



## **Our Future Health Protocol**

**Version 4.0**  
**15 DEC 2022**

[REDACTED]  
Chief Investigator & Chief Executive Officer

## Co-Principal Investigators

[REDACTED]  
Chief Medical Officer

[REDACTED]  
Executive Director of Epidemiology

[REDACTED]  
Chief Behavioural Scientist

[REDACTED]  
Chief Science Adviser

[REDACTED]  
Chief Medical Adviser

# Our Future Health Protocol

## Contents

<b>1. Background .....</b>	<b>6</b>
1.1. Overall aims.....	6
1.2. Genesis of Our Future Health .....	7
1.3. Initial planning of Our Future Health .....	7
<b>2. Scientific rationale.....</b>	<b>8</b>
2.1. Prospective study design.....	8
2.2. Translational research.....	9
2.3. Sample frame .....	10
2.3.1. Sample frame ambition.....	11
2.3.2. Common and rare phenotypes and diseases .....	12
2.3.3. Subpopulations .....	14
2.3.4. Pre-diagnostic bloods.....	16
2.4. Participant questionnaires .....	17
2.5. Physical measurements.....	18
2.6. Data linkages.....	19
2.7. Biological samples.....	19
<b>3. Recruitment.....</b>	<b>21</b>
3.1. Overall strategy .....	21
3.2. Engagement .....	21
3.2.1. Engagement through partnerships.....	22
3.2.2. Publicity campaigns and communications .....	23
3.3. Invitation.....	23
3.3.1. Invitation methods .....	23
3.3.2. Feasibility of invitation strategy.....	24
3.3.3. Invitation development.....	25
3.4. Reimbursement.....	25
3.5. Consent .....	27
3.5.1. Participants without capacity to consent.....	28
3.6. Baseline questionnaire .....	29
3.7. Short baseline questionnaire.....	30
3.8. Phlebotomy and physical measurements .....	31
3.8.1. Community routes.....	31
3.8.2. NHS blood donor route .....	32
3.8.3. NHS routes .....	33
<b>4. Blood sample logistics, processing and genotyping .....</b>	<b>35</b>
4.1. Repeat blood samples .....	35

5.	Data linkage.....	37
5.1.	NHS data linkages.....	37
5.2.	Other linked data sets.....	37
6.	Feedback of health-related information to participants .....	38
6.1.	Background to issues around feedback .....	38
6.2.	Rationale for participant feedback .....	38
6.3.	Framework for participant feedback .....	39
6.4.	Examples of participant feedback .....	39
6.5.	Assessment of outcomes of participant health-related feedback .....	40
7.	Re-contacting participants for future research .....	42
7.1.	Background on re-contact .....	42
7.2.	Rationale for re-contacting participants.....	42
7.3.	Planned approach for re-contact.....	43
7.3.1.	Inviting participants to third party research studies.....	43
8.	Participant communication and withdrawal.....	44
9.	Data access .....	45
9.1.	General approach and principles .....	45
9.1.1.	Key principles guiding access.....	45
9.2.	Researchers and organisations .....	46
9.3.	Dissemination of results .....	46
9.4.	Access agreements and fees .....	46
9.5.	Trusted research environments.....	47
9.6.	Data to be shared for research .....	49
9.7.	Accreditation of TREs.....	49
10.	Digital data and platform .....	50
10.1.	Overview .....	50
10.2.	Systems architecture.....	50
10.2.1.	Participant platforms .....	50
10.2.2.	Data platforms .....	52
10.2.3.	Researcher platforms .....	52
10.3.	Security, governance and operations systems.....	53
10.4.	Cloud infrastructure .....	53
10.5.	Primary data flows .....	53
10.6.	Consent and participant withdrawal .....	54
10.7.	Data linkage .....	55
10.8.	Security and resilience .....	55
11.	Public and participant involvement and engagement (PPIE) .....	57
11.1.	Rationale for PPIE.....	57
11.2.	PPIE strategy .....	57
11.3.	Summary of PPIE work .....	58
11.4.	Development of public-facing materials.....	58
11.5.	Participant-reported experiences.....	59

12. Governance.....	60
---------------------	----

# 1. Background

## 1.1. Overall aims

The overarching objective of Our Future Health is to help people live healthier lives for longer through better prevention, earlier detection and improved treatment of diseases. The Our Future Health research programme will speed up the discovery of new methods of early disease detection, and the evaluation of new diagnostic tools, to help identify and treat diseases early when outcomes are usually better.

To achieve these objectives, we will recruit up to 5 million adults from across the UK to create a diverse and inclusive cohort of people who have consented to participate in the research. In addition to being asked for permission to link their personal health data to other health-relevant data, participants will be asked to provide biological samples and complete questionnaires on recruitment; agree to re-contact for ongoing biological sampling and questionnaires and consider taking part in further research studies; and agree to being offered personal health information arising from the research. The specific aims are presented below.

- **Specific Aim 1:** Build a resource linking multiple sources of health and health-relevant information, including genetic data, on millions of people in the UK, to facilitate basic discovery research by academic and commercial researchers on early indicators of disease
- **Specific Aim 2:** Analyse the data in the resource to estimate personal disease risk information for participants, based on genetic and non-genetic information, and offer this estimated personal health information to participants
- **Specific Aim 3:** Re-contact sub-groups of participants generally for additional samples, and additional data collection including linkage to digital data sources
- **Specific Aim 4:** Re-contact participants on a risk-stratified basis (i.e. recall-by-genotype/phenotype or sociodemographic characteristics) over time specifically to enable secondary studies by academic and commercial researchers that are greatly enhanced by being able to identify highly enriched sub-populations/sub-cohorts of participants

Building this large resource with linkage, feedback and re-contact will facilitate a new generation of discovery and translational research that will advance the development and testing of early diagnostic technologies and preventive (or ‘personalised precision health’) interventions.

The UK is uniquely placed to deliver this programme. We have an exceptional track record in population research, and many outstanding research groups. Our diverse (ethnically/socioeconomically) population is willing to take part in research. Our government is committed to levelling up the major inequalities in health outcomes seen across the population. The NHS and our comprehensive disease registration systems provide a mechanism for invitation, recruitment and follow-up at an unprecedented size and scale. In addition, a programme of this scale and nature is made possible by the major advances in digital technologies over the past decade. Furthermore, the proportion of the UK public connected to data and devices (‘digital health’) is substantial and rapidly growing.

The Our Future Health research programme is intended to be both a prospective observational cohort and a platform for future discovery and translational research studies with consent for return of results, risk-stratification, and re-contact. Our Future Health will build on our national

strengths and complement existing prospective cohort resources and translational research efforts in the UK.

## **1.2. Genesis of Our Future Health**

The initial idea to set up a very large cohort in the UK to improve early detection of chronic diseases was first discussed in 2016, when several medical research charities (including Cancer Research UK), the Medical Research Council, and leading public health practitioners and academics started exploring the rationale. The concept was described in the 2017 Life Sciences Industrial Strategy and discussions progressed on possible government funding.

The proposal to establish a 5-million strong volunteer cohort enabling research intended to improve the early detection of chronic diseases was set out in the Accelerating Detection of Disease challenge of the government's Industrial Strategy Challenge Fund (ISCF).

An investment of £79 million was allocated from the ISCF by UK Research and Innovation to test the feasibility of, and establish, the Our Future Health research programme. This was expected to be matched by funding of at least £160 million from industry and charity partners who would work in partnership with Our Future Health to design and deliver the programme.<sup>1</sup>

## **1.3. Initial planning of Our Future Health**

In 2018, a Science Task & Finish Group was convened to make recommendations on the scientific design / scientific protocol of Our Future Health (Successive Chairs: [REDACTED]). The Science Task & Finish Group concluded its work in early 2020.

An Ethics and Feedback Advisory Group (EFAG) was established in Sept 2019 to provide strategic advice on the development of ethical guidelines and principles for Our Future Health, and to develop an Ethics and Governance Framework to guide its operations (Successive Chairs: [REDACTED]). The first draft of this Framework (completed in Oct 2020) provides advice to the Board and Executive, and will be publicly available for funders, partners, researchers, participants and the general public. Building on the work of EFAG, an Ethics Advisory Board has been established as part of the long-term governance of the cohort and is responsible for monitoring the implementation of the Ethics & Governance Framework, and for reviewing and updating it as appropriate.

An Industry Advisory Group and an NHS Advisory Group were established early on in the development of the programme. Further details regarding governance structures being put in place can be found at the end of this document.

In Sept 2019, a not-for-profit company was established to run Our Future Health. The company, initially named Early Disease Detection Research Project UK (EDDRP UK), was registered as a charity in May 2020. The Executive was established in April-May 2020 with [REDACTED] as the CEO. The programme was subsequently renamed from the placeholder EDDRP UK to Our Future Health in 2021.

---

<sup>1</sup> <https://www.ukri.org/innovation/industrial-strategy-challenge-fund/accelerating-detection-of-disease/>

## 2. Scientific rationale

### 2.1. Prospective study design

Prospective observational cohort studies are valuable because these real-world studies facilitate the identification of biomarkers and causative factors that contribute to future disease.<sup>2</sup> Prospective study designs are less prone to bias than case-control study designs.

Existing large prospective studies in the UK and Europe include:

- **UK Biobank:** 500,000 participants, UK<sup>3</sup>
- **Million Women Study:** 1.3 million participants, England and Scotland<sup>4</sup>
- **Whitehall I and II Studies:** 17,500 and 10,000 participants respectively, England<sup>5,6</sup>
- **ALSPAC:** Avon Longitudinal Study of Parents and Children, 14,000 families, England<sup>7</sup>
- **Understanding Society:** The UK Longitudinal Household Study, 40,000 households, UK<sup>8</sup>
- **GLAD Study:** Genetics Links to Anxiety and Depression, n=22,000 to date (aim is 40,000), UK<sup>9</sup>
- **EPIC:** European Prospective Investigation of Cancer and Nutrition, 500,000 participants, UK and 9 countries in Europe<sup>10</sup>
- **Genes and Health:** 50,000 participants (aim is 100,000), UK<sup>11</sup>

There are a number of large or influential prospective studies in the US, including:

- **All of Us** (currently have recruited ~350,000 participants; aiming to achieve a final sample size of 1 million, US)<sup>12</sup>
- **Million Veteran Programme** (1 million participants, US)<sup>13</sup>
- **Nurses Health Study** (275,000 participants, US)<sup>14</sup>
- **Framingham Heart Study** (originally 5,000 participants, US)<sup>15</sup>

Prospective studies with East Asian and Hispanic populations have also been set up, including the following which are collaborations with investigators at Oxford University:

- **Kadoorie Study** (500,000 participants, China)<sup>16</sup>
- **Mexico City Prospective Study** (150,000 participants, Mexico)<sup>17</sup>

---

<sup>2</sup> Manolio et al (*Nature Reviews Genetics*, 2006) Genes, environment and the value of prospective cohort studies. <https://pubmed.ncbi.nlm.nih.gov/16983377/>

<sup>3</sup> <https://www.ukbiobank.ac.uk/>

<sup>4</sup> <http://www.millionwomenstudy.org/introduction/>

<sup>5</sup> van Rossum et al (*J Epidemiol Community Health*, 2000) Employment grade differences in cause specific mortality. A 25 year follow up of civil servants from the first Whitehall study.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1731642/>

<sup>6</sup> <https://academic.oup.com/ije/article/34/2/251/746997>

<sup>7</sup> <http://www.bristol.ac.uk/alspac/about/>

<sup>8</sup> <https://www.understandingsociety.ac.uk/>

<sup>9</sup> <https://gladstudy.org.uk/about/>

<sup>10</sup> <https://epic.iarc.fr/>

<sup>11</sup> <https://www.genesandhealth.org/>

<sup>12</sup> <https://allofus.nih.gov/>

<sup>13</sup> <https://www.research.va.gov/mvp/>

<sup>14</sup> <https://www.nurseshealthstudy.org/>

<sup>15</sup> <https://framinghamheartstudy.org/>

<sup>16</sup> <https://www.ckbiobank.org/site/>

<sup>17</sup> <https://www.ctsu.ox.ac.uk/research/prospective-blood-based-study-of-150-000-individuals-in-mexico>



Other listings include the International HundredK+ Cohorts Consortium (IHCC; [www.ihccglobal.org](http://www.ihccglobal.org)) and the NCI Cohort Consortium (<https://epi.grants.cancer.gov/cohort-consortium>).

Prospective cohort studies provide important insights into disease aetiology, however they have not traditionally been designed to provide insights on whether or how these basic discoveries can be translated into actual health benefits for individuals and societies. For this, translational research is needed.

## 2.2. Translational research

Translational research studies are valuable because they aim to establish whether and how basic discoveries about disease aetiology can be translated into positive outcomes for populations.<sup>18</sup> Translational research has been defined as having four phases (T1-T4):<sup>19</sup>

- T1 involves processes that bring ideas from basic research through early testing in humans
- T2 involves the establishment of effectiveness in humans and clinical guidelines
- T3 primarily focuses on implementation and dissemination research  
Implementation research has been defined as the scientific inquiry into questions concerning implementation—the act of carrying an intention into effect, which in health research can be policies, programmes, or individual practices (collectively called interventions); the intent is to understand what, why, and how interventions work in “real world” settings and to test approaches to improve them<sup>20</sup>
- T4 focuses on outcomes and effectiveness in populations

Very recent examples of translational research include studies to develop and assess the outcomes and effectiveness of COVID19 vaccines.<sup>21</sup>

In the US, the NIH-organised and funded Electronic Medical Records and Genomics (eMERGE) Network consortium provides an example of discovery and translational research at a national scale. The eMERGE Network, founded in 2007, brings together US medical research institutions and researchers with a wide range of expertise in genomics, statistics, ethics, informatics, and clinical medicine. The primary goal of eMERGE is to develop, disseminate, and apply approaches to research that combine biorepositories with electronic medical record systems for genomic discovery and genomic medicine implementation research. In addition, the consortium includes a focus on social and ethical issues such as privacy, confidentiality, and interactions with the broader community.<sup>22</sup>

---

<sup>18</sup> Woolf SH. The Meaning of Translational Research and Why It Matters. JAMA. 2008;299(2):211–213. doi:10.1001/jama.2007.26 [https://jamanetwork.com/journals/jama/article-abstract/1149350?casa\\_token=udidtTeK55sAAAAA:hLwr1Fm0mlBmfbkjNrxXQCu3leSAYdhQo\\_XUZUMgPB7aDbwKvps80SSOqb6Ca86JYajafw7MEA](https://jamanetwork.com/journals/jama/article-abstract/1149350?casa_token=udidtTeK55sAAAAA:hLwr1Fm0mlBmfbkjNrxXQCu3leSAYdhQo_XUZUMgPB7aDbwKvps80SSOqb6Ca86JYajafw7MEA)

<sup>19</sup> Fort, D., Herr, T., Shaw, P., Gutzman, K., & Starren, J. (2017). Mapping the evolving definitions of translational research. Journal of Clinical and Translational Science, 1(1), 60–66. doi:10.1017/cts.2016.10 <https://doi.org/10.1017/cts.2016.10>

<sup>20</sup> Peters, Taghreed, Olakunle, Akua, Nhan. Implementation research: what it is and how to do it BMJ 2013; 347:f6753 <https://www.bmj.com/content/347/bmj.f6753.full>

<sup>21</sup> Sharpe et al (2020) The early landscape of coronavirus disease 2019 vaccine development in the UK and rest of the world <https://onlinelibrary.wiley.com/doi/full/10.1111/imm.13222>

<sup>22</sup> <https://www.genome.gov/Funded-Programs-Projects/Electronic-Medical-Records-and-Genomics-Network-eMERGE>

In the UK, examples of individual translational research studies that are embedded within public health programmes at a national scale include:

- A study examining uptake of colorectal cancer screening over three invitation rounds in the NHS Bowel Cancer Screening Programme among individuals aged 60-64yrs (n=62,000)<sup>23</sup>
- A randomised controlled trial (UK Age trial) investigating the effect of mammogram breast screening from age 40 years on breast cancer mortality involving 23 breast screening units across Great Britain (n=161,000)<sup>24</sup>
- A study of the effectiveness of NHS Health Check programme at reducing cardiovascular disease risk among patient aged 40-74 years after one year (n=3,172)<sup>25</sup>

The intent is that Our Future Health research programme will provide both a:

- (1) **Prospective observational dataset** for basic science / epidemiological, discovery and aetiological research e.g. on the causes and early signs of disease; and
- (2) **Translational research platform** comprising a cohort of people who can be re-contacted for translational/implementation research to develop and test new diagnostic technologies, prevention strategies and treatments.

Our ambition is to recruit an ethnically and socioeconomically diverse population. We expand further on this important point in the sample frame section below.

### 2.3. Sample frame

The ambition of Our Future Health is to recruit 5 million participants each of whom will provide consent, complete a baseline questionnaire, donate a blood sample and be eligible for linkage to health-related data. The 5 million number will provide a UK prospective cohort in which, with sufficient statistical precision, it will be possible to study:

- Both common and rare phenotypes and diseases
- Subpopulations based on ethnicity, index of multiple deprivation, geography, risk stratification, genotypes, and precursor conditions
- Statistical interactions between genotypes and environmental factors in relation to disease
- Pre-diagnostic/pre-interventional bloods for studies of subpopulations with specific diseases/phenotypes
- Sub-populations based on their disease-specific genotypes or polygenic risk scores
- Populations invited and consented to ancillary studies.

