Cohort Study (1970BCS)	Dr Alice Sullivan	Centre for Longitudinal Studies, Institute of Education http://www.cls.ioe.ac.uk	Economic and Social Research Council
Aberdeen Children of the 1950s (ACONF)	Professor Sally Macintyre Professor David Leon	University of Aberdeen http://www.abdn.ac.uk/aconf	Medical Research Council
Avon Longitudinal Study of Parents and Children (ALSPAC)	Professor George Davey-Smith	University of Bristol http://www.bristol.ac.uk/alspac	Medical Research Council Wellcome Trust
Born in Bradford (BiB)	Professor John Wright	Bradford Royal Infirmary http://www.borninbradford.nhs.uk	No core funding
Boyd Orr Cohort	Professor Richard Martin Professor David Gunnell, Professor George Davey Smith	University of Bristol http://www.bris.ac.uk/social-community-medicine/projects/boyd-orr/about/	British Heart Foundation
Breakthrough Generations Study	Professor Anthony Swerdlow Professor Alan Ashworth	Breakthrough Breast Cancer http://www.breakthrough.org.uk/ http://www.icr.ac.uk/	Breakthrough Breast Cancer
British Regional Heart Study (BRHS)	Professor Peter Whincup Professor S Goya Wannamethee Professor Richard Morris	University College London http://www.ucl.ac.uk/pcph/research-groups-themes/brhs-pub	British Heart Foundation
British Women's Heart & Health Study (BWHHS)	Professor Shah Ebrahim, Professor Juan P. Casas Professor Debbie Lawlor	London School of Hygiene and Tropical Medicine http://www.lshtm.ac.uk/eph/ncde/research/bwhhs/index.html	Department of Health British Heart Foundation
Cognitive Function and Ageing Studies I (CFAS I)	Professor Paul Ince	University of Sheffield CFAS http://www.cfas.ac.uk/	Medical Research Council
Cognitive Function and Ageing Studies II (CFAS II)	Professor Carol Brayne	University of Cambridge http://www.cfas.ac.uk/	Medical Research Council
Determinants of Adolescent Social well-being and Health (DASH)	Professor Seeromanie Harding	MRC Social and Public Health Sciences Unit http://dash.sphsu.mrc.ac.uk/ http://www.sphsu.mrc.ac.uk/research-programmes/eh/dash/	Medical Research Council

Cohort name	Principal Investigator(s)	Organisation	Current Core Facility Funders
English Longitudinal Study of Ageing (ELSA)	Professor Sir Michael Marmot Professor Andrew Steptoe	University College London, Institute for Fiscal Studies, University of Manchester, NatCen Social Research http://www.ifs.org.uk/ELSA	UK Government Consortium (Department for Communities and Local Government, Department for Transport, Department for Work and Pensions, HM Revenue and Customs, Office for National Statistics, Department of Health National Institute Health Research, Office for National Statistics) National Institute on Aging (NIH)
European Prospective Investigation of Cancer Norfolk (EPIC Norfolk)	Professor Kay-Tee Khaw	University of Cambridge, http://www.srl.cam.ac.uk/epic/	Medical Research Council Cancer Research UK
European Prospective Investigation of Cancer Oxford (EPIC Oxford)	Professor Tim Key	University of Oxford http://www.epic-oxford.org/home/	Cancer Research UK
Gemini: health and development in twins	Professor Jane Wardle	Department of Epidemiology and Public Health, University College London http://www.ucl.ac.uk/hbrc http://www.geministudy.co.uk	Cancer Research UK
Growing up in Scotland (GUS)	Dr Paul Bradshaw	ScotCen Social Research http://www.growingupinScotland.org.uk MRC Social and Public Health Sciences Unit http://www.sphsu.mrc.ac.uk Centre for Research on Families and Relationships http://www.cfr.ac.uk	Scottish Government
Hertfordshire cohort study (HCS)	Professor Cyrus Cooper	MRC Lifecourse Epidemiology Unit, http://www.mrc.soton.ac.uk/herts/index.asp?page=1	Medical Research Council
Lothian Birth Cohort 1936 (LBC1936)	Professor Ian Deary	University of Edinburgh; www.ed.ac.uk , http://www.lothianbirthcohort.ed.ac.uk	Age UK Medical Research Council
Millennium Cohort (MCS)	Professor Lucinda Platt	Centre for Longitudinal Studies, Institute of Education http://www.cls.ioe.ac.uk/	Economic and Social Research Council

Cohort name	Principal Investigator(s)	Organisation	Current Core Facility Funders
Million Women Study	Professor Dame Valerie Beral	University of Oxford http://www.millionwomenstudy.org/introduction/	Cancer Research UK Medical Research Council
MRC National Survey of Health and Development Cohort /1946 Birth Cohort (NSHD/1946BC)	Professor Diana Kuh	MRC Unit for Lifelong Health and Ageing http://www.nshd.mrc.ac.uk/nshd.aspx	Medical Research Council
National Child Development Study/1958 Birth Cohort (NCDS/1958BC)	Professor Jane Elliott Professor Paul Burton (Biomedical Resource) Professor Chris Power (Biomedical data collection)	Centre for Longitudinal Studies, Institute of Education http://www.cls.ioe.ac.uk/ncds	Economic and Social Research Council Medical Research Council Wellcome Trust
Newcastle 85+	Professor T Kirkwood Dr Joanna Collerton	Newcastle University http://www.ncl.ac.uk/iah/research/areas/biogerontology/85plus/	Medical Research Council Biotechnology and Biological Sciences Research Council Newcastle University
Rural Uganda General Population Cohort (GPC)	Professor Janet Seeley Dr Anatoli Kamali	MRC/UVRI Uganda Research Unit on AIDS http://www.mrcuganda.org/	Medical Research Council
Southall and Brent Revisited (SABRE)	Professor Nish Chaturvedi	http://www.sabrestudy.org/	Wellcome Trust British Heart Foundation
Southampton Women's Survey (SWS)	Professor Cyrus Cooper	MRC Lifecourse Epidemiology Unit http://www.mrc.soton.ac.uk/sws/index.asp	Medical Research Council British Heart Foundation European Union
Twin Early Development Study (TEDS)	Professor Robert Plomin	King's College London http://www.kcl.ac.uk/index.aspx	Medical Research Council
Twins UK	Professor Tim Spector	King's College London http://www.twinsuk.ac.uk/	Wellcome Trust
UK Biobank	Professor Rory Collins	University of Oxford http://www.ctsu.ox.ac.uk/research/mega-studies/uk-biobank	Medical Research Council Wellcome Trust Chief Scientist Office Department of Health National Institute for Health Research British Heart Foundation Northwest Regional Development Agency Welsh Government
UK Women's Cohort Study (UKWCS)	Professor Janet Cade	University of Leeds http://www.leeds.ac.uk/medicine/ceb/NutEp/ukwcs/	No core funding

Cohort name	Principal Investigator(s)	Organisation	Current Core Facility Funders
Understanding Society	Professor Nick Buck	Institute for Social and Economic Research, University of Essex https://www.understandingsociety.ac.uk	Economic and Social Research Council
West of Scotland Twenty-07 Study (Twenty-07)	Professor Anne Ellaway	MRC Social and Public Health Sciences Unit http://www.sphsu.mrc.ac.uk/	Medical Research Council
Whitehall II Study	Professor Mika Kivimaki	University College London http://www.ucl.ac.uk/whitehallII	Medical Research Council National Institute on Aging (NIH)

Highlighted cohorts receive core funding from the MRC

Annex 2. Cohort Questionnaire

As part of the data gathering for each cohort, a document pre-populated with data collated by the Office was sent to the cohort PIs when possible. PIs were asked to confirm the data provided and to fill in other information on cohort attrition rate as well as anticipated future plans for their cohort.

