THE UNIVERSITY
of EDINBURGH

Potentially serious incidental findings in the UK Biobank Imaging Study

Dr Lorna Mary Gibson

THE UNIVERSITY *of* EDINBURGH

Doctor of Philosophy

The University of Edinburgh

2018

Abstract

The increased use of imaging across research, clinical and commercial contexts has generated debate and calls for evidence on the benefits and harms of incidental findings (defined as those which are unrelated to the purpose of imaging) to inform policy and practice. Evidence on clearly non-serious incidental findings is of limited clinical usefulness; this thesis therefore focuses on potentially serious incidental findings (PSIFs), defined as those which may indicate the possibility of a condition which, if it was confirmed, would carry a real prospect of seriously threatening life span, or of having a substantial impact on major body functions or quality of life.

In 2014, the UK Biobank Imaging Study began performing brain, cardiac and body magnetic resonance imaging (MRI), dual-energy X-ray absorptiometry and carotid Doppler ultrasound, and aims to image 100,000 of its population-based participants. The imaging data can be combined with extensive sociodemographic, lifestyle, physical measures, biochemical, genetic and linked healthcare data, to generate a research resource which will facilitate studies into a wide range of diseases. Due to the scale of the UK Biobank Imaging Study, PSIFs are a particularly pertinent issue. UK Biobank therefore evaluates the impact of its protocol for handling PSIFs, the data from which form the basis of this thesis.

This thesis aims to provide empirical data on seven themes relating to PSIFs: their prevalence and nature; follow-up and final diagnoses; factors associated with PSIFs and with serious final diagnoses; participants' understanding of consent to feedback of PSIFs; non-medical impacts of feedback of PSIFs; opinions of receiving feedback of PSIFs; and the economic impact of feedback of PSIFs on hospital services.

Chapter 1 outlines the scale of the challenge of incidental findings, and summarises current literature and gaps in our knowledge relating to each of the seven themes on PSIFs. Chapter 2 reviews systematically and meta-analyses published studies of brain and body MRI of apparently asymptomatic adults. Chapter 3 introduces the UK Biobank, the UK Biobank Imaging Study, and the rationale behind and protocol used to handle PSIFs in 100,000 largely asymptomatic participants: radiographer flagging of concerning images for a radiologist to review. Chapter 4 presents a study comparing two protocols to handle PSIFs in the first 1,000 imaged UK Biobank participants: radiographer flagging versus systematic radiologist review of all images. Chapter 5 investigates the factors associated with PSIFs and with serious final diagnoses. Chapter 6 examines the economic impact of feedback of PSIFs on hospital services, using linked routinely collected healthcare data.

In the systematic review, pooled prevalences of PSIFs on brain, thorax, abdominal and brain and body MRI were: 1.4–1.7%; 1.3–3.0%; 1.9–4.5%; and 3.9–12.8% respectively, the upper estimates reflecting the inclusion of indeterminate findings. There was substantial heterogeneity, but few informative data on potential sources of this. Around half of PSIFs were suspected malignancies.

Based on the first 7,334 participants in the UK Biobank Imaging Study (283 of whom had PSIFs), the PSIFs protocol had the largest influence on the prevalence of PSIFs and serious final diagnoses of any of the investigated factors: systematic radiologist review resulted in around 13 times more PSIFs and around four times more serious final diagnoses compared to radiographer flagging. A lower proportion of PSIFs detected by radiologists were finally diagnosed as serious compared to radiographer flagging (12% and 32% [Chapter 4 and 5]).

Feedback of PSIFs resulted in substantial impacts in terms of: clinical assessments (all participants visited their general practitioner, and 90% underwent some form of other clinical assessment, mostly imaging or referral to a specialist [Chapter 4]); non-medical impacts on participants (including on emotional wellbeing, insurance and finances and work and activities in 17%, 9% and 6% respectively [Chapter 4]); and hospital service use and cost (81% of cases with PSIFs generated some hospital use and costs, which had increased compared to controls, and to cases' hospital use and costs during the year before feedback of a PSIF [Chapter 6]). Importantly, as around 80% of PSIFs turned out not to be serious (Chapters 2, 4 and 5), many of these impacts may be unnecessary.

Despite these negative impacts, the vast majority of participants were glad to have received feedback of a PSIF and to have taken part in the imaging study (98% and 99% respectively), although almost a quarter changed their minds over time about whether or not feedback should always be given. Around a quarter of participants incorrectly thought they could choose to receive feedback and UK Biobank has improved its consent materials accordingly (Chapter 4).

Feedback of PSIFs impacts on participants and publicly-funded health services (and in turn patients in need); most PSIFs turn out not to be serious and many of these impacts may be unnecessary. Researchers can substantially influence these impacts via IFs policies, which must be designed to minimise unnecessary harms, and be clearly explained to participants to facilitate informed consent. These, and other implications of this thesis are further described in Chapter 7, which also discusses the results in the context of the broader literature, outlines the strengths and limitations of this thesis, and suggests directions for future work.

Lay summary

Apparently asymptomatic people who volunteer to have a scan for a research study may be told that they have an abnormality which may impact on their health or quality of life. We call these abnormalities 'potentially serious incidental findings,' or PSIFs. This thesis uses published studies and new data from the world's largest scan project, the UK Biobank Imaging Study, to provide more information on PSIFs, and their impacts on people and hospital services.

From a review of published studies, we found that PSIFs occur on brain, chest, abdominal and brain and body magnetic resonance imaging in 1.4–1.7%, 1.3–3.0%, 1.9–4.5% and 3.9–12.8% of apparently asymptomatic people respectively. Until they were assessed further, around half of these PSIFs were thought to be cancers.

All UK Biobank participants with PSIFs saw their general practitioners, and 90% had some other tests or appointments. News of a PSIF impacted on some people's emotional wellbeing (17%), insurance and finances (9%) and work and activities (6%). We found that 81% of people with PSIFs generated hospital costs, which were greatly increased compared to controls, and to their costs before they had feedback of a PSIF. As most PSIFs do not turn out to be serious disease (around 80%), some of these impacts on people and our hospitals may be unnecessary. Despite this, almost everybody was glad to have been told about their PSIF (98%) and to have had a research scan (99%), although almost a quarter changed their minds over time about whether or not people should always be told about a PSIF. Around a quarter misunderstood that UK Biobank would always tell them about a PSIF (thinking that they could choose to be told or not), so UK Biobank improved its consent materials.

We found that by far the biggest influence on the detection of PSIFs, and of those which turn out to be serious, is the researchers' protocol for handling PSIFs. Therefore, it is essential that researchers carefully design protocols which minimise the potential negative impacts of PSIFs, whilst still enabling important research which is needed to understand health and disease; our results may help them to do this. Our results suggest that researchers must strive to explain their PSIFs protocols clearly, and may help them to estimate and explain the potential impacts of PSIFs to people who are considering having a research scan, so that those people can make an informed decision to take part in a research study.

Declaration

I declare that the thesis has been composed by myself and that the work has not been submitted for any other degree or professional qualification. I confirm that the work submitted is my own, except where work which has formed part of jointly-authored publications has been included. My contribution and those of the other authors to this work have been explicitly indicated and described below. I confirm that appropriate credit has been given within this thesis where reference has been made to the work of others.

The work presented in Section 2.2 was previously submitted to the BMJ as 'Potentially serious incidental findings on brain and body magnetic resonance imaging of apparently asymptomatic adults: systematic review and meta-analysis,' by LM Gibson, L Paul, FM Chappell, M Macleod, WN Whiteley, R Al-Shahi Salman, JM Wardlaw, CLM Sudlow. Author contributions (student's contributions in bold): study design (**LMG**, JMW, CLMS); data curation (**LMG**, LP); data analyses (**LMG**, FMC); data interpretation (**LMG**, MM, FMC, WNW, RA-SS, JMW, CLMS); creation of tables and figures (**LMG**, FMC, CLMS); writing – original draft preparation (**LMG**); writing – review and editing (**all authors**).

The work presented in Section 3.2 was originally published as 'Management of incidental findings on multimodal imaging in UK Biobank. LM Gibson, J Sellors, CLM Sudlow. In: Weckbach S. (eds) Incidental radiological findings. Medical Radiology. 2016. Springer, Cham.' Permissions were obtained from the publisher and my co-authors to include this work in this thesis. Author contributions (student's contributions in bold): creation of tables (**LMG**); writing – original draft preparation (**LMG**); writing – review and editing (**all authors**).

The work presented in Section 4.2 was originally published as 'Impact of detecting potentially serious incidental findings during multi-modal imaging [version 3; referees: 2 approved, 1 approved with reservations]. LM Gibson, TJ Littlejohns, L Adamska, S Garratt, N Doherty, UK Biobank Imaging Working Group, JM Wardlaw, G Maskell, M Parker, R Brownsword, PM Matthews, R Collins, NE Allen, J Sellors, CLM Sudlow. Wellcome Open Research 2018; 2:114.' This paper is available as open access and is appropriately cited within this thesis. In addition, I obtained permission from my co-authors to include this work in this thesis. Author contributions (student's contributions in bold): study design (**LMG**, TJL, PMM, RC, NEA, JS, CLMS); data curation (**LMG**, TJL, LA, SG, CLMS); data analysis (**LMG**, TJL); data interpretation (**LMG**, TJL, JMW, MP, RB, PMM, RC, NEA, JS,

CLMS); creation of tables and figures (**LMG**); writing – original draft preparation (**LMG**); writing – reviewing and editing (**all authors**); project administration (ND).

The work presented in Section 5.2 has been conducted in collaboration with, and edited by, my co-authors in preparation for submission to a journal as 'Factors associated with potentially serious incidental findings and with serious final diagnoses on multi-modal imaging in the UK Biobank Imaging Study: a prospective cohort study,' by LM Gibson, J Nolan, TJ Littlejohns, E Mathieu, S Garratt, N Doherty, S Petersen, NCW Harvey, J Sellors, NE Allen, JM Wardlaw, CA Jackson, CLM Sudlow. Author contributions (student's contributions in bold): study design (**LMG**, NEA, JMW, CAJ, CLMS); data curation (**LMG**, JN, TJL, EM, SG, CLMS); data analyses (**LMG**, JN); data interpretation (**LMG**, CAJ, CLMS); creation of tables and figures (**LMG**); writing – original draft preparation (**LMG**); writing – reviewing and editing (**all authors**); project administration (ND).

The work presented in Section 6.2 has been conducted in collaboration with, and edited by, my co-authors in preparation for submission to a journal as 'Use and cost of hospital services by UK Biobank participants with potentially serious incidental findings: a case-control study utilising linked English Hospital Episode Statistics data from 2013-2016,' by LM Gibson, J Nolan, E Mathieu, TJ Littlejohns, S Garratt, S Sheard, N Doherty, C Keerie, NE Allen, JM Wardlaw, J Leal, AM Gray, CLM Sudlow. Author contributions (student's contributions in bold): study design (**LMG**, CK, JMW, JL, AMG, CLMS); data curation (**LMG**, JN, EM, TJL, SG, SS, CLMS); data analyses (**LMG**, JN); data interpretation (**LMG**, TJL, JMW, JL AMG, CLMS); creation of tables and figures (**LMG**); writing – original draft preparation (**LMG**); writing – reviewing and editing (**all authors**); project administration (ND).

Dr Lorna M. Gibson

Acknowledgements

I would like to thank my supervisors, Cathie Sudlow and Joanna Wardlaw, for the opportunity to design and conduct the work presented in this thesis. Particular thanks are due to Rustam Al-Shahi Salman, chair of my thesis committee, for guidance and support throughout my clinical research training fellowship. Thank you also to the members of the UK Biobank Executive Management Committee and the Imaging Working Group, in particular, Paul Matthews and Rory Collins, for their support to conduct the projects which form the majority of this thesis.

I am grateful for all of the advice on and contributions to many aspects of the work within this thesis from my additional co-authors: Ligia Adamska, Naomi Allen, Roger Brownsword, Francesca Chappell, Nicola Doherty, Steve Garratt, Alastair Gray, Nick Harvey, Caroline Jackson, Catriona Keerie, Jose Leal, Thomas Littlejohns, Malcolm Macleod, Giles Maskell, Edouard Mathieu, John Nolan, Mike Parker, Laura Paul, Steffen Petersen, Jonathan Sellors, Simon Sheard and Will Whiteley.

I would like to extend a special note of thanks to the UK Biobank participants and Imaging Centre and Co-ordinating Centre staff, without whom the projects would not have been possible. Finally, I am grateful to the Wellcome Trust for providing the funding to enable me to conduct this work.

To acknowledge the contributions of my co-authors, ‘we’ will be used instead of ‘I’ throughout this thesis.

Publications and other contributions arising from work associated with this thesis

Peer reviewed paper

Gibson LM, Littlejohns TJ, Adamska L, Garratt S, Doherty N, UK Biobank Imaging Working Group, Wardlaw JM, Maskell G, Parker M, Brownsword R, Matthews PM, Collins R, Allen NE, Sellors J, Sudlow CLMS. Impact of detecting potentially serious incidental findings during multi-modal imaging [version 3; referees: 2 approved, 1 approved with reservations]. Wellcome Open Research 2018, 2:114.

Book chapters

Gibson LM, Sudlow CLM, Wardlaw JM. Incidental findings: current ethical debates and future challenges in advanced neuroimaging. In: Neuroethics: anticipating the future. Illes J (Editor), Hossain S (Associate editor). Second edition. Oxford University Press, Oxford, UK: 2017; 54-69.

Gibson LM, Sellors J, Sudlow CLM. Management of incidental findings on multi-modal imaging in UK Biobank. In: Incidental radiological findings. Weckbach S (Editor). First edition. Springer, Cham, Switzerland: 2016; 71-78.

Abstract

Gibson LM, Nolan J, Littlejohns TL, Mathieu E, Garratt S, Doherty N, Allen NE, Wardlaw JM, Jackson CA, Sudlow CLM. Risk factors associated with potentially serious incidental findings and with serious final diagnoses on multi-modal imaging in the UK Biobank Imaging Study. BMJ Evidence-based medicine (*In Press September 2018).*

Presentations

Incidental findings on imaging in the UK Biobank study. Clinical Neurosciences Seminar, Edinburgh, May 2017.

Impact of detecting potentially serious incidental findings during multi-modal imaging: experience from UK Biobank. Preventing Overdiagnosis Conference, Quebec City, August 2017.

Impact of detecting potentially serious incidental findings during multi-modal imaging: experience from UK Biobank. Incidental Findings in Neuroimaging Research Meeting, Oxford, December 2017.

Risk factors associated with potentially serious incidental findings and with serious final diagnoses on multi-modal imaging in the UK Biobank Imaging Study. Preventing Overdiagnosis Conference, Copenhagen, August 2018.

Protocols

Gibson L, Paul L, Wardlaw J, Sudlow C. Potentially serious incidental findings on brain and body magnetic resonance imaging conducted among apparently healthy adults: a systematic review. PROSPERO 2016:CRD42016029472. Available from http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42016029472

Dataset

Gibson L, Paul L, Sudlow CLM. Potentially serious incidental findings on brain and body magnetic resonance imaging conducted among apparently healthy adults: a systematic review and meta-analysis, [dataset]. 2017. Available from: https://datashare.is.ed.ac.uk/handle/10283/2773

Analyses code files

Gibson LM, Nolan J. Factors associated with potentially serious incidental findings and with serious final diagnoses on multimodal imaging in the UK Biobank Imaging Study: a prospective cohort study, [software]. 2018. Available from: https://datashare.is.ed.ac.uk/handle/10283/3112

Magazine article

Gibson L. Scans from ‘healthy’ volunteers reveal serendipitous findings: a blessing or a curse? eu:sci 2017;20:31-32.

Contents

Table of tables

Table of figures

Table of supplementary tables

Table of supplementary figures

Chapter 1 Introduction

1.1 The challenge of incidental findings

1.1.1 Definitions

In 2006, Illes et al. defined incidental findings (IFs) as "observations of potential clinical significance unexpectedly discovered in healthy subjects or in patients recruited to any imaging research study, and unrelated to the purpose or variables of the study" (Illes et al., 2006). However, an increasing awareness and experience of managing individuals with IFs led major research funding bodies to state that IFs should be anticipated, rather than defined as 'unexpected' (Medical Research Council and Wellcome Trust, 2014). As such, an IF may be defined as a finding "concerning an individual research participant that has potential health or reproductive importance and is discovered in the course of conducting research but is beyond the aims of the study" (Wolf et al., 2008). However, researchers vary in their application of this definition. Rather than focusing on only those findings of 'health or reproductive importance,' some studies consider normal anatomical variants and post-surgical appearances to be IFs (Sandeman et al., 2013), although some studies do then further sub-classify these as common asymptomatic findings (Hegenscheid et al., 2013).

Ethicists have argued that individuals should only receive feedback of IFs which are clinically significant. For example, IFs which are life-threatening (such as malignancies), or IFs which could be treated (such as large abdominal aortic aneurysms), or where feedback may inform reproductive decision making (such as cystic fibrosis carrier status) (Wolf et al., 2008). In practice, determining which IFs meet the definition of 'clinically significant' is difficult for two reasons: the limitations of research imaging, and the limits of current knowledge about particular IFs.

1.1.1.1 Limitations of research imaging

Research imaging may not facilitate firm diagnoses. Research imaging is tailored to generate imaging data in order to address a specific scientific question or questions. In contrast, clinical imaging is tailored to optimise the demonstration of a diagnosis, or diagnoses, which a doctor thinks may account for a patient's particular set of clinical symptoms and signs. As such, research imaging can differ enormously from the imaging performed within clinical practice, and may be performed in a very different group of individuals. Any form of structural disease may be detected and classified as an IF, if it is unrelated to the purposes of

the study. For example, a study which uses body magnetic resonance imaging (MRI) to investigate the distribution of body fat in a group of asymptomatic individuals may also demonstrate structural diseases such as tumours or aneurysms. Different types of clinical imaging are required to firstly optimally demonstrate tumours versus aneurysms, and secondly to provide the specific visual information which is needed to inform the very different treatment plans for these two conditions, and in clinical practice patients with tumours will undergo different imaging to those with aneurysms. IFs on research imaging are therefore usually not demonstrated optimally, making judgements of their clinical severity difficult in some cases; and research imaging may not provide all of the detailed visual information needed to inform the treatment plan.

Given the difficulties with optimal demonstration of IFs on research imaging, some authors have called for clinical imaging to be added to all research study protocols (Milstein, 2008). The potential benefit of this approach would be to make firm diagnoses within the research setting, saving publicly-funded healthcare systems from the service and cost burdens of investigating IFs. However, adding clinical imaging to a research protocol may simply not be feasible within a research context, due to additional costs, the limits of the available scanning time, or the tolerance of participants to lie still for further scans (Booth et al., 2010).

1.1.1.2 Limitations of current knowledge of particular IFs

The application of a definition of 'clinically significant' presupposes knowledge of the natural history of an IF; however, such data may be lacking. For example, a systematic review found that members of the general population with white matter hyperintensities had an approximately three times increased risk of incident stroke and of dementia compared to those without (Debette and Markus, 2010). From this, the authors concluded that participants with white matter hyperintensities should be investigated for other risk factors for stroke and dementia (Debette and Markus, 2010). However, the methods used to measure burden of white matter hyperintensities varied between studies, and it was not clear how the risk of stroke and dementia varied with either the volume or the presence versus absence of white matter hyperintensities. As such, informing participants about white matter hyperintensities may be of limited clinical value at present, but our judgements of the clinical severity of some IFs is likely to evolve as new evidence becomes available.

Image readers may have variable access to clinical information about research participants, which may also cause difficulties in judging the clinical significance of an IF. For example, the UK Biobank provides its radiologists with participants' self-reported medical history data

(along with brief data on their demographics and lifestyle risk factors such as smoking) (Gibson et al., 2018), but the radiologists do not have access to previous clinical imaging. In this context, an atypical-appearing kidney cyst will likely be fed back as this may represent malignancy. However, if previous imaging was available and could demonstrate no significant change over a reasonable time interval, this would enable a firmer judgement that such an IF is most likely non-serious, and may not warrant feedback.

Taking in to account the limitations of research imaging, our current knowledge of the prognosis of IFs, and variable availability of clinical information on research participants, the UK Biobank study proposed the term 'potentially serious IFs' (PSIFs), defined as those 'indicating the possibility of a condition which, if confirmed, would carry a real prospect of seriously threatening life span, or of having a substantial impact on major body functions or quality of life' (Gibson et al., 2018).

Multiple chapters presented in this thesis are based on UK Biobank data, and as such the term 'PSIFs' is used to refer specifically to IFs which meet the UK Biobank definition; otherwise, the more general term 'IFs' is applied. The next section will briefly describe the recent development of large imaging studies, including the UK Biobank Imaging Study, in order to summarise the scale of the challenge of IFs.

1.1.2 Scale of the challenge

Our estimates of the scale of the challenge of IFs are informed by the development of large population-based imaging research projects, and the increase in imaging within non-research contexts, including public health screening programmes, clinical imaging and direct-to-consumer (i.e. commercial) imaging. While this thesis focuses on PSIFs detected in apparently asymptomatic volunteers undergoing brain and body MRI, there is a large body of literature on IFs detected within non-imaging contexts, particularly genomics, of which the more recent and controversial developments will be summarised.

1.1.2.1 Large population-based imaging research projects

MRI is becoming increasingly popular as a research imaging tool, owing to the lack of ionising radiation. Worldwide, several large population-based studies are collecting single or multi-region MRI data on large subsets of – or their entire – cohorts. Data from the imaged cohorts will enable investigations of associations between imaging and other variables, so generating insights into health and disease (Bertheau et al., 2016; Collins, 2012; German National Cohort (GNC) Consortium, 2014; Hegenscheid et al., 2013; Ikram et al., 2015;

Matthews and Sudlow, 2015; Petersen et al., 2017; Smith et al., 2015; Stephan et al., 2015; Suzuki et al., 2017; Volzke et al., 2011).

The largest ongoing multi-region MRI studies are based in the UK and in Germany, and a further large study is being planned in Canada. The UK Biobank will generate the world's largest multi-modal imaging dataset by performing brain, cardiac, and body MRI, carotid Doppler ultrasound and dual-energy X-ray absorptiometry (DXA) in 100,000 of its 500,000 cohort (i.e. around one fifth of all the participants) (Matthews and Sudlow, 2015). As of September 2018, over 27,000 participants have been imaged (UK Biobank, 2018c). Participants were originally recruited between 2006 and 2010, when aged 40–69 years old, and underwent extensive phenotyping via questionnaire, blood sampling, physical measurements and cognitive testing (Sudlow et al., 2015). Their health is followed up via linkages to routinely collected healthcare data (Sudlow et al., 2015). In Germany, the Study of Health in Pomerania (SHIP) performed whole-body MRI in 2,500/4,416 (57%) of their population-based participants (Hegenscheid et al., 2013) which involved coronal plane whole-body MRI, with further sequences focused on imaging the spine, brain, neck, chest and abdomen and pelvis (Hegenscheid et al., 2009). The German National Cohort study has built on the experiences of the SHIP and collaborates with other ongoing national cohort studies including UK Biobank (German National Cohort (GNC) Consortium, 2014). They aim to conduct 3.0T MRI of the brain, heart, body and spine in 30,000 of their 200,000 (15%) participants (Bertheau et al., 2016). The Canadian Alliance for Healthy Hearts and Minds (CAHHM) are currently recruiting participants from existing Canadian cohorts (including from the Canadian Partnership for Tomorrow Project which itself consists of cohorts from five studies, totalling > 300,000 participants), and perform brain, cardiac and abdominal MRI, plus dedicated cerebrovascular MRI sequences (Anand et al., 2016).