---

<sup>23</sup> Lo SH, Halloran S, Snowball J, et al Colorectal cancer screening uptake over three biennial invitation rounds in the English bowel cancer screening programme. *Gut* 2015;64:282-291.  
<https://gut.bmj.com/content/64/2/282>

<sup>24</sup> Duffy et al (*Lancet Oncology*, 2020) Effect of mammographic screening from age 40 years on breast cancer mortality (UK Age trial): final results of a randomised, controlled trial.  
<https://www.sciencedirect.com/science/article/pii/S1470204520303983>

<sup>25</sup> Artac et al (2013) Effectiveness of a national cardiovascular disease risk assessment program (NHS Health Check): Results after one year.  
[https://www.sciencedirect.com/science/article/pii/S0091743513001473?casa\\_token=2N7-pXNXZBgAAAAA:3mJYFPV7jt7dbdA4LdvD4HNEwI8579DrVQzivVRuplw7wvr0oSRR-JYspHbmSa6QQOr8t3nC2w](https://www.sciencedirect.com/science/article/pii/S0091743513001473?casa_token=2N7-pXNXZBgAAAAA:3mJYFPV7jt7dbdA4LdvD4HNEwI8579DrVQzivVRuplw7wvr0oSRR-JYspHbmSa6QQOr8t3nC2w)

After describing the sample frame ambition, the sections that follow provide additional details of the above novel aspects of Our Future Health that are facilitated by the large sample size and the ability to recontact participants.

### 2.3.1. Sample frame ambition

The 5 million participants we aspire to recruit into Our Future Health will be reflective of the UK population according to the most recent census data available. To inform such, we have drawn on 2011 census data of population counts by age, sex, ethnicity, and index of multiple deprivation for England and Wales, Scotland, and Northern Ireland. We will update our sample frame with census 2021/22 when they become available in late 2022 and 2023.

**Table 1. Sample frame ambition**

<i>Sex &gt;</i>	<b>Female</b>				<b>Male</b>			
<i>Ethnicity &gt;</i>	<b>White</b>	<b>Black</b>	<b>Asian</b>	<b>Mixed/Other</b>	<b>White</b>	<b>Black</b>	<b>Asian</b>	<b>Mixed/Other</b>
<i>Age &gt;</i>								
<b>18–19</b>	68,090	3,360	6,755	3,859	69,687	3,308	7,068	3,968
<b>20–29</b>	354,596	16,074	45,872	16,666	353,073	14,744	47,995	18,184
<b>30–39</b>	347,493	17,760	44,405	12,902	343,187	15,893	44,993	14,818
<b>40–49</b>	416,477	18,376	28,215	9,369	406,695	16,686	27,690	10,168
<b>50–59</b>	356,575	9,130	19,971	4,972	350,377	8,091	19,132	5,337
<b>60–69</b>	331,220	4,114	11,255	2,609	317,787	2,982	9,516	2,639
<b>70–79</b>	229,045	3,419	6,365	1,433	195,353	3,075	6,536	1,384
<b>80–84</b>	85,797	799	1,479	421	59,129	702	1,332	372
<b>85+</b>	92,450	442	953	346	43,752	358	684	242
<b>Totals</b>	<b>2,281,742</b>	<b>73,474</b>	<b>165,270</b>	<b>52,577</b>	<b>2,139,040</b>	<b>65,840</b>	<b>164,946</b>	<b>57,110</b>
<b>%</b>	<b>45.6%</b>	<b>1.5%</b>	<b>3.3%</b>	<b>1.1%</b>	<b>42.8%</b>	<b>1.3%</b>	<b>3.3%</b>	<b>1.1%</b>

This sample frame ambition is reflective of the UK population in terms of age, sex, and ethnicity. It will provide large numbers of participants of the primary ethnic minority groups resident in the UK – Indian, Pakistani, Bangladeshi, Chinese, Black African, Black Caribbean, Arab and Mixed – populations that have been underrepresented in health research. Combined with a proportional representation of participants across the nations, this sample frame will provide for a diverse cohort that will be amenable to a variety of studies that have not been possible in a UK prospective cohort before.

Achieving a sample that is reflective of the UK population is an overarching aim of the Our Future Health programme for both ethical and scientific reasons. Ethically, it is important that we make substantial efforts to make participation in Our Future Health equitable and accessible to people regardless of their socioeconomic position, disability, physical health, mental health, sex, age, and ethnicity. Scientifically, it is important that the participants in Our Future Health are sufficiently diverse to facilitate a range of discovery and translational research the resource is intended to support. Related to this, the concept of “representativeness” has been debated at length in the

epidemiologic literature<sup>26,27</sup>. We recognise that our participant sample is unlikely to be fully representative of the UK population in a large range of demographics and risk factors that extend beyond those shown in Table 1. However, the important aspects of representativeness and selection biases are specific to any hypothesis being tested and the external population to which an inference is to be made<sup>28</sup>. By ensuring we recruit a diverse population, we will provide a resource that is amenable to a large range of hypotheses and potential inferences to improve the health of the UK population.

### 2.3.2. Common and rare phenotypes and diseases

The size of this sample frame ambition will provide the ability to prospectively assess, with strong statistical precision, a wide range of common and rare phenotypes and diseases in a UK population. In addition, it will provide large numbers of prevalent diseases for retrospective, statistically powered case-control and case-cohort studies.

We have estimated incident diagnoses of disease that would accrue in Our Future Health in the initial 2.5-years of follow-up using various population/subpopulation sizes (**Appendix A**), demonstrating the immediacy of the impact that this programme will have on health research.

To interpret the advantages of these estimated incident diagnoses that may accrue in the population that comprises the Our Future Health cohort, we have also calculated minimal detectable odds ratios for aetiologic studies using ranges of case numbers, alpha values (critical p values), and exposure prevalence (**Table 2**). Note that the colour shading of all tables in these sample frame ambition sections indicates 0 cases (pure green), 5,000 cases (yellow), and 10,000 or more cases (red), a scale based on aetiologic odds ratios of ~1.5 for mid-range exposure prevalence and mid-range alphas.

---

<sup>26</sup> Nohr, E. A. and J. Olsen (2013). "Commentary: Epidemiologists have debated representativeness for more than 40 years--has the time come to move on?" *International Journal of Epidemiology* 42(4): 1016-1017.

<sup>27</sup> Schooling, C. M. and H. E. Jones (2014). "Is representativeness the right question?" *International Journal of Epidemiology* 43(2): 631-632.

<sup>28</sup> Huang, J. Y. (2021). "Representativeness Is Not Representative: Addressing Major Inferential Threats in the UK Biobank and Other Big Data Repositories." *Epidemiology* 32(2): 189-193.

**Table 2. Minimal Detectable Odds Ratios by Case Count, Exposure Prevalence, and Critical p-value**

Exposure Prevalence <sup>a</sup>	Critical p-value	Minimum detectable odds ratio, by number of cases <sup>b</sup>						
		500	1,000	2,500	5,000	7,500	10,000	25,000
0.5	0.05	1.28	1.19	1.12	1.08	1.07	1.06	1.04
0.5	0.00005	1.62	1.40	1.24	1.16	1.13	1.11	1.07
0.5	0.00000005	1.89	1.56	1.32	1.22	1.17	1.15	1.09
0.25	0.05	1.31	1.22	1.13	1.09	1.08	1.07	1.04
0.25	0.00005	1.65	1.44	1.26	1.18	1.15	1.13	1.08
0.25	0.00000005	1.90	1.59	1.35	1.24	1.19	1.17	1.10
0.1	0.05	1.45	1.31	1.19	1.13	1.11	1.09	1.06
0.1	0.00005	1.92	1.62	1.38	1.26	1.21	1.18	1.11
0.1	0.00000005	2.25	1.83	1.50	1.35	1.28	1.24	1.15
0.025	0.05	1.88	1.60	1.37	1.26	1.21	1.18	1.11
0.025	0.00005	2.83	2.22	1.74	1.51	1.41	1.35	1.22
0.025	0.00000005	3.53	2.66	1.98	1.67	1.54	1.47	1.29
0.01	0.05	2.43	1.97	1.59	1.41	1.33	1.29	1.18
0.01	0.00005	4.08	3.02	2.19	1.82	1.66	1.56	1.35
0.01	0.00000005	5.35	3.78	2.61	2.09	1.87	1.75	1.46
0.0025	0.05	4.21	3.11	2.25	1.85	1.69	1.59	1.36
0.0025	0.00005	8.60	5.70	3.63	2.75	2.39	2.18	1.72
0.0025	0.00000005	12.32	7.76	4.65	3.38	2.88	2.59	1.95
0.001	0.05	6.80	4.67	3.10	2.41	2.12	1.96	1.59
0.001	0.00005	16.14	9.85	5.67	4.00	3.35	2.98	2.18
0.001	0.00000005	24.59	14.28	7.70	5.20	4.24	3.72	2.59

<sup>a</sup> Exposure prevalence among controls.

<sup>b</sup> Calculated at 80% power assuming 4 controls per case.

A variety of scenarios can be deduced from this flexible table set up. For example, the advantages of having a population of 1 million can be clearly seen for diagnoses such as transient ischaemic attack and ischaemic stroke, irritable bowel syndrome, cholecystitis, tinnitus, diabetic eye disease, uterovaginal prolapse, postmenopausal bleeding, septicaemia, migraine, peripheral neuropathy, dementia, and rosacea, all of which accrue 4,000–6,000 cases in this short term period of follow-up – a threshold that provides strong statistical precision for testing aetiologic hypotheses. An increase to 2 million participants sees many other diagnoses surpass a 4,000-case threshold in this short period of follow-up, including lung and bowel cancers, stroke, non-rheumatic mitral valve disorder, anal fissure, Barrett's oesophagus, blindness, obstructive and reflux uropathy, neuropathic bladder, agranulocytosis, spinal stenosis, rheumatoid arthritis, fibromatosis, polymyalgia rheumatica, chronic sinusitis, sleep apnoea, and pulmonary collapse. The diseases that can be researched with strong statistical precision is obviously increased further as the number of participants and thus case numbers increase. At 5 million participants, many rarer diseases accrue sufficient numbers to provide for strong statistical precision for a range of hypotheses to be investigated. These tables underscore the benefits of progressing towards and reaching 5 million participants in being able to study both common and rare phenotypes and diseases. They also underscore the unique opportunities that Our Future Health will provide to the world research community.

Another important use of the Our Future Health cohort will be in developing and validating predictive models of health and disease status. This includes polygenic/integrative risk scores and biomarkers, which will be generated and returned to consenting participants as a primary objective of the programme. However, with the rapid growth and development of machine learning methods, in step with advances in computing power, there will be a broad interest in using the cohort to develop new predictive models of different types. The case numbers shown in Appendix A demonstrate the potential of Our Future Health for research into predictive models of health and disease as they provide for high precision for estimates of validation statistics such as sensitivity and specificity (**Table 3**).

**Table 3. Margins of Error for Estimates of Sensitivity and Specificity by Case or Control Count.**

Sensitivity or Specificity (%)	Margin of error <sup>a</sup> (%), by number of cases or controls <sup>b</sup>						
	500	1,000	2,500	5,000	7,500	10,000	25,000
50	4.4	3.1	2.0	1.4	1.1	1.0	0.6
70	4.0	2.8	1.8	1.3	1.0	0.9	0.6
80	3.5	2.5	1.6	1.1	0.9	0.8	0.5
90	2.6	1.9	1.2	0.8	0.7	0.6	0.4
95	1.9	1.4	0.9	0.6	0.5	0.4	0.3

<sup>a</sup> Margin of error is half the width of a 95% confidence interval

<sup>b</sup> Cases for sensitivity; controls for specificity

### 2.3.3. Subpopulations

We will strive for diversity in our sample frame ambition by aspiring to reflect the UK population in terms of age, sex, ethnicity, socioeconomic status, and geography. Achieving this aim will deliver a variety of subpopulations of interest each of which will accrue sufficient incident disease as to offer aetiological and diagnostic insights. Subpopulations may be defined by the participant factors stated above, as well as minor allele frequencies, precursor diseases (e.g. colonic polyps, Barrett's oesophagus), genetic disease risk profiles, and blood group subtypes. These examples of potential subpopulations for study provide further underscore the benefits of our sample frame ambition. In terms of ethnic diversity, 5 million participants would be partly comprised of 149,000 Black participants and 330,000 Asian participants, primary ethnic-specific populations of which would include: 100,000 Indian, 68,000 Pakistani, 62,000 Black African, 45,000 Black Caribbean, 31,000 Chinese, and 26,000 Bangladeshi. This will provide a research platform to understand differences in disease risk by ethnicity<sup>29,30,31</sup>, providing a levelling-platform with the potential to improve health for all ethnicities in the UK. Statistical interactions

Our Future Health will create new opportunities to investigate how different factors interact to cause disease or alter treatment efficacy. Estimating these interaction effects has previously been challenging; in contrast to the individual effects of specific treatments, environmental factors or

<sup>29</sup> Ali, R., et al. (2021). "Life expectancy by ethnic group in England." *BMJ* 375: e068537.

<sup>30</sup> Maruthappu, M., et al. (2015). "Incidence of prostate and urological cancers in England by ethnic group, 2001-2007: a descriptive study." *BMC Cancer* 15: 753.

<sup>31</sup> Shirley, M. H., et al. (2014). "Incidence of breast and gynaecological cancers by ethnic group in England, 2001-2007: a descriptive study." *BMC Cancer* 14: 979.

genetic variants, estimating interactions requires considerably larger sample sizes<sup>32</sup>. Additional obstacles to the study of interactions include the availability of high-quality and wide-ranging exposure assessments, known temporality of exposures, and ethnically and geographically diverse populations<sup>33</sup>.

To examine the potential to estimate gene-environment interactions using the Our Future Health cohort, we have calculated minimum detectable odds ratios under a variety of scenarios. Here we assume a conservative scenario, such as might be found in a pharmacogenetic context, where the environmental exposure is a treatment with a modest effect ( $OR_E = 1.25$ ) and the genetic factor has no effect on the outcome ( $OR_G = 1.0$ ) except in treated individuals. We assume that the minor (risk) allele is dominant. We then estimate the minimum detectable odds ratio for the interaction between the treatment and genotype ( $OR_{GE}$ , the effect of the genetic factor in treated individuals) for a range of genotype minor allele frequencies (MAF), treatment prevalence, critical p-values for significance tests, and numbers of cases (**Table 4**).<sup>34</sup> These calculations ignore model misspecification, measurement error and other issues which reduce precision, further underscoring the need for the sample frame ambition of Our Future Health.

---

<sup>32</sup> Ritz, B. R., et al. (2017). "Lessons Learned From Past Gene-Environment Interaction Successes." *American Journal of Epidemiology* 186(7): 778-786.

<sup>33</sup> McAllister, K., et al. (2017). "Current Challenges and New Opportunities for Gene-Environment Interaction Studies of Complex Diseases." *American Journal of Epidemiology* 186(7): 753-761.

<sup>34</sup> Moore, C.M., et al. (2019). "Power and sample size calculations for genetic association studies in the presence of genetic model misspecification." *Human Heredity* 84:256-271.

**Table 4. Minimal Detectable Interaction Odds Ratios by Case Count, Risk Allele Frequency, Exposure Prevalence, and Critical p-value**

Risk allele frequency	Exposure prevalence <sup>a</sup>	Critical p-value	Minimum detectable interaction odds ratio, OR <sub>GE</sub> , by number of cases <sup>b</sup>			
			2,500	5,000	10,000	25,000
0.25	0.25	0.05	1.30	1.20	1.14	1.09
0.25	0.25	0.00005	1.57	1.38	1.25	1.16
0.25	0.1	0.05	1.44	1.30	1.20	1.12
0.25	0.1	0.00005	1.87	1.56	1.38	1.23
0.25	0.05	0.05	1.64	1.42	1.28	1.17
0.25	0.05	0.00005	2.33	1.83	1.54	1.32
0.1	0.25	0.05	1.38	1.26	1.18	1.11
0.1	0.25	0.00005	1.75	1.49	1.33	1.20
0.1	0.1	0.05	1.57	1.38	1.26	1.16
0.1	0.1	0.00005	2.18	1.74	1.49	1.29
0.1	0.05	0.05	1.84	1.55	1.37	1.22
0.1	0.05	0.00005	2.87	2.12	1.71	1.41
0.05	0.25	0.05	1.53	1.35	1.24	1.15
0.05	0.25	0.00005	2.06	1.68	1.45	1.27
0.05	0.1	0.05	1.81	1.53	1.35	1.21
0.05	0.1	0.00005	2.76	2.06	1.68	1.40
0.05	0.05	0.05	2.22	1.77	1.51	1.30
0.05	0.05	0.00005	4.02	2.67	2.02	1.57
0.01	0.25	0.05	2.39	1.86	1.56	1.33
0.01	0.25	0.00005	4.56	2.91	2.15	1.63
0.01	0.1	0.05	3.40	2.39	1.87	1.49
0.01	0.1	0.00005	9.81	4.59	2.92	1.99
0.01	0.05	0.05	5.50	3.27	2.32	1.72
0.01	0.05	0.00005	63.37	8.97	4.36	2.53

<sup>a</sup> Population prevalence of the exposure.