Question	Response	Guidance
General cohort information		
PI		If this Cohort builds on previous data please specify. This is the date the Cohort started and participants have been recruited
Organisation (including website)		Include Cohort core funding in the last 5 years excluding any research funding (including sample collection and any related sample processing)
Funders of Core Cohort funding within the last 5 years. Include dates of funding.		Include dates of the funding (start and end dates)
Current MRC funding/other funders		
Study population		
Recruitment start date		
Original sample size		
Current sample size: Estimated Sample Size based on last sweep/contact		If a sub-group of the Cohort is followed, please specify if you intend to go back and follow the original number.
Number of participants who have opted out or were unable to be contacted		
Number of participants who have died		
Age range at the start of the Cohort		
Age range of the Cohort in 2012		
Gender		
Ethnicity		
Date of last sweep/contact		
How many are you still following or intending to follow?		
Length of follow-up		
Geographical location		
Data collection sweeps		
Scientific focus - past and current focus (with link to IJE + website as relevant)		

Question	Response	Guidance
Variables collected		
Anthropometric measures including self-reported illness and measurements taken by researchers	Y/N	
Height		
Weight		
Waist circumference		
Hip circumference		
Blood pressure		
Please list any other Anthropometric measures collected		
Physical Measures	Y/N	
Cardiovascular		
Respiratory		
Musculoskeletal		
Hearing and Vision		
Reproductive		
Please list any other Physical Measurements collected		
Cognitive measures	Y/N	
Mental health		
Is this Self-rated or Tested or both?		Tested for a mental health variable only includes a clinical diagnosis made by a qualified practitioner or verification of clinically diagnosed mental illness via health records
Cognitive function		
Is this Self-rated or Tested or both?		
Socio-economic measures	Y/N	
Occupation and Employment		
Income and Finances		Includes state financial support and income benefits
Family circumstances		
Housing and accommodation		
Education		
Ethnic group		
Marital status		
Social support		Includes emotional, family and community social support

Question	Response	Guidance
Please list any other socio-economic measures collected		
Lifestyle measures	Y/N	
Smoking		
Physical activity		
Dietary habits		
Alcohol		
Please list any other lifestyle measures collected		
Biological Samples	Y/N	
Blood		
Urine		
Saliva		
Extracted DNA/RNA		
Cell Lines		
Please list any other Biological Samples collected		
Omics analyses	Y/N	Please provide details of the number of participants analysis has been carried out on and the area of interest
Genotyped		
Whole genome or exome sequencing		
Epigenetic studies		
Metabolomics studies		
If you haven't carried out any omics do you have plans to? Please provide details		
Please list any other relevant information		
Consent	Y/N	
Consent for re-contact		
Consent for linkage		
Data linkage	Y/N	Please indicate which datasets you link to
Health Records including registries		
Administrative data		Includes education, Ministry of Justice, pensions etc.
Census		

Question	Response	Guidance
Births and Deaths		
Environmental		
Please list any other data sets that you link to		
If you don't currently link to other data sets do you have plans to? Please provide details		
Cohort research output:		
Total number of publications		
Number of publications in last 12 months		
Top five publications over the life of the Cohort		List only the top 5 publications relevant to your Cohort. Please attach to this document.
Impact on policy		Provide up to 3 examples on how research generated from this Cohort has had impact on health policy and practice (250 words each). Include references to any relevant publications but do not include the publication. Please attach to this document.
Is the Cohort part of a national or international consortium?		
If so, which ones?		
Future plans		
Anticipated future scientific focus of the Cohort		
Future recruitment		List any plan to recruit new participants to the Cohort in the future and under which circumstances
Estimated size of the Cohort in 10 years based on current number of participants		
Estimated number of new recruits within the next 10 years (if relevant)		
Future primary data collection(s)		
When do you envisage there will be no further primary data collected from your cohort?		
Estimated size of the cohort at last anticipated primary data collection(s)		

Question	Response	Guidance
Plans for future funding: MRC or other funders		
Sustainability of the cohort: plans for future curation and archiving including data and sample storage following the last primary data collection		
Data access		
Data access: does your Cohort fully comply with the MRC Data access policy?		
If not, when do you envisage it would do so?		

Annex 3. Cohort Data Overview

Cohort	Start Date	Sex	Sample Size at Recruitment	Age at Recruitment (yr)	Anthropometric Variables	Physical Variables	Cognitive Measures	Lifestyle	Socio-economic position	Biological Samples	Data Linkage
11-16 and 16+ Study	1994	MF	2,586	11	✓	✓	✓	✓	✓		
1970 British Cohort Study (1970 BCS)	1970	MF	17,287	0	✓	✓	✓	✓	✓		✓
Aberdeen Children of the 1950s (ACONF)	1962 (1999)	MF	12,150	6-12 (43-49)	✓	✓	✓	✓	✓	✓	✓
Avon Longitudinal Study of Parents and Children (ALSPAC)	1991-2	MF	Children 14,062 Mothers 14,541	0 16-45	✓	✓	✓	✓	✓	✓	✓
Born in Bradford (BiB)	2006-11	MF	Children 13,857 Mothers 12,453	0 15-49	✓	✓	✓	✓	✓	✓	✓
Boyd Orr Cohort	1937 (1988)	MF	4,397	6-12 (51-70)	✓	✓		✓	✓	✓	✓
Breakthrough Generations Study	2004-09	F	112,798	16-100	✓	✓		✓	✓	✓	✓
The British Regional Heart Study (BRHS)	1978-80	M	7,735	40-59	✓	✓	✓	✓	✓	✓	✓

Cohort	Start Date	Sex	Sample Size at Recruitment	Age at Recruitment (yr)	Anthropometric Variables	Physical Variables	Cognitive Measures	Lifestyle	Socio-economic position	Biological Samples	Data Linkage
British Women's Heart & Health Study (BWHHS)	1999	F	4,286	60-79	✓	✓	✓	✓	✓	✓	✓
Cognitive Function and Ageing Studies I (CFAS I)	1989	MF	18,500	>65		✓	✓	✓	✓	✓	✓
Cognitive Function and Ageing Studies II (CFAS II)	2008	MF	7,524	>65		✓	✓	✓	✓	✓	✓
Determinants of Adolescent Social well-being and Health (DASH)	2002-03	MF	6,643	11-13	✓	✓	✓	✓	✓	✓	✓
English Longitudinal Study of Ageing (ELSA)	2002	MF	12,099	50-100	✓	✓	✓	✓	✓	✓	✓
European Prospective Investigation of Cancer Norfolk (EPIC Norfolk)	1993-97	MF	30,000	40-79	✓	✓	✓	✓	✓	✓	✓