In contrast to UK Biobank, the SHIP, the German National Cohort and CAHHM, other studies are performing (or have completed) MRI of single body regions, most commonly either of the brain or the heart. The Rotterdam Scan Study began performing brain MRI in a subset of participants in 1995 and 1999, and in 2005 extended its programme to image the entire cohort (Ikram et al., 2015; Ikram et al., 2011). As of July 2015, over 12,000 brain scans had been conducted in over 5,800 participants (Ikram et al., 2015). The Lothian Birth Cohort study acquired brain MRI in 700/1,091 (64%) members of its cohort three years after their initial recruitment (Wardlaw et al., 2011). Similarly, brain MRI was acquired in 803/135,335 (0.6%) Canadian participants enrolled in the multinational Prospective Urban Rural Epidemiological study (Smith et al., 2015) and in 1,923/4,931 (39%) Dijon-based

participants enrolled in the French Three City study (Stephan et al., 2015). Large cardiac MRI datasets have been, or are being, acquired in general populations (Schelbert et al., 2012; Tsao et al., 2011), and within populations with higher rates of cardiac disease (Marwick et al., 2013; Taylor, 2005; Victor et al., 2004). The Multi-Ethnic Study of Atherosclerosis, the Jackson Heart Study and the Dallas Heart Study aim to improve understanding of how cardiovascular disease could be prevented, particularly in black Americans, and to this end conducted cardiac MRI in 5,000/6,814 (73%) (Natori et al., 2006), around 2,000/5,301 (38%) (Marwick et al., 2013; Taylor, 2005) and 2,793/6,101 (46%) participants respectively (Victor et al., 2004). Also in the USA, 1,794/3,539 (50.7%) participants enrolled in the Framingham Heart Study Offspring Cohort underwent cardiac MRI (Tsao et al., 2011). The ICELAND MI study performed cardiac MRI in 936/5,764 (16%) population-based participants involved in the Age, Gene/Environment Susceptibility Reykjavik study (Schelbert et al., 2012).

Many other large epidemiological studies do not conduct imaging at present. The Medical Research Council funds 36 UK-based cohorts of over 1,000 participants each, the largest being the Million Women Study (N=1.3 million) (Medical Research Council, 2014). Three of the largest non-UK cohorts are in the magnitude of hundreds of thousands of adult participants: the China Kadoorie Biobank (N=512,891 adults aged 30–79) (Chen et al., 2011), CONSTANCES (N=200,000 French adults aged 18–69) (Zins and Goldberg, 2015b) and the Mexico City Prospective Study (N=150,000) (CTSU, 2018). The addition of imaging to these large cohorts in future would add further to the scale of the challenge of IFs.

1.1.2.2 Other imaging settings

The previous section described the scale of the challenge of IFs in terms of studies which perform MRI for research in large, population-based cohorts of apparently asymptomatic people. This section will describe the contributions of non-research imaging of apparently asymptomatic people and of clinical imaging of patients to the scale of the challenge of IFs.

Apparently asymptomatic people (who we define as ‘community-dwelling people not selected for imaging on the basis of symptoms, risk factors, or disease’) may access MRI directly (Lee et al., 2008; Tarnoki et al., 2015; Tsushima et al., 2005) via imaging services marketed direct-to-consumers. They may also access imaging via occupational health assessments (Weber and Knopf, 2006), private health insurance (Cieszanowski et al., 2014) or company health care programmes (Goehde et al., 2005; Tarnoki et al., 2015).

Direct-to-consumer marketing of prescription-only medications is not permitted in the UK (Magrini and Font, 2007), but there are no such limits on direct-to-consumer advertising of imaging services. In 2004, a study of 40 print advertisements for direct-to-consumer imaging services found that none mentioned the risks of having a scan (Illes et al., 2004a), such as the detection of IFs which, after clinical assessment, turn out not to be serious. In 2010, in a joint statement to the UK Health Secretary, the British Medical Association and the Academy of Medical Royal Colleges called for stronger regulation of the marketing of direct-to-consumer imaging tests, including mandatory provision of information on the risks of the test, and the implications of the results of the test and any follow-up that may be required (Meldrum and Douglas, 2010). To help inform people who were considering accessing direct-to-consumer imaging, the UK National Screening Committee and the National Health Service (NHS) published a leaflet, outlining the differences between these types of imaging, which are often referred to as 'screening', and public health screening services offered by the NHS (UK National Screening Committee and National Health Service, 2014). In brief, before implementation, proposed public health screening programmes are assessed against criteria to ensure they detect diseases which are important public health problems with good sensitivity and specificity at an early stage, for which an effective, acceptable treatment is available (Wilson et al., 1968). Screening programmes should only be implemented when they are deemed to provide net benefit over harm after consideration of the evidence, which should be reviewed at regular intervals (Harris et al., 2011). In contrast to the diseases targeted by public health screening programmes, direct-to-consumer imaging of asymptomatic people may not necessarily detect a disease at a stage where treatment will confer a survival benefit; this would depend on the exact imaging offered, and the disease of interest. However, public health screening programmes may also generate substantial volumes of IFs. Recent analyses of 17,309 participants' low-dose computed tomography (CT) chest scans for lung cancer screening found that this programme generated extra-pulmonary IFs requiring further evaluation in every fifth person (Nguyen et al., 2017), which will likely result in negative impacts, such as psychological distress, for some (Harris et al., 2014). Furthermore, unlike diseases which are the focus of public health screening programmes, there is no clear evidence on which to base decisions about treatment for many IFs. Early treatment of some disorders, such as asymptomatic unruptured intracranial aneurysms, may in fact cause harm (Mohr et al., 2014).

Rapidly increasing rates of clinical imaging threaten to overwhelm radiology services. Between 2010 and 2015, the numbers of different types of scans performed in Scotland all markedly increased: MRI by 48%; CT by 35%; ultrasound by 11%; and other types,

including radiographs, by 10% (The Royal College of Radiologists Standing Scottish Committee, 2016). In response, an emergency group convened to redesign Scottish radiology services at a national level (Scottish Radiology Transformation Programme and NHS National Services Scotland, 2017). Imaging is now commonly used to rule out concerning conditions, rather than being selectively applied to the smaller number of patients in whom there is a high probability of that specific diagnosis. In a study based in an American emergency department, staff requested 370 CT angiograms over 12 months for patients with suspected aortic dissection; 19 (5%) did have new dissections, but 30 (8%) had IFs which required follow-up (Prabhakar et al., 2015). Furthermore, clinical imaging for some particular conditions is now increasing, either due to increased coverage of tissue volumes (for example, whole-body CT for trauma patients who would have previously undergone targeted imaging (Sierink et al., 2014)) or new indications for so-called 'routine' imaging (for example, imaging prior to some operations (See et al., 2010)). These changes in the use of imaging in clinical practice contribute to the recent huge increases in demand for imaging services.

IFs have been documented in studies of patients undergoing many different types of clinical imaging of different body regions, including positron emission tomography (Shie et al., 2009), DXA (Bazzocchi et al., 2012), ultrasound (Choi et al., 2016), CT (James et al., 2017; Prabhakar et al., 2015; Sierink et al., 2014) and MRI (Sherrer et al., 2018). A systematic review of 44 studies of IFs detected in patients undergoing diagnostic imaging (N>100,000; one study did not report sample size) found that the mean frequency of IFs was around 24%. IFs were most common in patients undergoing CT (31%), and in patients with non-specific initial diagnoses (31%) (Lumbreras et al., 2010). However, the prevalence of IFs may vary widely between studies. A recent umbrella review of 20 systematic reviews of the prevalence of IFs on imaging of mixed patient and asymptomatic populations found evidence of substantial between-study heterogeneity in 15 of the included reviews (O'Sullivan et al., 2018). For example, data from a systematic review of eleven studies of patients undergoing cardiac MRI found the prevalence of extra-cardiac IFs to be 34%, with an I^2 of 99% (O'Sullivan et al., 2018).

The increase in clinical imaging will result in more IFs, the numbers of which are likely to vastly outstrip those generated in research imaging settings. As described earlier, IFs may require even more imaging to reach firm diagnoses and guide treatment plans, and as such IFs from clinical imaging place a substantial burden on health services, which are already overstretched (The Royal College of Radiologists Standing Scottish Committee, 2016).

1.1.2.3 Non-imaging contexts

While this thesis focuses on PSIFs detected on research MRI conducted in apparently asymptomatic volunteers, considerable attention has been given to IFs which arise from non-imaging tests conducted across the life-course, from prenatal testing through to autopsy (Bui et al., 2014; Kingsley-Loso et al., 2015; Parker et al., 2013; Thomas et al., 2017). Recently, there have been considerable developments and controversies surrounding genomic IFs, which will be briefly summarised here.

The human genome sequence was completed in 2003 and made available to researchers in order to further our knowledge of the role of genes in health and disease (National Human Genome Research Institute, 2012). Following a rapid period of technological advancement, the cost of sequencing fell from around $25 million in 2006, to less than $1,000 today (National Human Genome Research Institute, 2016). As such, genomics testing has become a much more accessible tool for research, as well as clinical testing, and is available direct-to-consumers. As with imaging research, several large scale population-based studies are generating, or already hold, genomic data on large subsets or their entire cohort, such as UK Biobank (N=502,000) (UK Biobank, 2018a), the China Kadoorie Biobank (N>100,000) (China Kadoorie Biobank, 2015), the 100,000 Genomes Project (Samuel and Farsides, 2017) and CONSTANCES (N=200,000) (Zins and Goldberg, 2015a; b).

As the use of genomics in research is expanding, so too is its role in clinical practice. Clinical genomics testing has a wide range of applications, from assessment of risk of complex diseases, identification of carriers of diseases such as cystic fibrosis, predictive screening in prenatal and new-borns, diagnostic testing for diseases such as Duchenne muscular dystrophy, predictive testing for diseases such as breast cancer and Huntington disease, and assessing likely responses to treatments such as immunosuppressive agents (Delaney et al., 2016). As genomics technologies reduced in price over recent years, companies began marketing genomics testing direct-to-consumers to enable them to assess health risks, carrier status, predict responses to medication and trace their ancestry (Roberts and Ostergren, 2013).

In contrast to imaging IFs which are visually apparent to viewers, genomic IFs are not immediately apparent, but must be deliberately sought out (Green et al., 2013). In 2013 the American College of Medical Genetics and Genomics (ACMG) appointed a working group to make recommendations on handling genomics IFs responsibly (Green et al., 2013). While acknowledging the lack of evidence on the clinical utility of many mutations, the ACMG

working group recommended that laboratories seek and report on 56 different mutations when performing any exome or genome sequencing, regardless of the initial indication for sequencing, patient preferences for feedback, or patient age (Green et al., 2013). The ACMG recommendations implied that the referring doctor would be required to disclose these to the patient, or the patient would have to decline being tested entirely (Kang et al., 2016). This situation presented 'novel ethical and legal issues' (Evans, 2013) such as where patient autonomy could be ignored and children may be given results about adult-onset illnesses (Kang et al., 2016). Furthermore, laboratories and clinicians may be liable whether or not they check for the recommended 56 mutations (Evans, 2013): on the one hand, failing to disclose any clinically actionable IFs may generate negligence claims against laboratories and physicians; on the other, disclosing IFs which result in emotional harm, such as the return of unwanted results, is considered an '[injury] caused by purposeful behaviour' which may generate an intentional tort lawsuit (Evans, 2013). The ACMG subsequently revised its position to suggest that patients should be able to opt out of testing for genes that were unrelated to the indication for sequencing (ACMG Board of Directors, 2015).

1.2 Current knowledge of IFs and PSIFs

The increasing use of imaging in research, clinical diagnostics, public health screening and direct-to-consumer commercial imaging enterprises will likely only increase the challenge of imaging IFs in the years to come. This section will summarise our current knowledge of IFs and PSIFs on research imaging, organised in to seven themes: their prevalence and nature (Section 1.2.1); follow-up and final diagnoses (Section 1.2.2); variation in prevalence of PSIFs and serious final diagnoses (Section 1.2.3); participants' understanding of consent to feedback of IFs (Section 1.2.4); non-medical impacts of feedback of PSIFs (Section 1.2.5); participants' and healthcare professionals' opinions on feedback of PSIFs (Section 1.2.6); and the economic impact of feedback of PSIFs (Section 1.2.7).

1.2.1 Prevalence and nature of IFs on MRI of apparently asymptomatic volunteers

Our current knowledge of the prevalence and types of IFs detected on brain MRI of apparently asymptomatic volunteers is largely informed by a systematic review (Morris et al., 2009). Meta-analyses of 16 studies of 19,559 apparently asymptomatic people found that the pooled prevalence of neoplastic IFs was 0.7% (95% confidence interval [CI] 0.5–1.0%),

and of 15 studies of 15,559 people the pooled prevalence of non-neoplastic IFs was 2.0% (95% CI 1.1–3.1%) (Morris et al., 2009).

The division of IFs into neoplastic and non-neoplastic categories is not equivalent to the categories of potentially serious versus non-serious IFs. The most common neoplastic IFs were meningiomata (pooled prevalence 0.29% [95% CI 0.13–0.51%]) (Morris et al., 2009). In many cases meningiomata may not cause symptoms or serious disease, and those < 2 cm diameter are no longer fed back to participants within the population-based Rotterdam Scan Study (Bos et al., 2016).

It is also important to distinguish whether or not prevalence estimates of IFs represent suspected conditions or final diagnoses. None of the studies included in this review systematically followed up unselected participants with an IF to gather data on final diagnoses. Therefore, while the pooled prevalence of intracranial aneurysms was 0.35% (95% CI 0.13–0.67%) (Morris et al., 2009) this may reflect the prevalence of suspected, rather than proven, aneurysms.

The prevalence estimates from this systematic review exclude certain findings, such as normal anatomical variants (Morris et al., 2009). As such their estimates of prevalence of IFs are not overinflated by such findings. In turn, this highlights one area of caution with interpreting the overall prevalence estimates of individual studies, which may count normal variants such as mega cisterna magna (Haberg et al., 2016) or post-surgical appearances as IFs (Sandeman et al., 2013). The review also excluded silent brain infarcts from the non-neoplastic pooled prevalence estimates after the authors considered how difficult it can be to accurately determine whether or not an infarct is truly asymptomatic, and the lack of evidence for benefits of primary prevention therapies in such patients (Morris et al., 2009). While silent brain infarcts were found in 7% of 2,000 participants from the Rotterdam Scan Study (Vernooij et al., 2007), the benefits of treating these remain unclear as there have not been any relevant randomised controlled trials (Smith et al., 2017). However, at least one such trial is planned (Sui, 2017), and as such, future studies may influence our understanding of the clinical significance of, and therefore measurement and prevalence of, IFs over time.

Changes in imaging technology may also influence the prevalence of IFs over time. Morris et al. included studies published up to and including 2008, and found that IFs were more prevalent in studies which used high compared to low resolution imaging (4.3% [95% CI 3.0–5.8%], 1.7% [95% CI 1.1–2.4%, $p<0.001$) (Morris et al., 2009). With advances in

imaging technologies and their increased use as research tools over the past decade since this review was published, the prevalence of IFs may have increased.

By contrast with our knowledge of brain IFs, there are no existing systematic reviews of the prevalence of IFs, PSIFs, or serious final diagnoses on body MRI conducted in healthy populations (O'Sullivan et al., 2018). Estimates of the prevalence of IFs on body MRI may be as great as 30% (The Royal College of Radiologists, 2011), but three of the four studies underpinning this estimate were conducted in patient, rather than apparently asymptomatic volunteer populations and used CT rather than MRI (Furtado et al., 2005; Machaalany et al., 2009; Siddiki et al., 2008) (the remaining study described their sample as both 'research participants' and 'patients' (Orme et al., 2010)). As such, this estimate of IFs may not be easily generalisable to apparently asymptomatic populations undergoing body MRI. The prevalence of IFs seems to be greater on body, compared to brain MRI. One of the largest studies of brain and body imaging included 2,500 healthy volunteers, whose images were first reviewed by radiologists and any potentially relevant IFs were escalated for review by an advisory board (Hegenscheid et al., 2013). This process resulted in an overall prevalence of IFs of 2% (n=46) on brain, 3% (n=76) on chest, and 22% (n=552) on abdominal and pelvic MRI (Hegenscheid et al., 2013).

In summary, IFs occur in approximately 1-2% of apparently asymptomatic volunteers who undergo brain MRI, but it is not entirely clear what proportion of these represent potentially serious IFs or serious final diagnoses. Furthermore, with advances in our knowledge of different conditions and of imaging technologies, these estimates may change over time. Conversely, relatively little is known of the prevalence of IFs on body MRI, but it is likely to be higher compared to brain MRI.

1.2.2 Follow-up and final diagnoses

As described in Section 1.1.1, the non-diagnostic nature of research imaging and our lack of knowledge of natural history and variable availability of clinical information about research participants will lead to IFs which are of uncertain clinical significance. Participants with such IFs will likely undergo further clinical assessments in order to resolve this uncertainty.

In 1972, Rang described Ulysses syndrome, as patients 'though healthy enough at the outset, make a long journey through the investigative arts and experience a number of adventures before reaching their point of departure once again' (Rang, 1972). Its cause: 'meritorious desire to investigate a patient fully,' secondary to coverage of a test by insurance, junior staff

ordering tests to avoid criticism by seniors, and the gamut of tests available (Rang, 1972). Rang describes Ulysses syndrome as short-lived, and that 'no mortality or permanent harmful effects have yet been noted.' But by 2003, opinion on the harmful effects of over-investigation had changed. Victims of Medical Imaging Technology (VOMIT) became a tongue-in-cheek acronym for the patients anxious as a result of tests, and the doctors whose time must be spent reassuring them (Hayward, 2003). By 2014, concerns about the causes of over-investigation and its effects on patients and health services led the Scottish Chief Medical Officer to question the fundamental principles of how we practise medicine today (Chief Medical Officer for Scotland, 2016).

First we must understand the magnitude of the burden of clinical assessments performed for participants with IFs, and secondly, we must know how many of these turn out to be serious disease. Investigations which result in a non-serious final diagnosis may be deemed unnecessary in retrospect (Gibson et al., 2018). Clinical assessments of participants with IFs may involve appointments with general practitioners (GPs) or hospital doctors, blood tests, imaging, or more invasive tests such as biopsies, before a final diagnosis is reached. However, little is known of the burden of this follow-up, or in contrast, of the natural course of IFs which are not investigated and left untreated (Wardlaw et al., 2015), or the net benefit or harms to participants. While there are many case reports which describe the follow-up of individual patients, systematically collected, long-term data on unselected participants with IFs are relatively limited. The majority of Rotterdam Scan Study participants with IFs were managed using either wait-and-see policies, or discharged after one hospital appointment (144/188, 77%), but the numbers of participants undergoing particular types of clinical assessment, for example, repeat imaging, were not reported (Bos et al., 2016). Furthermore, this study only performed brain MRI, and the results may not be generalisable to the management of IFs detected on imaging of other body regions. As mentioned in Section 1.2.1, data on final diagnoses, and of the prevalence of serious final diagnoses in particular, of IFs are lacking; none of the 16 studies included in a systematic review of IFs on brain MRI systematically followed up participants with IFs (Morris et al., 2009).

Until large, long-term studies of unselected participants with IFs report data on the clinical assessments performed and the final diagnoses, we will not be able to provide potential research participants with details of the burden of follow-up they may endure in the pursuit of a final diagnosis, what proportion of participants' journeys will lead them back to their point of departure, and what proportion will be given a diagnosis of serious disease. Currently, the best we can offer potential imaging research participants is general advice

only: there is a possibility of demonstrating an IF, and if that occurs, some further tests and referrals may be required (Illes, 2006; Illes and Chin, 2008).

1.2.3 Variation in the prevalence of PSIFs and of serious final diagnoses

Best practice in handling IFs from research imaging encourages researchers to determine the likely prevalence of IFs which may be detected when studying a particular population (Medical Research Council and Wellcome Trust, 2014). Estimates of prevalence may vary depending on a number of different factors, which researchers may need to take into account when designing their IFs handling policies and creating participant consent materials. This section summarises our knowledge of the participant and study factors which may influence the prevalence of PSIFs and of serious final diagnoses.

1.2.3.1 Participant factors

Participant factors, such as age, sex, ethnicity, deprivation levels, lifestyle, body mass index (BMI) and medical history may result in variation of the prevalence of PSIFs or serious final diagnoses.

Studies offer conflicting results on the variation of prevalence of IFs by age and sex. The prevalence of asymptomatic brain infarcts increased with the age of participants in the Rotterdam Scan Study, from 30/750 (4%) in those aged 45–59 years, to 47/257 (18%) in those aged 75–97 years. Conversely, there was no variation in the prevalence of incidental intracranial aneurysms with age in this same sample (Vernooij et al., 2007). The prevalence of IFs on brain MRI was higher in men compared to women all aged 73 enrolled in the Lothian Birth Cohort (134/368 [36%] versus 89/332 [27%] respectively, p=0.007) (Sandeman et al., 2013), whereas another study found no difference in the prevalence of IFs on brain MRI between men and women aged 22–84 years (N=1,113) (Hoggard et al., 2009).

There are limited data on the associations between other factors and PSIFs and serious final diagnoses in apparently asymptomatic volunteers, and we can only speculate on the direction of any possible effects. Prevalence of PSIFs and of serious final diagnoses may vary by ethnicity or deprivation; if these result in differential access to healthcare, conditions are less likely to have been previously diagnosed, thus increasing the risk of PSIFs. The associations of harmful lifestyle factors such as high alcohol intake, smoking, poor diet, low physical activity and high BMI with structural diseases such as solid malignancies (Danaei et al., 2005) may mean that these factors also confer increased risks of PSIFs and serious final

diagnoses. In a study of 148 volunteers who underwent whole-body MRI, those with IFs had a higher median BMI than those without (27 kg/m^2 (interquartile range [IQR] 24–32 kg/m^2) versus 24 kg/m^2 (IQR 22–28 kg/m^2), p=0.002) (Morin et al., 2009). However, the vast majority of IFs in this study would not have been considered as PSIFs. Similarly, pre-existing morbidity may confer increased risks of structural diseases which may be detected incidentally. For example, a history of hypertension may conceivably increase the risk of incidentally detected large vessel aneurysms.

1.2.3.2 Study factors

The participant factors described above may affect the prevalence of PSIFs by influencing the underlying prevalence of diseases within a population. Study factors, relating to the imaging and the readers of those images, instead will affect the prevalence of PSIFs by influencing rates of disease demonstration and detection respectively.

1.2.3.2.1 Imaging factors

As described in Section 1.1.1.1, both research and clinical imaging are tailored to answer a particular question. This 'tailoring' occurs by varying parameters such as the field-of-view, the image resolution, use or omission of contrast media, and with regards to MRI, the strength of the magnet and the sequences performed. Studies from apparently asymptomatic volunteer and patient populations demonstrate how these factors may result in variation of the prevalence of IFs.

The prevalence of IFs appears to increase with increasing field-of-view and with increasing image resolution, but does not seem to change with magnet strength. In a study of 254 patients undergoing cardiac CT angiography, all had a low-dose whole-thorax CT before coned angiographic cardiac CT. The former, with a larger field-of-view, resulted in 66 clinically significant extra-cardiac IFs in 52 (20%) patients; the latter, with smaller field-of-view around the heart, resulted in four such findings in four (1.6%) patients (Kim et al., 2009). IFs were detected more frequently on brain MRI of apparently asymptomatic volunteers imaged using high resolution sequences, compared to standard resolution sequences (4.3% versus 1.7%, p<0.001) (Morris et al., 2009). There was no difference in the prevalence of IFs on brain MRI of 'normal volunteers' imaged on a 1.5T compared to a 3.0T scanner (8% versus 10%, p=0.6) (Hoggard et al., 2009).

