Project Details	
Project Code	MRC22PHSBr Hughes
Title	Developing, evaluating and applying weighting methods for handling selection, dropout and collider bias.
Research Theme	Population Health Sciences
Summary	Most studies in medical research may suffer bias due to selection (participants are different to non-participants), or dropout (those who leave the study are different to those who remain). One way to investigate and address this bias is via inverse probability weighting (IPW), but barriers to its widespread usage remain. This project will use simulation and applied examples to develop IPW methods and software that will have a major impact across medical research.
Description	Selection bias (where people who choose to be in a study differ from those who do not) and dropout bias (where people who leave a study are different from those who stay in) affect most studies in medical research. For example, participants in UK Biobank (UKBB, one of the largest and most-used studies in medical research) are healthier and wealthier than the general population of the UK (1). We have shown that those who participate in optional aspects of UKBB differ from those who do not (2), and similarly for ALSPAC those who remain in the study differ in health, education and access to GP care from those who do not. (3) The most common approach to address these biases is multiple imputation – but an alternative approach, inverse probability weighting (IPW), may often be more appropriate, in particular, where there is information about the marginal distribution of covariates in the wider population (e.g. we know the age, sex distribution in the UK from census data). Another area of application is in reducing index event bias, a particular form of selection bias that occurs when examining risk factors for prognosis of a given disease (4). IPW is of particular relevance in this context when studying the epidemiology of COVID-19 (5), by developing models for the likelihood of being tested, and for the likelihood of being infected. However, IPW has been less studied than imputation, and there remain challenges in its application. This project will use simulations to address some of the current issues with IPW. First, the student will investigate the best functional form for the weighting model. The current approach is to use logistic regression and include main effects only. Other choices include modelling the risk ratio (rather than the odds ratio), which allows the inclusion of an interaction on the log probability scale. The robustness of IPW to the form of the weighting model has not been investigated. The weighting model must include variables to make the analysed cases representative of the whole sample. Selection of variables to include in the weighting model is not straightforward. For example, it is not necessarily the case that all predictors of selection should be included in the weighting model. Poor selection of variables can lead to increased variability of the weights and hence larger standard errors than necessary. The weighting model can also include variables to improve efficiency (i.e., variables not predictive of selection but associated with the variables of the analysis of interest). Again careful selection of the variables is required to avoid reducing efficiency. The student will examine strategies for selection of variables for a weighting model, and develop diagnostics for assessing the fit of

	the weighting model. Finally, the student will develop methods to carry out sensitivity analyses to examine robustness of results (e.g., sensitivity to the specification of the weighting model or stability of the weights). Alongside this methodological work, the student will apply the methods in an applied area which can be tailored to their interests. The supervisory team have experience of selection bias/index event bias in a range of areas, including selection in UKBB and ALSPAC, and index-event bias in COVID-19 and other disease prognosis studies. This PhD will make a major contribution by developing methods that will be of great relevance to all studies with incomplete data, especially much-used resources such as UK Biobank. The PhD student will be able to take advantage of the supervisory team's links to a wide range of cohort studies, to examine generalisability of the PhD results and encourage uptake of the methods by other researchers. 1.AJE 2017; 186(9) 2. Nature Communications 2021; 12 (886) 3. IJE 2018; 47: 1207-1216. 4. PLoS Genet. 2017 Oct; 13(10): e1006944. 5. Nat Commun 2020;11(1):5749.
Supervisory Team	
Lead Supervisor	
Name	Dr Rachael Hughes
Affiliation	Bristol
College/Faculty	Bristol Medical School
Department/School	Population Health Sciences
Email Address	rachael.hughes@bristol.ac.uk
Co-Supervisor 1	
Name	Dr Jonathan Bartlett
Affiliation	Bath
College/Faculty	Science
Department/School	Department of Mathematical Sciences
Co-Supervisor 2	
Name	Dr Rhian Daniel
Affiliation	Cardiff
College/Faculty	Medicine
Department/School	Population Sciences
Co-Supervisor 3	
Name	Dr Jon Heron
Affiliation	Bristol
College/Faculty	Bristol Medical School
Department/School	Department of Population Health Sciences
Co-Supervisor 4	
Name	Professor Kate Tilling
Affiliation	Bristol
College/Faculty	Bristol Medical School
Department/School	Department of Population Health Sciences