Cohort	Start Date	Sex	Sample Size at Recruitment	Age at Recruitment (yr)	Anthropometric Variables	Physical Variables	Cognitive Measures	Lifestyle	Socio-economic position	Biological Samples	Data Linkage
European Prospective Investigation of Cancer Oxford (EPIC Oxford)	1993-2000	MF	65,000	17-98	✓	✓		✓	✓	✓	✓
Gemini	2008	MF	4,808	4-20 months	✓			✓	✓	✓	
Growing up in Scotland (GUS)	2005 2005 2011	MF	2,858 5,217 6,127	34 months 10 months 10 months	✓		✓	✓	✓		✓
Hertfordshire cohort study (HCS)	1931 (1990)	MF	3,225	0 (59-72)	✓	✓	✓	✓	✓	✓	✓
Lothian Birth Cohort 1936 (LBC1936)	1947 (2004)	MF	1,091	11 (70-72)	✓	✓	✓	✓	✓	✓	✓
Millennium Cohort Study (MCS)	2000-02	MF	Children 19,519 Families 19,244	0 All ages	✓	✓	✓	✓	✓	✓	✓
Million Women Study	1996-2001	F	1,360,000	50-64	✓	✓	✓	✓	✓	✓	✓
MRC National Survey of Health and Development Cohort /1946 Birth Cohort	1946	MF	5,362	0	✓	✓	✓	✓	✓	✓	✓

Cohort	Start Date	Sex	Sample Size at Recruitment	Age at Recruitment (yr)	Anthropometric Variables	Physical Variables	Cognitive Measures	Lifestyle	Socio-economic position	Biological Samples	Data Linkage
MRC National Survey of Health and Development Cohort /1946 Birth Cohort (NSHD) /1946BC)	1958	MF	17,416	0	✓	✓	✓	✓	✓	✓	✓
Newcastle 85+	2006	MF	851	85	✓	✓	✓	✓	✓	✓	✓
Rural Uganda General Population Cohort (GPC)	1989	MF	10,000	All	✓	✓		✓	✓	✓	✓
Southall and Brent Revisited (SABRE)	1988	MF	4,858	40-69	✓	✓	✓	✓	✓	✓	✓
Southampton Women's Survey (SWS)	1998	MF	Children 3,159 Mothers 12,583	0 20-34	✓	✓	✓	✓	✓	✓	✓
Twin Early Development Study (TEDS)	1994-96	MF	13,690	0	✓		✓	✓	✓	✓	✓
Twins UK	1992	MF	346	49-69	✓	✓	✓	✓	✓	✓	✓
UK Biobank	2006-10	MF	503,316	40-69	✓	✓	✓	✓	✓	✓	✓
UK Women's Cohort Study (UKWCS)	1995	F	35,372	35-69	✓	✓	✓	✓	✓	✓	✓

Cohort	Start Date	Sex	Sample Size at Recruitment	Age at Recruitment (yr)	Anthropometric Variables	Physical Variables	Cognitive Measures	Lifestyle	Socio-economic position	Biological Samples	Data Linkage
Understanding Society	2009	MF	40,000 household 100000 individuals	All	✓	✓	✓	✓	✓	✓	✓
West of Scotland Twenty-07 Study (Twenty-07)	1986	MF	4,510	15, 35, 55	✓	✓	✓	✓	✓		✓
Whitehall II Study	1985	MF	10,308	35-55	✓	✓	✓	✓	✓	✓	✓

Highlighted cohorts receive core funding from the MRC

Italicised figures show data from when cohorts received sustained funding

Annex 4. Cohort Data by Age and Number

Cohort	Start Date	Sex	Sample Size at Recruitment	Age at Recruitment (yr)	Estimated Sample Size based on last sweep/contact	Approximate age range in 2013 (yr)	Length of follow-up (yr)	Comments
11-16 and 16+ Study	1994	MF	2,586	11	1,258	30	12	Last sweep in 2002-2004. Sub-set of 558 contacted in 2006 No further follow-up of cohort is planned
1970 British Cohort Study (1970 BCS)	1970	MF	17,287	0	8,874	43	43	Last sweep 2012
Aberdeen Children of the 1950s (ACONF)	1962 (Sustained cohort funding from 1999)	MF	12,150	6-12 (43-49 when cohort received sustained funding)	7,000	57-63	51	Cohort is based on historical data from the Aberdeen Child Development Survey (ACDS) which collected data on individuals born 1950-1956; conducted in 1962 Last contact in 2001-03
Avon Longitudinal Study of Parents and Children (ALSPAC) - children	1991-2	MF	14,062	0	11,264	22	22	Numbers include those contactable
Avon Longitudinal Study of Parents and Children (ALSPAC) - mothers	1991-2	F	14,541	16-45	11,264	38-67	22	Numbers include those contactable
Born in Bradford (BiB) – children	2007-11	MF	13,857	0	13,500	2-6	2-6	Follow-up is primarily through routine data except when sub-group contacted
Born in Bradford (BiB) – mothers	2006-10	F	12,453	15-49	12,000	17-52	2-6	Follow-up is primarily through routine data except when sub-group contacted

Cohort	Start Date	Sex	Sample Size at Recruitment	Age at Recruitment (yr)	Estimated Sample Size based on last sweep/contact	Approximate age range in 2013 (yr)	Length of follow-up (yr)	Comments
Boyd Orr Cohort	1937 (Sustained cohort funding from 1988)	MF	4,397	0-19 (51-70 when cohort received sustained funding)	7,99	76-95	76	Cohort is based on historical data from the Carnegie United Kingdom Trust's study of Family Diet in Pre-War Britain (1937-39) of 4999 children aged 0-19 Last clinical sweep 2002/3; last vital status update from NHS Information Centre: 7th March 2007
Breakthrough Generations Study	2004-09	F	112,798	16-100	110,000	16-102	9	Last sweep 2010-2012
British Regional Heart Study (BRHS)	1978-80	M	7,735	40-59	3,054	75-94	35	
British Women's Heart & Health Study (BWHHS)	1999	F	4,286	60-79	3,236	72-93	14	
Cognitive Function and Ageing Studies I (CFAS I)	1989	MF	18,500	>65	150	>85	24	Numbers include those currently followed
Cognitive Function and Ageing Studies II (CFAS II)	2008	MF	7,524	>65	7,524	>67	5	Last follow-up 2012/13
Determinants of Adolescent Social well-being and Health (DASH)	2002-03	MF	6,643	11-13	4,779	22-24	11	Last sweep in 2005/6
English Longitudinal Study of Ageing (ELSA)	2002	MF	12,099	50-100	10,317	50-100	11	Last sweep in 2012