The use of intravenous contrast agents may increase the overall prevalence of IFs, but may not change the prevalence of clinically significant IFs (Kim et al., 2009). However, the

influence of contrast agents on the prevalence of PSIFs may not be important to large population-based imaging studies. Immediate adverse reactions to gadolinium-based contrast occur 0.2–3.3 times per 1,000 injections, and severe reactions (such as cardiac arrest) occur in approximately 1 in 40,000 injections (Prince et al., 2011). Even though these reactions are rare, it is difficult to justify these risks of harm to apparently asymptomatic volunteers who will undergo research imaging with no benefit to themselves. As such, the larger population-based imaging studies (including UK Biobank which aims to image 100,000 people (Matthews and Sudlow, 2015)), do not perform contrast-enhanced imaging.

1.2.3.2.2 Imaging reader factors

Many different groups of readers may look at images collected for research (including scientists, radiographers, radiologists, and other medical specialists) and their level of experience may vary (The Royal College of Radiologists, 2011).

It might be assumed that medically trained, experienced readers (such as consultant radiologists) are the most appropriate people to read images for IFs in order to detect and characterise them accurately. There is no evidence that the prevalence of IFs detected by subspecialist versus general radiologists differs. A systematic review of the prevalence of IFs on brain MRI conducted in apparently asymptomatic volunteers found no significant differences between the prevalence of IFs detected in 11 studies using neuroradiologists, compared to three studies using general radiologists (3.5% versus 2.3%, p=0.3) (Morris et al., 2009).

However, shortages of radiology staff limit their availability for reporting research imaging (The Royal College of Radiologists, 2011). There was no significant difference in the proportion of 14 CT brain scans which were correctly classified as either normal or abnormal between a group of experienced readers (consultant neuroradiologists, consultant neurologists with an interest in stroke and consultant stroke physicians) compared to other readers (consultant neurologists, general practitioners and trainee physicians) (68% versus 63%, p=0.3) (Wardlaw et al., 1999). Owing to the shortage of radiologists, further work is needed to compare the prevalence of PSIFs detected by different readers, and on stratified protocols whereby radiologists review only those scans which have been ‘flagged’ as needing their attention by a first-line staff member.

Finally, the ability of readers to detect and characterise an IF may vary with the provision of information about participants, although there are currently no head-to-head comparisons to inform the likely direction of an association (if any) between the provision of clinical

information to readers and the prevalence of either PSIFs or serious final diagnoses. A lack of information about participants could reduce prevalence of PSIFs if readers' attention is not directed toward looking for particular abnormalities. Providing information on medical history may either result in 'hypervigilance' for related abnormalities, such as tumour recurrence in a research participant with a history of malignancy (thus increasing the prevalence of PSIFs), or conversely reduce the prevalence of PSIFs if readers can see that such findings have already been documented in medical records or are likely to be explained by the medical history.

To summarise, there are limited data on the associations between participant factors and PSIFs, and that which is available on age and sex is inconsistent. The prevalence of IFs may increase with increasing field-of-view and image resolution, but may not vary with magnet strength. Finally, while radiologists may be assumed to be the most appropriate people to read images for IFs, staff shortages mean this is impractical to implement for all research imaging, and there are no head-to-head comparisons of different readers to inform this aspect of IFs policy design.

1.2.4 Participants' understanding of consent to feedback

The principles of ethical medical research which involves human subjects, their data and other materials are stated in the Declaration of Helsinki (World Medical Organisation, 2013). With regards to facilitating informed consent for research:

> *"Each potential subject must be adequately informed of the aims, methods, sources of funding, any possible conflicts of interest, institutional affiliations of the researcher, the anticipated benefits and potential risks of the study and the discomfort it may entail, post-study provisions and any other relevant aspects of the study" (World Medical Organisation, 2013).*

The research staff should check that the participant has understood the information, before seeking the participant's voluntary consent (World Medical Organisation, 2013). The process of informed consent aims to respect the autonomy of potential research participants, and to protect them from potential harm (Jefford and Moore, 2008). Informed consent is also crucial to maintaining public trust in medical research (Farrar and Savill, 2014; Roache, 2014), to enable studies of human volunteers to continue and thus generate insights which benefit human health.

Regarding IFs, to facilitate informed consent for imaging research, study staff should inform participants of the potential for the detection of IFs, and explain how these will be handled (Farrar and Savill, 2014; Illes and Chin, 2008). The Netherlands Epidemiology of Obesity (NEO) population-based cohort study conducted MRI of the abdomen, plus either the brain or heart, in 2,580 individuals (de Mutsert et al., 2013). With regards to feedback of IFs, potential participants were informed that:

> *"In principle, you do not receive the result of the MRI scan. The images of the MRI scan will be interpreted by a radiologist. When unexpected abnormalities are found that are likely to have serious health consequences when left undiagnosed, we will contact you and your GP within four weeks after the MRI scan. However, when no unexpected abnormalities are identified, this will not completely exclude medical abnormalities, as the quality of the images of the MRI scan performed for the NEO study may be not as good as an MRI scan for medical diagnostics" (de Boer et al., 2018).*

To assess participants' understanding of this information, focus groups were conducted with 23/56 (41%) NEO study participants who had received feedback of an IF (de Boer et al., 2018). Despite the information on the limitations of the research imaging to completely exclude medical abnormalities, participants thought that a lack of IFs meant that they were healthy, and that the research MRI could detect diseases early (de Boer et al., 2018). These misperceptions had practical implications for participants, as they were seen as potential benefits of taking part in the study, and influenced their decisions to participate (de Boer et al., 2018).

The NEO study participants also thought that all IFs would be fed back, rather than only those with likely 'serious health consequences' (de Boer et al., 2018). This may be due to participants' underlying wishes for feedback of IFs. A study of 104 participants who had undergone neuroimaging for research found that 90% wished to be informed of any IF, regardless of its clinical significance (Kirschen et al., 2006). Similarly, qualitative interviews with 45 participants in the National Child Development Study cohort (N=17,000) who had not previously undergone MRI for research found that 41% would only participate if all IFs were fed back (Brown and Knight, 2010).

These expectations of feedback of IFs may be related to beliefs around public health screening programmes (McCaffery et al., 2016), or participants' motivations for undergoing research imaging. Research participants may view the detection of conditions before

symptoms arise as being advantageous to their health, allowing early treatment, and the opportunity to make decisions around family planning or enable relatives to undergo prompt screening (Opinion Leader, 2012; Ransohoff et al., 2002), benefits which are emphasised with regards to public health screening, but, as discussed above, may not necessarily occur in research imaging. The most common reason for taking part in imaging research appears to be the opportunity to get information about health. In SHIP, 394/405 (97%) participants said they took part to 'know whether I am healthy' (Schmidt et al., 2013). This sentiment is echoed by the early results of qualitative work with participants from the Rotterdam Scan Study (Bos and Vernooij, 2016) and participants from the NEO study, who took part in order to get results on either the presence or absence of diseases (de Boer et al., 2018). Feasibly, people in regions with overstretched publicly-funded healthcare systems may consider research imaging as a faster route to access a health service. Alternatively, people who are unable to afford healthcare in regions with private healthcare systems may look to research imaging as a 'free scan' (Wardlaw et al., 2015).

These studies demonstrate participants' misperceptions of the abilities of research imaging to make clinical diagnoses, and that their motivations to participate in studies may be based on these misperceptions. It is concerning that these misperceptions occur despite carefully worded consent materials which aim to give a realistic account of what participants can expect with regards to feedback. Evaluating participants' understanding of consent is therefore key to developing a truly informed consent process.

1.2.5 Non-medical impacts of feedback on participants

Section 1.1.1 described how the limitations of research imaging may lead to a lack of firm diagnoses. In Section 1.2.2, we saw how feedback of such IFs will tend to generate medical impacts in the form of clinical assessments which expose participants to harm, and that these medical impacts will not be offset by clinical benefits in the event of a non-serious final diagnosis. This section summarises our current knowledge on the non-medical impacts of feedback of IFs on three domains: participants' emotional wellbeing, insurance, and their work and activities.

1.2.5.1 Emotional wellbeing

Feedback of IFs may result in negative impacts on emotional wellbeing by generating distress and anxiety, the levels of which may change over time.

During the SHIP study, 405/471 (86%) of participants with IFs responded to surveys about the impact of feedback on their emotional wellbeing. Almost half (190/405, 47%) reported that they experienced some distress while waiting to receive feedback of an IF, and almost one in ten (40/405, 10%) reported feeling strongly distressed during this period (Schmidt et al., 2013). After an IF had been disclosed, a quarter had concerns about their health (90/405, 24%), and between 9–15% of respondents reported having sleeping difficulties, feelings of unrest, uncertainty or a depressed mood (Schmidt et al., 2013). Despite these reports, there were no differences in mental health component summary scores calculated from the Short Form Health Survey, or in depressive symptoms assessed using the Patient Health Questionnaire for groups of participants with IFs versus those without (Schmidt et al., 2016). This apparent contradiction may simply indicate the limited sensitivity of these instruments to detect the impact of IFs on emotional wellbeing (Schmidt et al., 2016). Similar to SHIP, focus groups conducted with participants who received feedback of an IF in the NEO study found that the period following disclosure was a 'worrying' and 'uncertain' time (de Boer et al., 2018).

These anxiety levels may wane or fluctuate over longer periods. Preliminary results from qualitative work with participants who received feedback of IFs in the Rotterdam Scan Study indicated that this had not resulted in any long-term psychological harm (Bos and Vernooij, 2016). Similarly, qualitative interviews with six US veterans who had pulmonary nodules fed back after clinical imaging found that their anxiety generally reduced over time (Sullivan et al., 2015). However, the veterans' anxieties could also spike, particularly around the time of their follow-up scans, or after seeing media coverage of health topics such as cancer (Sullivan et al., 2015).

1.2.5.2 Insurance

Disclosure of an IF to an insurance company may result in withdrawal or refusal of coverage or increased premiums, whereas failing to disclose an IF may render insurance void (Apold and Downie, 2011). Research participants, scientists and doctors recognise the potential for feedback of IFs to impact on insurance (Booth et al., 2010; Murphy and Thompson, 2009), but the story of a neuroscientist in the US shows that this is not always the case (Anon, 2005). After volunteering to help test a new research MRI scanner, he discovered he had an incidental brain tumour. He accepted a referral to a local neurosurgeon 'without proper consideration,' not realising the financial implications of his decision. After disclosing his diagnosis to his insurance company, they refused to cover him, leaving him in 'the uneasy position of facing surgery that could cost [him and his] family everything' (Anon, 2005).

1.2.5.3 Work and activities

Very little is known of the impact of feedback of IFs on participants' work and activities beyond data from the SHIP study. Impairments to work or leisure activities were reported by 13/405 (4%) and 17/405 (5%) of SHIP participants with IFs respectively (Schmidt et al., 2013). Participants with life-threatening IFs reported impairments to their work and leisure activities more frequently than those with non-life-threatening IFs (28% versus 20%; 4% versus 3% respectively) (Schmidt et al., 2013).

In summary, feedback of IFs generates distress and anxiety, which may fluctuate over time. Feedback of IFs impacts on participants' work and leisure activities regardless of whether or not the IF is life threatening, but the impact on insurance is not known.

1.2.6 Opinions of receiving feedback

Given the potential for feedback of IFs to result in clinical assessments and negative impacts on emotional wellbeing, insurance, and work and activities, the opinions of participants who receive feedback, and their healthcare providers, are paramount to informing judgements on the net benefit and harms of feeding back PSIFs. While some studies ask for people's views on hypothetical scenarios (Brown and Knight, 2010; Kirschen et al., 2006; Opinion Leader, 2012), this section will describe studies of the opinions of participants who actually receive feedback, and of healthcare providers who have experience of managing patients with IFs.

1.2.6.1 Participants' opinions

Despite reporting negative effects on emotional wellbeing (Section 1.2.5.1), qualitative interviews with 23 participants with IFs detected during the NEO study found that all were happy to have taken part in the study, and grateful to have been informed about their IFs (de Boer et al., 2018). A preliminary report of a qualitative study of participants from the Rotterdam Scan Study also suggested that participants were content with knowing about their IF, although the full report is not available as of September 2018 (Bos and Vernooij, 2016).

This apparent contradiction between participants' gratitude for receiving feedback despite the negative effects on their wellbeing seems to reflect patients' opinions of screening (de Boer et al., 2018). There is seen to be no 'downside' to screening: patients are grateful for positive results which confer the benefit of early detection of disease; patients are also grateful for negative results which confer reassurances about health (Ransohoff et al., 2002).

While this may explain the apparent contradiction in the research participants' responses, it also highlights potential misperceptions surrounding feedback of IFs detected on research imaging. As mentioned in Section 1.1.2.2, in contrast to screening, a 'positive result' (i.e. an IF) on research imaging may not indicate early detection of disease (but rather, a non-serious condition), and a 'negative result' (i.e. no IF) cannot confer reassurance of health due to the limitations of the imaging.

1.2.6.2 Healthcare professionals' opinions

Many different healthcare professionals may be involved in caring for people with IFs, due to the range of potential different types of IFs. To our knowledge, there are no studies of healthcare professionals' opinions on the benefits or harms of feedback of IFs to particular individuals, but three studies sought the general opinions of professionals who had previously managed patients with IFs.

Semi-structured interviews with 30 family doctors found that they sometimes felt compelled to follow-up IFs from clinical imaging due to low tolerance of diagnostic uncertainty, local healthcare culture, and fear of missing serious diseases (Zafar et al., 2016). These doctors described the frustration that resulted from performing costly clinical assessments of IFs which they thought would not result in clinical benefit to their patient (Zafar et al., 2016).

Similarly, within secondary care, a qualitative study of eight neurologists found that they also reported feeling compelled to follow up IFs when they too thought this was unnecessary (Booth and Boyd-Ellison, 2015). The neurologists described feeling pressure from patients to follow up IFs, particularly from patients seen within the private healthcare setting whom the neurologists perceived as feeling entitled to more tests (Booth and Boyd-Ellison, 2015). The neurologists also described an increased in their workload due to IFs detected on neuroimaging, and having to spend more time with patients who were anxious about their IFs (Booth and Boyd-Ellison, 2015).

1.2.7 Economic impact of feedback

Feedback of IFs will likely generate further clinical assessment (Sections 1.1, 1.2.2), and non-medical impacts for participants (Section 1.2.5). An unknown proportion of IFs will turn out not to represent serious disease, and any related clinical assessments and associated costs may have been unnecessary. The economic impacts of managing IFs are largely unknown; (Sandeman et al., 2013; Wardlaw et al., 2015) and this section will outline why measuring such costs is challenging.

Due to the range of IFs which may be demonstrated on brain and body imaging, the range of services which may be needed to care for patients with IFs is huge. As a group, people with IFs will likely be managed by primary care physicians, and will also come in to contact with a broad range of specialist secondary care physicians, surgeons, and diagnostic services such as imaging. The care provided may comprise of various tests and treatments, which may vary between different patients, and between different healthcare providers, healthcare services, and healthcare systems. Taking into account that the costs of these services will vary by system, and over time, estimating the cost of care for patients with IFs is complex.

Unsurprisingly, no published study takes in to account all of the potential sources of healthcare costs of managing IFs. Two UK studies (one of elderly volunteers undergoing research MRI [N=29], and one of patients undergoing CT colonography [N=225]) found mean costs of follow-up between £153 and £433 (Pinato et al., 2012; Xiong et al., 2006). However our knowledge of the cost of managing IFs is predominantly informed by studies of patients undergoing CT of a single body region in non-UK health systems (Bendix et al., 2011; Flicker et al., 2008; Gluecker et al., 2003; Goehler et al., 2014; Lee et al., 2010; Machaalany et al., 2009; Maizlin et al., 2007; Mutneja et al., 2017; Pickhardt et al., 2008; Schramm et al., 2016; Veerappan et al., 2010; Wagner et al., 2002) which unfortunately may not be generalisable to apparently asymptomatic volunteers undergoing multi-region research MRI within the UK. Furthermore, these studies may be limited due to methodological issues. Some studies confined their measurement to costs of radiological follow-up only (Gluecker et al., 2003; Maizlin et al., 2007; Priola et al., 2013; Veerappan et al., 2010; Wagner et al., 2002), resulting in underestimates of the total cost of managing IFs. The duration of follow-up is unclear in some studies, making interpretations of their results, and comparisons between them difficult (Bromage et al., 2012; Maizlin et al., 2007; Mutneja et al., 2017; Pinato et al., 2012; Schramm et al., 2016; Wagner et al., 2002). Few studies provide measures of the precision of their estimates of cost, such as 95% CIs (Lee et al., 2010; Pickhardt et al., 2008; Pinato et al., 2012).

People who receive feedback of IFs may already be using healthcare services to manage existing conditions. Teasing apart the health service uses and costs generated during participants' usual medical care from those generated during the clinical assessment of IFs from either patient records or self-reported diaries of care may be tedious, and may still underestimate resource use (Ridyard and Hughes, 2010). Data on the total cost of clinical assessment of people with PSIFs is of limited usefulness without the context of their usual healthcare costs to inform judgements on the clinical significance of any cost differences.

The differences in overall health service use and costs between groups with PSIFs versus those without (or after versus before feedback of PSIFs) is not known, but this could be feasibly measured using routinely collected healthcare data.

1.2.8 Efforts to manage IFs in imaging research

With the developments in research and clinical imaging, and concerns about the potential impacts of IFs, the challenge of how best to manage IFs in order to avoid harm to research participants and wider society has been the focus of recent reports and studies' policies.

In 2009, representatives from the Royal College of Radiologists, research imaging centres, professional societies, funders, regulatory and ethics bodies and patient organisations convened at the Wellcome Trust in London. This meeting generated a report on the management of IFs on research imaging, published in 2011, which summarised the limitations of the knowledge base, the variability in centres' IFs handling policies and access to radiology staff, and which highlighted that existing guidelines provided conflicting or ambiguous advice, and provided guidance for areas of future research (The Royal College of Radiologists, 2011).

These uncertainties about handling IFs were further acknowledged in the 'Framework on the feedback of health-related findings in research' proposed in 2014 by two of the largest medical research funders in the UK (Medical Research Council and Wellcome Trust, 2014). They defined health-related findings as those with potential health or reproductive importance. In contrast to earlier calls for a uniform approach to handling IFs (Brown and Hasso, 2008), this framework offers researchers flexible guidance on what to consider when developing an appropriate IFs policy. It encourages researchers to recognise the uncertainties around IFs, to tailor feedback policies to the specific context of their study, and to manage participants' expectations about the way that IFs will be handled (Medical Research Council and Wellcome Trust, 2014). Both the Wellcome Trust and the Medical Research Council mandate their researchers to submit an IFs policy along with any study application, and also have committed to covering the costs of feedback pathways for their funded studies (Farrar and Savill, 2014).

The Wellcome Trust and Medical Research Council's framework, or rather, its approach, has not yet been adopted internationally. The relevant guidance in Canada does not specifically advise researchers to design an IFs policy in advance of starting a study (Canadian Institutes of Health Research et al., 2014). Rather, they advise researchers to consult colleagues or

existing guidance on how to handle particular IFs when they arise. The guidelines are ambiguous regarding the IFs which should be disclosed, advising on one hand that any IF which may have welfare significance for a participant should be disclosed, but on the other hand that researchers should exercise caution in disclosing findings which may have implications for insurance and employability (Canadian Institutes of Health Research et al., 2014). This may cause difficulties in practice; how to disclose a finding of potential significance to a participant who works as a driver, or pilot, for example.

Dutch researchers developed a list of minimum standards for handling IFs in imaging research after analysing qualitative interviews with 20 researchers (Bunnik et al., 2017). The Dutch minimum standards were further developed and refined following a two-day meeting of 14 representatives of large European population-based imaging studies. However, some of the proposed minimum standards may not be feasible. For example, they recommend that 'in studies in which diagnostic-quality images are acquired, some form of routine review of research scans should be arranged' (Bunnik et al., 2017). Given that the majority of studies will generate images with some level of anatomical detail on which a large lesion such as a tumour or aneurysm may well be visible, implementing this standard in the UK may overwhelm already overstretched clinical radiology services.

In the same year that the Wellcome Trust and Medical Research Council's Framework was made available, American guidance published by the Presidential Commission for the Study of Bioethical Issues also acknowledged the lack of empirical evidence to inform the management of IFs. The latter addressed this issue by calling on professional bodies to develop guidance for the management of IFs, and for empirical data on the detection, feedback and management of IFs, and for such research to be funded by federal agencies and other parties (Presidential Commission for the Study of Bioethical Issues, 2013).

Following this report, several groups published guidance on the management of IFs on different imaging modalities, and of specific types of IFs (Heller et al., 2013; Hoang et al., 2015; Khosa et al., 2013; MacMahon et al., 2017; Patel et al., 2013; Sebastian et al., 2013). Recently, one of the largest European population-based imaging studies, the German National Cohort, conducted an extensive literature review, sought radiological opinions about IFs, and developed lists of IFs that they consider to be for feedback or not (Bertheau et al., 2016). Further lists have been developed by professional bodies and other research studies (Bos and Vernooij, 2016; The Royal College of Radiologists, 2011). We need more information on the impact on participants and health services of receiving feedback of

different types of IFs, and what the final diagnoses turn out to be, in order to judge the usefulness of these lists.

The UK Biobank Imaging Study subsequently adapted the lists generated by the German National Cohort to inform its feedback policy for IFs detected on multi-modal imaging of 100,000 of its 500,000 existing participants (Gibson et al., 2018). UK Biobank developed a pragmatic IFs policy that aimed to minimise harm to this large number of apparently asymptomatic people (radiographer flagging of concerning scans for a radiologist to review (Gibson et al., 2018)). The UK Biobank Imaging Study only feeds back IFs that are 'potentially serious' (defined as those 'indicating the possibility of a condition which, if confirmed, would carry a real prospect of seriously threatening life span, or of having a substantial impact on major body functions or quality of life' (Gibson et al., 2018)). This decision was informed by the nature of UK Biobank participants' existing consent, and is consistent with other studies' policies (Bertheau et al., 2016; Bos and Vernooij, 2016; UK Biobank Ethics and Governance Council, 2015). Before participant recruitment began, UK Biobank acknowledged that its staff may have professional or moral obligations to disclose abnormalities that may impact on participants' health (e.g. during the initial recruitment visit, a research nurse may notice a hypertensive blood pressure reading or a skin melanoma) and participants joined UK Biobank after consenting to receive feedback of such findings (UK Biobank, 2013a). Within UK Biobank, it was deemed important to maintain a consistent approach to feedback of health-related information in order to promote participants' understanding of the feedback policies and facilitate their informed consent to participate in different aspects of the UK Biobank study (UK Biobank, 2013a). Therefore for the imaging study, UK Biobank maintained the position of feeding back only findings which were likely to impact on health, and formalised its approach by defining (and adapting lists of) PSIFs in anticipation of the prevalence and clinical implications of IFs which would be demonstrated on the research imaging (UK Biobank, 2013a).