<sup>b</sup> Calculated at 80% power assuming 4 controls per case.

#### 2.3.4. Pre-diagnostic bloods

The lack of large prospective cohort studies with pre-diagnostic bloods available to researchers has stifled disease interception research<sup>35</sup>. There has been a distinct lack of progression of biomarkers from nested case-control studies to the pre-diagnostic arena for insights on diagnosis and prognosis. The sample size ambition of Our Future Health will reduce pressures of biospecimen retention requirements and enable the possibility of boutique subpopulation research studies using pre-diagnostic blood specimens. Although the Access Board will devise the rules for biospecimen access, any such study proposal will undoubtedly require strong preliminary

<sup>35</sup> Pepe, M. S., et al. (2001). "Phases of biomarker development for early detection of cancer." Journal of the National Cancer Institute 93(14): 1054-1061.



evidence<sup>36,37</sup> that will form the basis of a conservative selection strategy<sup>38,39</sup> to ensure the resource is used appropriately and preserved for the long term. Nevertheless, Our Future Health has the possibility to be the only large-scale UK prospective cohort study to offer the possibility of access to pre-diagnostic biologic samples for biomarker validation.

Broad assays (-omics) are no longer confined to small sample sets – high throughput efficiencies are enabling the ability to deep phenotype the totality of samples in a given resource<sup>40,41,42,43</sup>. Continued gained efficiencies will likely spur the continued movement to big, layered -omics data and Our Future Health will provide an ideal platform for such cohort-wide deep phenotyping which greatly increases the potential for insights into the determinants and causes of disease.

## 2.4. Participant questionnaires

Participant questionnaires comprise an essential component of any health-related population study. Through questionnaires, we can elicit important health-related information that is not available, is incomplete, or is potentially incorrect in medical records and other health-related linked data that participants consent to donating to Our Future Health when joining the programme. Questionnaire data is also more rapidly obtainable than many health-related linked data and can complement health-record linkage by obtaining repeated measurements or more detailed self-reports than might be possible via the healthcare system. Examples of questionnaire derived health-related information that can supplement and complement health-related linked data include:

- up-to-date health status, exposures, and outcomes (e.g. general health, current cigarette smoker, mental health)
- long-term health or exposure histories that pre-date electronic medical records (e.g. any surgical procedure before 1990, lifetime cigarette smoking history)
- exposures or outcomes that are either poorly or not captured by electronic medical records (e.g. typical alcohol consumption, recent physical activity, over-the-counter medications, mental health)
- Measures that are not regularly repeated or typically available until later in life in medical records (e.g. body weight, alcohol consumption, anxiety, physical activity)
- Health relevant lifestyle factors and personal characteristics that might be unavailable in medical records (e.g. employment status, marital status, type of housing)

---

<sup>36</sup> Wentzensen, N. and R. C. Eldridge (2015). "Invited Commentary: Clinical Utility of Prediction Models for Rare Outcomes-The Example of Pancreatic Cancer." *American Journal of Epidemiology* 182(1): 35-38.

<sup>37</sup> Wentzensen, N. and S. Wacholder (2013). "From differences in means between cases and controls to risk stratification: a business plan for biomarker development." *Cancer Discov* 3(2): 148-157.

<sup>38</sup> Pepe, M. S., et al. (2008). "Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design." *Journal of the National Cancer Institute* 100(20): 1432-1438.

<sup>39</sup> Pepe, M. S., et al. (2015). "Improving the quality of biomarker discovery research: the right samples and enough of them." *Cancer Epidemiology, Biomarkers and Prevention* 24(6): 944-950.

<sup>40</sup> He, K. Y., et al. (2017). "Big Data Analytics for Genomic Medicine." *Int J Mol Sci* 18(2).

<sup>41</sup> Rappaport, S. M. (2012). "Biomarkers intersect with the exposome." *Biomarkers* 17(6): 483-489.

<sup>42</sup> Siroux, V., et al. (2016). "The exposome concept: a challenge and a potential driver for environmental health research." *Eur Respir Rev* 25(140): 124-129.

<sup>43</sup> Wild, C. P. (2012). "The exposome: from concept to utility." *International Journal of Epidemiology* 41(1): 24-32.

Thus, self-report questionnaires significantly contribute to a comprehensive assessment of health, enhancing research insights into how we may improve the nation's health.

Questionnaires also allow us to serve certain populations that would otherwise continue to be misclassified and underrepresented in health research. For example, asking about sexual orientation and gender identity will allow us to ensure we are recruiting a sample representative of all peoples in the UK. This in turn will enable research that supports the provision of a health-care system that serves the needs of all, rather than over-generalising research from nonrepresentative populations.

Delivering regular, repeated questionnaires also facilitates the ongoing involvement of participants – reminding them of the important research programme they have consented to be part of, and retaining their interest and attention. We will have the opportunity to further engage participants by using questionnaire responses to formulate feedback and advice that might be of interest and support individual health choices.

In addition to our core participant questionnaire, we will develop a roadmap of future questionnaires built from these principles that underscore the scientific and participant rationales of our programme.

## 2.5. Physical measurements

We will assess physical measurements when participants provide blood samples for those metrics and in locations where this is feasible and cost-effective. Although height and weight are typically accurately self-reported within a population<sup>44</sup>, in-person assessment of these metrics provides individual accuracy which is important for a variety of disease risk estimations as well as participant feedback that incorporates such information<sup>45</sup>. Weight is associated with many diseases including cardiovascular disease and cancer. Height can be used to calculate body mass index (BMI) from weight which has greater predictive accuracy for disease, and height also an independent predictor of certain cancers, vascular disease and all-cause mortality.

Capturing height and weight in-person can easily be extended to measurement of other physical characteristics such as waist circumference<sup>46</sup>, bioimpedance<sup>47</sup>, and blood pressure. Waist circumference is highly correlated with intra-abdominal fat mass while bioimpedance analysis is a non-invasive, low-cost analysis of body composition. Waist circumference and bioimpedance analysis have each been shown to be associated with a higher risk of diabetes and vascular events independently of BMI. Excessive intra-abdominal fat may be more harmful to health than fat elsewhere due to higher release of free fatty acids into the portal bloodstream which lowers the body's sensitivity to insulin, and alters the balance of blood lipids.

---

<sup>44</sup> Celis-Morales et al (*Genes Nutr*, 2015) How reliable is internet-based self-reported identity, socio-demographic and obesity measures in European adults? <https://pubmed.ncbi.nlm.nih.gov/26143178/>

<sup>45</sup> Newell et al (*Am J Prev Med*, 2017) The accuracy of self-reported health behaviors and risk factors relating to cancer and cardiovascular disease in the general population 1: A critical review. <https://pubmed.ncbi.nlm.nih.gov/10987638/>

<sup>46</sup> Ross et al (*Nat Rev Endocrinol*, 2020) Waist circumference as a vital sign in clinical practice: a Consensus Statement from the IAS and ICCR Working Group on Visceral Obesity. <https://pubmed.ncbi.nlm.nih.gov/32020062/>

<sup>47</sup> Böhm et al (*Eur J Clin Nutr*, 2017) The use of bioelectrical impedance analysis for body composition in epidemiological studies. <https://pubmed.ncbi.nlm.nih.gov/23299875/>

Elevated blood pressure or hypertension is a well-established cause of coronary heart disease, stroke and several other vascular diseases. In addition, blood pressure accounts for a large proportion of the effects of obesity on health, such that a proper understanding of the effects of obesity is not possible without a proper understanding of the effects of blood pressure.

Thus, each of these baseline physical measurements provide significant contributions to disease risk predictions and have a strong rationale for being included in the Our Future Health research programme.

## **2.6. Data linkages**

Health-related data linkages are a core component of the UK research infrastructure, made possible by routine data collection that can be safely and securely linked to participants. Our Future Health participants consent to data linkages when joining the programme, and their donation of these data provide important additional information on individual and geographic disease-related exposures as well as individual health outcomes. Examples of each of these are shown below:

- Individual disease-related exposures
  - medication prescriptions
  - coronavirus infection
  - surgical implants
  - radiation therapy
- Geographic disease-related exposures
  - particulate matter from combustion and other sources
  - food choices including distance and accessibility
  - meteorological information including flooding and extreme weather
  - geographic deprivation metrics
- Individual health outcomes
  - diagnoses captured in primary care records, secondary care records, and cancer registration databases
  - survival time following a serious diagnosis or clinical intervention
  - date of death including underlying and contributory causes of death

These participant-level geotemporal health-related data will greatly enrich the Our Future Health programme enabling researchers to assess the success of interventions, how and in who new therapies extend survival from acute disease, and how social determinants contribute to health inequalities. These examples highlight the strong rationale for data linkages in this research programme.

## **2.7. Biological samples**

A broad variety of biological specimen types could theoretically be collected in a given research study, but most prospective cohort studies have decided to collect blood at baseline on all of their participants. This is because it is a minimally-invasive, cost-effective, and a participant-accepted specimen type that can provide systemic insights on an individual's health-related exposures, disease risk, and disease status. For example, blood can provide information on viral exposures,

pesticide exposures, polycyclic aromatic hydrocarbon exposure, lipid profile including high density and low density lipoproteins, genetic susceptibility, DNA adducts, diabetes metrics, circulating tumour DNA (ctDNA), and circulating proteins indicative of disease. These examples of scientific insights that can be derived from blood provide a strong rationale for the collection of this biospecimen from all Our Future Health participants.

From these baseline blood samples, we plan to assess cholesterol and HbA1c.

Hypercholesterolaemia is a well-established cause of coronary heart disease, stroke and several other vascular diseases and its measurement is required to provide an integrated risk score for ischaemic heart disease. HbA1c provides an estimate of blood glucose (sugar) levels and will be conducted for participants who have a diabetes risk score above the recommended test threshold. Vascular diseases and diabetes are major causes of morbidity and mortality in the UK providing the strong rationale for measuring these biomarkers at baseline.

Moreover, blood collection is highly feasible and cost-effective given the facts that phlebotomy services are readily available and deployable, and that high-throughput automated laboratories and biobanks exist for processing and storage. These facts support the selection of blood as the central biological sample that we will collect when a participant joins the programme. However, this does not preclude the collection of additional biological specimen types, and we will continuously monitor the feasibility and cost-effectiveness of such as the programme progresses.

## 3. Recruitment

### 3.1. Overall strategy

As described in our sample frame section, our ambition is to recruit up to 5 million people from diverse backgrounds. To achieve this ambition, we plan to engage the public (our potential participants), invite eligible participants by post or email, attain digital consent to participate in the programme, attain a digital baseline health questionnaire, and attain physical measurements and a blood sample at an in-person appointment.

We will deliver local and national communication strategies to make the public aware of the programme, its primary aims and what and how a participant's time, data, and blood sample will be used.

In addition to using post and email for invitations, we will also design and deploy in-person settings to advertise and consent individuals into the programme.

To ensure we do not preclude participation of individuals who cannot or prefer not to use digital tools, we will enable hardcopy and telephone completion of consent and questionnaires.

We will provide flexibility in venues for attaining physical measurements and blood from consented participants, primarily focusing on:

1. Community strategies such as pharmacy collaborations, mobile units, and 'pop-up' clinics.
2. Partnerships with the NHS to use existing appointments such as blood donations and health care phlebotomy.

Schematic summaries of these recruitment workflows are shown in Figures [1](#), [2](#), and [3](#) below.

This primary overarching plan has a strong rationale in that it is flexible, scalable, cost-efficient, and feasible. It will be built with the widely used COM-B Model (Capability, Opportunity, Motivation, Behaviour)<sup>48</sup> of human behaviour in mind, which encourages careful consideration of barriers when designing activities relating to behaviour (in this case, becoming a participant of Our Future Health). Our recruitment strategy will also be aligned with the highest ethical principles, as detailed in our Ethics & Governance Framework (**Appendix B**).

Our recruitment strategy will enable rapid, large-scale recruitment into the programme while simultaneously allowing us to adapt, tailor, and target our methods to ensure inclusion of populations that have been underrepresented in health research. In the following sections, we describe our detailed plans for engagement, invitation, consent, baseline questionnaire, and phlebotomy.

### 3.2. Engagement

The goals of engagement are to raise awareness of Our Future Health, generate interest in taking part, and to provide opportunities for the public and other stakeholders to share their views.

---

<sup>48</sup> COM-B Model (Michie et al, 2011) <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3096582/>

To recruit a cohort that is reflective of the UK population, specific engagement strategies will be designed, tested and deployed. Engagement strategies will be designed to promote equality, diversity, and inclusion in Our Future Health. These engagement strategies will be designed to support our local and region-specific recruitment plans.

### **3.2.1. Engagement through partnerships**

Identifying and building relationships with individuals and organisations who can support participant recruitment to Our Future Health is an essential component of this programme. We will aim to work with partners that can support recruitment by publicising the programme to target audiences and through existing networks, and with partners that are able to support our plans to send invitations to people to take part. This will include local opinion formers in public health, large businesses, community organisations, academia and local government.

Collaborative engagement with delivery partners – such as community pharmacy networks – has the potential to increase awareness and engagement via well-known organisations that have strong local and national footprints. In hospitals, we will work with research ambassadors (volunteers who are Good Clinical Research Practice trained), who will directly support engagement of potential participants. In primary care we will work with patient participation groups and for blood donation recruitment we will be working with donor carers.

To overcome hesitancy within seldom heard/underrepresented groups, we will aim to develop partnerships with trusted voices and community leaders or representatives, including local public health teams, community groups and those who have strong networks in each area.

Engagement strategies will be designed with the support of our Diversity and Inclusion Advisory Board which comprises people with expertise in engaging communities locally and nationally. We will also leverage the expertise on our Ethics Advisory Board, Scientific Advisory Board and our Public Advisory Board, each of which provide additional expertise of the challenges of engaging and recruiting participants into health research studies from diverse backgrounds across the country in culturally appropriate ways.

To achieve engagement with and maintain ongoing support from a range of partners, we are planning to focus on a combination of:

- Clear, motivating partnership proposals, backed up by credible voices from a range of domains (e.g. science, healthcare, charities, politics, celebrities) and working with existing partners to engage others
- Providing feedback on partners' support in driving recruitment, so they can be credited for their efforts and achievements
- Strong, enduring relationships with leaders in partner organisations, particularly in the third sector, and ongoing engagement and outreach
- High quality tools, content and campaign resources to make it easy for partners to promote Our Future Health
- Tailored campaigns, co-created with partners to improve impact
- Supporting national, local and digital PR, to provide a positive ongoing context for partners choosing and continuing to help us

### **3.2.2. Publicity campaigns and communications**

A comprehensive programme of publicity activities will increase awareness and understanding of Our Future Health among the target population. We will deliver regional publicity campaigns in targeted recruitment locations from early Summer 2022. The campaigns will be designed to increase responses to invitations sent to members of the public, and so enable recruitment into the cohort. Activities will include advertising, public relations, social media and community-based events that can effectively reach the target population in each area.

Following the series of regional engagement activities, as we expand the scope of recruitment across the UK from 2023 onwards, we are planning national level awareness engagement and campaign activities. These activities will aim to increase levels of awareness across the target population as a whole, particularly those groups that we anticipate are less likely to engage and respond. We will explore a range of messaging, channels and methods of engagement designed to increase motivations to participate in the programme, including:

- Participant referrals – supported by tools, simple processes and (non-financial) incentives to encourage people to recruit others within their family and communities
- Membership bodies/groups – tapping into the scale of organisations that have a strong connection to or presence within our target audiences. These could be place-based or interest-based opportunities, or a combination of the two (e.g. faith organisations)
- Patients as advocates – exploring routes through patient charities/groups to encourage people with diseases to make the case within their communities and families for participation
- Partnership marketing – working with organisations with significant reach (e.g. charities, consumer brands, sports orgs, large employers) to promote participation across their customers, staff, audiences and supporters
- Social campaigning – adopting a networking approach to increase reach using social media in particular, drawing on the potential to tap into motivations around specific disease areas with the support of relevant charities
- Influencers – engaging and enlisting the support of high profile advocates who have reach and influence across target populations, both geographically and digitally.

We are also exploring ways to influence national and local healthcare decisions by engaging with credible stakeholders/advocates at the regional and national level. These may be a combination of those within the NHS healthcare system (e.g. high profile clinicians, NHS Health Check commissioners, NHS leadership) and across the broader public health arena e.g. medical charities, local authorities, patient groups and researchers.

## **3.3. Invitation**

### **3.3.1. Invitation methods**

We will use NHS DigiTrials as the primary route for postal invitation. NHS DigiTrials provides unparalleled scale, the ability to target invitations based on demographic information, and will help us build trust with potential participants. The NHS DigiTrials application process includes Section 251 support which will enable NHS DigiTrials – on our behalf – to select eligible individuals and send named invitations for participation in Our Future Health. We will dynamically adjust the number of invites sent to specific population groups based on conversion rates and our ambition to recruit a diverse cohort that is reflective of the UK population.

Invitations may also be sent by collaborations with community pharmacy networks using their existing customer databases. We are also exploring other ways to send invitations, including potential partnerships with existing cohorts such as REACT.