Cohort	Start Date	Sex	Sample Size at Recruitment	Age at Recruitment (yr)	Estimated Sample Size based on last sweep/contact	Approximate age range in 2013 (yr)	Length of follow-up (yr)	Comments
European Prospective Investigation of Cancer Norfolk (EPIC Norfolk)	1993-97	MF	30,000	40-79	20,025	59-93	20	Last sweep in 2006-2011
European Prospective Investigation of Cancer Oxford (EPIC Oxford)	1993-2000	MF	65,000	17-98	50,810	32-100	20	
Gemini	2008	MF	4,808	4-20 months	4,440	5-6	5	
Growing up in Scotland (GUS)	2005	MF	2,858	34 months	2,200	9-10	8	Last sweep 2009/10
	2005	MF	5,217	10 months	4,000	7-8	8	Last sweep 2012/13
	2011	MF	6,127	10 months	6,127	2-3	2	Last sweep 2011/12
Hertfordshire cohort study (HCS)	1931 (Sustained cohort funding from 1990)	MF	3,225	0 (59-72 when cohort received sustained funding)	1,700-1,800	74-81	82	Cohort is based on historical data collected from individuals born in Hertfordshire between 1931 and 1939
Lothian Birth Cohort 1936 (LBC1936)	1947 (Sustained cohort funding from 2004)	MF	1,091	11 (70-72 when cohort received sustained funding)	700	77	66	Cohort is based on historical data collected from individuals born in 1936 who took part in the The Scottish Mental Survey in 1947

Cohort	Start Date	Sex	Sample Size at Recruitment	Age at Recruitment (yr)	Estimated Sample Size based on last sweep/contact	Approximate age range in 2013 (yr)	Length of follow-up (yr)	Comments
The Millennium Cohort Study (MCS) – children	2000-02	MF	19,519	0	13,469	12	13	Last sweep in 2008
The Millennium Cohort Study (MCS) – families	2000-02	MF	19,244	all ages	13,287	14-69	13	Last sweep in 2008
Million Women Study	1996-2001	F	1,360,000	50 -64	1,240,000	66–80	17	Numbers include those still contactable but nearly all recruited women remained in the study dataset Sweep underway in 2013
MRC National Survey of Health and Development Cohort /1946 Birth Cohort (NSHD/1946BC)	1946	MF	5,362	0	3,116	67	67	Last sweep in 2006-11
The National Child Development Study/1958 Birth Cohort (NCDS/1958BC)	1958	MF	17,416	0	9790	55	55	Last sweep in 2013
Newcastle 85+	2006	MF	851	85	342	>91	5-7	Last sweep in 2012-13
Rural Uganda General Population Cohort (GPC)	1989	MF	10,000	All	18,000	16-100	24	This is an open cohort Last sweep 2012
Southall and Brent Revisited (SABRE)	1988	MF	4,858	40-69	2,572	63-93	25	

Cohort	Start Date	Sex	Sample Size at Recruitment	Age at Recruitment (yr)	Estimated Sample Size based on last sweep/contact	Approximate age range in 2013 (yr)	Length of follow-up (yr)	Comments
Southampton Women's Survey (SWS) – children	1998	MF	3,159	0	2,500	5-14	5-14	Last sweep 2012/13
Southampton Women's Survey (SWS) – mothers	1998	F	12,583	20-34	2,500	34-48	11-15	Only mothers of the SWS children are being followed up
Twin Early Development Study (TEDS)	1994-96	MF	13,690 families	0	21,800	19	19	Last sweep in 2012
Twins UK	1992	MF	346	49-69	9,170	16-90	21	This is an open cohort Last sweep in 2012
UK Biobank	2006-10	MF	503,316	40-69	502,713	42-75	2-7	Last sweep in 2012/13
UK Women's Cohort Study (UKWCS)	1995	F	35,372	35-69	12,453	53-87	18	Last sweep in 2002
Understanding Society	2009	MF	40,000 household 100000 individuals	All	31,000 households 78,000 individuals	0-103	4	Last sweep in 2012
West of Scotland Twenty-07 Study (Twenty-07)	1986	MF	4,510	15, 35, 55	3,174	42, 62, 82	27	Numbers include those contactable in 2012 Last sweep in 2007/8 No plans for further follow-up, but flagged for mortality
Whitehall II Study	1985	MF	10,308	35-55	7,735	63-83	28	Last sweep in 2012/13

Annex 5. Cohort Data by Consent for Re-contact and Linkage

Cohort	Consent for re-contact	Consent for linkage
11-16 and 16+ Study	✓	
1970 British Cohort Study (1970 BCS)	✓	✓
Aberdeen Children of the 1950s (ACONF)	✓	✓
Avon Longitudinal Study of Parents and Children (ALSPAC)	✓	✓
Born in Bradford (BiB)	✓	✓
Boyd Orr Cohort	✓	✓
Breakthrough Generations Study	✓	✓
British Regional Heart Study (BRHS)	✓	✓
British Women's Heart & Health Study (BWHHS)	✓	✓
Cognitive Function and Ageing Studies I (CFAS I)	✓	✓
Cognitive Function and Ageing Studies II (CFAS II)	✓	✓
Determinants of Adolescent Social well-being and Health (DASH)	✓	✓
The English Longitudinal Study of Ageing (ELSA)	✓	✓
European Prospective Investigation of Cancer Norfolk (EPIC Norfolk)	✓	✓
European Prospective Investigation of Cancer Oxford (EPIC Oxford)	✓	✓
Gemini	✓	
Growing up in Scotland (GUS)	✓	✓
Hertfordshire cohort study (HCS)	✓	✓
Lothian Birth Cohort 1936 (LBC1936)	✓	✓
Millennium Cohort Study (MCS)	✓	✓
Million Women Study	✓	✓
MRC National Survey of Health and Development Cohort / 1946 Birth Cohort (NSHD)/1946BC)	✓	✓
The National Child Development Study/ 1958 Birth Cohort (NCDS/1958BC)	✓	✓
Newcastle 85+	✓	✓
Rural Uganda General Population Cohort (GPC)	✓	✓
Southall and Brent Revisited (SABRE)	✓	✓
Southampton Women's Survey (SWS)	✓	✓
Twin Early Development Study (TEDS)	✓	✓
Twins UK	✓	✓

Cohort	Consent for re-contact	Consent for linkage
UK Biobank	✓	✓
UK Women's Cohort Study (UKWCS)	✓	✓
Understanding Society	✓	✓
West of Scotland Twenty-07 Study (Twenty-07)	✓	✓
Whitehall II Study	✓	✓

Annex 6. Cohort Data by Anthropometric and Blood Pressure Variables

Cohort	Height	Weight	Waist circumference	Hip circumference	Blood pressure
11-16 and 16+ Study	✓	✓			
1970 British Cohort Study (1970 BCS)	✓	✓			✓
Aberdeen Children of the 1950s (ACONF)	✓	✓			✓
Avon Longitudinal Study of Parents and Children (ALSPAC)	✓	✓	✓	✓	✓
Born in Bradford (BiB)	✓	✓			✓
Boyd Orr Cohort	✓	✓	✓	✓	✓
Breakthrough Generations Study	✓	✓	✓	✓	
British Regional Heart Study (BRHS)	✓	✓	✓	✓	✓
British Women's Heart & Health Study (BWHHS)	✓	✓	✓	✓	✓
Cognitive Function and Ageing Studies I (CFAS I)					
Cognitive Function and Ageing Studies II (CFAS II)					
Determinants of Adolescent Social well-being and Health (DASH)	✓	✓	✓	✓	
The English Longitudinal Study of Ageing (ELSA)	✓	✓	✓		✓
European Prospective Investigation of Cancer Norfolk (EPIC Norfolk)	✓	✓	✓	✓	✓
European Prospective Investigation of Cancer Oxford (EPIC Oxford)	✓	✓	✓	✓	✓
Gemini	✓	✓			
Growing up in Scotland (GUS)	✓	✓			
Hertfordshire cohort study (HCS)	✓	✓	✓	✓	✓
Lothian Birth Cohort 1936 (LBC1936)	✓	✓			✓