Defining IFs which are for feedback or not may generate a tension between studies' IFs and participants' rights to know (and rights not to know) about their health. Such rights are endorsed by both the Convention on Human Rights and Biomedicine, and the United Nations Educational, Scientific and Cultural Organization's Declaration on the Human Genome and Human Rights (Council of Europe, 1997; United Nations Educational Scientific and Cultural Organization, 1997). In addition, research participants' rights not to know about their health may come into conflict with researchers' professional and ethical obligations to disclose finding which may impact on health (e.g. a participant who is a lorry driver who is

found to have a brain tumour). UK Biobank take a pragmatic approach, by informing participants of what they can and cannot reasonably expect of the study (in light of the non-diagnostic nature of the imaging and that images will not routinely be reviewed by medical staff), emphasising that only potentially serious abnormalities will be fed back, and that lack of feedback does not confer 'health' (UK Biobank, 2013a; 2014a). Participants consent to take part on this basis, and are not offered the option to 'opt-out' of feedback (UK Biobank, 2013a). However, such an approach may not be suitable for all studies, and while it is likely that there is no 'one-size-fits-all' policy for handling IFs in research (UK Biobank, 2013a), empirical evidence is needed to inform appropriate policy design for studies conducted across a range of different contexts.

1.3 Summary

IFs occur in approximately 1–2% of apparently asymptomatic volunteers who undergo brain MRI, and some evidence suggests that the prevalence is likely to be higher on body MRI. Due to the limitations of research imaging, and of current knowledge about the natural history of particular IFs, the clinical significance of some IFs may not be clear, which may prompt further clinical assessments to resolve the diagnostic uncertainty. However, the numbers, types and costs of clinical assessments performed in pursuit of final diagnoses of IFs is not known, and are potentially challenging to study. Furthermore, an unknown proportion of these IFs will represent serious disease, so that any clinical assessments performed, non-medical impacts on participants' emotional well-being, insurance, work and activities or uses and costs of health services may be deemed unnecessary in the event of non-serious final diagnoses. While a small number of studies suggest that participants may not hold negative opinions about feedback of IFs, despite experiencing negative impacts on emotional wellbeing, studies of healthcare professionals highlight their concerns about inappropriate uses of healthcare services to manage IFs.

The increasing popularity of imaging as a research tool, particularly in large, population-based cohorts, will only increase the scale of the challenge of imaging IFs in future. In response, major research funders have called for scientists to develop IFs policies before starting their studies, which should be underpinned by evidence and clearly communicated to participants. Researchers should estimate the likely prevalence of IFs which may be detected when imaging a particular population, but there is inconsistent evidence on the associations of IFs with age and sex, and the influence of other participant, imaging, and imaging reader factors on IFs is relatively unknown. Due to the current shortage of radiologists, data on the

prevalence of PSIFs and serious final diagnoses detected by other types of imaging readers are urgently needed to inform IFs policy design. Despite clear descriptions of IFs policies, some evidence suggests that participants may misperceive research imaging as being able to provide firm clinical diagnoses or firm reassurance of health, and that they misunderstand consent materials with regards to what will be fed back.

More data are needed to inform on the prevalence and types of IFs on brain and body MRI conducted in apparently asymptomatic people. Long-term follow-up studies of unselected participants with IFs are needed to generate evidence on the clinical assessment, final diagnoses, non-medical impacts on participants, economic impact on health services, and participants' and healthcare professionals' opinions on handling IFs. These empirical data are crucial to informing judgements of the benefits and harms of feedback of IFs. A greater understanding of the factors which influence prevalence of IFs, including head-to-head comparisons of different imaging readers, will help researchers estimate the likely scale of the problem of IFs in their future studies. Evaluating participants' understanding of consent with regards to IFs feedback policies will enable consent processes to be improved. Collectively, these data will help to inform the design of appropriate IFs policies which minimise harm to research participants and publicly-funded healthcare services.

1.4 Aims of this thesis

Generating new knowledge about IFs which are clearly non-serious would be of limited potential value to individuals considering undergoing imaging, and the researchers and healthcare providers managing them. Therefore, the remainder of this thesis focuses on PSIFs, which we define as IFs indicating the possibility of a condition which, if confirmed, would carry a real prospect of seriously threatening life span, or of having a substantial impact on major body functions or quality of life (Gibson et al., 2018).

The remaining chapters of this thesis aim to address gaps in knowledge relating to at least one of the following themes relating to PSIFs: prevalence and nature; follow-up and final diagnoses; factors associated with PSIFs and with serious final diagnoses; participants' understanding of consent to feedback of PSIFs; non-medical impacts of feedback of PSIFs on participants; opinions of receiving feedback of PSIFs; economic impact of feedback of PSIFs.

Chapter 2 presents a systematic review and meta-analysis of published studies of brain and body MRI which aims to determine the prevalence and types of PSIFs in apparently

asymptomatic adults, describe factors associated with PSIFs, and summarise what is known on follow-up and final diagnoses.

All of the remaining research analyses described in this thesis use data generated by the UK Biobank Imaging Study. Chapter 3 provides a summary of the UK Biobank, the UK Biobank Imaging Study, and the rationale behind and protocol for handling PSIFs on multi-modal imaging in 100,000 largely asymptomatic UK Biobank participants.

Chapter 4 presents a study which compares two PSIFs handling protocols which were applied to the first 1,000 imaged UK Biobank participants: radiographer flagging of concerning images for a radiologist to review, versus systematic radiologist review of all images. Chapter 4 addresses multiple themes of this thesis, by presenting empirical data on the prevalence and nature of PSIFs on multiple body regions and imaging modalities, the follow-up generated and the resulting final diagnoses, and how these vary by each of the PSIFs protocols. Chapter 4 also presents data on: an initial exploration of participant factors associated with PSIFs; non-medical impacts of feedback of PSIFs on participants; participants' understanding of consent to feedback of PSIFs; and participants' and their GPs' opinions on receiving feedback of a PSIF.

Chapter 5 builds on aspects of the work presented in Chapter 4. By using data from a larger cohort of UK Biobank participants (N=7,334), Chapter 5 aims to improve the accuracy of our estimates of the prevalence of both PSIFs and serious final diagnoses, and investigate the associations of these with a wider range of participant factors.

Chapter 6 furthers our knowledge of the impact of feedback of PSIFs. While Chapter 4 will describe the medical and non-medical impacts on participants, the case-control study presented in Chapter 6 aims to explore the economic impact of feedback of PSIFs on hospital services using linked routinely collected healthcare data.

The results of these studies are summarised and discussed in the context of each other and the broader literature in Chapter 7, which will also describe the strengths, limitations and implications of this thesis and potential directions for future research.

Chapter 2 Potentially serious incidental findings on brain and body magnetic resonance imaging of apparently asymptomatic adults: systematic review and meta-analysis

2.1 Introduction

Chapter 1 described the current literature and highlighted areas where robust, empirical data on incidental findings (IFs) are needed to inform the design of feasible IFs management policies and improve informed consent processes. The usefulness of the existing estimates of prevalence is limited by variation between studies' definitions of IFs, some of which included clearly non-serious conditions (e.g. simple renal cysts (Morin et al., 2009)), or normal variant anatomy (e.g. mega cisterna magna (Haberg et al., 2016)). Data on the factors associated with potentially serious IFs (PSIFs) and serious final diagnoses would facilitate researchers' estimations of the expected prevalence of both of these outcomes in future studies, and inform their design of consent materials, but there is conflicting evidence on the associations with participants' age and sex, and limited informative data on the associations with other factors. While there are some data on the follow-up and final diagnoses of participants with IFs detected on brain magnetic resonance imaging (MRI) (Bos et al., 2016), more information about the burden of follow-up (such as the numbers undergoing particular types of follow-up) and final diagnoses of PSIFs detected on brain and other body regions would help facilitate the informed consent of potential research participants.

The work presented in this chapter was originally developed to inform the ongoing UK Biobank Imaging Study, which aims to conduct brain, cardiac and body MRI, carotid Doppler ultrasound and dual-energy X-ray absorptiometry (DXA) in 100,000 largely asymptomatic people (Matthews and Sudlow, 2015). As such, this chapter focuses on apparently asymptomatic people (defined as community-dwelling people not selected for imaging on the basis of symptoms, risk factors, or disease) undergoing MRI, as this is the UK Biobank imaging modality which is most likely to generate PSIFs due to the large volume of imaged tissue and the range of pathologies which may be demonstrated therein.

This chapter aims systematically to review studies of brain, thorax, abdomen and of brain and body MRI conducted among apparently asymptomatic adults to 1) determine the

prevalence and types of PSIFs, 2) describe factors associated with PSIFs, and 3) determine what is known about the follow-up and final diagnoses of people with PSIFs.

This study was submitted to the BMJ; the manuscript is included in full in Section 2.2 and the supplementary materials are included in full in Section 2.3.

2.2 Potentially serious incidental findings on brain and body magnetic resonance imaging of apparently asymptomatic adults: systematic review and meta-analysis (submitted)

Lorna M Gibson, Laura Paul, Francesca M Chappell, Malcolm Macleod, William N Whiteley, Rustam Al-Shahi Salman, Joanna M Wardlaw, Cathie LM Sudlow.

Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh BioQuarter, 9 Little France Road, Edinburgh, EH16 4UX, UK

Lorna M Gibson

Clinical research training fellow

Cathie LM Sudlow

Professor

Department of Clinical Radiology, Ground Floor, Queen Elizabeth Building, Glasgow Royal Infirmary, Alexandra Parade, Glasgow, G31 2ER, UK

Laura Paul

Specialty trainee in clinical radiology

Centre for Clinical Brain Sciences, University of Edinburgh, Chancellor's Building, 49 Little France Crescent, Edinburgh, EH16 4SB, UK

Francesca M Chappell

Medical statistician

Malcolm Macleod

Professor

William N Whiteley

Clinician scientist

Rustam Al-Shahi Salman

Professor

Joanna M Wardlaw

Professor

Correspondence to: Cathie LM Sudlow cathie.sudlow@ed.ac.uk

2.2.1 Structured abstract

Objectives

To 1) determine prevalence and types of potentially serious incidental findings (PSIFs) on magnetic resonance imaging (MRI) in apparently asymptomatic adults, 2) describe factors associated with PSIFs, and 3) summarise information on follow-up and final diagnoses.

Design

Systematic review and meta-analyses.

Data sources

Medline and Embase (inception to 25th April 2017), citation searches of relevant articles and authors' files.

Review methods

We included published studies reporting prevalence and types of incidental findings (IFs) detected among apparently asymptomatic adults undergoing MRI of brain, thorax, abdomen or brain and body. We extracted data on study population and methods, prevalence and types of IFs, and final diagnoses. We estimated pooled prevalence using random effects meta-analysis, and heterogeneity using tau-squared statistics.

Main outcome measures

Prevalence of PSIFs on MRI of brain, thorax, abdomen, and brain and body.

Results

Among 5,905 retrieved studies, 32 (0.5%) met the inclusion criteria (n=27,643 participants), pooled prevalence of PSIFs on brain and body MRI was: 3.9% (95% confidence interval [CI] 0.4–27.1%; brain 1.4% [95% CI 1.0–2.1%]; thorax 1.3% [95% CI 0.2–8.1%], abdomen 1.9% [95% CI 0.3–12.0%]); and 12.8% (95% CI 3.9–34.3%) when including IFs of uncertain potential seriousness, with generally substantial heterogeneity among included studies. Around half of PSIFs were suspected malignancies (brain 0.6% [95% CI 0.4–0.9%]; thorax 0.6% [95% CI 0.1–3.1%]; abdomen 1.3% [95% CI 0.2–9.3%]; brain and body 2.3% [95% CI 0.3–15.4%]). There were few informative data on potential sources of between-study

variation or factors associated with PSIFs. Limited data suggested that relatively few PSIFs had serious final diagnoses (48/234, 20.5%).

Conclusions

A substantial proportion of apparently asymptomatic adults will have PSIFs on MRI, but little is known of their health consequences. Systematic, long-term follow-up studies are needed to better inform on these and the implications for policies on feedback of PSIFs.

Systematic review registration

PROSPERO CRD42016029472.

2.2.2 What this paper adds

What is already known on this topic

Estimates of prevalence of IFs vary widely, and may be of limited value to practice as they often include non-serious IFs.

Previous systematic reviews have focused on IFs detected on MRI of a single body region, patient populations undergoing MRI, or apparently asymptomatic people imaged using another modality.

These estimates are not generalisable to brain and body MRI of apparently asymptomatic people, i.e. imaging which is increasingly conducted within large-scale imaging research and screening settings.

What this study adds

In meta-analyses of published studies, pooled prevalence of PSIFs on MRI of apparently asymptomatic people was 3.9% (1.4% brain, 1.3% thorax, 1.9% abdomen), and 12.8% (1.7% brain, 3.0% thorax, 4.5% abdomen) when including IFs of uncertain potential seriousness.

Around half of PSIFs were suspected malignancies.

Limited follow-up data suggest that most PSIFs may not be clinically serious on follow-up, and further research is needed.

2.2.3 Introduction

Brain and body (i.e. brain, thorax and abdomen) magnetic resonance imaging (MRI) is increasingly used for clinical and commercial screening and for research, with several large-scale population-based imaging initiatives ongoing around the world (Icelandic Heart Association, 1999; Ikram et al., 2015; Nationale Kohorte, 2018; Post, 2014; UK Biobank, 2018c). The detection of incidental findings (IFs) unrelated to the purpose of the imaging (Wolf et al., 2008) is an inevitable consequence. Clinicians and researchers should therefore anticipate IFs and develop appropriate policies for managing them, taking into account their expected prevalence and clinical severity (Medical Research Council and Wellcome Trust, 2014). Existing data on the prevalence of IFs from systematic reviews of MRI of a single body region (Morris et al., 2009), patient populations undergoing MRI (Dunet et al., 2016), or apparently asymptomatic people imaged using another modality (Jacobs et al., 2008), are not generalisable to brain and body MRI of apparently asymptomatic people (defined here as community-dwelling people not selected for imaging on the basis of symptoms, risk factors, or disease).

The clinical severity of IFs ranges from non-serious (e.g. simple renal cyst) to potentially life-threatening (e.g. some malignancies), but their nature and severity are often unclear. Diagnostic radiological imaging is tailored optimally to demonstrate (or exclude) pathologies relevant to a patient's presentation. By contrast, since IFs are, by definition, unrelated to the imaging's purpose (Wolf et al., 2008), no imaging protocol is specifically designed to optimise firm diagnoses of these. Further specific clinical follow-up is therefore often needed to permit final clinical diagnoses of IFs.

Given that knowing about clearly non-serious IFs would be of limited potential benefit, we focus here on 'potentially serious' IFs (PSIFs), defined as those indicating the possibility of a condition which, if confirmed, would carry a real prospect of seriously threatening life span, or of having a substantial impact on major body functions or quality of life (Petersen et al., 2013). The development of well-informed approaches to the management of such PSIFs on brain and body MRI in apparently asymptomatic adults requires data on their prevalence and types, factors associated with these, and on the resulting final diagnoses.

We therefore aimed systematically to review studies of brain, thorax, abdomen and of brain and body MRI to 1) determine the prevalence and types of PSIFs among apparently asymptomatic adults, 2) describe factors associated with PSIFs, and 3) determine what is

known about the follow-up and final diagnoses of people with PSIFs. This study was motivated by - and mainly conducted during preparations for - the ongoing UK Biobank multi-modal imaging study (including brain and body MRI) of 100,000 people (UK Biobank, 2018c).

2.2.4 Methods

We registered the protocol for this review with PROSPERO (Gibson et al., 2016a), and archived data online (Gibson et al., 2017a).

2.2.4.1 Patient involvement

Patients were not involved in the development or design of this study.

2.2.4.2 Data sources

We searched Medline and Embase from inception to 25th April 2017 for references to studies in any language which reported the prevalence of IFs in apparently asymptomatic adults undergoing cardiac, abdominal or brain and body (i.e. brain and thorax and abdomen) MRI (Supplementary Table 2-1). For brain MRI, we screened studies included in a published systematic review of IFs in apparently asymptomatic volunteers (Morris et al., 2009) and updated the search to 25th April 2017 (Supplementary Table 2-2). We searched authors' files and forward and backward citations of retrieved studies for further relevant studies.

2.2.4.3 Study selection

One author (LG) screened all references for potentially eligible studies. A second author (LP) independently screened a random sample of 10% of references to assess the reliability of this process. Disagreements were resolved through discussion between these authors, with arbitration by a senior author (CLMS) if necessary. We retrieved full text articles of potentially eligible studies. One author (LG) assessed articles for inclusion, and discussed uncertainties with a senior author (CLMS).

We defined apparently asymptomatic people as those who were not selected on the basis of any symptoms, risk factors, or disease, and who attended for population-based research imaging studies, commercial or occupational screening, or as research controls. We excluded studies of: patients (i.e. people selected for a study based on symptoms, risk factors or disease, or those admitted to or attending a health care facility for clinical diagnostic imaging); magnetic resonance angiography which only reported vascular IFs (due to limited

generalisability); pre-specified subgroups of IFs (which would underestimate the prevalence of other IFs); children (<18 years old). We excluded studies which were not published in full. If multiple publications arose from a study, we prioritised the primary review question of prevalence, and included data from the largest cohort.

2.2.4.4 Data extraction

One author (LG) extracted data from all included studies on study population, study methods, and prevalence and types of all IFs using a pre-piloted, standardised data-extraction spreadsheet. To assess the reliability of this process, a second author (LP) independently extracted data from a 10% random sample of studies. Disagreements were resolved through discussion between these authors.

2.2.4.4.1 Study and population characteristics

We extracted data on: sample size; numbers of men and women; mean age and age range of participants; country in which the imaging was conducted (or, if this was not reported, the country of the first author's institution); body region(s) imaged; imaging setting (classified as either research [if participants were imaged during research studies], or non-research settings [imaging was performed in other contexts, including occupational imaging, or commercial imaging]).

2.2.4.4.2 Study imaging and IF reporting methods

We extracted data on: whether prevalence of IFs was assessed by reviewing MR images or reports; the specialist field and number of those reporting images; blinding of reporters to information about the participants; the MRI sequences performed; the dates that MRI was performed.

2.2.4.4.3 Data on IFs

We extracted data on: the total number of participants with IFs, the total number of IFs, or both if available; the number of participants with multiple IFs; the prevalence of IFs by age, sex, imaging sequence, reporter or any other factor assessed for association with IFs; all available data on follow-up investigations, treatment and final diagnoses for studies in which all participants with IFs or a specified subtype or severity of IFs were followed-up systematically.

2.2.4.4.4 Classification of IFs and final diagnoses

To determine which IFs were potentially serious according to our definition (Petersen et al., 2013), we referred to a list of potentially serious and non-serious IFs developed by UK Biobank, based on consultations with radiologists, published literature and the German National Cohort's methods (Bertheau et al., 2016) (Supplementary Tables 2-3, 2-4 and 2-5). For any IFs not on this list, we directly applied our definition of a PSIF; where there was insufficient published information to apply our definition, we used study definitions of severe IFs, accepting that these vary somewhat between studies (Gibson et al., 2017a). We sub-classified PSIFs as suspected malignancy (e.g. masses), non-malignant, or possible indicators of malignancy (IFs which were not masses, but could be related to malignancy, e.g. pleural effusions [Section 2.3.1.3]). We classified final diagnoses as serious if they were likely to significantly threaten lifespan or have a major impact on quality of life or major body functions, and not serious if this was not the case. We described IFs or final diagnoses that could not be classified as 'indeterminate.'

2.2.4.5 Risk of bias assessment

In the absence of a validated quality assessment tool for studies of the prevalence of IFs, we extracted data on study characteristics which may influence risk of bias (sample selection methods, blinding of reporters to information about the participants, the specialty and number of image readers, and whether data on IFs were generated from reads of images or extracted from reports), and planned to consider their potential influence on the results through a series of subgroup analyses.

2.2.4.6 Data synthesis

We meta-analysed studies with a random effects model (Borenstein et al., 2010), using maximum likelihood estimation methods (Hamza et al., 2008) and modelling within-study variance as binomial, to calculate pooled prevalence of PSIFs, and of suspected malignant IFs, separately for MRI of brain, thorax, abdomen, and brain and body. For the pooled estimates, we calculated both 95% confidence intervals (CI) and 95% prediction intervals; the latter indicate the range of true prevalence values expected in future studies (Riley et al., 2011). We used t-scores (rather than the usual z score) to calculate 95% CIs, generating conservative estimates and allowing comparison with our prediction intervals (which also use t-scores). We included region-specific data from studies of brain and body MRI in the brain, thoracic and abdominal MRI meta-analyses. We derived data on thoracic IFs from studies of either cardiac or brain and body MRI or both. To obtain upper estimates of the

prevalence of PSIFs and of suspected malignant IFs, we performed sensitivity meta-analyses by adding the indeterminate IFs to the PSIFs, and possible indicators of malignancy to the suspected malignant IFs. We calculated 95% CIs for individual studies' prevalence estimates using Clopper Pearson exact methods. We assessed statistical heterogeneity using tau-squared statistics, which provide a logit scale measure of between-study variance, represented in a more readily interpretable way by the 95% prediction intervals. We initially considered all study-level characteristics as potential candidates for subgroup analyses to explore reasons for heterogeneity of the prevalence of PSIFs. However, we chose not to conduct subgroup analyses that were likely to be un-informative (e.g. due to missing data for a large proportion of studies or substantial imbalance in subgroup sizes). We performed subgroup analyses by including study characteristics as covariates in the meta-analyses (Petitti, 2001). We decided not to perform formal statistical tests for possible publication bias since their application is limited in meta-analyses where outcome is expressed as a proportion (Bland, 2006; Hunter et al., 2014). We further decided not to conduct formal meta-analysis of data on the percentage of PSIFs that resulted in serious final diagnoses (i.e., the positive predictive value of PSIFs), to avoid undue emphasis on the limited data available. Instead, we described available findings and calculated a rough estimate of this percentage by summing numerators and denominators across the few studies with relevant data.

We used Microsoft Excel 2013 for descriptive statistical analyses, StatsDirect 3.0.177 for calculating 95% CIs for individual studies, and SAS 9.4 PROC NLMIXED (www.sas.com) for meta-analyses.

We obtained all data for this study from existing publications, and so did not need ethical approval.

2.2.5 Results

Two authors agreed on 99% of the duplicate screened reference selections, and 100% of the duplicate extracted data.

2.2.5.1 Included studies

We included 32 studies (Alphs et al., 2006; Baumgart and Egelhof, 2007; Bos et al., 2016; Boutet et al., 2017; Brugulat-Serrat et al., 2017; Cieszanowski et al., 2014; Goehde et al., 2005; Haberg et al., 2016; Hartwigsen et al., 2010; Hegenscheid et al., 2013; Hoggard et al.,

2009; Illes et al., 2004b; Katzman et al., 1999; Kumar et al., 2008; Lee et al., 2008; Li et al., 2015; Lo et al., 2008; Loy et al., 2015; Lubman et al., 2002; Menzler et al., 2010; Morin et al., 2009; Reneman et al., 2012; Sandeman et al., 2013; Saya et al., 2017; Sommer et al., 2013; Tarnoki et al., 2015; Trufyn et al., 2014; Tsushima et al., 2005; Vogel-Claussen et al., 2009; Wahlund et al., 1989; Weber and Knopf, 2006; Yue et al., 1997) of 27,643 participants (range 2 to 5,800 participants, mean/median age range 21 to 75 years, 14,037/27,643 [50.8%] male) imaged between 1985 and 2016 (Supplementary Figure 2-1, Supplementary Table 2-6). These 32 studies comprised eight of brain and body MRI (Baumgart and Egelhof, 2007; Cieszanowski et al., 2014; Goehde et al., 2005; Hegenscheid et al., 2013; Lo et al., 2008; Morin et al., 2009; Saya et al., 2017; Tarnoki et al., 2015), 22 of brain MRI (Alphs et al., 2006; Bos et al., 2016; Boutet et al., 2017; Brugulat-Serrat et al., 2017; Haberg et al., 2016; Hartwigsen et al., 2010; Hoggard et al., 2009; Illes et al., 2004b; Katzman et al., 1999; Kumar et al., 2008; Lee et al., 2008; Li et al., 2015; Lubman et al., 2002; Menzler et al., 2010; Reneman et al., 2012; Sandeman et al., 2013; Sommer et al., 2013; Trufyn et al., 2014; Tsushima et al., 2005; Wahlund et al., 1989; Weber and Knopf, 2006; Yue et al., 1997) and two of cardiac MRI (Loy et al., 2015; Vogel-Claussen et al., 2009). No abdomen-only studies were identified (Supplementary Table 2-6).