In partnership with NHS bodies, invitations to Our Future Health will be sent by email to blood donors by NHSBT, and through existing patient communication systems – which may include post, email and text message – in primary and secondary care. Text messages are attractive because they are inexpensive, already sent in high volume by the NHS, and can include a link to our participant information sheet and consent process that can be accessed via a patient's smartphone or computer prior to a planned appointment. Our Public Advisory Board members and secondary care PPIE work in 2021 revealed that this was an acceptable and viable format, but with the caveat that it needs to come from a known and trusted source.

NHS primary and secondary care settings offer the opportunity to recruit a diverse cross-section of the UK population in terms of age, ethnicity and deprivation, countering concerns of the healthier, less diverse population that can typically be recruited from the blood donor population. For example, NHS primary care conducts the Health Check programme<sup>49</sup>, which is offered every 5-years via postal invite to GP registrants aged 40 to 74 years. The Health Check programme has consistently higher uptake within higher deprived populations as well as within underrepresented ethnicities in health research including Indian, Caribbean and Chinese.<sup>50</sup>

### 3.3.2. Feasibility of invitation strategy

The invitation methods described above have high feasibility based on past use of the infrastructure by prior studies.

The largest UK example of a successful postal recruitment strategy is UK Biobank. Individuals registered with the NHS were invited by post and able to respond via post, internet, or phone to arrange an appointment at an assessment centre. Consent, questionnaire, baseline measurements, and phlebotomy were all conducted at assessment centres that were specially designed and fitted out for this purpose.

NHS DigiTrials is a similar, more formalised process to invite NHS-registered individuals to research studies, which is already demonstrating success with recruitment to the Randomised Evaluation of COVID-19 Therapy (RECOVERY) trial and the Platform Randomised trial of Interventions against COVID-19 In older people (PRINCIPLE) Trial.<sup>51</sup>

We have already demonstrated the feasibility of NHSBT email invitations sent to blood donors in joining Our Future Health and this pilot study is summarised in **Appendix D**. In addition, INTERVAL<sup>52</sup> and COMPARE<sup>53</sup> studies have already successfully demonstrated recruitment of blood donors for research studies that use the first 35 ml of blood that would otherwise be discarded for

---

<sup>49</sup> <https://www.nhs.uk/conditions/nhs-health-check/>

<sup>50</sup> Cook et al (*Int J Equity Health*, 2016) Who uses NHS health checks? Investigating the impact of ethnicity and gender and method of invitation on uptake of NHS health checks.  
<https://pubmed.ncbi.nlm.nih.gov/26791963/>

<sup>51</sup> <https://digital.nhs.uk/features/nhs-digitrials-already-saving-lives>

<sup>52</sup> <https://www.intervalstudy.org.uk/>

<sup>53</sup> <https://www.comparestudy.org.uk/>



infection control, while the STRategies to Improve Donor ExperienceS (STRIDES) study<sup>54</sup> is also currently successfully recruiting blood donors.

With regards to NHS primary and secondary care, we are conducting discovery work to establish optimal ways in which to recruit from existing hospital outpatient lists, while Born in Bradford<sup>55</sup> successfully recruited participants in NHS GP practices. Our work to date has included PPIE work comprising a series of multi-disciplinary consultations with patients, public, health care professionals and healthcare research delivery teams to understand how Our Future Health would be received and adopted alongside existing research portfolios. We have also had discussions with Genes and Health to identify commonalities in recruitment strategies and optimise invitations to attract a wider cohort of participants.

### 3.3.3. Invitation development

We will take a theory- and evidence-based approach to the development of our invitation content. Specifically, we will conduct user research to assess comprehension and acceptability of invitations and use randomised online experiments to identify content that maximises response.

For digital invitations, where possible, we will send a limited number of pre-invitation notifications and invite reminders to optimise response rates. This has been validated in the findings from our PPIE work in primary and secondary care and via input from our Public Advisory Board.

To avoid excluding individuals who do not own a smartphone or have lower digital literacy, in addition to postal invites, we will develop the ability to enable hardcopy and telephone completion of the consent form and questionnaires.

## 3.4. Reimbursement

**Background:** Our Future Health is committed to the principle of equity in participation and widening access as substantially as possible and we are exploring several alternative solutions to achieve these aims.

Given the current economic climate, we recognise that participants may incur costs and will be offering their time to attend an Our Future Health appointment, and this may be a barrier to participation among those who cannot afford to incur such costs. Previous studies have shown a significant impact on response rates following the offer of reimbursement – for example during COVID the REACT study saw response rates more than double among younger age groups.

To test if providing reimbursement will reduce barriers to entry and help create a more representative programme, we are proposing to run a pilot in early 2023. This pilot will help us understand the impact of reimbursement on enhancing equity and participation in the programme.

**Hypothesis:** We are testing if the offer of reimbursement increases the following three conversion rate metrics:

---

<sup>54</sup> <https://www.strides-study.org.uk/>

<sup>55</sup> <https://borninbradford.nhs.uk/>

1. Proportion of people receiving an invitation consenting to join the programme
2. Proportion of people receiving an invitation booking an appointment to donate blood
3. Proportion of people receiving an invitation completing a blood donation appointment

**Design:** Our Future Health is recruiting in 4 regional locations (West Yorkshire, Greater Manchester, Birmingham & Black Country and London) until the end of March 2023, when we are planning to expand recruitment to a greater number of regional locations. From mid-December, we are commencing a new “test set” of 6 invitation letter variants via NHS Digitrials to test the relative performance of each variant. Each of these letter variants has a unique invitation code so that when participants register on our system we can track which version of the letter they received. To enable the appropriate evaluation of the pilot of reimbursement we plan to:

1. Send letters using these 6 variants for a 6 week period (mid Dec-end Jan) to all invited participants in all regions
2. After 6 weeks, for invitations being distributed to potential participants in the Greater Manchester area (as identified by postcode selection) for the next 6 weeks (Feb-mid March) include the below wording offering the opportunity for reimbursement and change the invitation codes enabling tracking of these new variants in our systems
3. Make no change to the letters in other 3 regions.

The wording that will be included in the letters is as follows: “At the end of your appointment, you will have the option to claim a £10 voucher to cover any costs of attending the clinic. You do not need to show any receipts.”

We plan to share this wording with our Public Advisory Board in the new year to test the language and check the acceptability and comprehension of the language.

Based on our current operational plan we expect to distribute around 400,000 invitations in the Greater Manchester area between mid-December and mid-March and hence around 200,000 participants will receive the letter variants that contain the reimbursement offer.

We are offering a £10 reimbursement which we believe ensures that the offer of reimbursement does not interfere with the voluntariness of participants’ consent by acting as an inducement.

**Evaluation:** To evaluate the impact of the offer of reimbursement for each of the three conversion rate metrics noted above we will evaluate them by a before and after comparison. We will estimate for all letter variants the conversion rate metrics using the “before” period tracking code and the “after” period tracking code. We will do this for Greater Manchester and for all three other regions combined. This will enable us to populate the following Table.

Region	Pre-Period	Post Period	Performance Increase

Greater Manchester (Test)	Consent: XX Booking: XX Donation: XX	Consent: XX Booking: XX Donation: XX	Consent: XX Booking: XX Donation: XX
West Yorks, Birm, London (Control)	Consent: XX Booking: XX Donation: XX	Consent: XX Booking: XX Donation: XX	Consent: XX Booking: XX Donation: XX

Our **Overall Measure of Performance** will be the relative performance increase comparing Greater Manchester to the Control Area.

We will assess the sensitivity of the result by letter variant type and by control region.

**Success Criteria:** We will consider the pilot successful in increasing equity and participation if we achieve at least a doubling of the overall measure of performance from invitation to completed blood appointment among the following groups:

1. Socioeconomic Status: Lowest two quintiles of IMD
2. Young People: under 40 years old
3. Major Ethnic Subgroups: Asian, Black English (who are on average both more deprived and younger.)

Following this pilot, and all the other pilots testing ways to increase equity in participation, we will review and adopt the options which offer the most suitable solutions for the programme in the future.

### 3.5. Consent

The primary method of consent in Our Future Health will be digital. The consent process starts with information provision which comprises the consent form and the participant information sheet, opportunities for potential participants to have their questions answered, and a formal recording that the individual consents to participate in the Our Future Health research programme. Participants register an account with their contact details either before or at the time of consent so we know who the consent belongs to, and so we can contact that person as part of their involvement in the programme.

We designed the consent and participant information sheet in alignment with the principles set out in our Ethics & Governance Framework, namely that valid consent comprises three components: information, comprehension, and voluntariness. For further details, please see the Ethics & Governance Framework (**Appendix B**).

The consent form and participant information sheet were rigorously co-developed with members of the public in the following forums:

- 18 focus groups with 82 members of the public.

- 4 meetings with a co-design group comprised of 8 members of the public (a different group to the two that informed the design of the leaflet and video scripts).
- 21 user testing interviews with members of the public (who had previously participated in one of the 18 focus groups).

We worked with Claremont and digital agency Kainos on three rounds of user testing (total n=36) of the participant information sheet (along with other parts of the digitally-delivered process including registration form, consent form, and questionnaire completion). We also had input to the first version of the participant information sheet from the Ethics & Feedback Advisory Group as well as other external stakeholders and advisers including national and international experts in consent from academia. The consent form and participant information sheet were approved by the REC and used in the pilots studies. We recently revised these forms again, after feedback from our participants and with consultation and re-review by our governance Boards, including the Public Advisory Board and the Ethics Advisory Board. We also conducted individual user testing interviews with nine members of the public.

The consent includes the ability to recontact participants, so we can invite them to complete additional questionnaires, provide further samples, receive personal health-related information, and consider invitations to enrol in future (stage 2) studies that will have separate REC-approved study protocols with their own consents and participant information sheets. Some stage 2 studies may recruit participants based on their risk of specific diseases calculated from their self-reported, genetic and/or other health-related information.

We will continue to use our Participant-Reported Experiences Survey (**Appendix C**), which over 900 participants have completed to date, to obtain feedback on the ease and acceptability of the consent process, and as part of our evaluation/analytics and insights.

Individuals with queries can call or email the Our Future Health support centre, which will expand in capacity as our recruitment scales. The support centre will be operated by specially trained staff using an integrated computer system developed for, and dedicated to, Our Future Health. The main functions of the support centre are to:

- Answer questions about consent procedures and the scope of Our Future Health
- Allow questions from potential participants (and their GPs) to be addressed either by the trained call centre staff or, if not possible, by more senior members of the Our Future Health team
- Administer the questionnaire to visually impaired participants, and those who do not or cannot access the questionnaire via the website/digitally

Continued interactions with our participants and the public through the Participant-Reported Experiences Survey and the support centre, respectively, will ensure we can iterate the consent and participant information sheet further, if required, with updates sent to the REC for review and approval prior to deployment.

### **3.5.1. Participants without capacity to consent**

After a participant has consented to be part of Our Future Health, they may subsequently lose capacity (e.g. due to dementia). In such situations we need to balance the ethical requirements to enable access to research participation, as well as protecting vulnerable participants from intrusive or interventional research.

There are two situations where we may become alert to potential changes in capacity:

1. Directly contacted by a representative of the participant
2. From our regular updating of the health information through linkage with national data controllers

For option 1, we will establish a process where we verify the identity of the participant and the legal status of the reporting individual. Once verified we will proceed to our loss of capacity process detailed below.

Option 2 is complex since there are no codes in the medical record which indicate loss of capacity either temporarily or permanently. There are some medical conditions in which loss of capacity is more common and these include: strokes, dementias, and traumatic brain injury. Using only diagnosis codes can be misleading since this is not the same as loss of capacity. Studies have reported that individuals diagnosed with stroke, mild-moderate dementia, and traumatic brain injury results in a 2%<sup>56</sup>, 26%<sup>57</sup> and 47%<sup>58</sup> likelihood, respectively, of lacking capacity to consent within a year of diagnosis.

We do not yet know if it is possible to develop an algorithm that can accurately predict who has lost capacity using clinical codes due to the following points:

1. Discovery and validation of an algorithm requires formal assessments of capacity in identified participants.
2. Capacity is a temporal and decision-specific assessment with considerable heterogeneity between patients with the same diagnosis, which increases the risk of any algorithm of inappropriately excluding individuals from research.

We therefore propose that, once we have obtained linked healthcare data, we collaborate with other research programmes to propose a research study to develop a loss of capacity algorithm. This algorithm would be based on broad clinical codes (ICD, Read, SNOMED) and would require formal capacity assessments to be conducted. If a validated algorithm with high positive predictive value and sensitivity in identifying individuals who have lost capacity can be developed, we will explore the appropriateness of using it to identify participants who should proceed through our loss of capacity process. This will require careful interpretation of the Mental Capacity Act, and engagement with participant representatives.

### **3.6. Baseline questionnaire**

We have developed a baseline questionnaire to capture health and lifestyle factors that would be difficult or impossible to obtain from data linkages. This questionnaire was developed largely by a core scientific advisory team and was intended to align closely to existing large epidemiological

---

<sup>56</sup> Pendlebury ST. Stroke-related dementia: Rates, risk factors and implications for future research. *Maturitas*. 2009;64(3):165-71.

<sup>57</sup> Guarino PD, Vertrees JE, Asthana S, Sano M, Llorente MD, Pallaki M, et al. Measuring informed consent capacity in an Alzheimer's disease clinical trial. *Alzheimers Dement (N Y)*. 2016;2(4):258-66.

<sup>58</sup> Thornhill S, Teasdale GM, Murray GD, McEwen J, Roy CW, Penny KI. Disability in young people and adults one year after head injury: prospective cohort study. *BMJ*. 2000;320(7250):1631-5.

cohorts, such as the UK Biobank. We went through a process of cognitive testing with Our Future Health participants to refine how questions are asked and understood. The baseline questionnaire collects information on demographics, socioeconomics, physical activity, lifestyle exposures, family history, medical history, depression and anxiety, medications, and supplements. It takes, on average, 35 minutes to complete. This initial version of the baseline questionnaire was previously approved by the REC and was used in pilots conducted in 2021, in which we collected complete questionnaire data on over 1,500 people (report of pilots attached as **Appendix D**).

Piloting, additional cognitive testing, and internal scientific review has enabled us to identify and design improvements to the baseline questionnaire as well as expand slightly on health and family history sections. The changes have been reviewed by the Scientific Advisory Board and are included in **Appendix E**.

From our pilot studies, we know that 95% of participants who start the baseline questionnaire complete it. During early phase recruitment, we will test whether deploying the questionnaire before or after phlebotomy has significant effects on attaining full recruitment of an individual, defined as provision of consent, completion of the baseline questionnaire, and donation of a blood sample.

### **3.7. Short baseline questionnaire**

We will develop a short version of the questionnaire. This will be a shortened list of the already-approved questions from our baseline questionnaire. There will be no additional items or new question types.

The items that comprise the short questionnaire will be selected by:

1. Data reduction approaches

Where several items have been included to measure the same underlying phenotype, we will identify cardinal items for inclusion in short questionnaire using data reduction approaches.

We will use pilot data to examine the internal validity (Cronbach alpha) of any related items. We will establish inter-item correlations and perform factor analyses to establish the overall fit, factor scores and item loadings. We will iteratively drop less well performing items and re-evaluate internal validity using Cronbach alpha and factor analyses until we find the minimum number of items that can be used to index the underlying outcome.

2. Stakeholder input

We will review the results from this data driven approach with stakeholders including our scientific advisory board and participant advisory board. We will collect input from stakeholders on priorities for retention and outcomes or items that do not represent immediate priorities for healthcare research.

3. Testing and piloting

We will pilot the resultant shorter questionnaire with members of the public to establish whether the time taken to complete meets our length criteria (<10minutes) and to ensure it is being understood and is generally acceptable to participants.

When we have a satisfactory short questionnaire, we will test it against the longer baseline questionnaire to establish whether rates of conversion to full participant (consent + blood sample + questionnaire) are improved. The short form questionnaire may enable a streamlined, in-person, full participant recruitment model to be deployed in the field, if desired or required. Any participants who complete a short core questionnaire will be asked to provide responses to the remaining questions from the primary baseline questionnaire after recruitment.

### **3.8. Phlebotomy and physical measurements**

We will record physical measurements and collect a blood sample from each consented participant at an in-person appointment in one of the following settings:

1. Community covering pharmacies, mobile units, and 'pop-up' clinics
2. Existing NHSBT appointments for blood donations
3. Existing NHS appointments for health care phlebotomy

Participants will have a choice in where they donate their blood and undergo a brief physical assessment. Location choice will largely be within the community route (pharmacy collaborations, Our Future Health mobile units and 'pop-up' clinics). If the participant has been recruited through an NHS route, then they will likely automatically be in an existing phlebotomy route such as blood donation, or primary or secondary health care.

Blood samples will be used to extract DNA and conduct genotyping, as well as conduct baseline assessments of cholesterol and HbA1c and will be sent to our biobank for long-term storage. At all opportunities the blood collection time and date, and time of last significant meal will be collected at the time of blood draw. Physical measurements will also be taken at this time including blood pressure, height, weight, and waist circumference.