Cohort	Height	Weight	Waist circumference	Hip circumference	Blood pressure
Millennium Cohort Study (MCS)	✓	✓	✓		
Million Women Study	✓	✓	✓	✓	✓
MRC National Survey of Health and Development Cohort / 1946 Birth Cohort (NSHD)/1946BC)	✓	✓	✓	✓	✓
The National Child Development Study/ 1958 Birth Cohort (NCDS/1958BC)	✓	✓	✓	✓	✓
Newcastle 85+	✓	✓	✓	✓	✓
Rural Uganda General Population Cohort (GPC)	✓	✓	✓	✓	✓
Southall and Brent Revisited (SABRE)	✓	✓	✓	✓	✓
Southampton Women's Survey (SWS)	✓	✓	✓	✓	✓
Twin Early Development Study (TEDS)	✓	✓	✓		
Twins UK	✓	✓	✓	✓	✓
UK Biobank	✓	✓	✓	✓	✓
UK Women's Cohort Study (UKWCS)	✓	✓	✓	✓	
Understanding Society	✓	✓	✓		✓
West of Scotland Twenty-07 Study (Twenty-07)	✓	✓	✓	✓	✓
Whitehall II Study	✓	✓	✓	✓	✓

Annex 7. Cohort Data by Physical Health Variables

Cohort	Cardiovascular	Respiratory	Musculoskeletal	Hearing and Vision	Reproductive
11-16 and 16+ Study	✓	✓	✓	✓	
1970 British Cohort Study (1970 BCS)	✓	✓	✓	✓	✓
Aberdeen Children of the 1950s (ACONF)		✓			✓
Avon Longitudinal Study of Parents and Children (ALSPAC)	✓	✓	✓	✓	✓
Born in Bradford (BiB)		✓		✓	
Boyd Orr Cohort	✓	✓	✓		
Breakthrough Generations Study			✓	✓	✓
British Regional Heart Study (BRHS)	✓	✓			
British Women's Heart & Health Study (BWHHS)	✓	✓	✓	✓	✓
Cognitive Function and Ageing Studies I (CFAS I)	✓	✓		✓	
Cognitive Function and Ageing Studies II (CFAS II)	✓	✓	✓	✓	
Determinants of Adolescent Social well-being and Health (DASH)	✓	✓			
The English Longitudinal Study of Ageing (ELSA)	✓	✓	✓	✓	✓
European Prospective Investigation of Cancer Norfolk (EPIC Norfolk)	✓	✓	✓	✓	✓
European Prospective Investigation of Cancer Oxford (EPIC Oxford)	✓	✓	✓		✓
Gemini					
Growing up in Scotland (GUS)					
Hertfordshire cohort study (HCS)	✓	✓	✓		
Lothian Birth Cohort 1936 (LBC1936)	✓	✓	✓	✓	✓

Cohort	Cardiovascular	Respiratory	Musculoskeletal	Hearing and Vision	Reproductive
Millennium Cohort Study (MCS)	✓	✓	✓	✓	
Million Women Study	✓	✓	✓		✓
MRC National Survey of Health and Development Cohort / 1946 Birth Cohort (NSHD)/1946BC)	✓	✓	✓	✓	✓
The National Child Development Study/ 1958 Birth Cohort (NCDS/1958BC)	✓	✓	✓	✓	✓
Newcastle 85+	✓	✓	✓	✓	
Rural Uganda General Population Cohort (GPC)	✓	✓	✓	✓	✓
Southall and Brent Revisited (SABRE)	✓	✓	✓	✓	✓
Southampton Women's Survey (SWS)	✓	✓	✓		✓
Twin Early Development Study (TEDS)					
Twins UK	✓	✓	✓	✓	✓
UK Biobank	✓	✓	✓	✓	✓
UK Women's Cohort Study (UKWCS)					✓
Understanding Society		✓			
West of Scotland Twenty-07 Study (Twenty-07)	✓	✓	✓	✓	✓
Whitehall II Study	✓	✓	✓		

Includes self-reported illness and measurements taken by researchers

Annex 8. Cohort Data by Mental Health and Cognitive Measures

Cohort	Mental health			Cognitive Function		
	Y/N	Self-rated	Tested*	Y/N	Self-rated	Tested
11-16 and 16+ Study	✓	✓				
1970 British Cohort Study (1970 BCS)	✓	✓		✓	✓	✓
Aberdeen Children of the 1950s (ACONF)				✓		✓
Avon Longitudinal Study of Parents and Children (ALSPAC)	✓	✓		✓	✓	✓
Born in Bradford (BiB)	✓	✓				
Boyd Orr Cohort						
Breakthrough Generations Study						
British Regional Heart Study (BRHS)				✓		✓
British Women's Heart & Health Study (BWHHS)	✓	✓				
Cognitive Function and Ageing Studies I (CFAS I)	✓	✓	✓	✓	✓	✓
Cognitive Function and Ageing Studies II (CFAS II)	✓	✓	✓	✓	✓	✓
Determinants of Adolescent Social well-being and Health (DASH)	✓	✓		✓	✓	
The English Longitudinal Study of Ageing (ELSA)	✓	✓		✓	✓	✓
European Prospective Investigation of Cancer Norfolk (EPIC Norfolk)	✓	✓	✓	✓	✓	✓
European Prospective Investigation of Cancer Oxford (EPIC Oxford)						
Gemini						
Growing up in Scotland (GUS)	✓	✓		✓		✓
Hertfordshire cohort study (HCS)	✓	✓		✓		✓
Lothian Birth Cohort 1936 (LBC1936)	✓	✓		✓		✓

Cohort	Mental health			Cognitive Function		
	Y/N	Self-rated	Tested*	Y/N	Self-rated	Tested
Millennium Cohort Study (MCS)	✓	✓		✓		✓
Million Women Study	✓	✓		✓	✓	
MRC National Survey of Health and Development Cohort / 1946 Birth Cohort (NSHD)/1946BC)	✓	✓	✓	✓	✓	✓
The National Child Development Study/ 1958 Birth Cohort (NCDS/1958BC)	✓	✓		✓	✓	✓
Newcastle 85+	✓		✓	✓		✓
Rural Uganda General Population Cohort (GPC)						
Southall and Brent Revisited (SABRE)	✓	✓		✓		✓
Southampton Women's Survey (SWS)	✓	✓	✓	✓		✓
Twin Early Development Study (TEDS)	✓	✓	✓	✓		✓
Twins UK	✓	✓		✓	✓	✓
UK Biobank	✓	✓		✓		✓
UK Women's Cohort Study (UKWCS)				✓		✓
Understanding Society	✓	✓		✓		✓
West of Scotland Twenty-07 Study (Twenty-07)	✓	✓		✓		✓
Whitehall II Study	✓	✓		✓		✓

*Tested only includes a clinical diagnosis made by a qualified practitioner or verification of clinically diagnosed mental illness via health records.