Studies were performed in Europe (20 studies, 17,702 participants (Baumgart and Egelhof, 2007; Bos et al., 2016; Boutet et al., 2017; Brugulat-Serrat et al., 2017; Cieszanowski et al., 2014; Goehde et al., 2005; Haberg et al., 2016; Hartwigsen et al., 2010; Hegenscheid et al., 2013; Hoggard et al., 2009; Loy et al., 2015; Menzler et al., 2010; Morin et al., 2009; Reneman et al., 2012; Sandeman et al., 2013; Saya et al., 2017; Sommer et al., 2013; Tarnoki et al., 2015; Wahlund et al., 1989; Weber and Knopf, 2006)), North America (six studies, 5,789 participants (Alphs et al., 2006; Illes et al., 2004b; Katzman et al., 1999; Trufyn et al., 2014; Vogel-Claussen et al., 2009; Yue et al., 1997)), Asia (four studies, 3,576 participants (Lee et al., 2008; Li et al., 2015; Lo et al., 2008; Tsushima et al., 2005)), and Australia (two studies, 576 participants (Kumar et al., 2008; Lubman et al., 2002)) (Supplementary Table 2-6). All but three assessed images for IFs; one assessed imaging reports (Lubman et al., 2002), and two did not report on this (Lee et al., 2008; Wahlund et al., 1989). All studies involved radiologists, except one in which a cardiologist reported IFs on cardiac MRI (Supplementary Table 2-6) (Loy et al., 2015); in two studies, radiologists were involved in confirming IFs detected by others (trained readers [defined as researchers with training to doctor of medicine-level or training in neuropsychiatry] in one study (Bos et al., 2016) and MRI scan operators [not further defined] in another (Li et al., 2015)).

2.2.5.2 Imaging sequences

The vast majority of participants were imaged using scanners of 1.5T or less (19 studies, 23,809/27,643 [86.1%] participants (Baumgart and Egelhof, 2007; Bos et al., 2016; Boutet et al., 2017; Cieszanowski et al., 2014; Goehde et al., 2005; Haberg et al., 2016; Hegenscheid et al., 2013; Kumar et al., 2008; Lee et al., 2008; Lubman et al., 2002; Menzler et al., 2010; Morin et al., 2009; Sandeman et al., 2013; Saya et al., 2017; Sommer et al., 2013; Tsushima et al., 2005; Vogel-Claussen et al., 2009; Weber and Knopf, 2006; Yue et al., 1997)). However, seven studies (1,556/27,643 [5.6%] participants) used 3.0T scanners (Brugulat-Serrat et al., 2017; Hartwigsen et al., 2010; Hoggard et al., 2009; Lo et al., 2008; Loy et al., 2015; Tarnoki et al., 2015; Trufyn et al., 2014), two studies (370/27,643 [1.3%] participants) used 1.5T in some participants and 3.0T in others (Li et al., 2015; Reneman et al., 2012), and four studies (1,908/27,643 [6.9%] participants) did not report magnet strength (Alphs et al., 2006; Illes et al., 2004b; Katzman et al., 1999; Wahlund et al., 1989) (Supplementary Table 2-7). All but three brain MRI studies (Cieszanowski et al., 2014; Sommer et al., 2013; Wahlund et al., 1989) used T1-weighted imaging; one further study used T1-weighted imaging in an unknown subset of participants (Hoggard et al., 2009). Of the ten thoracic MRI studies, eight used non-contrast whole thorax imaging (n=4,817) (Baumgart and Egelhof, 2007; Cieszanowski et al., 2014; Goehde et al., 2005; Hegenscheid et al., 2013; Lo et al., 2008; Morin et al., 2009; Saya et al., 2017; Tarnoki et al., 2015) and five used cardiac-specific sequences (n=4,099) (Baumgart and Egelhof, 2007; Goehde et al., 2005; Hegenscheid et al., 2013; Loy et al., 2015; Vogel-Claussen et al., 2009). All abdominal MRI studies used T1-weighted imaging (Supplementary Table 2-7).

2.2.5.3 Risk of bias assessment

Only one study appeared to have imaged an unselected, random population sample (n=2,500) (Hegenscheid et al., 2013). The majority of the remainder imaged selected samples or did not clearly report sampling methods. At least one radiologist reported all images in almost all studies; 14 studies had more than one reader for each set of images (8,199/27,643 [29.7%] participants (Baumgart and Egelhof, 2007; Cieszanowski et al., 2014; Goehde et al., 2005; Haberg et al., 2016; Hartwigsen et al., 2010; Hegenscheid et al., 2013; Illes et al., 2004b; Lo et al., 2008; Menzler et al., 2010; Sandeman et al., 2013; Saya et al., 2017; Tarnoki et al., 2015; Tsushima et al., 2005; Vogel-Claussen et al., 2009)) (Supplementary Table 2-6). Data on blinding of readers to participants' characteristics were incomplete, with only 16 studies (19,617/27,643 [71.0%] participants) clearly reporting blinding of image readers to participant characteristics (Alphs et al., 2006; Bos et al., 2016;

Boutet et al., 2017; Cieszanowski et al., 2014; Goehde et al., 2005; Haberg et al., 2016; Hegenscheid et al., 2013; Li et al., 2015; Lubman et al., 2002; Menzler et al., 2010; Reneman et al., 2012; Sandeman et al., 2013; Saya et al., 2017; Sommer et al., 2013; Weber and Knopf, 2006; Yue et al., 1997) (Supplementary Table 2-6). There were no direct within-study comparisons of radiologist versus non-radiologist readers, of single versus multiple readers, or of blinding versus non-blinding of readers to participants' characteristics to reliably inform on any potential biases such methods may have on the prevalence of PSIFs.

2.2.5.4 Prevalence and types of PSIFs

Although 14 studies reported data on multiple IFs per participant, none provided the number of participants with >1 PSIF, or data to enable calculations of this (Alphs et al., 2006; Boutet et al., 2017; Goehde et al., 2005; Haberg et al., 2016; Hartwigsen et al., 2010; Hegenscheid et al., 2013; Lee et al., 2008; Morin et al., 2009; Sandeman et al., 2013; Saya et al., 2017; Sommer et al., 2013; Trufyn et al., 2014; Vogel-Claussen et al., 2009; Weber and Knopf, 2006). We therefore based prevalence estimates on the assumption that no participant had >1 PSIF, recognizing that a very small number of participants may have more than one. The pooled prevalences of PSIFs on brain, thoracic, abdominal and brain and body MRI were 1.4% (95% CI 1.0–2.1%), 1.3% (95% CI 0.2–8.1%), 1.9% (95% CI 0.3–12.0%) and 3.9% (95% CI 0.4–27.1%) respectively. When indeterminate IFs were included, pooled prevalence estimates increased to 1.7% (95% CI 1.1–2.6%), 3.0% (95% CI 0.8–11.3%), 4.5% (95% CI 1.5–12.9%) and 12.8% (95% CI 3.9–34.3%) respectively. Study-specific prevalence estimates ranged widely, with correspondingly wide prediction intervals, and tau-squared values ranging from 0.8 to 5.7 (indicative of substantial variance between studies) (Figures 2-1 and 2-2, Supplementary Figure 2-2, Supplementary Table 2-8).

Region imaged and study name	Sample size	Potentially serious incidental findings only[a]		Potentially serious incidental findings and indeterminate incidental findings[a]	
		n	Prevalence, % (95% CI)	n	Prevalence, % (95% CI)
Brain					
Bos 2016	5800	364	6.3 (5.7 to 6.9)		
Yue 1997	3672	17	0.4 (0.2 to 0.7)		
Weber 2006	2536	18	0.7 (0.4 to 1.1)	24	1.0 (0.6 to 1.4)
Hegenscheid 2013	2500	53	2.1 (1.6 to 2.8)	55	2.2 (1.7 to 2.9)
Lee 2008	2164	20	0.9 (0.6 to 1.4)		
Tsushima 2005	1113	12	1.1 (0.6 to 1.9)		
Baumgart 2007	1007	13	1.3 (0.7 to 2.2)		
Haberg 2016	1006	28	2.8 (1.9 to 4.0)		
Katzman 1999	1000	9	0.9 (0.4 to 1.7)		
Sommer 2013	722	13	1.8 (1.0 to 3.1)	21	2.9 (1.8 to 4.4)
Sandeman 2013	700	14	2.0 (1.1 to 3.3)		
Cieszanowski 2014	666	0	0.0 (0.0 to 0.5)		
Alphs 2006	656	34	5.2 (3.6 to 7.2)	61	9.3 (7.2 to 11.8)
Brugulat-Serrat 2017	545	23	4.0 (2.6 to 5.9)	37	6.4 (4.6 to 8.8)
Hoggard 2009	525	14	2.7 (1.5 to 4.4)		
Boutet 2016	503	15	3.0 (1.7 to 4.9)		
Kumar 2008	478	10	2.1 (1.0 to 3.8)	12	2.5 (1.3 to 4.3)
Goehde 2005	298	5	1.7 (0.6 to 3.9)		
Hartwigsen 2010	206	16	7.8 (4.5 to 12.3)		
Reneman 2012	203	1	0.5 (0.01 to 2.7)	9	4.4 (2.0 to 8.2)
Li 2015	167	1	0.6 (0.02 to 3.3)		
Illes 2004	151	3	2.0 (0.4 to 5.7)	10	6.6 (3.2 to 11.8)
Morin 2009	148	0	0.0 (0.0 to 2.4)		
Lo 2008	132	0	0.0 (0.0 to 2.7)		
Wahlund 1989	101	0	0.0 (0.0 to 3.6)		
Menzler 2010	100	2	2.0 (0.2 to 7.0)		
Lubman 2002	98	1	1.0 (0.03 to 5.6)		
Trufyn 2014	56	2	3.6 (0.4 to 12.3)	3	5.4 (1.1 to 14.9)
Saya 2017	44	0	0.0 (0.0 to 8.0)		
Tarnoki 2015	22	0	0.0 (0.0 to 15.4)		
Combined	**27349**	**688**	**1.4 (1.0 to 2.1)**	**763**	**1.7 (1.1 to 2.6)**
Heterogeneity			$T^2 = 0.8$		$T^2 = 1.0$

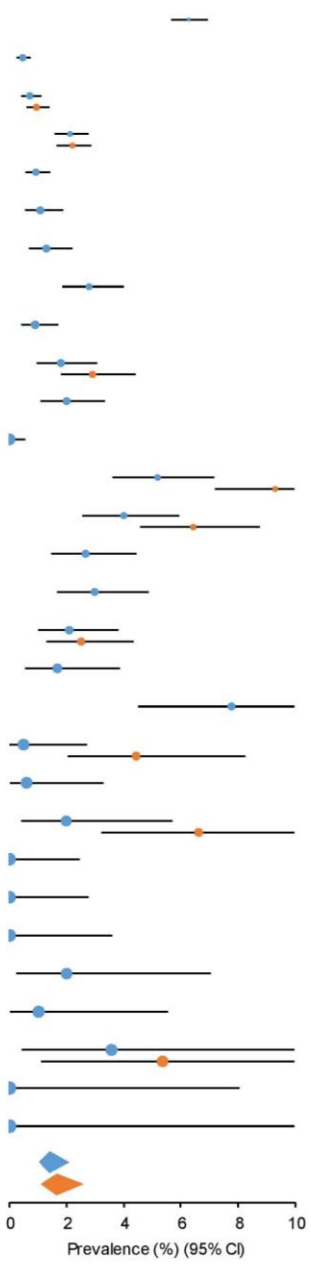


Figure 2-1: Forest plots of the per-study prevalence and pooled prevalence estimates of potentially serious incidental findings (PSIFs), and of PSIFs plus indeterminate incidental findings (IFs), detected on brain magnetic resonance imaging (MRI)

CI = confidence interval

Tau-squared is an estimate of between-study variance on the logit scale. Zero represents no variance, and increasing values of tau-squared indicate increasing heterogeneity.

Blue = Per-study point prevalence and pooled prevalence estimate of PSIFs on brain MRI

Orange = Sensitivity analyses which include IFs classified as indeterminate in the per-study point prevalence and pooled prevalence estimate of PSIFs on brain MRI. Details of the types and numbers of PSIFs are provided in Figure 2-3 and Supplementary Table 2-9, while details of indeterminate findings are available online (Gibson et al., 2017a).

a. **We excluded 138 vascular IFs detected in six studies that used MR angiography, from pooled analyses (Alphs et al., 2006; Goehde et al., 2005; Haberg et al., 2016; Lee et al., 2008; Tsushima et al., 2005; Weber and Knopf, 2006).**

Region imaged and study name	Sample size	Potentially serious incidental findings only[a]		Potentially serious incidental findings and indeterminate incidental findings[a]	
		n	Prevalence, % (95% CI)	n	Prevalence, % (95% CI)
Thorax					
Hegenscheid 2013	2500	163	6.5 (5.6 to 7.6)		
Baumgart 2007	1007	1	0.0 (0.0 to 0.5)		
Cieszanowski 2014	666	0	0.0 (0.0 to 0.5)	69	10.4 (8.2 to 12.9)
Goehde 2005	298	38	12.8 (9.2 to 17.1)	39	13.1 (9.5 to 17.5)
Vogel-Claussen 2009	254	26	10.2 (6.8 to 14.6)	29	11.4 (7.8 to 16.0)
Morin 2009	148	1	0.6 (0.0 to 3.7)		
Lo 2008	132	5	3.8 (1.2 to 8.6)	8	6.1 (2.7 to 11.6)
Saya 2017	44	0	0.0 (0.0 to 8.0)		
Loy 2015	40	0	0.0 (0.0 to 8.8)		
Tarnoki 2015	22	4	18.2 (5.2 to 40.3)		
Combined	**5111**	**238**	**1.3 (0.2 to 8.1)**	**314**	**3.0 (0.8 to 11.3)**
Heterogeneity			$T^2 = 5.1$		$T^2 = 2.9$
Abdomen					
Hegenscheid 2013	2500	380	15.2 (13.8 to 16.7)	481	19.2 (17.7 to 20.8)
Baumgart 2007	1007	5	0.5 (0.2 to 1.2)		
Cieszanowski 2014	666	0	0.0 (0.0 to 0.5)	32	4.8 (3.3 to 6.7)
Goehde 2005	298	11	3.7 (1.9 to 6.5)		
Morin 2009	148	2	1.4 (0.2 to 4.8)	3	2.0 (0.4 to 5.8)
Lo 2008	132	8	6.1 (2.7 to 11.6)		
Saya 2017	44	0	0.0 (0.0 to 8.0)	1	2.3 (0.06 to 12.0)
Tarnoki 2015	22	6	27.3 (10.7 to 50.2)		
Combined	**4817**	**412**	**1.9 (0.3 to 12.0)**	**547**	**4.5 (1.5 to 12.9)**
Heterogeneity			$T^2 = 4.2$		$T^2 = 1.6$
Brain and body					
Hegenscheid 2013	2500	685	27.4 (25.7 to 29.2)	793	31.7 (29.9 to 33.6)
Baumgart 2007	1007	19	1.9 (1.1 to 2.9)		
Cieszanowski 2014	666	0	0 (0 to 0.5)	133	20.0 (17.0 to 23.2)
Goehde 2005	298	54	18.1 (13.9 to 23.0)	55	18.5 (14.2 to 23.3)
Morin 2009	148	3	2.0 (0.4 to 5.8)	4	2.7 (0.7 to 6.8)
Lo 2008	132	15	11.4 (6.5 to 18.0)	18	13.6 (8.3 to 20.7)
Saya 2017	44	0	0.0 (0.0 to 8.0)	2	4.6 (0.6 to 15.5)
Tarnoki 2015	22	10	45.5 (24.4 to 67.8)	15	68.2 (45.1 to 86.1)
Combined	**4817**	**786**	**3.9 (0.4 to 27.1)**	**1039**	**12.8 (3.9 to 34.3)**
Heterogeneity			$T^2 = 5.7$		$T^2 = 2.0$

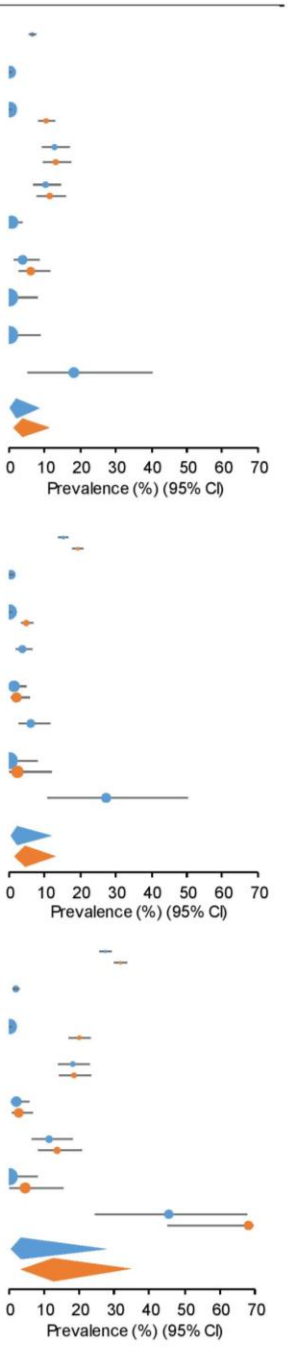


Figure 2-2: Forest plots of the per-study prevalence and pooled prevalence estimates of potentially serious incidental findings (PSIFs), and of PSIFs plus indeterminate incidental findings (IFs), detected on thoracic, abdominal and brain and body magnetic resonance imaging (MRI)

CI = confidence interval

Tau-squared is an estimate of between-study variance on the logit scale. Zero represents no variance, and increasing values of tau-squared indicate increasing heterogeneity.

Blue = Per-study point prevalence and pooled prevalence estimate of PSIFs on thoracic, abdominal and brain and body MRI

Orange = Sensitivity analyses which include IFs classified as indeterminate in the per-study point prevalence and pooled prevalence estimate PSIFs on thoracic, abdominal and brain and body MRI. Details of the types and numbers of PSIFs are provided in Figure 2-3 and Supplementary Tables 2-10 and 2-11, while details of indeterminate findings are available online (Gibson et al., 2017a).

a. **We excluded 200 IFs detected in studies that used specialist imaging sequences (97 breast lesions in a study including MR mammography (Hegenscheid et al., 2013), 87 colonic polyps in two studies which included MR colonography (Baumgart and Egelhof, 2007; Goehde et al., 2005), 15 vascular findings such as stenosis or plaque in four studies which included MR angiography (Baumgart and Egelhof, 2007; Goehde et al., 2005; Hegenscheid et al., 2013; Tarnoki et al., 2015), and one myocardial infarction in a study which included post-contrast cardiac imaging (Goehde et al., 2005)) from pooled analyses.**

Across body regions, suspected malignancies were the most common types of PSIFs (accounting for roughly half of all such findings), with vascular findings also common on brain MRI (Figure 2-3 and Supplementary Tables 2-9, 2-10 and 2-11). Pooled prevalences of suspected malignant PSIFs were: brain 0.6% (95% CI 0.4–0.9%; thorax 0.6% (95% CI 0.1–3.1%); abdomen 1.3% (95% CI 0.2–9.3%); and brain and body 2.3% (95% CI 0.3–15.4%). When possible indicators of malignancy were included, these were 0.6% (95% CI 0.4–0.9%), 1.0% (95% CI 0.2–5.4%), 1.6% (95% CI 0.2–10.9%) and 3.0% (95% CI 0.4–20.4%) respectively (Supplementary Figure 2-2).

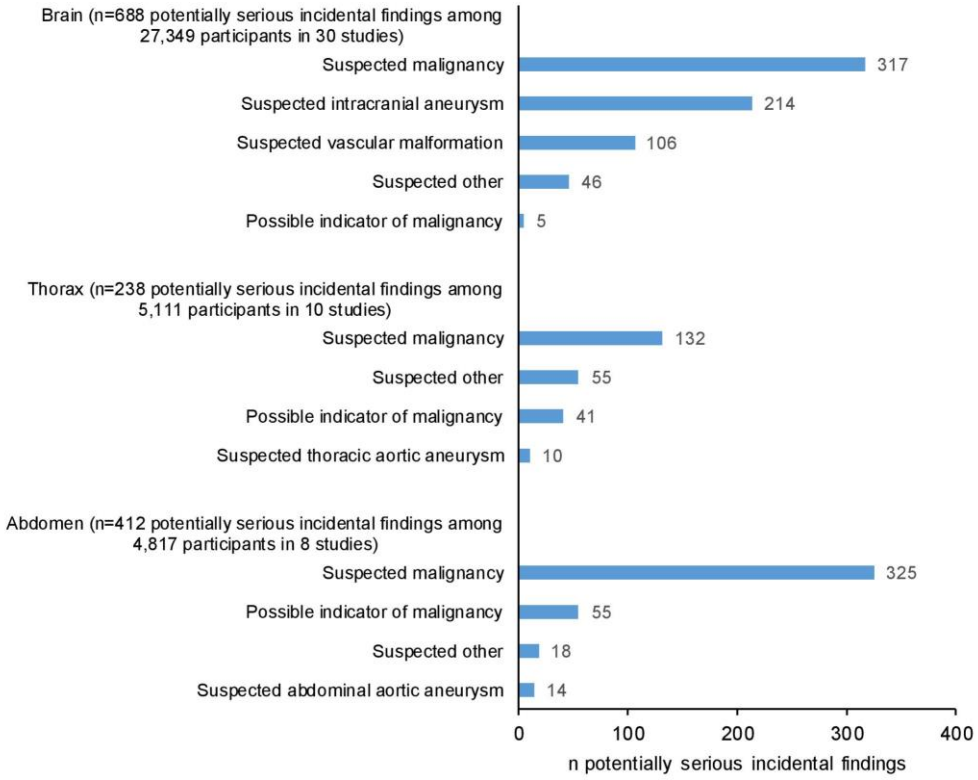


Figure 2-3: Numbers and types of potentially serious incidental findings (PSIFs) on magnetic resonance imaging (MRI) by body region

Further details of the types of PSIFs are provided in Supplementary Tables 2-9, 2-10 and 2-11. We sub-classified PSIFs as suspected malignancy (e.g. masses), possible indicators of malignancy (incidental findings (IFs) which were not masses, but could be related to malignancy, e.g. pleural effusions) or non-malignant (Section 2.3.1.3). For the purposes of this figure, PSIFs which were not suspected malignancies, possible indicators of malignancies, or suspected vascular findings were grouped as 'suspected other.'