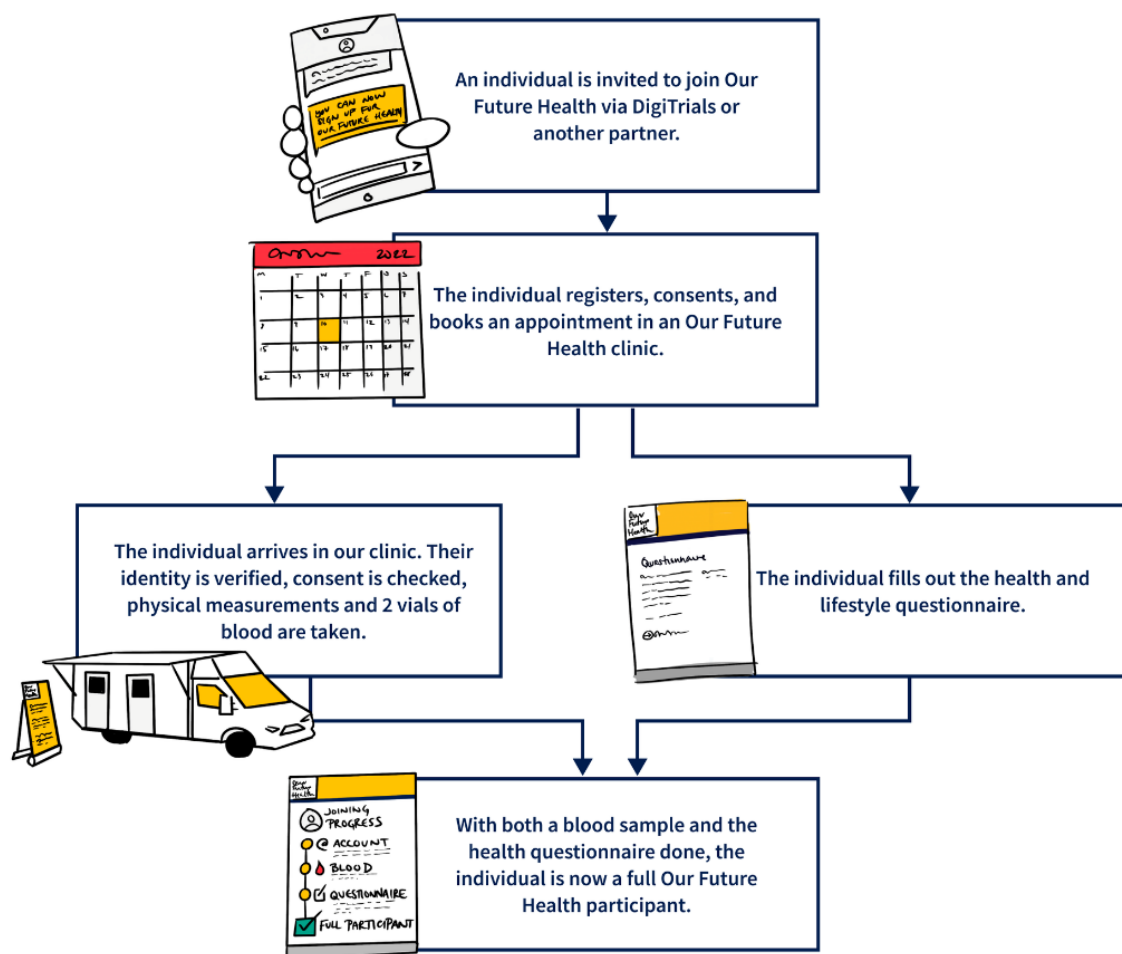
#### **3.8.1. Community routes**

We have conducted extensive market research and a viability assessment from which we are confident of being able to conduct phlebotomy and physical measurements outside of the NHS in a cost-efficient manner using pharmacy collaborations, mobile units, or 'pop-up' clinics.

Through a booking system, consented participants will be able to book, online or by telephone, an appointment location and time for phlebotomy and physical measurements. By establishing Our Future Health community collection sites, we will have greater control over locations, the participant experience and how we manage consent relative to limitations of working solely within the NHS.

We have benefited from information shared by GRAIL which has recently operationalised the Galleri study, and this has allowed us a greater understanding of, for example, cancellation rates, no shows, and other aspects of user behaviour. Based on our research, our approach is to phase the roll out of the venues, with a lower capacity for the first two months of deployment, which will allow us to learn and adapt the service according to behaviour.

Figure 1. Flowchart of recruitment in community settings



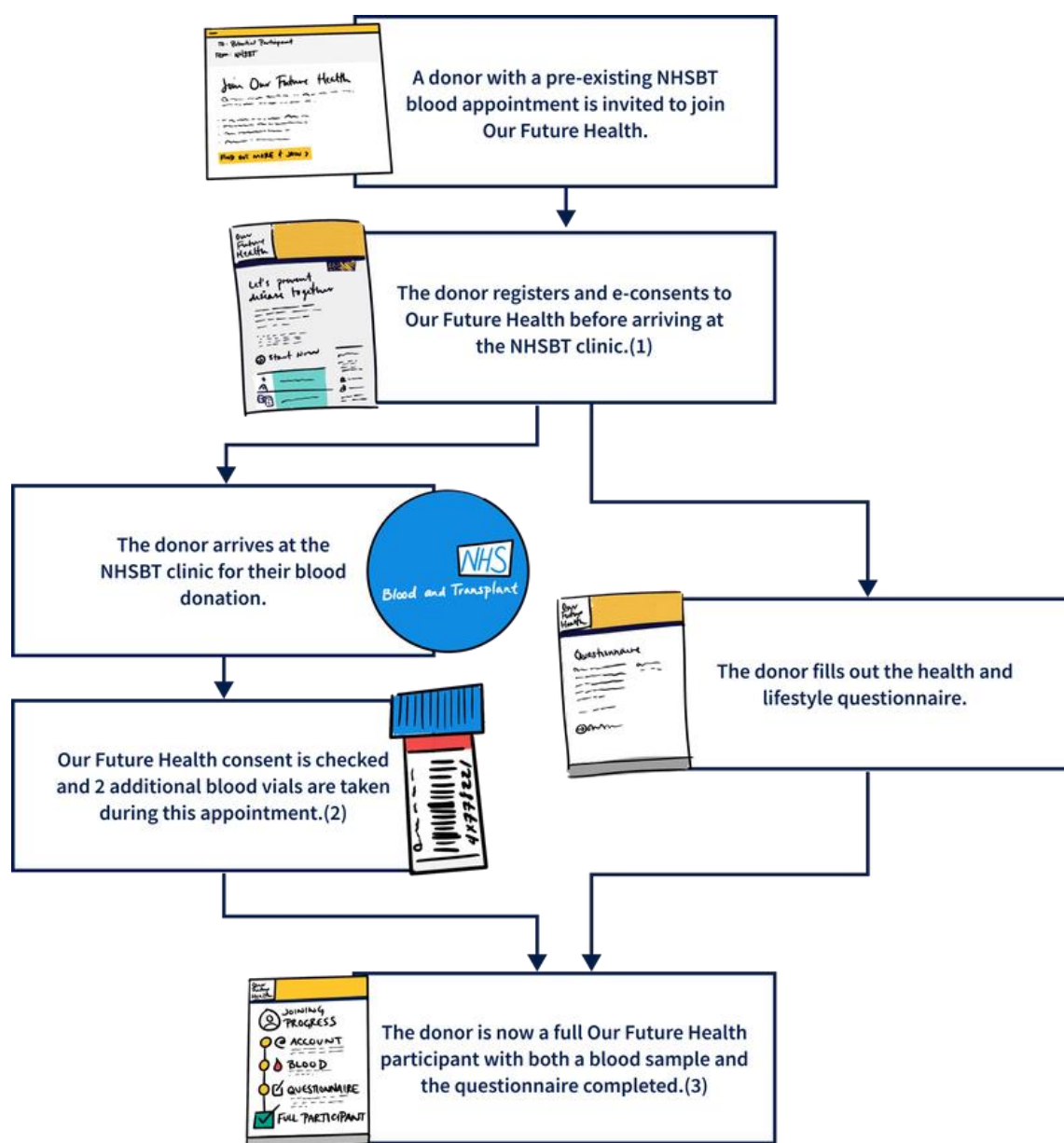
We are also working with community pharmacy groups and the NIHR ‘research ready community programme for community pharmacies.’ Community pharmacies are an attractive venue with 89% of the UK population being able to access a community pharmacy within a 20-minute walk, with access being greater in areas of highest deprivation. Community pharmacists can also access harder to reach patient populations.

### 3.8.2. NHS blood donor route

For the NHS blood donor route, following a donor consenting online to Our Future Health, we are now working with NHSBT to optimise the process by which participants are linked with their future blood samples. The goal is to maximise efficiency and minimise friction for both NHSBT and Our Future Health. A flag is created in the NHSBT system to notify phlebotomists of blood donors who have consented to participate in Our Future Health and wish to provide a blood sample at their next donor appointment. Our Future Health regional leads (senior NHSBT research nurses) will support donor carers in implementing this phlebotomy route in the main phase of the programme.



Figure 2. Flowchart of NHSBT recruitment



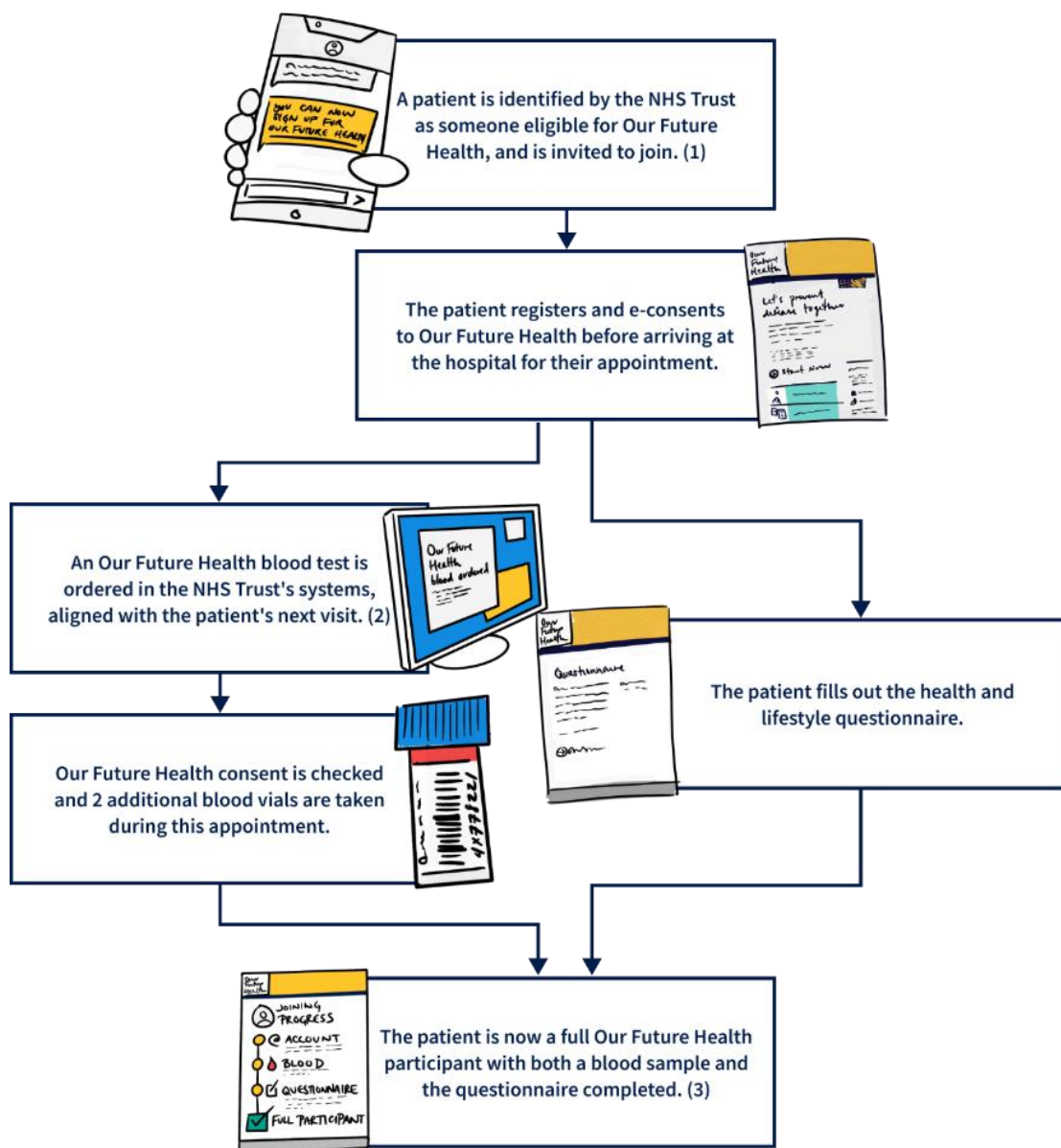
### 3.8.3. NHS routes

For NHS hospitals, we will utilise existing phlebotomy service infrastructure as a cost-effective and scalable solution for Our Future Health participants to provide a blood sample. For many outpatient department appointments, there is a requirement for patients to give a blood sample for diagnostic or monitoring purposes. There is an opportunity to invite patients scheduled to attend phlebotomy to consider becoming a participant of Our Future Health and then donating their blood sample during their scheduled medical phlebotomy appointment.

In conjunction with the hospitals, we will develop a specific Our Future Health blood test request within the pathology ordering system. This has been done for other research projects within NHS trusts such as Leeds Teaching Hospital Trust (LTHT). Our Future Health bloods will be requested

for all consenting participants and will be collected at the same time as their next routine clinical blood draw. This will minimise any additional burden on NHS phlebotomy services.

Figure 3. Flowchart of the recruitment steps in the context of working with NHS hospitals



We will also have pop-up clinics in NHS hospitals whereby staff, visitors and patients who are not having a routine blood draw can take part in the programme. This is effectively 'the community route in hospitals' and participants will also be able to have physical measurements taken in these pop-up clinics.

## 4. Blood sample logistics, processing and genotyping

Each participant blood sample will have a unique barcode label which will be linked in Our Future Health's system to participant ID and personal data. Blood samples will be sent overnight at ambient temperature and centrifuged within 24–30 hours at the country-specific processing facility to fractionate the blood. DNA will be extracted from buffy coat before being genotyped using a custom genotype array. Aliquots of plasma, buffy coat and residual DNA will be sent to our ultra-low temperature biobank for long-term storage.

We will design the blood sample logistics, processing, and genotyping pipeline with flexibilities, redundancies, and integrations throughout. Quality management, quality control and data security will be embedded in this pipeline with a variety of security standards and ISOs requiring certification or self-declared conformity. This pipeline will deliver the real-time throughput needs of Our Future Health and will provide quality-checked, called genotype data that will flow back to Our Future Health via a secure API.

We will conduct imputation on the genotyping data to expand the number of genetic markers using UK reference panels. Imputed genotypes will help us to deliver each of our programme's four specific aims, providing a resource for basic and translational research, as well as the basis for generating integrated risk scores and other genetic information which may be returned to participants. To ensure that this information can be generated, offered and delivered to participants who choose to receive it in a timely manner, this bioinformatics pipeline will include the calculation of polygenic/integrated risk scores. This will enable the programme to be able to offer such information to participants as well part of the ways in which participants can be identified and subsequently invited to join additional studies, each of which will have their own REC approvals and materials.

### 4.1. Repeat blood samples

There are strong scientific rationales for collecting repeat blood samples. These include the ability to investigate:

- Age-specific biomarker thresholds to improve risk prediction or provide individualised baseline levels
- Age-specific biomarker changes to improve risk prediction or provide for proxy endpoints
- Specific time-windows proximal to disease when the likelihoods of earlier detection and improved intervention are high
- More accurate classification through correction of regression-dilution bias
- Random variation and patterns of variation (e.g. seasonal, menstrual, etc.) of biomarkers in healthy, asymptomatic and symptomatic patients
- Natural history of disease
- Discover biomarkers in a disease course that are predictive of treatment response

We are committed to collecting repeat blood samples to cover this research gap. Repeat blood samples will enhance the resource, broaden scientific opportunities, and help future-proof the scientific utility of the Our Future Health programme by enabling diversity and evolution of the types of blood samples collected.

We will continue to work with all stakeholders – including our participants and all of our boards outlined in the attached Governance Manual (**Appendix F**) – to design and deliver a programme of repeat blood samples.

## 5. Data linkage

Linkage to health-related data is a central component of the Our Future Health programme, forming part of the core cohort dataset. We will link to and/or store data that is controlled by third parties, and provide that data in de-identified form to researchers. Agreeing to link to data held by third parties will be a requirement of joining the programme.

### 5.1. NHS data linkages

We will link all Our Future Health participants to the country-specific central demographic register. This will confirm electronic identification of the participant and enable the first high priority set of data linkages that will include primary care, secondary care, cancer data, and death data.

Demographic registers for the UK include the Personal Demographics Service (PDS; England), the NHS Central Register (NHSCR; Scotland), the Welsh Demographic Service Dataset (WDSD; Wales), and the Health and Social Care Northern Ireland (HSCNI; Northern Ireland). Once a participant is linked with the demographic register this will enable us to search priority linkage datasets for any records each participant may have.

High-priority NHS datasets for the Our Future Health programme include:

- a. **Primary care data:** General practice data are an essential component of our programme to provide a detailed picture of a participant's health. These data include exposures, phenotypes, diagnoses, and prescriptions/dispensing of medicines.
- b. **Secondary care data including hospital admissions:** Secondary care data provide detailed records of hospital outpatient and inpatient visits, surgeries, and procedures that are an essential component of understanding a participant's health status, diagnoses, and progression/regression of disease.
- c. **Cancer registration data:** Cancer registration provides almost complete capture of cancer diagnoses in the UK. These datasets provide patient and tumour level information including pathology reports, molecular testing results, treatment records, and hospital activity records.
- d. **Death registration data:** Vital status, date of death, underlying cause of death, and contributory causes of death are essential data for any study within the Our Future Health programme.