Annex 9. Cohort Data by Lifestyle Variables

Cohort	Smoking	Physical activity	Dietary habits	Alcohol
11-16 and 16+ Study	✓	✓	✓	✓
1970 British Cohort Study (1970 BCS)	✓	✓	✓	✓
Aberdeen Children of the 1950s (ACONF)	✓			✓
Avon Longitudinal Study of Parents and Children (ALSPAC)	✓	✓	✓	✓
Born in Bradford (BiB)	✓	✓	✓	✓
Boyd Orr Cohort	✓	✓	✓	✓
Breakthrough Generations Study	✓	✓	✓	✓
British Regional Heart Study (BRHS)	✓	✓	✓	✓
British Women's Heart & Health Study (BWHHS)	✓	✓	✓	✓
Cognitive Function and Ageing Studies I (CFAS I)	✓	✓		✓
Cognitive Function and Ageing Studies II (CFAS II)	✓	✓	✓	✓
Determinants of Adolescent Social well-being and Health (DASH)	✓	✓	✓	✓
The English Longitudinal Study of Ageing (ELSA)	✓	✓	✓	✓
European Prospective Investigation of Cancer Norfolk (EPIC Norfolk)	✓	✓	✓	✓
European Prospective Investigation of Cancer Oxford (EPIC Oxford)	✓	✓	✓	✓
Gemini	✓	✓	✓	✓
Growing up in Scotland (GUS)	✓	✓	✓	✓
Hertfordshire cohort study (HCS)	✓	✓	✓	✓
Lothian Birth Cohort 1936 (LBC1936)	✓	✓	✓	✓

Cohort	Smoking	Physical activity	Dietary habits	Alcohol
Millennium Cohort Study (MCS)	✓	✓	✓	✓
Million Women Study	✓	✓	✓	✓
MRC National Survey of Health and Development Cohort / 1946 Birth Cohort (NSHD)/1946BC)	✓	✓	✓	✓
The National Child Development Study/ 1958 Birth Cohort (NCDS/1958BC)	✓	✓	✓	✓
Newcastle 85+	✓	✓	✓	✓
Rural Uganda General Population Cohort (GPC)	✓	✓	✓	✓
Southall and Brent Revisited (SABRE)	✓	✓	✓	✓
Southampton Women's Survey (SWS)	✓	✓	✓	✓
Twin Early Development Study (TEDS)	✓		✓	✓
Twins UK	✓	✓	✓	✓
UK Biobank	✓	✓	✓	✓
UK Women's Cohort Study (UKWCS)	✓	✓	✓	✓
Understanding Society	✓	✓	✓	✓
West of Scotland Twenty-07 Study (Twenty-07)	✓	✓	✓	✓
Whitehall II Study	✓	✓	✓	✓

Annex 10. Cohort Data by Socio-Economic Position

Cohort	Occupation	Finances*	Family circumstances	Accommodation	Education	Ethnic group	Marital status	Social support**
11-16 and 16+ Study	✓	✓	✓	✓	✓	✓		✓
1970 British Cohort Study (1970 BCS)	✓	✓	✓	✓	✓	✓	✓	✓
Aberdeen Children of the 1950s (ACONF)	✓			✓	✓		✓	
Avon Longitudinal Study of Parents and Children (ALSPAC)	✓	✓	✓	✓	✓	✓	✓	✓
Born in Bradford (BiB)	✓	✓	✓	✓	✓	✓	✓	✓
Boyd Orr Cohort	✓	✓	✓	✓	✓			
Breakthrough Generations Study	✓		✓		✓	✓	✓	
British Regional Heart Study (BRHS)	✓	✓	✓	✓	✓	✓	✓	✓
British Women's Heart & Health Study (BWHHS)	✓	✓	✓	✓	✓	✓	✓	✓
Cognitive Function and Ageing Studies I (CFAS I)	✓		✓	✓	✓	✓	✓	✓
Cognitive Function and Ageing Studies II (CFAS II)	✓		✓	✓	✓	✓	✓	✓
Determinants of Adolescent Social well-being and Health (DASH)	✓	✓	✓	✓	✓	✓	✓	
The English Longitudinal Study of Ageing (ELSA)	✓	✓	✓	✓	✓	✓	✓	✓
European Prospective Investigation of Cancer Norfolk (EPIC Norfolk)	✓				✓		✓	
European Prospective Investigation of Cancer Oxford (EPIC Oxford)	✓				✓	✓	✓	
Gemini	✓	✓	✓	✓	✓	✓	✓	
Growing up in Scotland (GUS)	✓	✓	✓	✓	✓	✓	✓	✓

Cohort	Occupation	Finances*	Family circumstances	Accommodation	Education	Ethnic group	Marital status	Social support**
Hertfordshire cohort study (HCS)	✓		✓	✓	✓		✓	✓
Lothian Birth Cohort 1936 (LBC1936)	✓		✓	✓	✓	✓	✓	✓
Millennium Cohort Study (MCS)	✓	✓	✓	✓	✓	✓	✓	✓
Million Women Study	✓		✓	✓	✓	✓	✓	✓
MRC National Survey of Health and Development Cohort /1946 Birth Cohort (NSHD)/1946BC)	✓	✓	✓	✓	✓		✓	✓
The National Child Development Study/ 1958 Birth Cohort (NCDS/1958BC)	✓	✓	✓	✓	✓	✓	✓	✓
Newcastle 85+	✓	✓	✓	✓	✓	✓		✓
Rural Uganda General Population Cohort (GPC)	✓		✓	✓	✓	✓	✓	
Southall and Brent Revisited (SABRE)	✓	✓	✓	✓	✓	✓	✓	✓
Southampton Women's Survey (SWS)	✓	✓	✓	✓	✓	✓	✓	
Twin Early Development Study (TEDS)	✓	✓	✓		✓	✓		
Twins UK	✓	✓	✓		✓	✓	✓	
UK Biobank	✓	✓		✓	✓	✓		✓
UK Women's Cohort Study (UKWCS)	✓				✓		✓	
Understanding Society	✓	✓	✓	✓	✓		✓	✓
West of Scotland Twenty-07 Study (Twenty-07)	✓	✓	✓	✓	✓	✓	✓	✓
Whitehall II Study	✓	✓	✓	✓	✓	✓	✓	✓