2.2.5.5 Subgroup analyses

Examination of the available data (Supplementary Tables 2-6 and 2-7) showed that several potential subgroup analyses would be uninformative due to very imbalanced subgroups or non-reporting of the relevant data for a large subset of studies. One or both of these reasons precluded subgroup analyses with respect to magnet strength (almost all 1.5T), contrast use (incomplete data), data source (almost all studies used images rather than reports of these), image reader specialty (almost all studies had reporting by radiologists), and sample selection method (only one study randomly selected participants) (Hegenscheid et al., 2013). We did not conduct subgroup analyses by: age or sex, because we did not have individual participant data to allow meaningful comparisons; study country, since there was no clear a priori reason for variation in PSIFs prevalence by country; or body region because studies of brain and body MRI contributed data on different body regions from the same participants, violating the assumption that data within different subgroups are independent. We conducted brain and body and region-specific MRI subgroup analyses for imaging setting (research versus non-research) and for several factors which may inform on risks of bias (blinding of readers to participant characteristics and number of image readers) where sufficient data allowed. There was no evidence of any clinically meaningful or statistically significant difference in prevalence of PSIFs following the inclusion of subgroups as covariates (Supplementary Figures 2-3 to 2-11, Supplementary Table 2-12).

2.2.5.6 Study-specific reports of factors associated with PSIFs

Eight studies reported factors associated with PSIFs (Bos et al., 2016; Brugulat-Serrat et al., 2017; Haberg et al., 2016; Hoggard et al., 2009; Morin et al., 2009; Saya et al., 2017; Sommer et al., 2013; Yue et al., 1997), while a further five reported factors associated with IFs requiring follow-up, which we considered an approximate proxy for PSIFs (Hartwigsen et al., 2010; Illes et al., 2004b; Loy et al., 2015; Sandeman et al., 2013; Tsushima et al., 2005) (Supplementary Tables 2-13, 2-14 and 2-15). Two studies found significant associations between IFs requiring follow-up and increasing age (Hartwigsen et al., 2010; Illes et al., 2004b), while a further two studies found a consistently higher prevalence of IFs requiring follow-up (Tsushima et al., 2005) and cavernomata (Brugulat-Serrat et al., 2017) in older age groups, albeit not statistically significant (Supplementary Table 2-13). There was no clear variation in prevalence of PSIFs by sex (Supplementary Table 2-14). Too few data were available on other factors (including medical history, symptoms, lifestyle factors and genetics) to demonstrate any clear associations with PSIFs (Supplementary Table 2-15). No

data were available on the associations between imaging sequence or reporter specialty with prevalence of PSIFs.

2.2.5.7 Follow-up and final diagnoses

Only five studies systematically followed-up and reported data on the final clinical diagnoses of selected subsets of participants with IFs (total number of such participants followed up = 234), representing 1.4 to 18.2% of all imaged participants in these studies (Table 2-1) (Bos et al., 2016; Lo et al., 2008; Morin et al., 2009; Sandeman et al., 2013; Saya et al., 2017). Summing arithmetically across these studies, overall only 48 of these 234 participants (i.e. about one fifth) had clinically serious final diagnoses (although half had indeterminate final diagnoses, mostly from one study of brain MRI (Bos et al., 2016), in which participants were managed under 'wait and see' policies). No study reported follow-up in a manner which enabled enumeration of the clinical assessments (e.g. further imaging examinations, specialty referrals, biopsies etc.) performed to clarify final diagnoses.

Table 2-1: Methods of follow up of 234 people with potentially serious incidental findings and severity of their final diagnoses

Study variables			Methods of follow-up of incidental findings			Severity of final diagnoses (n)		
First author surname and year of publication	Imaged body regions	n followed-up/ N total imaged (%)	Subset of participants followed-up[a]	Data type (source)	Duration of follow-up	Serious	Non-serious	Indeter-minate
Bos 2016	Brain	188/5800 (3.2)	All those with an incidental finding who were referred to specialists[b]	Clinical management (medical records)	Until last clinical follow-up or death	39	34	115
Sandeman 2013	Brain	10/700 (1.4)	All those with an incidental finding who were referred to family doctors[c]	Resulting action (medical records)	-	5	5	0
Morin 2009	Brain and body	5/148 (3.4)	All those with highly significant findings[d]	Investigations and treatments (contact with general practitioner or participant)	-	0	3	2
Lo 2008	Brain and body	24/132 (18.2)	All those with an incidental finding which required further work-up[e]	-	-	4	20	0
Saya 2017	Brain and body	7/44 (15.9)	All with incidental findings deemed to require follow-up[f]	Investigations (-)	-	0	7	0
Total n (% of 234 followed up)			.	.	.	48 (20.5)	69 (29.5)	117 (50)

- = not specified, . = not applicable

a. This could be considered as a study-specific proxy for potentially serious incidental findings (PSIFs) but is not identical to the consistent definition that we applied in meta-analyses of prevalence of PSIFs. Hence study-specific n here differs from study specific numbers of PSIFs in meta-analyses.

b. Decision for referral depended on the IF and consultation with clinicians.

c. Decision for referral depended on discussion between radiologists and a geriatrician and other clinicians as necessary.

d. Highly significant findings were defined as those requiring prompt medical follow-up, such as indeterminate masses in solid organs, enlarged lymph nodes and ovarian masses/cysts, as judged by consensus of two radiologists. Participants' family doctors were informed of the finding.

e. Definition of IFs requiring further work-up, or processes for judging this are not reported.

f. As determined by study radiologists, follow-up was discussed by a multi-disciplinary team including principal investigators, radiologists and other study staff (not otherwise specified).

2.2.6 Discussion

2.2.6.1 Principal findings

We performed meta-analyses of published studies of the prevalence of PSIFs among apparently asymptomatic adults undergoing MRI of brain, thorax, abdomen or brain and body. The pooled prevalence of PSIFs was 3.9% (1.4% brain, 1.3% thorax, 1.9% abdomen). When additionally including IFs of uncertain potential seriousness, the pooled prevalence increased to 12.8% (1.7% brain, 3.0% thorax, 4.5% abdomen). There was wide variation among studies in their prevalence estimates, likely reflecting variation between studies in participants' characteristics, imaging setting, sample selection methods, and methods of detecting IFs, as well as the challenges of applying a consistent definition of PSIFs to the available descriptions of IFs in published papers. Suspected malignant IFs accounted for around half of all PSIFs on brain, thoracic, abdominal and brain and body MRI (0.6%, 0.6%, 1.3% and 2.3% respectively). The very limited systematic follow-up data available (mainly from brain MRI studies) demonstrated that only about 1/5 people with a PSIF had a serious final clinical diagnosis.

2.2.6.2 Strengths and limitations of this study

By including all identified published data on the prevalence of PSIFs on brain, thoracic, abdominal and brain and body MRI, and by applying as consistent as possible a definition of PSIFs across studies, we have provided data on the prevalence of those IFs which may have an important impact on health. This is the first review to include data on PSIFs from different body regions, enabling comparisons of prevalence between regions. As such, our results are informative to people undergoing, or staff conducting, brain and body or region-specific MRI in apparently asymptomatic adult volunteers. As most studies comprised selected apparently asymptomatic populations, our results are directly applicable to imaging performed for research and non-research settings such as screening.

While we have not shown evidence of a statistically significant difference in the prevalence of PSIFs between body regions, the pooled point prevalences were generally higher on abdomen MRI, and on brain and body MRI compared to either brain or thorax MRI, particularly so when indeterminate findings were included in sensitivity analyses. This pattern is biologically plausible and was also seen in data from some primary studies (Gibson et al., 2018; Hegenscheid et al., 2013; Lo et al., 2008; Morin et al., 2009; Tarnoki et al., 2015). It is possible that the heterogeneity between included studies, the relative rarity of PSIFs, methods of meta-analyses and conservative calculation of 95% CIs may have obscured true differences in the prevalence of PSIFs between regions. Results on IFs from ongoing large population-based imaging studies (including the UK Biobank imaging sub-study, which by September 2018 had imaged >27,000 of an intended 100,000 participants) should be able to confirm or refute this pattern in future (Bertheau et al., 2016; German National Cohort (GNC) Consortium, 2014; Gibson et al., 2016b; UK Biobank, 2018c).

There was no evidence of any meaningful differences in the prevalence of PSIFs between studies conducted in research or imaging settings for any body region, or between studies using readers blinded to participant characteristics versus not blinded or not stated, or for brain MRI studies using one versus >1 reader. Further subgroup analyses which may inform on factors influencing variation in prevalence in different body regions were limited, as data on relevant variables were either lacking for a large subset of studies, or resulted in very imbalanced subgroups.

Data were included in the review after screening and extraction by one, rather than multiple authors. While this may limit the accuracy of the data extraction, it is unlikely to have substantially impacted on our results given the very good agreement with a second reviewer on a 10% subset of the studies. Due to the lack of data on the participants with >1 PSIF, prevalence estimates were based on the assumption that only one PSIF occurs per participant; however, it is unlikely that a substantial proportion of participants had >1 PSIF. The prevalence of IFs deemed 'potentially serious' may vary with opinion and over time as evidence of their natural history accrues.

We could not explore the influence of technical imaging factors (e.g. image resolution, magnet strength) on the prevalence of PSIFs, due to limited data availability and reporting consistency, but these are unlikely to substantially influence the detection of the most common PSIFs (suspected malignancies and aneurysms). The vast majority of included studies involved systematic radiologist reviews of images to detect IFs. No study directly compared radiologist to non-radiologist readers, although other policies to detect IFs may produce very different results, such as radiographer flagging of concerning examinations for a radiologist to review (Gibson et al., 2016b).

2.2.6.3 Comparison with other studies

A recently published umbrella review of IFs arising from a range of imaging modalities (including MRI) found no existing systematic reviews of the prevalence of IFs in apparently asymptomatic volunteers on cardiac, abdominal or brain and body MRI for comparison with our findings (O'Sullivan et al., 2018).

Our update of an existing systematic review by Morris et al. (Morris et al., 2009) of IFs on brain MRI resulted in similar prevalence of suspected malignant IFs. The aforementioned recent umbrella review reported a prevalence of IFs on brain MRI of 22% (95% CI 14–31%), around ten times higher than our pooled prevalence estimate for brain MRI (Morris et al., 2009; O'Sullivan et al., 2018; Takashima et al., 2017). The majority of this difference is likely to be due to the umbrella review's inclusion of all reported IFs, regardless of their potential clinical significance, whereas we focused on PSIFs. Some of the difference may also be due to different study inclusion criteria (reflecting the different focus of the umbrella review, which had broader inclusion criteria, including studies of patients as well as apparently asymptomatic people), as well as a difference in meta-analytic methodologies. Prevalence data, as proportions, will have a binomial distribution. The umbrella review used an arcsine transformation in its analyses of prevalence data, which avoids the challenge of

directly modelling binomial data, whereas we used an exact method, which does model the within-study variance as binomial to generate unbiased estimates (Hamza et al., 2008).

The recent umbrella review also reported far more final diagnosis data from studies derived from Morris et al. than we have here (Morris et al., 2009; O'Sullivan et al., 2018). In order to calculate the proportion of IFs resulting in known final diagnoses, the participants who form the denominator should all undergo systematic follow-up in order to generate an accurate numerator. We therefore scrutinised reports of all our included studies and found that only five reported such systematic methods; we did not consider diagnosis data from other studies to be robust, since they may represent suspected, rather than final diagnoses.

2.2.6.4 Implications of this study

Apparently asymptomatic people may undergo brain and body MRI by participating in research, or access non-research MRI via referral from a doctor (Tarnoki et al., 2015), or directly (Lee et al., 2008; Tarnoki et al., 2015; Tsushima et al., 2005) (e.g. as part of occupational screening (Weber and Knopf, 2006), private health insurance (Cieszanowski et al., 2014), or company health care programmes (Goehde et al., 2005; Tarnoki et al., 2015)). Our prevalence data could be used to inform consent for MRI in both research and non-research settings. Such data could also help researchers calculate anticipated numbers of participants with PSIFs in future studies, to inform the design of appropriate IFs handling policies.

Our review highlights the limited data available on the follow-up and final diagnoses of PSIFs. Such data would inform judgements about the benefits versus harms of feeding back PSIFs, an issue which warrants further investigation with systematic, long-term follow-up of participants with PSIFs. Unlike public health screening programmes, which fulfill specific criteria to ensure net benefit (Wilson et al., 1968), identification of a PSIF does not always lead to detection of disease at a stage where intervention will confer benefit. Many PSIFs will turn out to be clinically non-serious, but require potentially anxiety-provoking follow-up and potentially uncomfortable or harmful investigations to discover this. Even for those PSIFs that do turn out to be clinically serious, for most there is no clear evidence base to inform decisions about treatment, and early treatment of some disorders may confer harm (Mohr et al., 2014). Our prevalence data could inform power calculations for future clinical trials of conservative or active treatments of PSIFs, in order to develop good medical practices which minimise harm to people with PSIFs, and ensure appropriate use of health services.

2.2.7 Additional information

2.2.7.1 Acknowledgements

Dr Christian Schnier and Dr Hanna Johnsson translated articles. This work was conducted on behalf of the UK Biobank Imaging Working Group; members are listed at the end of this article. We thank the BMJ's statistical reviewer, Professor Richard Riley, for his advice on previous versions of this manuscript.

2.2.7.2 Data sharing statement

The full dataset is available at http://dx.doi.org/10.7488/ds/2100 with open access.

2.2.7.3 Copyright and licensing

2.2.7.4 Competing interests

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare: Dr. Gibson reports grants from Wellcome Trust, during the conduct of the study, and personal fees from UK Biobank, outside the submitted work; Professor Sudlow is Chief Scientist of UK Biobank; the remaining authors have no financial relationships with any organisations that might have an interest in the submitted work in the previous three years or other relationships or activities that could appear to have influenced the submitted work**.**

2.2.7.5 Contributors and guarantor

Dr Lorna M Gibson designed and conducted the study, collected, managed, analysed and interpreted data, and prepared, reviewed and approved the manuscript.

Dr Laura Paul collected data, and reviewed and approved the manuscript.

Professor Malcolm Macleod advised on methods, interpreted data, reviewed and approved the manuscript.

Dr Francesca M Chappell analysed and interpreted data, and prepared, reviewed and approved the manuscript.

Dr William N Whiteley interpreted data, reviewed and approved the manuscript.

Professor Rustam Al-Shahi Salman interpreted data, reviewed and approved the manuscript.

Professor Joanna M Wardlaw designed and supervised the study, interpreted data, and reviewed and approved the manuscript.

Professor Cathie LM Sudlow designed and supervised the study, interpreted data, reviewed, approved and decided to submit the manuscript for publication.

Professor Cathie LM Sudlow is the guarantor.

The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

2.2.7.6 Transparency declaration

Prof. Cathie LM Sudlow (the manuscript's guarantor) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

2.2.7.7 Ethical approval

We obtained all data for this study from existing publications, and so did not need ethical approval.

2.2.7.8 Sources of funding

Dr Lorna M Gibson is funded by a Wellcome Trust Clinical Research Training Fellowship (107190/Z/15/Z).

2.2.7.9 Role of the study sponsors

None of the authors' funding organisations contributed to the design and conduct of the study; collection, management, analysis, or interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

2.2.7.10 Statement of independence of the researchers from the funders

All authors are independent from the funders.

2.2.7.11 Patient involvement

Patients were not involved in the development or design of this study. The results of this study will be disseminated to the public by the investigators where possible.

2.2.7.12 Author access to data

All authors had full access to all of the data (including statistical reports and tables) in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis.

2.3 Supplementary materials

2.3.1 Supplementary methods

2.3.1.1 Search strategies

Search strategy 1 and 2 were designed/adapted respectively by Dr Lorna Gibson (MBChB) and Professor Cathie Sudlow (DPhil).

Databases: Embase

Ovid MEDLINE

Supplementary Table 2-1: Search strategy 1: Incidental findings on cardiac and abdominal MRI

Line number	Search term
1	exp Magnetic Resonance Imaging/
2	exp Nuclear Magnetic Resonance Imaging/
3	(magnetic resonance or MRI or MR or NMR).tw.
4	1 or 2 or 3
5	(abdom$ or cardiac or heart or cardio$ or whole-body or (whole adj2 body)).tw.
6	*Abdomen/
7	*Heart/
8	5 or 6 or 7
9	4 and 8
10	Incidental findings/
11	(incidental$ or subclinical or serendipit$ or unexpected or asymptomatic).tw.
12	10 or 11
13	Humans/
14	9 and 12 and 13
15	(Conference abstract or Conference report or Conference paper or Conference review or Case reports or comment or editorial or letter or news).pt.
16	14 not 15

Supplementary Table 2-2: Search strategy 2: Update for review of incidental findings on brain MRI

Line number	Search term
1	(MR or MRI or magnetic resonance imaging or neuroimaging).tw.
2	(screen$ or incidental$ or healthy or asymptomatic or volunteer or control).tw.
3	(cranial or brain$ or neuro$).tw.
4	*Magnetic Resonance Imaging/
5	*Brain/ or exp Brain Diseases/
6	Humans/
7	1 and 2 and 3 and 4 and 5 and 6
8	(Conference abstract or Conference report or Conference paper or Conference review or Case reports or comment or editorial or letter or news).pt.
9	7 not 8
10	limit 9 to yr="2008 -Current"
11	remove duplicates from 10

Source: **Adapted from Morris et al. 2009**

2.3.1.2 Lists used to classify incidental findings (IFs) as potentially serious or non-serious

UK Biobank developed lists of findings which would be considered potentially serious, and findings not considered serious for use by radiographers and reporting radiologists. These lists were based on lists generated by the German National Cohort (Bertheau et al., 2016), and are subject to ongoing review. These lists were used to classify incidental findings (IFs) reported in included studies as potentially serious, or non-serious.

Supplementary Table 2-3: Incidental findings on brain magnetic resonance imaging

Potentially serious	Not serious
Acute brain infarction	Asymmetrical ventricles
Acute hydrocephalus	Chiari malformation[d]
Acute intracranial haemorrhage[a]	Chronic hydrocephalus
Arachnoid cyst[b]	Developmental anomalies (including venous anomalies)
Colloid cyst of third ventricle	Lipoma of corpus callosum
Intracranial mass lesion[c]	Non-acute brain infarction
Mastoiditis	Non-specific white matter hyperintensities
Suspected intracranial aneurysm or vascular malformation	Regional or global atrophy
	Suspected demyelination

a. Not old bleeds, or microbleeds only detected on gradient recalled echo sequences.

b. Only if large and considered likely to increase the risk of developing a subdural haematoma.

c. Except meningiomata in locations considered highly unlikely to cause problems.

d. Descent of part of the cerebellum +/- brainstem below the foramen magnum.

Supplementary Table 2-4: Incidental findings on thoracic magnetic resonance imaging

Potentially serious	Not serious
Aortic dissection	Atelectasis
Cardiac mass (including thrombus)	Calcified pleural plaque
Central pulmonary embolus	Calcified pulmonary nodule
Haemodynamically relevant pericardial effusion > 2 cm	Emphysema
Heart valve defects[a]	Right sided descending aorta
Hilar, mediastinal, axillary or cervical lymphadenopathy[b]	
Lobar pneumonia or lung consolidation	
Lung mass > 2 cm	
Mediastinal mass > 2 cm	
Pleural effusion	
Pleural mass > 2 cm	
Pneumothorax	
Severe left or right ventricular dilation or dysfunction	
Severe left ventricular hypertrophy > 2 cm thick wall	
Thoracic aortic aneurysm > 5 cm	

a. Severe regurgitation jet of any valve or severe turbulence (suggesting valve stenosis).

b. > 1.5 cm and > 3 lymph nodes grouped in a circumscribed region.

Supplementary Table 2-5: Incidental findings on abdominal magnetic resonance imaging

Potentially serious	Not serious
Abdominal aortic aneurysm > 5 cm	Abdominal wall hernia
Acute exudative pancreatitis	Bladder diverticulum
Adrenal lesion > 2 cm	Chronic cholecystitis
Ascites	Chronic pancreatitis
Cholestasis (intra- or extra-hepatic)[a]	Fatty liver
Deep vein thrombosis	Fibroids
Hepatomegaly	Gallstones
Ileus	Hiatus hernia
Intra-abdominal mass > 3 cm	Left sided inferior vena cava
Irregular/nodular liver margin	Liver cyst
Lymphadenopathy[b]	Renal calculus
Multiple small non-cystic, liver lesions (non haemangioma-like)	Simple renal cyst
Pneumoperitoneum	Single kidney
Portal vein occlusion	
Pyelonephritis	
Renal artery stenosis > 80% or bilateral	
Solid/cystic pancreatic tumour	
Solid gallbladder lesion	
Solid liver lesion	
Solid/semi-solid renal tumour > 2 cm	
Spleen infarction	
Splenomegaly > 15 cm	
Urinary obstruction	
Urinary tract mass > 2 cm	

a. Common bile duct > 15 mm (or > 20 mm post-cholecystectomy).

b. > 1.5 cm and > 3 lymph nodes grouped in a circumscribed region.

2.3.1.3 Sub-classification of potentially serious incidental findings (PSIFs)

Suspected malignancy

We sub-classified PSIFs as suspected malignancy if they described tumours, masses, complex cysts or lesions.

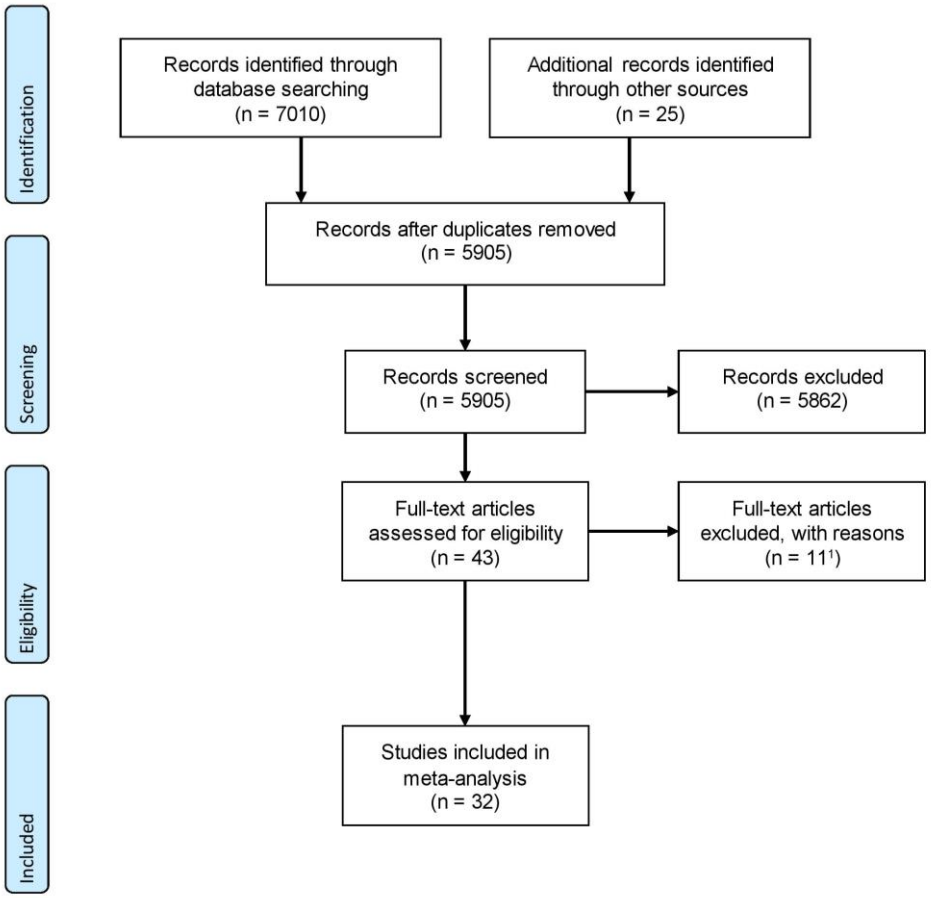
Possible indicator of malignancy

Other PSIFs which may be indicative of malignancy are those which are not masses, and may be related to either malignancy, or another aetiology. These included pleural and pericardial effusions, enlarged lymph nodes, hydronephrosis, splenomegaly, biliary dilatation, ascites, hydrocephalus and severe bone oedema.