We anticipate that following these high priority NHS data we will also explore linking to disease/service specific registries (e.g. cardiac disease, kidney disease, intensive care), coronavirus infection, coronavirus vaccination, imaging (e.g. Diagnostic Imaging Data Set [DID]), costings, pre-cancers (e.g. UK National Barrett's Oesophagus Registry [UKBOR]), and maternity (e.g. Maternity Services Data Set [MSDS]) amongst others.

### 5.2. Other linked data sets

Once the initial high priority set of linkages have been completed, we will further explore additional linkages to datasets such as census, education, welfare, employment, environment etc. Our ambition is to build a complete picture of health in our programme to enable comprehensive research studies to be conducted.

## 6. Feedback of health-related information to participants

### 6.1. Background to issues around feedback

In any new longitudinal cohort study, decisions need to be made about whether and how any personal results or data are to be made available to participants. These may include lifestyle-based findings, results from physical measurements, imaging, analysis of biological samples or genetic results, or indeed all the (raw) data that is obtained about a participant.

Debates around whether and how personal genetic results should be returned to research participants is not new. In 2013, the American College of Medical Genetics (ACMG) released its recommendation that secondary findings from a list of (at the time) 56 genes should be offered to research participants having clinical exome or genome sequencing.<sup>59</sup> In 2014, the Clinical Sequencing Exploratory Research (CSER) Consortium and the Electronic Medical Records and Genomics (eMERGE) Network in the US agreed that, in most circumstances, if results meet an actionability threshold for return and the research participant has consented to return, genomic results, along with referral for appropriate clinical follow-up, should be offered to participants.<sup>60</sup>

In recent years, it has also been argued, in the context of clinical trials, that research participants should be empowered to have ownership of their own data.<sup>61</sup>

### 6.2. Rationale for participant feedback

We will offer participants feedback of personal health-related results (including lifestyle and genetic results) if they wish to receive them and consent to feedback. The rationale for offering feedback is that:

1. It is useful for participants to have received some specific personal disease risk results if they are to be approached for risk-stratified enrolment in deep phenotyping and/or risk-stratified prevention trials.
2. It will provide the opportunity to provide much-needed empirical evidence on the delivery and outcomes of novel types of risk information that will provide valuable insights to inform healthcare delivery and policy.
3. It may be a reasonable 'value exchange' for at least some participants who might see feedback as a personal benefit from taking part.

---

<sup>59</sup> Green et al (AJHG, 2013) ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. <https://www.nature.com/articles/gim201373>.

<sup>60</sup> Jarvik et al (AJHG, 2014) Return of genomic results to research participants: the floor, the ceiling, and the choices in between. <https://www.sciencedirect.com/science/article/pii/S0002929714001815>

<sup>61</sup> S Terry & P Terry (Science Translational Medicine, 2011) Power to the people: participant ownership of clinical trial data. [https://stm.sciencemag.org/content/3/69/69cm3.short?casa\\_token=Fxcu6zimlKcAAAAA:BJdgYy7NtgZc18chJ-nTpFOhfW7KYaR5sCcV49j8XB7eLv5qO9FLr-jJlVmlZhHEwLvEWe9xr3X\\_iw](https://stm.sciencemag.org/content/3/69/69cm3.short?casa_token=Fxcu6zimlKcAAAAA:BJdgYy7NtgZc18chJ-nTpFOhfW7KYaR5sCcV49j8XB7eLv5qO9FLr-jJlVmlZhHEwLvEWe9xr3X_iw)

### 6.3. Framework for participant feedback

The Ethics and Feedback Advisory Group (EFAG) considered the issue of feedback in considerable detail. They highlighted a number of important considerations which form a framework for how we will develop a feedback policy:

1. It is good practice that there should be immediate feedback of key results from measurements at recruitment, for example, BMI or blood pressure. However, while the concept of providing clinically significant feedback on an ongoing basis to participants may initially be attractive, the practicalities should not be underestimated.
2. Providing personal health information of uncertain clinical validity or utility should be approached with caution, and recommends it should only be provided if participants give additional, specific consent to receive it.
3. There needs to be consideration of any feedback offered to participants on the resource implications for the NHS (GPs in particular) and that even a few cases of confusion or anxiety could lead to a damaging public impression of Our Future Health.
4. Providing access to the raw data about themselves (either on request, or routinely) can appear attractive, though can be problematic and care should be taken in considering interpretation. Providing access to vast amounts of uninterpreted information creates a risk that erroneous medical implications will be deduced, and leave participants overwhelmed and vulnerable.
5. It should not be assumed that everyone will be motivated by receiving individual feedback and therefore the offering should be carefully piloted and revisited and refined over the course of the cohort.

Guided by the Ethics and Feedback Advisory Group, our founding principles for how we approach all individual-level information access, including genomic information, include: providing participants with a choice; assessing the benefits and harms; having an explicit purpose for providing feedback which can be clearly explained to participants; careful communication; and ensuring adequate long-term clinical support for those receiving individual-level genomic results.

We recognise the challenges and complexities associated with providing information to participants. It will require extensive consultation, exploration, and discovery with our governance boards and members of the public to formulate a policy and procedure. It will need to be pilot tested before being widely deployed and be shown to contribute to the health maintenance of participants, while not contributing unsustainably to primary or secondary care workloads.

We will submit a future amendment detailing our approach to providing individual-level genomic information to participants.

### 6.4. Examples of participant feedback

Below are some examples of the types of feedback that are being considered:

- (a) Questionnaire insights are under development and may include comparisons to healthy recommendations, comparisons to the population, as well as resources participants may wish to consult.
- (b) Baseline physical measurements (blood pressure, height, weight, waist circumference) as well as baseline blood measures of cholesterol. These measurements are not intended to be a 'health check' and are not, for now, to be shared with their GP or the NHS. However, we are working with the NHS and the Office for Health Improvement and Disparities to explore potential options for sharing these data with the NHS in the future. We will provide written feedback to participants about their blood pressure and heart rate on a proforma, with advice to contact NHS 111, a Pharmacist or GP if they are abnormal or if they have any questions about their results. This is necessary as there will rarely be measurements that warrant medical attention within days. We will also direct participants to the NHS website on blood pressure, where they can obtain further information.
- (c) Being approached by NHSBT if their genotypes suggest that they have less common minor blood group antigens that are underrepresented among NHSBT blood donors. This may be particularly beneficial for minority groups for whom NHSBT sometimes has difficulty providing optimally matched blood products.
- (d) Providing participants with increased access to information stored in their NHS health care records.
- (e) Integrated risk scores that provide disease risk estimates. Leading integrated risk score candidates that may initially be feedback to participants include cardiovascular disease,<sup>62</sup> age-related macular degeneration, glaucoma, and type II diabetes. Each of these integrated risk scores provides good risk prediction and are clinically actionable.
- (f) Other examples include risk predictions based on common genetic variants associated with risk of iron overload, deep venous thrombosis, or that influence the efficacy or side effects of prescription drugs.

## 6.5. Assessment of outcomes of participant health-related feedback

One of the key rationales providing participants with health-related feedback is to provide valuable insights into the delivery and outcomes that can inform healthcare delivery and policy. As the Ethics and Feedback Advisory Group noted in the Ethics & Governance Framework (page 27), there has been very little research to explore the implications or value of receiving research findings, and Our Future Health has an opportunity to provide evidence and set best practice.

In the genomics field, examples of the limited research in this area include the Mi-Genes study in the US<sup>63</sup> and the FinnGen study in Finland.<sup>64</sup>

---

<sup>62</sup> Natarajan et al (*Circulation*, 2017) Polygenic Risk Score Identifies Subgroup With Higher Burden of Atherosclerosis and Greater Relative Benefit From Statin Therapy in the Primary Prevention Setting <https://www.ahajournals.org/doi/full/10.1161/circulationaha.116.024436>

<sup>63</sup> Kullo et al (2016) Incorporating a Genetic Risk Score Into Coronary Heart Disease Risk Estimates Effect on Low-Density Lipoprotein Cholesterol Levels (the MI-GENES Clinical Trial). <https://www.ahajournals.org/doi/full/10.1161/CIRCULATIONAHA.115.020109>

<sup>64</sup> Widen et al... Ripatti (18 September 2020, medRxiv) Communicating polygenic and non-genetic risk for atherosclerotic cardiovascular disease - An observational follow-up study. <https://www.medrxiv.org/content/10.1101/2020.09.18.20197137v1>



We are developing a plan for the assessment of participant-reported outcomes and experiences of receiving individual-level information (e.g. distress, anxiety, comprehension, medication initiation/adherence, 'lifestyle' behaviours) as well as for assessment of health care recorded outcomes (e.g. medication prescribing, hospital appointments, cholesterol levels).

## 7. Re-contacting participants for future research

### 7.1. Background on re-contact

Many previous prospective cohorts, including UK Biobank, allow for participants to be re-contacted for further sampling, questionnaires and other activities such as additional studies. However, most previous cohorts have not been set up in a way that allow for participants to be re-contacted on a risk-stratified basis.

For example, although 1% of UK Biobank access applications are to re-contact participants into third-party studies, UK Biobank participants “consented on the understanding that no results would be fed back to them following their assessment visits” and so “care is taken to ensure that re-contact does not represent implicit feedback of which participants are not aware”.<sup>65</sup> UK Biobank further states that “recruitment based on genotype or on phenotype that is not explicitly self-reported by the participant is highly restricted”.<sup>66</sup>

There is however some precedent for risk-stratified recontact (recall-by-genotype/phenotype) approaches to re-contact, for example in East London Genes and Health (ELGH)<sup>67</sup> (for specific example see ELGH familial hypercholesterolemia ‘genotype-first recall’ study<sup>68</sup>), and in NIHR BioResource<sup>69</sup> (for specific example see the IBD BioResource Protocol<sup>70</sup>).

### 7.2. Rationale for re-contacting participants

The development of new diagnostic tests and approaches is often limited by the lack of availability of samples from people with and without a specific disease. Different technologies require different types of samples; standardised, quality-assured processes for sample collection and storage can be critical to evidence generation, yet are rarely assured. In addition, samples collected and processed for one application or technology are often not suitable for others.

The need to establish studies at the scale required, tracking participants and data over long timeframes, whilst maintaining adequate levels of follow-up, limits industry and academic engagement in early detection research. Disease-specific cohorts are difficult to scale and are often not available for a range of other conditions. This is highly inefficient in its use of participants and samples.

---

<sup>65</sup> Conroy et al (2019) The advantages of UK Biobank's open-access strategy for health research. <https://onlinelibrary.wiley.com/doi/full/10.1111/joim.12955>

<sup>66</sup> UK Biobank's re-contact procedures for third party researchers. <https://www.ukbiobank.ac.uk/wp-content/uploads/2018/05/ukb-recontactprocs-14.3.2018-item-5b-2.pdf>

<sup>67</sup> Finer et al... & van Heel (2019) Cohort Profile: East London Genes & Health (ELGH), a community-based population genomics and health study in British Bangladeshi and British Pakistani people. <https://academic.oup.com/ije/article/49/1/20/5555939>

<sup>68</sup> ELGH genotype-first recall to study extreme genetic risk of atherosclerotic cardiovascular disease. <http://www.genesandhealth.org/research/research-studies-approved/s00013-genotype-first-recall-study-extreme-genetic-risk-atherosclerotic>

<sup>69</sup> <https://bioresource.nihr.ac.uk/>

<sup>70</sup> <https://www.ibdbioresource.nihr.ac.uk/wp-content/uploads/2017/01/protocol-V6.pdf>

There are two key reasons why a risk-stratified approach to re-contacting participants based on genotype or phenotype will substantially increase the value of Our Future Health as a research resource:

1. Having the cohort risk-stratified for a wide variety of diseases and consented for re-contact based on risk profiles will open up the possibility of enrolling participants at high risk in more detailed secondary protocols involving additional biological samples collection, imaging or other detection methods. This will substantially help improve the potential success of research on early detection, diagnosis, prevention and treatment of disease.
2. Being able to re-contact cohort participants on a risk stratified basis will allow researchers to test hypotheses regarding the benefits of early detection and early diagnosis. Enabling researchers to re-contact participants based on their risk will support research on whether focusing early detection methods such as screening on those at high risk increases their efficiency and cost-effectiveness. Research will also be supported exploring whether preventive measures could be applied before early signs, symptoms or diagnoses of advanced disease.

### **7.3. Planned approach for re-contact**

Our Future Health will recontact participants to invite them to further sampling, questionnaires and other activities such as additional studies which enhance the cohort and are being organised by the programme.

#### **7.3.1. Inviting participants to third party research studies**

We are currently developing our Access Process and this includes the specifics of our strategy for re-contact of participants with particular reference to the invites to third party studies. We are working with our Public Advisory Board members to write and define the re-contact policy.

The Access Board will manage applications for researchers (academic or industry) to re-contact participants eligible for further studies e.g. deep phenotyping or prevention trials. A model for this for some genetic disease predisposition variants has been pioneered through the NIHR BioResource with respect to referral to Clinical Pathways.

## 8. Participant communication and withdrawal

Following recruitment, participants will be able to set preferences related to how they are contacted by the research programme. Participants will be able to decide how they are contacted: by email, by text message or by letter. Participants will be able to choose:

- “Yes, you may contact me”
  - Participants decide how to be contacted (text message, email, letter).
- “No, you may not contact me”
  - Our Future Health will no longer contact participants with the exception of essential service messages (a partial withdrawal).

At the time of consenting, participants will be informed that they can withdraw from the programme at any time without providing a reason. If a user chooses to withdraw from Our Future Health, they will be given two options:

- “No further access”
  - Our Future Health will no longer contact the participant and delete all identifiable data from the participant record
  - Our Future Health will retain permission to use de-identified information and samples provided previously
  - Our Future Health will unlink external datasets, but will retain historical data in a de-identified form
- “No further use”
  - Our Future Health will no longer contact the participant
  - Our Future Health will not obtain any further information on the participant and will destroy all data and samples related to this participant (with the exception of existing models or analyses that were created using their de-identified records)\*
  - Data and samples for the participant will not be made available for new research projects.\*\*
  - It won’t be possible to remove their data from any research that took place before the withdrawal.

\*Participants are told that it may not be possible to trace all distributed sample remnants for destruction. Such a withdrawal will prevent information about them from contributing to further analyses, but it would not be feasible to remove their data from analyses that had already been done.

\*\* Participants are also told that this may not happen immediately as data is made available to researchers in a series of releases; we anticipate that new releases will be made on a quarterly basis. Their data will not be included in new releases but there could be a period of up to three months when their data are included in the current release of the data.

Participants who do not have internet access will contact the Our Future Health support team by telephone to request a withdrawal form. Once received and verified, an approved member of the team will make the required changes in the database and, where appropriate, initiate data deletion and sample destruction.

## 9. Data access

### 9.1. General approach and principles

Researchers will apply to the Our Future Health Access Board to access data, samples or to re-contact participants recruited through the research programme. The principles detailed below have been developed with public members of our Access Board, the Ethics Advisory Board and the Founders Board.

Further information about the Access Board, including all of the relevant documents (once approved by the Access Board when it is appointed) will be available on the Our Future Health website.

#### 9.1.1. Key principles guiding access

1. The access procedures will be as simple as possible, and the decisions will emerge in a timely fashion. The objective is to maximise responsible use of the dataset, not to unduly guard it for the benefit of a restricted user group.
2. The key principles underpinning these Access Procedures are the granting of data, samples and/or re-contacting of participants i.e. the Our Future Health 'Resource' to suitable research projects and to ensure that in this manner, the Resource is used extensively, in a responsible and useful way to benefit society as widely as possible.
3. Access to the Resource will be underpinned by the principles of the "[Five Safes](#)": safe data, safe projects, safe people, safe settings, safe outputs.
4. The Our Future Health Access Board is responsible for access to the data, samples, and participants. All data policies, access policies, and information relating to how the data is managed and accessed has been made publicly available to ensure transparency and disclosure.
5. These Access Procedures reflect the value of the Resource and the undertakings given to the participants when they joined the programme.
6. Our Future Health will continue to interact with participants, researchers, and society in general to maximise engagement and interest throughout the Resource's lifetime (which is intended to be some decades) and ensure that the research projects that are taking place as well as the findings that result from those projects are publicised with a view to generating further interest and maintaining the initiative's momentum.
7. Researchers who are granted access to the Resource for an approved research project will be required to return their results to Our Future Health and to publish their findings so that other researchers can use and build on this knowledge to further benefit the public interest (public health benefit). Full details will be included in the Our Future Health Publication Policy.
8. In order for a research project to be approved, the Researcher will need to demonstrate that their research will provide knowledge, further scientific understanding and that it meets our definition of public health benefit.
9. The process for applying to use the Resource has been designed to be efficient but robust. Data and/or samples will be provided in an expeditious manner once projects are awarded and the required documentation has been signed and approved, to enable research to begin in a timely manner.
10. Our Future Health will maintain an up-to-date list of Registered Researchers and their affiliations. Organisations conducting research studies will ensure compliance with security and information governance accreditations, as determined by Our Future Health over time.