* Finances includes state support and benefits

**Social support includes family and community support, social networks and emotional support

Annex 11. Cohort by Data Linkage

Cohort	Health Records including registries	Administrative data	Census	Births and Deaths	Environmental
11-16 and 16+ Study					
1970 British Cohort Study (1970 BCS)		✓		✓	
Aberdeen Children of the 1950s (ACONF)	✓		✓*	✓	
Avon Longitudinal Study of Parents and Children (ALSPAC)	✓	✓		✓	✓
Born in Bradford (BiB)	✓	✓			✓
Boyd Orr Cohort	✓				
Breakthrough Generations Study	✓			✓	
British Regional Heart Study (BRHS)	✓			✓	
British Women's Heart & Health Study (BWHHS)	✓	✓	✓		✓
Cognitive Function and Ageing Studies I (CFAS I)	✓			✓	
Cognitive Function and Ageing Studies II (CFAS II)	✓			✓	
Determinants of Adolescent Social well-being and Health (DASH)	✓				
The English Longitudinal Study of Ageing (ELSA)	✓	✓		✓	
European Prospective Investigation of Cancer Norfolk (EPIC Norfolk)	✓			✓	
European Prospective Investigation of Cancer Oxford (EPIC Oxford)	✓			✓	
Gemini					
Growing up in Scotland (GUS)	✓	✓			
Hertfordshire cohort study (HCS)	✓			✓	

Cohort	Health Records including registries	Administrative data	Census	Births and Deaths	Environmental
Lothian Birth Cohort 1936 (LBC1936)	✓			✓	✓
Millennium Cohort Study (MCS)	✓	✓	✓	✓	✓
Million Women Study	✓			✓	
MRC National Survey of Health and Development Cohort /1946 Birth Cohort (NSHD)/1946BC)	✓		✓	✓	✓
The National Child Development Study/ 1958 Birth Cohort (NCDS/1958BC)		✓	✓	✓	✓
Newcastle 85+				✓	
Rural Uganda General Population Cohort (GPC)					
Southall and Brent Revisited (SABRE)	✓			✓	
Southampton Women's Survey (SWS)	✓	✓			
Twin Early Development Study (TEDS)	✓	✓	✓	✓	
Twins UK	✓	✓		✓	
UK Biobank	✓				
UK Women's Cohort Study (UKWCS)	✓			✓	
Understanding Society		✓	✓		✓
West of Scotland Twenty-07 Study (Twenty-07)				✓	✓
Whitehall II Study	✓			✓	✓

* Only linked to 1962 Census

Annex 12. Cohort Data by Biological Samples

Cohort	Blood	Urine	Saliva	Other sample (including Buccal cell, post mortem brain, placenta, hair, teeth)
11-16 and 16+ Study				
1970 British Cohort Study (1970 BCS)				
Aberdeen Children of the 1950s (ACONF)	✓*			
Avon Longitudinal Study of Parents and Children (ALSPAC)	✓	✓	✓	✓
Born in Bradford (BiB)	✓	✓	✓	
Boyd Orr Cohort	✓			
Breakthrough Generations Study	✓	✓		✓
British Regional Heart Study (BRHS)	✓	✓		
British Women's Heart & Health Study (BWHHS)	✓			
Cognitive Function and Ageing Studies I (CFAS I)	✓			✓
Cognitive Function and Ageing Studies II (CFAS II)			✓	✓
Determinants of Adolescent Social well-being and Health (DASH)	✓**		✓**	
The English Longitudinal Study of Ageing (ELSA)	✓		✓	✓
European Prospective Investigation of Cancer Norfolk (EPIC Norfolk)	✓	✓		
European Prospective Investigation of Cancer Oxford (EPIC Oxford)	✓			✓
Gemini				✓
Growing up in Scotland (GUS)				
Hertfordshire cohort study (HCS)	✓	✓		✓

Cohort	Blood	Urine	Saliva	Other sample (including Buccal cell, post mortem brain, placenta, hair, teeth)
Lothian Birth Cohort 1936 (LBC1936)	✓	✓		
Millennium Cohort Study (MCS)			✓	
Million Women Study	✓			
MRC National Survey of Health and Development Cohort /1946 Birth Cohort (NSHD)/1946BC)	✓	✓	✓	✓
The National Child Development Study/ 1958 Birth Cohort (NCDS/1958BC)	✓		✓	
Newcastle 85+	✓			
Rural Uganda General Population Cohort (GPC)	✓			
Southall and Brent Revisited (SABRE)	✓	✓		
Southampton Women's Survey (SWS)	✓	✓		✓
Twin Early Development Study (TEDS)				✓
Twins UK	✓	✓	✓	
UK Biobank	✓	✓	✓	
UK Women's Cohort Study (UKWCS)	✓		✓	✓
Understanding Society	✓			
West of Scotland Twenty-07 Study (Twenty-07)	✓		✓	
Whitehall II Study	✓	✓	✓	

* A sub-set of 576 subjects as part of the Generation Scotland study

** A sub-set of subjects as part of a feasibility study

Annex 13. Cohort Data by Omics Analysis

Cohort	Genotyped	Whole genome or exome sequencing	Epigenetic studies	Metabolomics studies
11-16 and 16+ Study				
1970 British Cohort Study (1970 BCS)				
Aberdeen Children of the 1950s (ACONF)				
Avon Longitudinal Study of Parents and Children (ALSPAC)	✓	✓	✓	✓
Born in Bradford (BiB)	✓	✓	✓	✓
Boyd Orr Cohort	✓			
Breakthrough Generations Study	✓			✓
British Regional Heart Study (BRHS)	✓			✓
British Women's Heart & Health Study (BWHHS)	✓	✓		✓
Cognitive Function and Ageing Studies I (CFAS I)	✓		✓	
Cognitive Function and Ageing Studies II (CFAS II)			✓	
Determinants of Adolescent Social well-being and Health (DASH)				
The English Longitudinal Study of Ageing (ELSA)	✓	✓		
European Prospective Investigation of Cancer Norfolk (EPIC Norfolk)	✓	✓		
European Prospective Investigation of Cancer Oxford (EPIC Oxford)	✓		✓	✓
Gemini	✓			
Growing up in Scotland (GUS)				
Hertfordshire cohort study (HCS)			✓	✓
Lothian Birth Cohort 1936 (LBC1936)	✓	✓	✓	

Cohort	Genotyped	Whole genome or exome sequencing	Epigenetic studies	Metabolomics studies
Millennium Cohort Study (MCS)				
Million Women Study	✓			
MRC National Survey of Health and Development Cohort /1946 Birth Cohort (NSHD)/1946BC)	✓		✓	
The National Child Development Study/1958 Birth Cohort (NCDS/1958BC)	✓	✓	✓	
Newcastle 85+	✓		✓	
Rural Uganda General Population Cohort (GPC)	✓	✓		
Southall and Brent Revisited (SABRE)			✓	✓
Southampton Women's Survey (SWS)	✓		✓	
Twin Early Development Study (TEDS)	✓	✓	✓	
Twins UK	✓	✓	✓	✓
UK Biobank				
UK Women's Cohort Study (UKWCS)	✓			
Understanding Society	✓			
West of Scotland Twenty-07 Study (Twenty-07)				
Whitehall II Study	✓			✓