Non-malignant

If a PSIF could not be classified as either suspected malignancy, or as a possible indicator of malignancy, we sub-classified it as non-malignant.

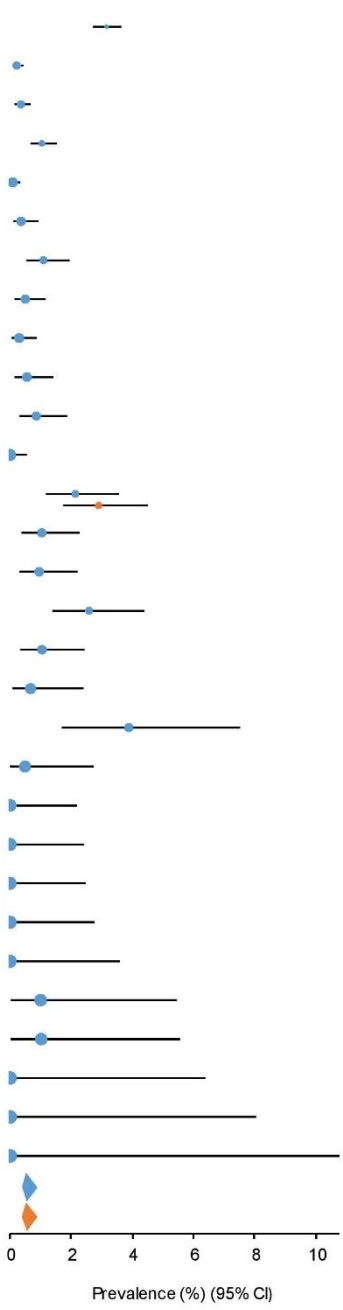
2.3.2 Supplementary figures



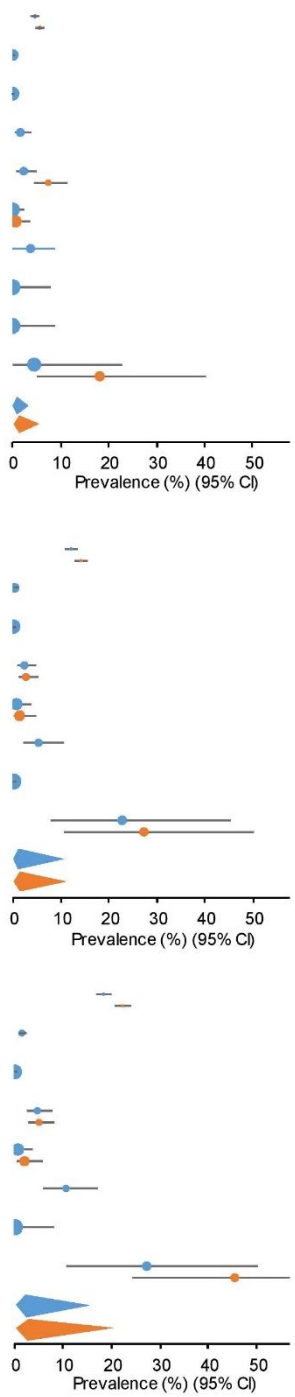
Supplementary Figure 2-1: Selection of included studies

1. **Eleven studies were excluded after reviewing the full-text article: two were superseded by larger cohorts reported in other articles; two included patients and apparently asymptomatic volunteers and did not report data separately for apparently asymptomatic volunteers; two did not report the age of participants; two included adults and children and did not report data separately for adults; one was a study of children; one investigated a single type of incidental finding; one did not report the age of participants undergoing brain imaging, and it was not clear if participants were apparently asymptomatic volunteers or not.**

Region imaged and study name	Sample size	Suspected malignant IFs[a,c]		Suspected malignant IFs and possible indicators of malignancy[b,c]	
		n	Prevalence, % (95% CI)	n	Prevalence, % (95% CI)
Brain					
Bos 2016	5800	183	3.2 (2.7 to 3.6)		
Yue 1997	3672	8	0.2 (0.1 to 0.4)		
Weber 2006	2536	9	0.4 (0.2 to 0.7)		
Hegenscheid 2013	2500	26	1.0 (0.7 to 1.5)		
Lee 2008	2164	2	0.1 (0.0 to 0.3)		
Tsushima 2005	1113	4	0.4 (0.1 to 0.9)		
Baumgart 2007	1007	11	1.1 (0.5 to 1.9)		
Haberg 2016	1006	5	0.5 (0.2 to 1.2)		
Katzman 1999	1000	3	0.3 (0.1 to 0.9)		
Sommer 2013	722	4	0.6 (0.2 to 1.4)		
Sandeman 2013	700	6	0.9 (0.3 to 1.9)		
Cieszanowski 2014	666	0	0.0 (0.0 to 0.6)		
Alphs 2006	656	14	2.1 (1.2 to 3.6)	19	2.9 (1.8 to 4.5)
Brugulat-Serrat 2017	575	6	1.0 (0.4 to 2.3)		
Hoggard 2009	525	5	1.0 (0.3 to 2.2)		
Boutet 2016	503	13	2.6 (1.4 to 4.4)		
Kumar 2008	478	5	1.0 (0.3 to 2.4)		
Goehde 2005	298	2	0.7 (0.1 to 2.4)		
Hartwigsen 2010	206	8	3.9 (1.7 to 7.5)		
Reneman 2012	203	1	0.5 (0.0 to 2.7)		
Li 2015	167	0	0.0 (0.0 to 2.2)		
Illes 2004	151	0	0.0 (0.0 to 2.4)		
Morin 2009	148	0	0.0 (0.0 to 2.5)		
Lo 2008	132	0	0.0 (0.0 to 2.8)		
Wahlund 1989	101	0	0.0 (0.0 to 3.6)		
Menzler 2010	100	1	1.0 (0.0 to 5.4)		
Lubman 2002	98	1	1.0 (0.0 to 5.6)		
Trufyn 2014	56	0	0.0 (0.0 to 6.4)		
Saya 2017	44	0	0.0 (0.0 to 8.0)		
Tarnoki 2015	22	0	0.0 (0.0 to 15.4)		
Combined	**27349**	**317**	**0.6 (0.4 to 0.9)**	**322**	**0.6 (0.4 to 0.9)**
Heterogeneity			$T^2 = 0.9$		$T^2 = 0.9$



Region imaged and study name	Sample size	Suspected malignant IFs[a,c] n	Suspected malignant IFs[a,c] Prevalence, % (95% CI)	Suspected malignant IFs and possible indicators of malignancy[b,c] n	Suspected malignant IFs and possible indicators of malignancy[b,c] Prevalence, % (95% CI)
Thorax					
Hegenscheid 2013	2500	114	4.6 (3.8 to 5.5)	138	5.5 (4.7 to 6.5)
Baumgart 2007	1007	1	0.1 (0.0 to 0.6)		
Cieszanowski 2014	666	0	0.0 (0.0 to 0.6)		
Goehde 2005	298	5	1.7 (0.5 to 3.9)		
Vogel-Claussen 2009	254	6	2.4 (0.9 to 5.1)	19	7.5 (4.6 to 11.4)
Morin 2009	148	0	0.0 (0.0 to 2.5)	1	0.7 (0.0 to 3.7)
Lo 2008	132	5	3.8 (1.2 to 8.6)		
Saya 2017	44	0	0.0 (0.0 to 8.0)		
Loy 2015	40	0	0.0 (0.0 to 8.8)		
Tarnoki 2015	22	1	4.5 (0.1 to 22.8)	4	18.2 (5.2 to 40.3)
Combined	**5111**	**132**	**0.6 (0.1 to 3.1)**	**173**	**1.0 (0.2 to 5.4)**
Heterogeneity			$T^2 = 2.8$		$T^2 = 3.9$
Abdomen					
Hegenscheid 2013	2500	302	12.1 (10.8 to 13.4)	354	14.2 (12.8 to 15.6)
Baumgart 2007	1007	3	0.3 (0.1 to 0.9)		
Cieszanowski 2014	666	0	0.0 (0.0 to 0.6)		
Goehde 2005	298	7	2.3 (0.9 to 4.8)	8	2.7 (1.2 to 5.2)
Morin 2009	148	1	0.7 (0.0 to 3.7)	2	1.4 (0.2 to 4.8)
Lo 2008	132	7	5.3 (2.2 to 10.6)		
Saya 2017	44	0	0.0 (0.0 to 8.0)		
Tarnoki 2015	22	5	22.7 (7.8 to 45.4)	6	27.3 (10.7 to 50.2)
Combined	**4817**	**325**	**1.3 (0.2 to 9.3)**	**380**	**1.6 (0.2 to 10.9)**
Heterogeneity			$T^2 = 4.4$		$T^2 = 4.5$
Brain and body					
Hegenscheid 2013	2500	459	18.4 (16.9 to 21.9)	558	22.3 (20.7 to 24.0)
Baumgart 2007	1007	15	1.5 (0.8 to 2.4)		
Cieszanowski 2014	666	0	0 (0 to 0.6)		
Goehde 2005	298	14	4.7 (2.6 to 7.8)	15	5.0 (2.8 to 8.2)
Morin 2009	148	1	0.7 (0.0 to 3.7)	3	2.0 (0.4 to 5.8)
Lo 2008	132	14	10.6 (5.9 to 17.2)		
Saya 2017	44	0	0.0 (0.0 to 8.0)		
Tarnoki 2015	22	6	27.3 (10.7 to 50.2)	10	45.5 (24.4 to 67.8)
Combined	**4817**	**509**	**2.3 (0.3 to 15.4)**	**615**	**3.0 (0.4 to 20.4)**
Heterogeneity			$T^2 = 4.5$		$T^2 = 5.0$



Supplementary Figure 2-2: Forest plots of the per-study prevalence and pooled prevalence estimates of suspected malignant incidental findings (IFs), and of suspected malignant IFs plus possible indicators of malignancy, detected on brain, thoracic, abdominal and brain and body magnetic resonance imaging (MRI)

IFs = incidental findings, CI = confidence intervals

Tau-squared is an estimate of between-study variance on the logit scale. Zero represents no variance, and increasing values of tau-squared indicate increasing heterogeneity.

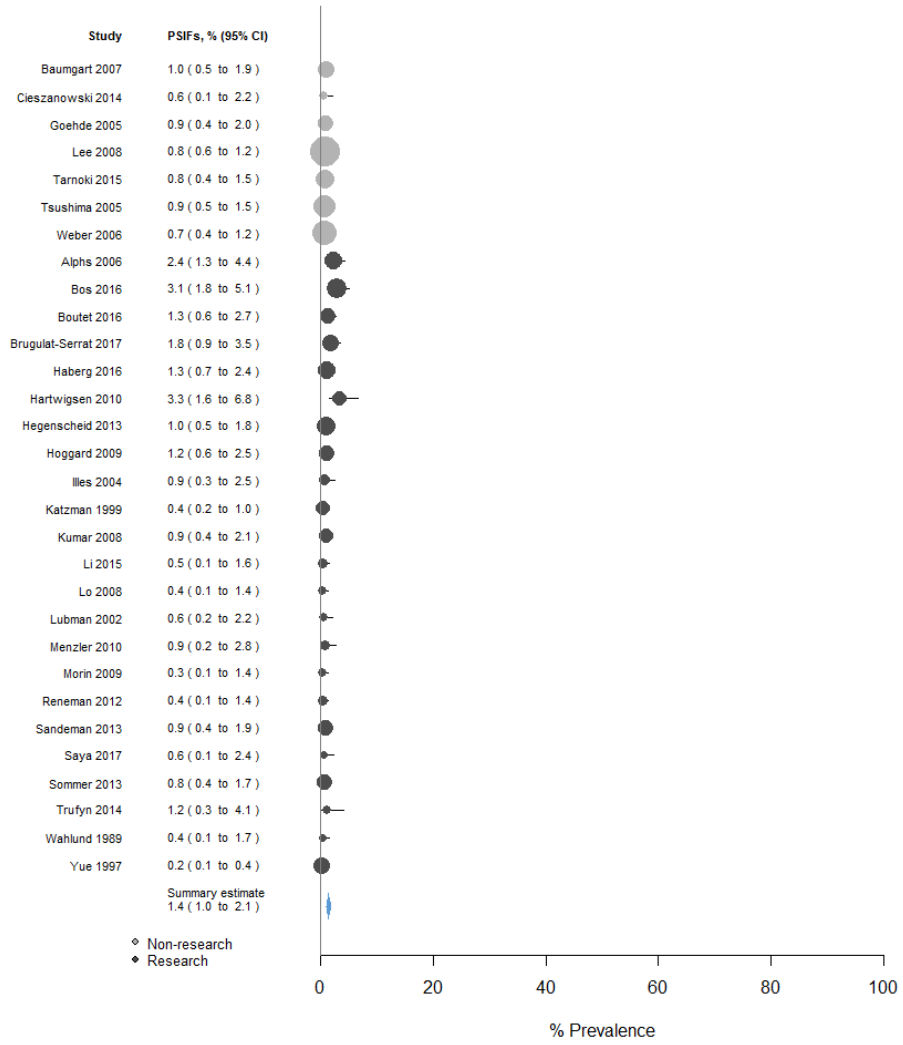
Blue = Per-study point prevalence and pooled prevalence estimate of suspected malignant IFs on brain, thoracic, abdominal and brain and body magnetic resonance imaging (MRI)

Orange = Sensitivity analyses which include IFs classed as possible indicators of malignancy in the per-study point prevalence and pooled prevalence estimate of suspected malignant IFs on brain, thoracic, abdominal and brain and body MRI

a. **Suspected malignant IFs include tumours, masses, complex cysts and lesions.**

b. **Possible indicators of malignancy include pleural and pericardial effusions, enlarged lymph nodes, hydronephrosis, splenomegaly, biliary dilatation, ascites, hydrocephalus and severe bone oedema.**

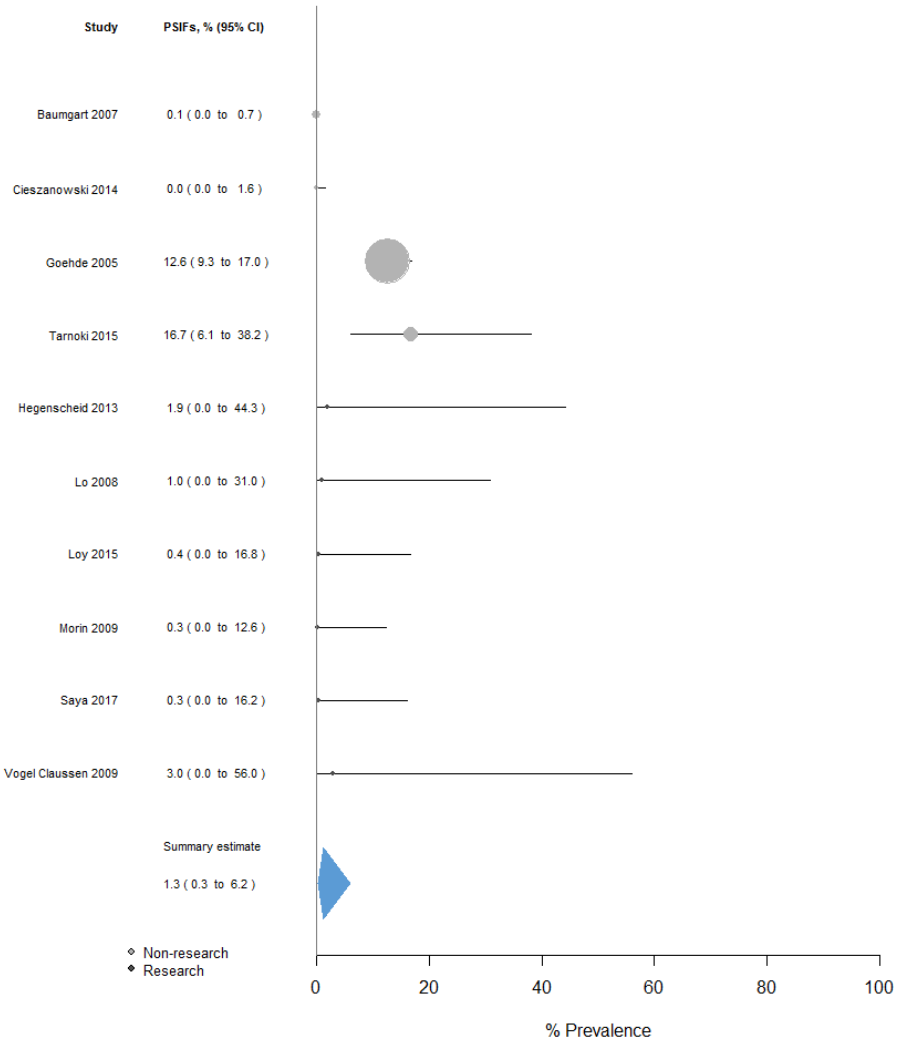
c. **As per Figure 2-2, we excluded IFs detected in studies that used specialist imaging sequences (97 breast lesions in a study including MR mammography (Hegenscheid et al., 2013), and 87 colonic polyps in two studies which included MR colonography (Baumgart and Egelhof, 2007; Goehde et al., 2005)) from pooled analyses.**



Supplementary Figure 2-3: Forest plot showing study-specific prevalence of potentially serious incidental findings, ordered by imaging setting – brain studies

PSIFs = potentially serious incidental findings, CI = confidence interval

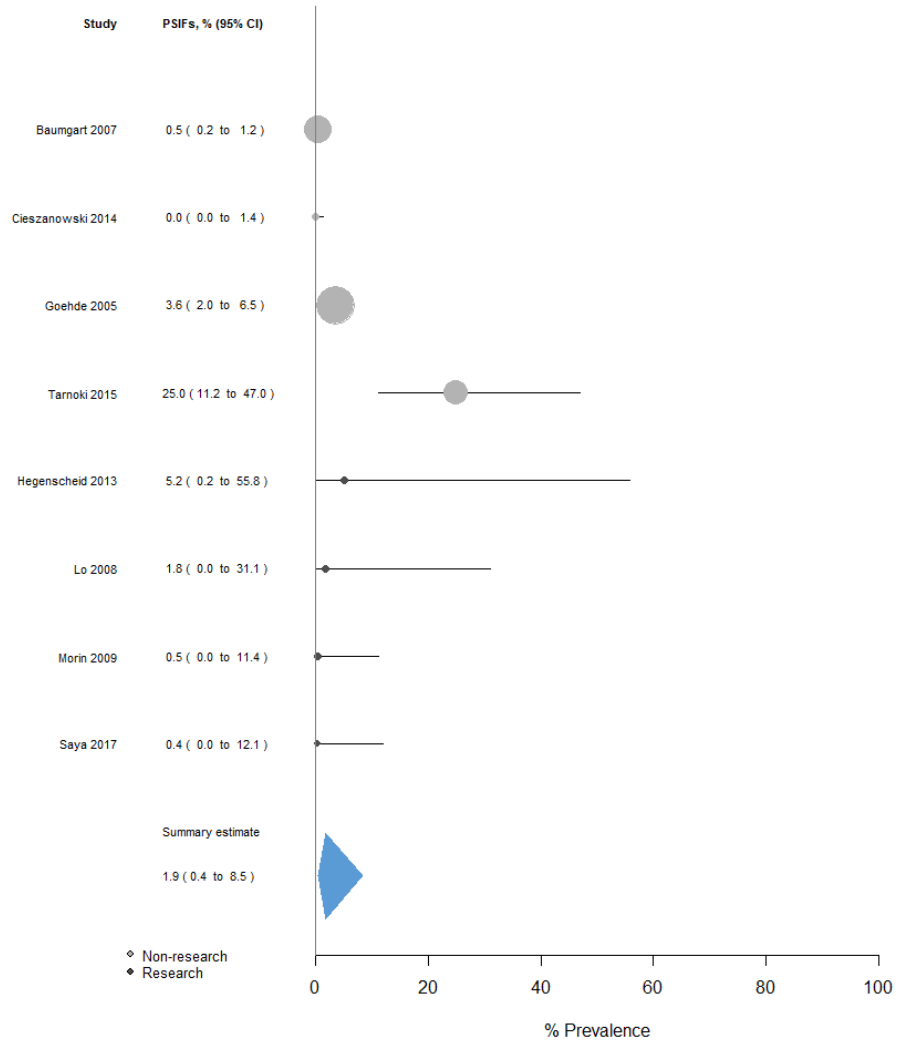
We defined research imaging as that performed as part of a research study. We classified other imaging of apparently asymptomatic participants as non-research (i.e. studies of occupational screening, commercial screening [i.e. paid for by the participant] or medical screening [whether referred by a doctor, or self-referred, or provided by health insurance]).



Supplementary Figure 2-4: Forest plot showing study-specific prevalence of potentially serious incidental findings, ordered by imaging setting – thorax studies

PSIFs = potentially serious incidental findings, CI = confidence interval

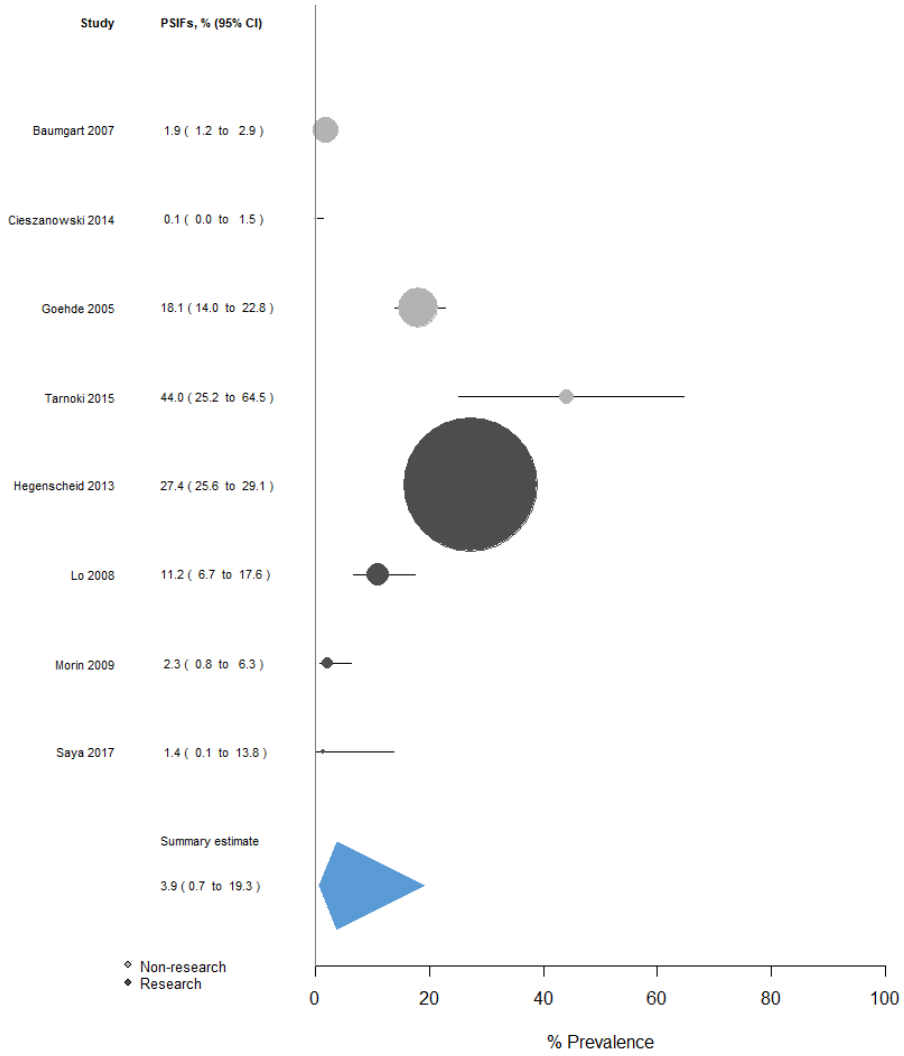
We defined research imaging as that performed as part of a research study. We classified other imaging of apparently asymptomatic participants as non-research (i.e. studies of occupational screening, commercial screening [i.e. paid for by the participant] or medical screening [whether referred by a doctor, or self-referred, or provided by health insurance]).