## **9.2. Researchers and organisations**

The registered researcher process is modelled on ‘safe people’: researchers are trained and authorised to use data safely. Our Future Health will maintain an up-to-date list of registered researchers and their affiliations. Organisations conducting research studies will ensure compliance with security and information governance accreditations as determined by Our Future Health over time, such as the NHS Data Security and Protection Toolkit. Researchers must not attempt to re-identify individual participants.

Researchers can only undertake their research in a trusted research environment. The results of studies run using Our Future Health will be returned to the Resource.

We will maintain an up-to-date list of registered researchers and their affiliation. Organisations conducting research studies will ensure compliance with security and information governance accreditations as determined by Our Future Health over time, such as the NHS Data Security and Protection Toolkit

## **9.3. Dissemination of results**

Researchers who use Our Future Health will be required to disseminate the results of their research as rapidly and widely as possible, subject to ethics and confidentiality considerations. They will be encouraged to discuss their research findings with other scientists and the public, and to share relevant data and materials as openly as possible. Researchers who have had access to samples will be required to provide details of the assay techniques used and return the results to the Our Future Health resource within 9 – 36 months of approval. A limited delay prior to the return of findings to the Our Future Health TRE will be permitted (depending on the party’s membership status, the exclusivity period will be between 9-36 months and will require review by the Access Board) in order to e.g. enable a paper to be published; a patent to be filed; or other competitive advantage to be pursued. Users will be required to undertake to notify Our Future Health in advance of publishing such findings, to acknowledge the contribution of the resource, and to provide a copy of any published reports. In addition, researchers will be required to provide Our Future Health with a copy of all of the results of their research based on the resource (including any negative findings and relevant supporting data) for incorporation into the central database.

## **9.4. Access agreements and fees**

As a condition of access to relevant data (i.e. assay results, physical measures, or questionnaire responses) from the resource, the approved researcher would be required to enter into an access agreement with Our Future Health. Researchers will need to detail the purposes for which they want to use the data, this includes hypothesis-driven and non-hypothesis driven research studies, and standard terms relating to exploitation and dissemination of results.

Similarly, when samples are provided to a laboratory for assays, a materials transfer agreement will require that the samples are used for the agreed purposes only and that the results of the assays are returned to the Our Future Health platform within specified time limits.

Information identifying participants will be removed before any data or samples are released, and the agreements will include an undertaking not to attempt to identify participants. We will generally permit exclusive use of the relevant data set for a limited period from its release in order

to allow time for the approved researcher to conduct and report the agreed analyses. Subsequently, the results will be incorporated into the resource database for use by other approved researchers.

Access to the resource will not be permitted for police use, except where required by court order, and the Our Future Health platform will resist access for this use (in particular by seeking to be represented in all court applications for such access). A system for monitoring compliance with the terms of the access agreement will be put in place before the resource becomes available for access, and a policy developed for dealing with non-compliance (e.g. restrictions on future access).

It is anticipated that a data access fee will generally be charged for access to the Our Future Health platform. This fee will differ between commercial users and academics or charities. The chief aim of this fee will be to cover the costs of any sample and/or data retrieval, preparation and analysis required for the particular research use and to help cover the costs of maintaining the resource for future users. This fee may be waived for initial funders of the programme and will be scaled to ensure those who have contributed are at a financial advantage when accessing the resource compared to those who have not participated in the funding round. The Board will determine a fee structure which, is in keeping with Our Future Health's charitable status. In addition, researchers will be billed their own cloud computing and storage costs within the TRE, as is happening with new TREs set up by UK Biobank and Genomics England. This allows for more intensive research techniques such as machine learning but leaves the choice to the researcher as to how much to spend.

## 9.5. Trusted research environments

The success of Our Future Health rests on the research and science that is conducted with the data and the cohort, but this must be balanced with strict security and confidentiality commitments we will be making. The best way to balance these needs will be to provide a trusted research environment within which to access de-identified data.

The evolving policy landscape in the UK is moving towards the use of trusted research environments (TREs) as safe shared spaces for data analytics that prevent the data from leaving. The UK Health Data Research Alliance (representing 33 major health data organisations in the UK including UK Biobank, the NIHR BioResource, NHS England, NHS Digital, NHSX, Public Health England, Genomics England and many key hospitals and charities) "is committed to an approach to data access based primarily around trusted (trustworthy) research environments; with appropriate robust and independent TRE accreditation, monitoring and auditing,"<sup>71</sup> and this commitment is echoed in the new UK genomics strategy.<sup>72</sup> NHS Digital have committed to shifting 80% of data access to their Data Access Environment (or other NHS Digital accredited secure data environments) by 2022,<sup>73</sup> and Genomics England have launched a trusted environment for all their researchers to use.<sup>74</sup>

All research will therefore be conducted within an appropriate accredited TRE, and each study will have a separate allocated workspace within the TRE. During the pilot phases we have been

---

<sup>71</sup> Trusted Research Environments (TRE), A strategy to build public trust and meet changing health data science needs, Health Data Research UK Green Paper, July 2020

<sup>72</sup> Genome UK: the future of healthcare, UK Government Office for Life Sciences, Sep 2020

<sup>73</sup> Improving our Data Processing Services (DPS), NHS Digital, Feb 2020

<sup>74</sup> Genomics England launches next-generation research platform central to UK COVID-19 response, Jun 2020

working with the NHS Digital Data Access Environment as a TRE accessible just to a small number of Our Future Health staff, but we aim to launch our main TRE for approved studies at the end of 2022, with a beta period beforehand. The TRE is the subject of a procurement in 2022, and will be a significant part of our technology stack, providing a wide variety of computation and data storage resources as well as analysis tools for clinical, genetic and other data types, to serve the needs of researchers from many disciplines.

The technical architecture and systems are explained in following sections, but here we illustrate the main approach to enabling research access to data. This must be read in conjunction with the Ethics and Governance Framework.

- Research data will be accessible **within accredited trusted research environments**. A trusted research environment is a secure online environment that allows researchers to access data and perform analysis or computation. No participant-level data can be exported from a trusted research environment.
- Accreditation of a trusted research environment will be **reviewed periodically** by Our Future Health. The accreditation criteria will apply **equally to all research environments**, and will include security, data governance, operational, technical, confidentiality and access control requirements as well as data licensing provisions from data controllers. Trusted research environment accreditation will include requirements and controls from accreditations such as ISO27001 Annex A, and from relevant UK data protection regulations such as UK General Data Protection Regulation (UK GDPR). Controls will assure the environment and its boundary; operational, technical and scientific staff; and organisational processes.
- Access to data held in trusted research environments will be limited to **registered researchers**
- Data will be made available within accredited trusted research environments for research projects that are approved for **fixed and agreed periods of time**
- Trusted research environments will maintain end-to-end verifiable logs of all aspects of their operation, including researcher access and activity; cloud and data usage; and data imports and exports.
- The Our Future Health trusted research environment will be **flexible** in providing a wide range of researcher-friendly tools for data science and analysis, including the ability for researchers to bring **additional datasets, libraries and code**, and scale their use of **chargeable cloud computing and storage**
- Researchers may export **results** from trusted research environments such as aggregate data, summary level statistics, visualisations, parameters and trained models.
- The TRE will have controlled ways to import code, tools, libraries and additional datasets, and to export code, tools and libraries. The export of individual level data is not permitted, but aggregate data, summary level statistics, visualisations, parameters and trained models can be exported. In order to manage this export in a safe way (the “safe outputs” principle from the “Five Safes”), we will initially create an “airlock” process with an airlock manager role to perform manual review, and during 2023 we will be investigating how to create a process that can scale up appropriately.
- We will accredit registered research or healthcare organisations which will enable researchers who are employed or contracted by these institutions to become registered researchers subject to them being verified as a bona fide researcher.



## 9.6. Data to be shared for research

- All data made available will be robustly **de-identified** to protect the privacy of participants while maintaining its scientific and research value
- Comprehensive **data dictionaries** and metadata will be maintained by Our Future Health and made available for researchers to use. **Provenance** & traceability of all data items will be recorded and made available for researchers
- Wherever possible data will be structured and coded using commonly used **standards to allow for broadest use of the data, in the public interest**
- Participants can choose to **withdraw** from the programme at any time. Ongoing use of their data will be dependent on their individual consent at the point of withdrawal.

## 9.7. Accreditation of TREs

We are undertaking detailed design work to inform the procurement of our own TRE and the accreditation process for all TREs that will be authorised to hold a copy of the Our Future Health data. As part of this, we are:

- Researching existing accreditation standards and frameworks
- Receiving expert input from data governance, cyber security, and legal experts
- Conducting a consultation exercise with: relevant industry and charity organisations who may wish to apply for accreditation; organisations such as HDRUK and the NHSX Data Strategy team to ensure that we are informed by emerging policy and standards for TREs; other research programmes such as Genomics England and UK Biobank; and other NHS organisations
- Engaging with the public via our Public Advisory Board and through a public consultation
- Working with regulatory and statutory bodies such as the Information Commissioner's Office (ICO) as part of our involvement in the "regulatory sandbox" programme
- Working with our advisory and governance boards
- Engaging with a range of researchers to co-develop functionality and usability to ensure the Our Future Health TRE and tooling delivers against their needs and requirements

We have committed to publishing an initial version of our accreditation criteria in June 2022, and publishing a final version in September 2022.

## 10. Digital data and platform

### 10.1. Overview

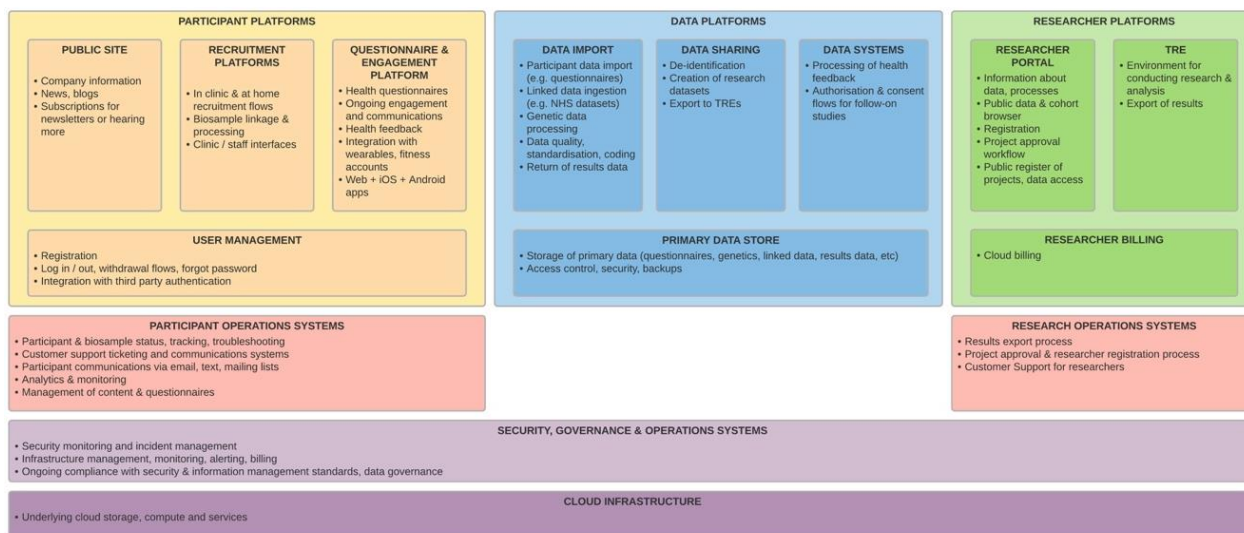
The data that are to be used by Our Future Health are of the highest sensitivity and, as such, need to be handled with the greatest care. Security is a prime concern. It is essential that Our Future Health is compliant with the requirements of relevant legislation, such as the Data Protection Act and the UK GDPR. We will not be able to gain access to the broad range of third-party datasets required, or be able to provide validated research data, if these external requirements are not considered. Key aspects of the controls required include identity and identifier management, ensuring the accuracy of the data collected, inclusion of comprehensive audit data (such as the staff and equipment involved in data collection) and strict controls on data access.

The following sections describe the first post-pilot iteration of the system as envisaged, but do not include the technical design for providing feedback to participants or for recruitment into stage 2 studies. The functionality for these future iterations will be developed during discovery and design phases over 2022-2023.

### 10.2. Systems architecture

The Our Future Health system is designed to meet the needs of participants, researchers and operational Our Future Health staff, and as such it comprises a number of primary components. The diagram below (**Figure 4**) shows a high-level system map of the components within the Our Future Health architecture (it does not include any partner or external components).

**Figure 4. High level system architecture – primary components**



#### 10.2.1. Participant platforms

The participant platforms are all the systems facing our participants. They will be scaled as participant numbers grow, and will need to serve the needs of a diverse community. We will aim for an AA standard of accessibility according to the Web Content Accessibility Guidelines (WCAG 2.1) to

ensure we can serve the needs of people with disabilities, and we will be adding an internationalisation layer to allow translations into common UK languages. To quote from the gov.uk Service Manual<sup>75</sup>, the WCAG guidelines explain how to make digital services, websites and apps accessible to everyone, including users with impairments to their:

- Vision - like severely sight impaired (blind), sight impaired (partially sighted) or colour blind people
- Hearing - like people who are deaf or hard of hearing
- Mobility - like those who find it difficult to use a mouse or keyboard
- Thinking and understanding - like people with dyslexia, autism or learning difficulties

As per the prior section on freephone information service, we will also offer the ability to complete the questionnaire via telephone.

The recruitment platforms manage the first contact with participants including registration, consent, booking of appointments and sample processing. This will be a flexible set of components to manage multiple recruitment routes with differing flows and requirements (for example, consent could be in a clinic or on a participant's device prior to a visit; there will be different ways to link a sample to a participant in a hospital compared to an NHSBT centre). As well as managing interactions with participants, the Recruitment Platforms will include interfaces for nurses, health assistants and staff in clinics. There will be multiple integrations with recruitment partner systems as well as sample processing labs. These platforms will continue to evolve building on the work we have done in our pilot studies during 2021, and will remain a bespoke set of web applications and outsourced services.

Our public site is our main organisational website and will serve a variety of needs for participants, researchers and stakeholders, providing information about the programme as well as news, and pointers to services for participants or researchers.

The questionnaire & engagement platform will be the long-term system for participants to engage with Our Future Health. It will include comprehensive health questionnaire functionality, ongoing communication with participants via web or app interfaces, the delivery of information or feedback. Any front-end integrations, such as with services for wearable or fitness tracking data, will be via this platform. Over time more modules will be added such as cognitive function testing. This platform is likely to be a customisation of a system we procure during 2022. We expect this platform to include both iOS and Android apps as well as an accessible mobile web experience.

Serving the needs of the platforms above will be a range of user management services including authentication and authorisation, common flows such as email verification and forgotten passwords. This includes the main store of participant login information. It will be connected to other systems using standard API protocols such as OAuth.

The participant operations systems include all of the platforms used by our staff in monitoring and operating the programme. Currently this comprises:

- The Customer Relationship Management (CRM) system can be used to communicate with participants and interested parties, to track the status of sample processing, and to provide operational analytics (it does not contain health data). Currently this is Microsoft Dynamics.

---

<sup>75</sup> <https://www.gov.uk/service-manual/helping-people-to-use-your-service/understanding-wcag>

- The Customer Support system is used by the support team to handle incoming questions and requests from participants via phone and email. Currently this is Freshdesk.
- There will be multiple analytics systems to understand metrics and progress. We use Matomo for web analytics at the moment, that maintains participant privacy as information does not leave our cloud environment.
- The Questionnaire & Engagement Platform it will include interfaces for staff to manage content and questionnaires. In addition, our public site includes a Content Management System (CMS) that is currently Wordpress.

### **10.2.2. Data platforms**

The data platforms form the core of our systems, where data is ingested, processed, stored and shared out. None of these systems are accessible to participants or researchers or our own analytics staff – all interfaces are via other platforms. We anticipate that these systems will remain under our control rather than being outsourced, although we will likely procure specific tools or systems as needed.

The primary data store is the storage for participant data: both the data we originate (like questionnaires and genetic data) and linked data we bring in from elsewhere. This component will see a great focus on security and resilience. The primary data store will separately store identifiable participant data and the participant health data, so that the identity of a participant is not stored together with their health data. A unique Participant ID will be allocated for every participant, although this will be purely for internal usage. There will be no access to this data store other than time-limited and audited access by technical systems administration. All data will be stored encrypted and securely backed up to storage considered off-line, within the UK, using public cloud services.

The data import systems will be a number of pipelines that process incoming data from our participant platforms (such as newly consented participants joining the programme, or questionnaire completions), genetic and operational / quality data from our lab partners, connected to linked NHS datasets, results returned from research studies. Our data engineering and software teams will be working on data cleaning, quality, standardisation, coding within these systems.

The data sharing systems will be created to manage important processes used to provision datasets for research for approved studies within TREs, including robust de-identification. A freshly generated Research Participant ID will be created for each participant, with the linkage back to the original Participant IDs only stored within the secure Our Future Health data store and only available to secure internal systems. Further data systems manage participant withdrawals, and will handle flows such as the processing of health feedback, polygenic risk scores or other personalized information for participants, the authorisation and consent needed for stage 2 studies.