Annex 14. Cohort Workshop Agenda and Attendees

MRC Population Cohort Strategy Workshop
6th March 2013, 10am – 4:30pm

AGENDA

09:30am	Coffee and Registration
10:00am	Welcome and Introduction – Prof Jill Pell
10:10am	Cohorts data analysis – Prof Jill Pell and Prof Hazel Inskip
10:45am	Theme 1: Beyond Epidemiology – Prof Frank Kelly
10:50am	Prof Anne Ferguson-Smith
11:00am	Dr Jules Griffin
11:10am	Dr Anders Malarstig
11:20am	Discussion
11:50am	Coffee and Tea Break
12:00pm	Theme 2: Cross-Cohort Collaborations – Prof Andrew Steptoe
12:05pm	Prof Diana Kuh
12:15pm	Prof James Nazroo
12:25pm	Prof Mika Kivimaki
12:35pm	Discussion
1:05pm	Lunch
1:45pm	Theme 3: Impact on Public Health Policy – Prof Hazel Inskip
1:50pm	Prof Mike Kelly
2:00pm	Prof Andrea Manca
2:10pm	Prof Eileen Kaner
2:20pm	Discussion
2:50pm	Theme 4: Portfolio Balance – Prof Jill Pell
2:55pm	Prof Albert Hoffman
3:05pm	Prof George Davey-Smith
3:15pm	Discussion
3:45pm	Coffee and Tea Break
4:00pm	General Discussion and Recommendations
4:30pm	Close of meeting

List of Attendees

Surname	First Name	Salutation	Institute
Aitman	Tim	Professor	Professor of Clinical and Molecular Genetics, MRC Clinical Sciences Centre
Akinwale	Bola	Ms	Principal Research Officer, Department for Work and Pensions
Brayne	Carol	Professor	Professor of Public Health Medicine, University of Cambridge
Brennan	Alan	Professor	Professor of Health Economics and Decision Modelling, University of Sheffield
Buck	Nick	Professor	Director of the Understanding Society Project, University of Essex
Capewell	Simon	Professor	Professor of Clinical Epidemiology, University of Liverpool
Chaturvedi	Nishi	Professor	Professor of Clinical Epidemiology, Imperial College London
Conner	Rachel	Ms	Principal Research Analyst, Social Science, Health Improvement Analysis Team, Department of Health
Cuthill	Vanessa	Miss	Head of Longitudinal Studies, ESRC
Danesh	John	Professor	Professor of Epidemiology and Medicine, University of Cambridge
Davey-Smith	George	Professor	Professor of Clinical Epidemiology, University of Bristol
Deary	Ian	Professor	Director Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh
Dezateux	Carol	Professor	Professor of Paediatric Epidemiology, University College London
Elliott	Jane	Professor	Director of the ESRC Resource Centre, Institute of Education, University of London
Ferguson-Smith	Anne	Professor	Professor of Developmental Genetics, University of Cambridge
Gallacher	John	Dr	Programme Lead: Healthy Ageing, University of Cardiff
Goodman	Alissa	Professor	Professor of Economics, University College London
Gray	Linsay	Dr	Senior Investigator Scientist, MRC Social and Public Health Sciences Unit
Griffin	Jules	Dr	Head of Lipid Profiling and Signalling, MRC Human Nutrition Research
Gupta	Sunjai	Dr	Deputy Director: Head of Health Related Behaviour and Senior Advisor to the Public Health Policy and Strategy Unit, Department of Health
Hayes	Richard	Professor	Professor of Epidemiology and International Health, London School of Hygiene and Tropical Medicine
Hingorani	Aroon	Professor	Chair of Genetic Epidemiology, University College London
Hofman	Albert	Professor	Professor of Epidemiology, Harvard University
Hunt	Kate	Professor	Head of Gender and Health, Social and Public Health Sciences Unit
Inskip	Hazel	Professor	Professor of Statistical Epidemiology, MRC Lifecourse Epidemiology Unit
Jacobsen	Sten Eirik	Professor	Professor of Stem Cell Biology, University of Oxford

Surname	First Name	Salutation	Institute
Jadeja	Nidhee	Dr	Science Portfolio Adviser in Pathogens, Immunology and Population Health, Wellcome Trust
Jebb	Susan	Dr	Head of Diet and Population Health, MRC Human Nutrition Unit
Jimenez	Michelle	Dr	Senior Portfolio Developer, Wellcome Trust
Kaner	Eileen	Professor	Institute Director, Institute of Health & Society, Newcastle University
Kaye	Paul	Professor	Professor of Immunology, University of York
Kee	Frank	Professor	Clinical Professor and Director of UKCRC Centre of Excellence for Public Health Research (NI), Centre for Public Health, University of Belfast
Kelly	Frank	Professor	Professor of Environmental Health, Kings College, London
Kelly	Mike	Professor	Director of the Centre of Public Health Excellence, NICE
Kivimaki	Mika	Professor	Professor of Epidemiology & Public Health, University College London
Kuh	Diana	Professor	Director of the MRC Unit for Lifelong Health and Ageing, University College London
Malarstig	Anders	Dr	Director of Human Genetics, Pfizer
Manca	Andrea	Professor	Professor of Health Economics, University of York
McNamara	Joe	Dr	Head of Population and Systems Medicine, MRC
Moody	Catherine	Dr	Programme Manager, MRC
Mulkeen	Declan	Dr	Chief Science Officer, MRC
Nazroo	James	Professor	Professor of Sociology, University of Manchester
Newland	Claire	Dr	Programme Manager, MRC
O'Donnell	Valerie	Professor	Professor of Biochemistry, University of Cardiff
Pell	Jill	Professor	Professor of Public Health, University of Glasgow
Phanwises	Jess	Mrs	Panel Manager, MRC
Ramsay	Mary	Dr	Head of Immunisation, Health Protection Agency
Reddington	Fiona	Dr	Head of Clinical and Population Research Funding, Cancer Research UK
Roddam	Andrew	Dr	International Head, Center for Observational Research, Amgen
Rodrigues	Laura	Professor	Professor of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine
Rooney	Lauren	Miss	RPG Administrator, MRC
Rossor	Martin	Professor	Professor of Neurology, University College London
Sattar	Naveed	Professor	Professor of Metabolic Medicine, University of Glasgow
Steptoe	Andrew	Professor	Professor of Epidemiology and Public Health, University College London
Sudlow	Cathie	Dr	Clinical Senior Lecturer, University of Edinburgh

Surname	First Name	Salutation	Institute
Vale	Luke	Professor	Health Foundation Chair in Health Economics, Deputy Director, University of Newcastle
Valentine	Janet	Dr	Head of Public Health and Ageing, MRC
Wareham	Nick	Professor	Director of the MRC Epidemiology Unit, MRC Epidemiology Unit
White	Michael	Professor	Professor of Systems Biology, University of Manchester
Witt	Stephen	Dr	Team Leader, Department for Education
Zoubiane	Ghada	Dr	Programme Manager for Public Health Partnerships, MRC

MEDICAL RESEARCH COUNCIL (Swindon office)
2nd Floor David Phillips Building
Polaris House
North Star Avenue
Swindon SN2 1FL

MEDICAL RESEARCH COUNCIL (London office)
14th Floor
One Kemble Street
London WC2B 4AN

www.mrc.ac.uk