### **10.2.3. Researcher platforms**

The researcher platforms are the systems for the research community to learn about Our Future Health, manage registration and study approvals, and gain access and use the TRE. The researcher portal will be a public web site and set of services with information about the data we have available, documentation about the TRE, the processes for registering and project approval. We anticipate there will be a public version of a high-level cohort browser, to allow researchers to see how many participants could be in a defined cohort (only showing aggregate

data such as participant counts). The process of registering as a researcher and managing the workflow of project approval will happen within this portal. Researchers will be able to log in to this portal to manage their project and gain access to the TRE. The workflow around stage two studies will likely be managed from within this portal. In addition here is where we will be able to show information about approved projects and how data is being used. Finally we will have the environment to support a growing community of researchers, and ways for them to interact with and learn from one another.

Within the TRE, researchers will pay for their own usage of cloud compute and any additional storage they choose to use within their own workspaces. The cloud consumption of all of the users of the TRE will likely dwarf that of the Our Future Health central systems, so this creates a sustainable way for the community of researchers and scientists to grow, and use whichever combination of tools they need to. The researcher billing service will be procured, and we require the supplier to manage the tracking of researcher cloud usage, invoicing, payment processing, and the resulting liability.

The research operations systems are used by Our Future Health staff to operate the research platforms and processes. These will include the management of researcher registration, the approval workflow for studies working with the Access Board, the airlock. There will be a customer support system for researchers, likely through the same platform as used for participant queries, and likely further systems as the research usage of Our Future Health scales.

### **10.3. Security, governance and operations systems**

This layer of the architecture represents systems used by Our Future Health technical staff to operate the various platforms. These will include monitoring, logging, alerting, incident management, billing. In governance terms, this will include policies and processes ensuring compliance with security and data governance standards. We describe our security posture below (section 10.8).

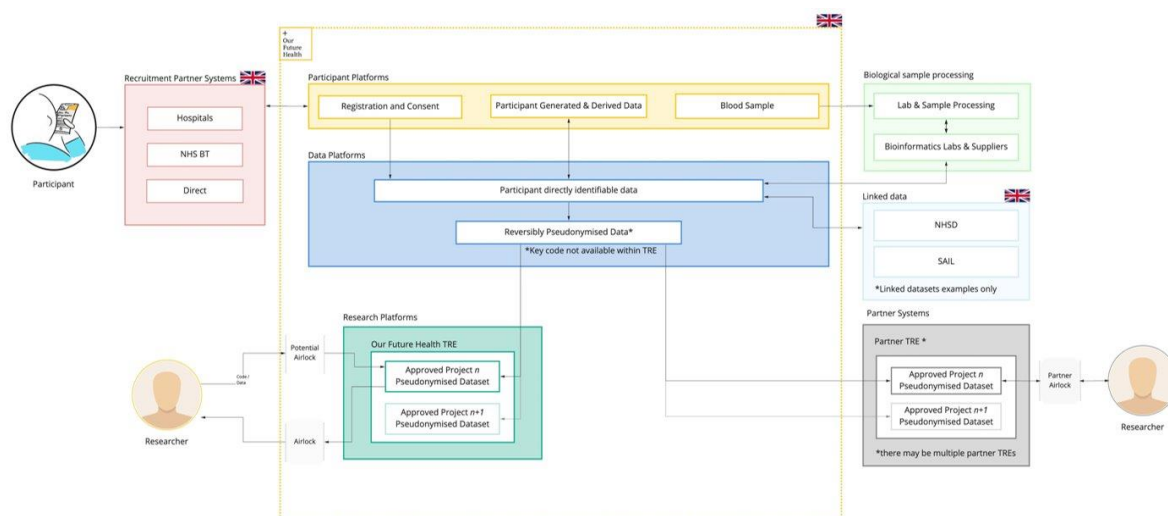
### **10.4. Cloud infrastructure**

This is the cloud compute and storage infrastructure that underlies all of the other services, including the hosting of the primary data store. We will be procuring a provider of public cloud infrastructure during 2022, that will become one of our major technology partners. There will be additional services here such as domain name registration, email sending, certificate management.

### **10.5. Primary data flows**

The figure below (**Figure 5**) illustrates the primary flows of data between the various system components described above and including the journey of blood samples.

**Figure 5. High level data flows**



## 10.6. Consent and participant withdrawal

Participant consent is described above (Section 3.4). From a technical perspective the consent functionality is designed to collect and store:

- Date and time of consent
- Site of consent, if at a physical location
- Details of the device, if known (device type, operating system, browser version, IP address)
- Version of the Participant Information Sheet (PIS) provided to the participant at the time of consenting
- Identifying and contact information provided by the individual
- Level of ID verification performed, if at a physical location

A copy of the consent form and participant information sheet will be sent to the participant. Once consent is provided, the participant will be assigned a unique participant ID which will be stored within the system.

Participants will be able to withdraw from the programme at any time as explained in Section 8 above. In the case of the fullest withdrawal option, where a participant has asked that their data and samples be removed, the system will be able to process deletion of participant data from all relevant systems across the architecture. We will commit to remove their pseudonymised data from all future research dataset releases, but it will be retained in research datasets that have already been distributed for approved studies.

System and operational logs will not contain identifiable information and web analytics will be anonymous.

An operational record will be retained of a participant's previous consent and subsequent withdrawal, for audit purposes, only accessible to very limited technical staff.

The deletion process will take place in batches on at least a monthly basis.

## 10.7. Data linkage

Section 5 above describes the data sets that Our Future Health will be linking with.

From a technical viewpoint, the linking component comprises multiple systems:

- Linkage to central NHS data from England will likely happen within an NHS Digital environment, as part of a pipeline of processes prior to data being loaded into a trusted research environment
- Linkage to other government or NHS data will happen either directly in the Our Future Health cloud environment, within the NHS Digital environment or even within other NHS or government environments, depending on the data sharing agreements and restrictions in place

In either case there will be a partial and iterative process to attempt matching where there is incomplete data, and a procedure for manual intervention and quality control. Over time there will be significant data cleaning and standardisation effort required to deal with differing identity systems across the UK, both different government data sources and the NHS across the devolved nations.

## 10.8. Security and resilience

Information security is a key concern for the programme. The Our Future Health baseline for cyber security utilises the current ISO27001, Annex A controls across the following domain areas:

- **Information security policies**  
Information security is led from upper management, and clearly communicated across the organisation through policy and process documentation that is regularly updated.
- **Organisation of information security**  
Information security is led and championed by Senior Management.
- **Human resource security**  
Robust Joiners, Movers and Leavers processes, that are linked to appropriate levels of vetting and contractual controls for all staff.
- **Asset management**  
Detailed understand of all applicable business information assets and their owners. Clear guidance in place showing appropriate classification and handling of assets based on sensitivity levels.
- **Access control**  
All access management processes, are based on the concept of least privilege, with systems including Role Based Access Control and detailed audit trails and access review processes in place.
- **Cryptography**  
Where sensitive data is processed, transmitted or stored, appropriately strong cryptography is in place to protect data in transit and at rest.
- **Physical and environmental security**  
Appropriate physical security to protect operational environments, and environmental protection to mitigate threats such as power failure and flooding. Where responsibility is outsourced to third parties such as hosting partners, there should be clear minimum requirements set.

- **Operations security**  
Security controls are built into standard operational processes and Business as Usual, across end user computing, back-office systems and management of online services.
- **Communication security**  
All internal and public network communications are appropriately protected based on their sensitivity, including email and collaboration services.
- **System acquisition, development and maintenance**  
Where services include bespoke development, scripting or customisation a robust development lifecycle is used that includes security at all stages.
- **Supplier relationships**  
Use of supply chain is carefully managed and monitored for all vendors, partners, contractors that are utilised in service delivery.
- **Information security incident management**  
All security incidents are managed through formal process, that includes investigation, root cause analysis and remediation where needed. Learning from incidents will be fully shared to enhance the security program.
- **Business continuity management**  
Business continuity that considers people, processes, systems and locations will be designed, implemented and tested regularly to ensure it is fit for purpose.
- **Compliance**  
All applicable compliance requirements will be monitored and updated based on contractual and regulatory changes, and monitoring of the threat landscape that may identify new requirements.

Our Future Health will achieve Cyber Essentials Plus certification and meet NHS Data Security and Protection Toolkit standards in 2022.



## **11. Public and participant involvement and engagement (PPIE)**

### **11.1. Rationale for PPIE**

Involving the public and participants of Our Future Health in the design and delivery of the programme is vitally important to its success. PPIE is important to Our Future Health as it helps us to:

- a) Design recruitment pathways that are feasible, relevant, accessible, and inclusive for the UK public
- b) Improve the process of informed consent for participants
- c) Improve the experience of participating in Our Future Health
- d) Improve the communication of research findings and discoveries to participants, the public and stakeholders and our funding partners
- e) Include the voice of people and communities that have traditionally been excluded or omitted from health research.

### **11.2. PPIE strategy**

We are committed to upholding best practice in PPIE, in accordance with the UK Standards for Public Involvement (NIHR 2021). We will do this by adopting the following principles as part of our PPIE strategy:

Clarity:

- Being clear about our definition of PPIE and what it means to Our Future Health.

Oversight:

- Ensuring that important decisions about the participants of Our Future Health are overseen by our PPIE representatives through our Public/Participant Advisory Boards and wider PPIE contributor networks.
- Ensuring that the public/participant voice is represented throughout Our Future Health, with nominated PPIE representatives on advisory boards with additional public/participant representatives invited to help shape our public/participant facing processes.

Transparency:

- Publishing our PPIE charter on our website including details of our PPIE Advisory Board representatives.
- Regularly measuring and evaluating the scale, impact, and value of PPIE activity within Our Future Health.

Opportunity:

- Adopting an inclusive approach to PPIE as an organisation by sharing PPIE opportunities widely on a variety of accessible platforms.
- Supporting our PPIE representatives by providing training and encouraging continuous feedback on our PPIE activities.

### 11.3. Summary of PPIE work

In 2020-2021, we involved over 200 members of the public in the design and development of the public-facing materials as well as in other aspects of the programme design. Much of this work was carried out with Claremont, a behaviour change communications agency with considerable experience in the health care and health research sectors. Activities included:

- 4 focus groups, 2 co-design meetings and 21 interviews with the public to develop the scientific protocol
- 12 interviews with a variety of stakeholders from charities and existing cohort studies
- 2 focus groups with 11 NHS primary care staff
- 18 focus groups, 10 co-design meetings, 21 interviews with the public to co-develop the participant information sheet, consent form & other public-facing videos & materials
- 21 interviews to understand the role of industry in health research
- 4 focus groups to explore insights around recruitment methods
- 3 focus groups to explore public motivators and feedback preferences
- 1 member of the public attended a REC meeting with a member of Our Future Health staff

Outputs from this work included:

- Co-designed and REC approved PIS and Consent form
- Co-designed explainer videos for Our Future Health (YouTube)
- Public Engagement Strategy (Claremont 2021)

During 2021 we have continued to demonstrate our commitment to PPIE with the following activities:

- 3 focus groups and 4 interviews with members of the public and health care professionals, providing insights into research recruitment in hospitals.
- In-person observations and interviews with patients and staff working in primary care and blood donation sites at NHSBT to learn more about their lived experiences and preferences for invitation into the programme. This work also helped to identify key national PPIE groups and representatives that we needed to build relationships with.
- We have appointed and trained 22 members of the public to our advisory boards and working groups (11 PPIE representatives for the Public Advisory Board, 2 representatives for the Secondary Care Working Group, 2 representatives for the Primary Care Working Group, 2 representatives for the Ethics Advisory Board, 2 representatives for the Technology Advisory Board, and 3 representatives for the Access Board), with a further list of over 40 volunteers for future user testing and PPIE activities.
- Commissioning of a public consultation (with research agency Kohlrabi Consulting) with 34 members of the public. This work helped to provide insights for our communication of access and storage of participant data.
- A PPIE training package for new public representatives

### 11.4. Development of public-facing materials

We will ensure that public-facing materials for Our Future Health are developed and co-designed with members of the public. As noted above, the written text and scripts for our existing materials were tested and reviewed as part of our PPIE programme. We previously sought REC approval for all the first versions of these written materials (January 2021):

- A short introductory video designed to reach a broad range of mainstream audiences
- Posters to increase awareness of Our Future Health in public spaces including in NHS sites and across a range of community locations
- A short leaflet about Our Future Health designed to be distributed to members of the public
- Invitation letters
- Five 'explainer' videos to support the consent process and make the information provided in the participant information sheet more accessible to potential participants, particularly those with lower reading ages or lower interest in reading written text
- The participant information sheet

### 11.5. Participant-reported experiences

We will administer a quantitative survey instrument/questionnaire to assess participant-reported experiences (PREs) to understand key indicators such as satisfaction, attitudes and understanding. The participant-reported experience measure (PREM) survey instrument will include the following measures:

- Satisfaction
  - e.g. with the information provided, website, consent process
- Informed choice
  - Attitudes/values
  - Knowledge/understanding
  - Decision
- Communication about the programme to others
  - e.g. family members, friends, GP, other healthcare professional
- Motivation
  - Reasons for taking part and completing various elements of the programme
  - Reasons for booking blood appointments

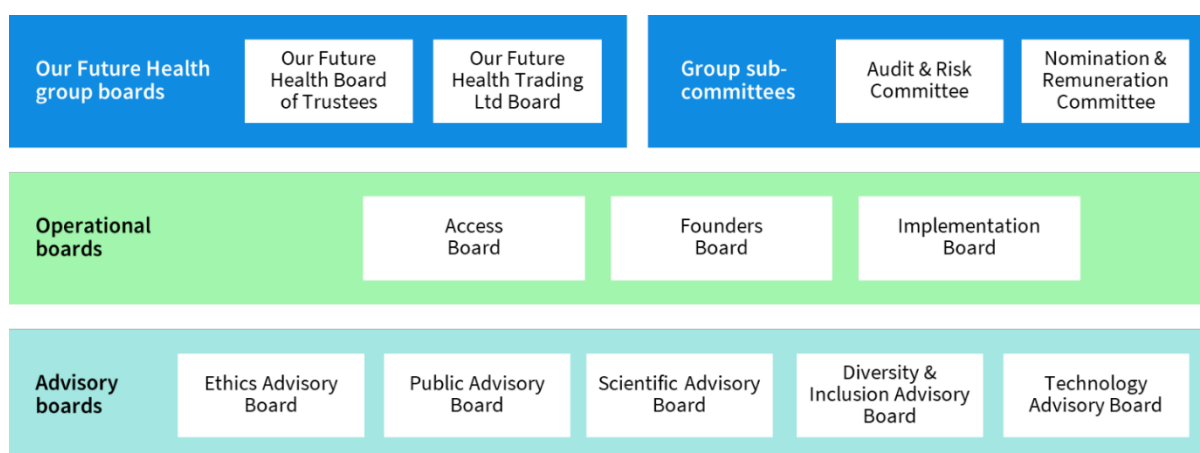
The PREM questionnaire was developed in 2021. We collected participant-reported experiences using the PREM during our 2021 pilot phase. We also conducted in-depth interviews with a subset of participants and active/passive decliners to provide complementary rich qualitative insights on their experiences. The current version of this measure is attached (**Appendix C**). Minor additions may be made to ensure we are accurately capturing participant reported experiences during 2022 recruitment activities. We will continue to conduct qualitative interviews on subsets of PREM questionnaire responders to ensure we complement quantitative findings with a deeper understanding of the participant experience.

## 12. Governance

The Our Future Health research programme is organisationally complex, receives funding and income through a range of sources and is required to work in partnership with various aspects of the NHS and across all nations of the UK. As such, it is critical that we have an appropriate governance and management structure that enables the programme to take effective and timely decisions, but also provides opportunities for consultation with stakeholders, which requires us to have agile approaches to programme governance. Given the pace of developments both in the genomic, digital, data and analytics areas, the programme must have the flexibility to respond to emerging opportunities and capitalise on innovations and novel research, as well as respond rapidly to new and emergent threats to data security, integrity and participants' wishes.

The programme also requires robust governance to ensure legal compliance, the security of data and privacy of participants, and to meet and exceed the expectations of the participants in order that we create long term relationships with the programme which are built on a bedrock of trust. It is also vital that we build into our plans the ability for participants to have meaningful interactions through the digital interfaces and provide effective support, information and reassurance where required. See **Figure 6**, below:

**Figure 6. Our Future Health Governance structure**



The governance model comprises a number of advisory and implementation boards. The function of these groups is to advise the Our Future Health Executive Team and the Our Future Health Board. The Access Board will be responsible for access to data, samples and participants and will report to the main Board. From time to time the programme may also establish task-focussed and time-limited working groups, as required. An overview of the governance structure can be found in the Governance Manual (**Appendix F**).

To promote coordination there will be some cross membership between the different governance structures, where this is appropriate; for example, the views of the public and participants will be intrinsic to the skills profile of a number of advisory boards and the Access Board.

Our Future Health is a company limited by guarantee registered in England and Wales (number 12212468) and a charity registered with the Charity Commission for England and Wales (charity number 1189681) and OSCR, Scottish Charity Regulator (charity number SC050917). Registered office: 2 New Bailey, 6 Stanley Street, Manchester M3 5GS.