Supplementary Figure 2-5: Forest plot showing study-specific prevalence of potentially serious incidental findings, ordered by imaging setting – abdomen studies

PSIFs = potentially serious incidental findings, CI = confidence interval

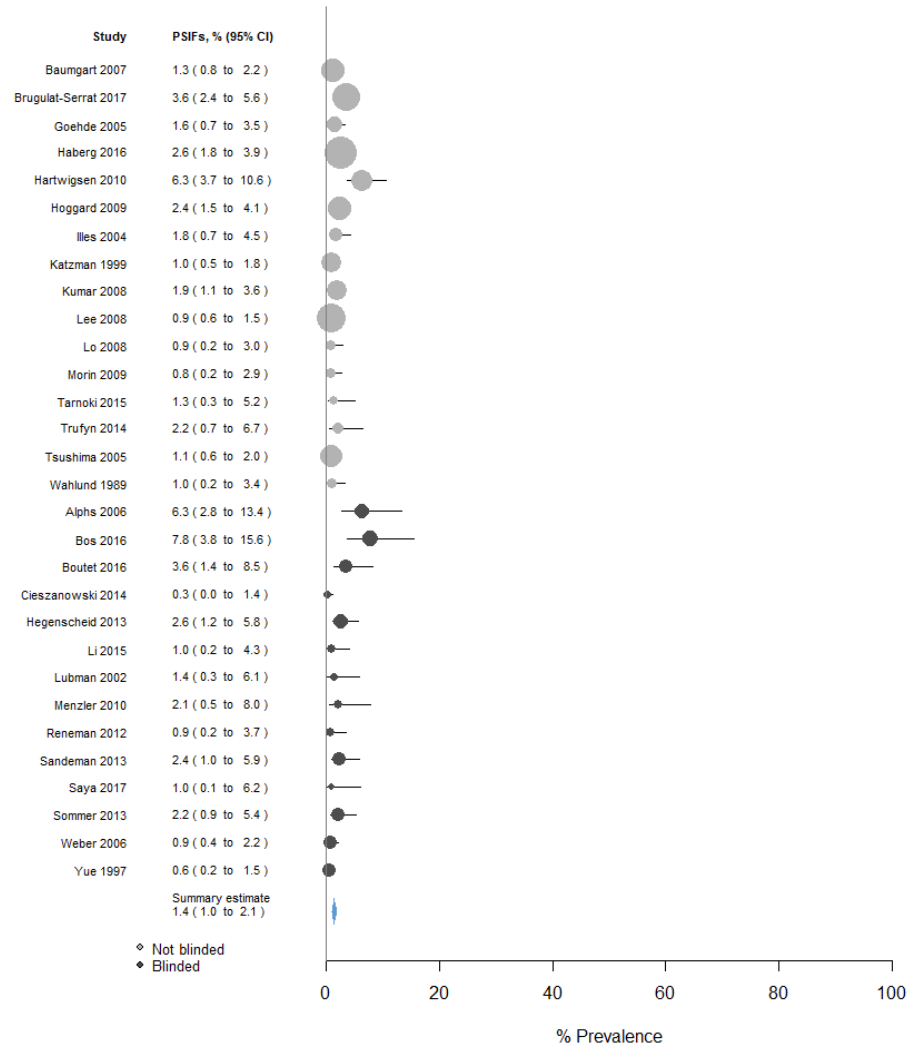
We defined research imaging as that performed as part of a research study. We classified other imaging of apparently asymptomatic participants as non-research (i.e. studies of occupational screening, commercial screening [i.e. paid for by the participant] or medical screening [whether referred by a doctor, or self-referred, or provided by health insurance]).



Supplementary Figure 2-6: Forest plot showing study-specific prevalence of potentially serious incidental findings, ordered by imaging setting – brain and body studies

PSIFs = potentially serious incidental findings, CI = confidence interval

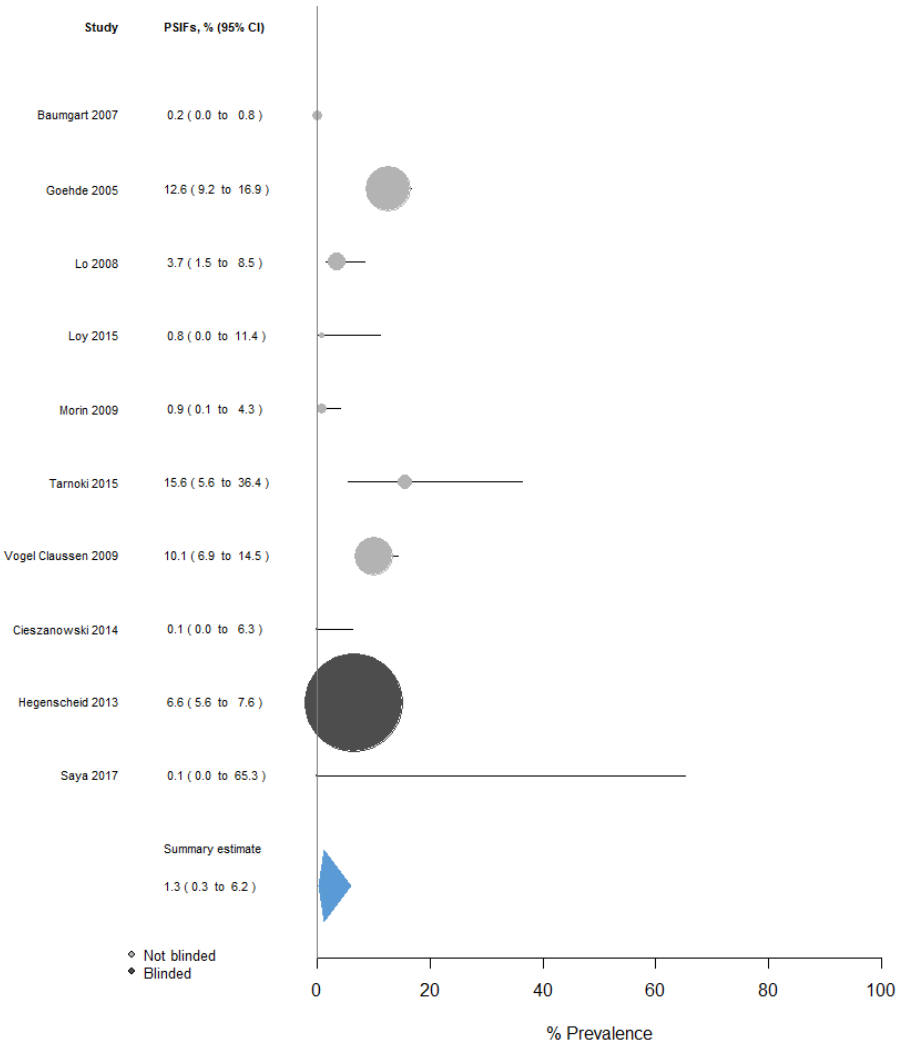
We defined research imaging as that performed as part of a research study. We classified other imaging of apparently asymptomatic participants as non-screening (i.e. studies of occupational screening, commercial screening [i.e. paid for by the participant] or medical screening [whether referred by a doctor, or self-referred, or provided by health insurance]).



Supplementary Figure 2-7: Forest plot showing study-specific prevalence of potentially serious incidental findings, ordered by blinding of readers to participants' characteristics imaging setting – brain studies

PSIFs = potentially serious incidental findings, CI = confidence interval

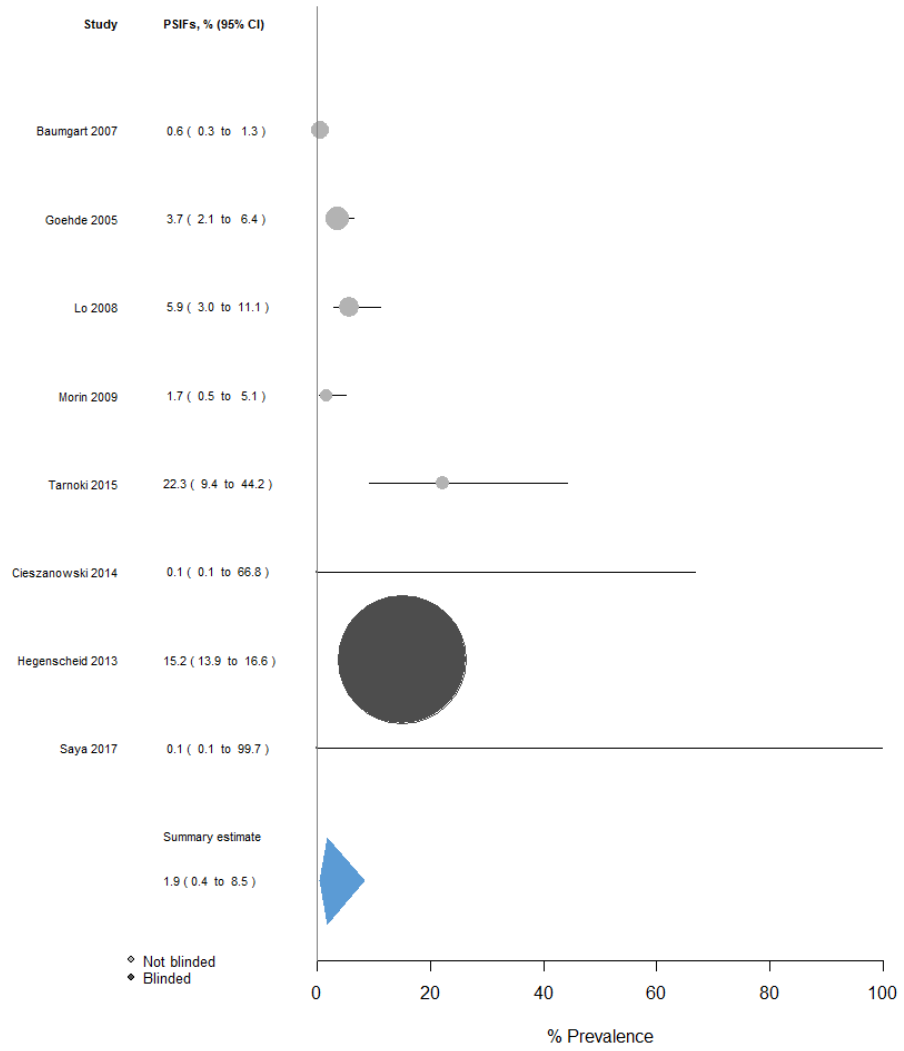
We classified studies which reported they did not blind readers, or studies who did not provide information on blinding, as not blinded for the purposes of these subgroup analyses.



Supplementary Figure 2-8: Forest plot showing study-specific prevalence of potentially serious incidental findings, ordered by blinding of readers to participants' characteristics imaging setting – thorax studies

PSIFs = potentially serious incidental findings, CI = confidence interval

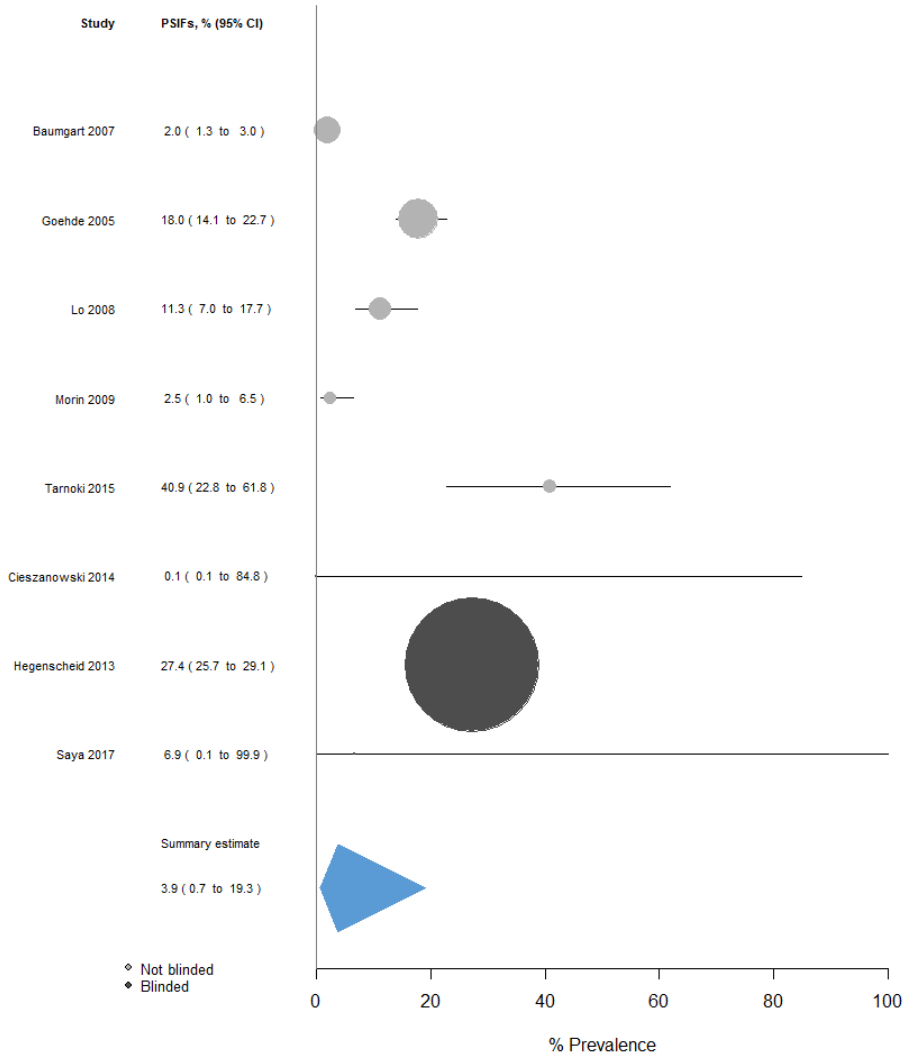
We classified studies which reported they did not blind readers, or studies who did not provide information on blinding, as not blinded for the purposes of these subgroup analyses.



Supplementary Figure 2-9: Forest plot showing study-specific prevalence of potentially serious incidental findings, ordered by blinding of readers to participants' characteristics imaging setting – abdomen studies

PSIFs = potentially serious incidental findings, CI = confidence interval

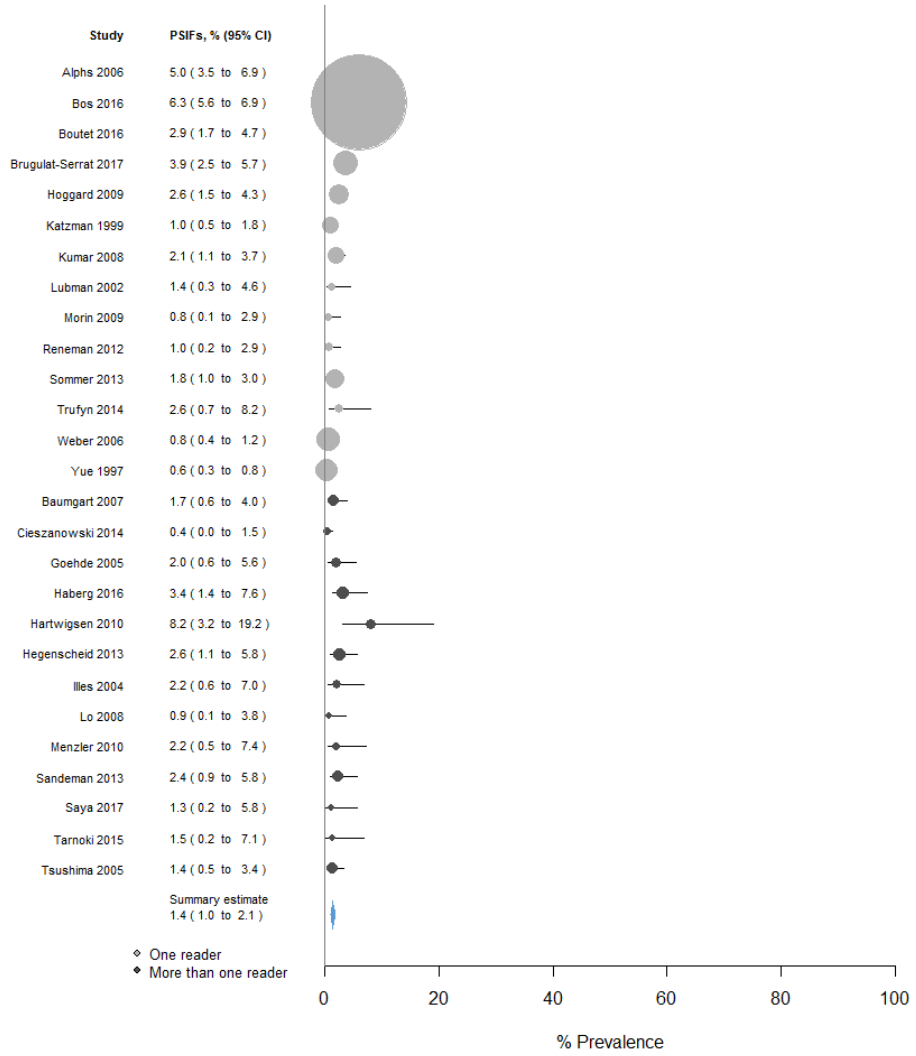
We classified studies which reported they did not blind readers, or studies who did not provide information on blinding, as not blinded for the purposes of these subgroup analyses.



Supplementary Figure 2-10: Forest plot showing study-specific prevalence of potentially serious incidental findings, ordered by blinding of readers to participants' characteristics imaging setting – brain and body studies

PSIFs = potentially serious incidental findings, CI = confidence interval

We classified studies which reported they did not blind readers, or studies who did not provide information on blinding, as not blinded for the purposes of these subgroup analyses.



Supplementary Figure 2-11: Forest plot showing study-specific prevalence of potentially serious incidental findings, ordered by number of image readers – brain studies

PSIFs = potentially serious incidental findings, CI = confidence interval

Number of image readers refers to the numbers of readers who assessed all images for IFs. Three studies of brain PSIFs (2,432/27,349 [8.9%] participants) (Lee et al., 2008; Li et al., 2015; Wahlund et al., 1989) did not provide data on number of image readers and were excluded from this subgroup analysis. There were sparse data within subgroups of studies involving one image reader for thorax, abdomen, and brain and body studies (PSIFs were present in only 1/188, 2/248 and 3/148 participants respectively); therefore these subgroup analyses were not performed.

2.3.3 Further supplementary tables

Supplementary Table 2-6: Details of included studies ordered by region imaged and descending sample size

Study first author and publication year	Study population variables					Imaging variables				
								Readers		
	N[a]	n (%) male	Mean age (range)	Random sampling[b]	Country	Imaging setting	Data source[c]	Specialities	N	Blinded[d]
Brain and body MRI										
Hegenscheid 2013	2500	1229 (49)	53 (21-88)	Yes	Germany	Longitudinal cohort	I	Radiologist	>1	Yes
Baumgart 2007	1007	715 (71)	55 (40-67)	No	Germany	Screening	I	Radiologist	>1	-
Cieszanowski 2014	666	465 (70)	46 (20-77)	No	Poland	Screening	I	Radiologist	>1	Yes
Goehde 2005	298	247 (83)	50 (31-73)	No	Germany	Screening	I	Radiologist	>1	No
Morin 2009	148	94 (64)	36[e] (21-69)	No	UK	Other research	I	Radiologist	1[f]	-
Lo 2008	132	111 (84)	56 (38-82)	No	Hong Kong	Other research	I	Radiologist	>1	-
Saya 2017[g]	44	17 (39)	38[e] (22-59)	No	UK	Other research	I	Radiologist	>1	Yes
Tarnoki 2015	22	18 (82)	47 (-)	No	Germany	Screening	I	Radiologist	>1	-

Study first author and publication year (cont.)	Study population variables					Imaging variables				
								Readers		
	N[a]	n (%) male	Mean age (range)	Random sampling[b]	Country	Imaging setting	Data source[c]	Specialities	N	Blinded[d]
Brain MRI only										
Bos 2016	5800	2606 (45)	65 (-)	-	Netherlands	Longitudinal cohort	I	Trained readers[h]	1	Yes
Yue 1997	3672	1531 (42)	- (≥65)	No	USA	Longitudinal cohort	I	Neuroradiologist	1	Yes
Weber 2006	2536	2536 (100)	21 (17-35)[i]	No	Germany	Screening	I	Radiologist	1	Yes
Lee 2008	2164	1234 (57)	52 (17-89)[i]	No	Taiwan	Screening	-	Neuro- or general radiologist	-	-
Tsushima 2005	1113	761 (68)	53 (22-84)	No	Japan	Screening	I	Radiologist and a neurosurgeon	>1	-
Haberg 2016	1006	476 (47)	59[d] (51-67)	No	Norway	Longitudinal cohort	I	Neuroradiologist	>1	No
Katzman 1999	1000	546 (55)	31 (3-83)[i]	No	USA	Other research	I	Neuroradiologist	1	-
Sommer 2013[g]	722	407 (56)	34 (-)	No	Netherlands	Other research	I	Neuroradiologist	1	Yes

Study first author and publication year (cont.)	Study population variables					Imaging variables				
								Readers		
	N[a]	n (%) male	Mean age (range)	Random sampling[b]	Country	Imaging setting	Data source[c]	Specialities	N	Blinded[d]
Brain MRI only continued										
Sandeman 2013	700	368 (53)	73 (-)	No	UK	Longitudinal cohort	I	Neuroradiologist	>1	Yes
Alphs 2006	656	656 (100)	61 (35-82)	No	USA	Longitudinal cohort	I	Neuroradiologist	1	Yes
Brugulat-Serrat 2017	575	227 (40)	59 (45-75)	No	Spain	Longitudinal cohort	I	Neuroradiologist	1	-
Hoggard 2009	525	330 (63)	35 (-)	No	UK	Other research	I	Neuroradiologist	1	-
Boutet 2016	503	208 (41)	75 (71-79)	No	France	Longitudinal cohort	I	Neuroradiologist	1	Yes
Kumar 2008	478	252 (53)	- (60-64)	-	Australia	Longitudinal cohort	I	Radiologist[j]	1[j]	-
Hartwigsen 2010	206	117 (57)	26 (9-50)[i]	No	Germany	Other research	I	Neuroradiologist	>1	-
Reneman 2012	203	92 (45)	22 (18-35)	No	Netherlands	Other research	I	Neuroradiologist	1	Yes

Study first author and publication year (cont.)	Study population variables					Imaging variables				
								Readers		
	N[a]	n (%) male	Mean age (range)	Random sampling[b]	Country	Imaging setting	Data source[c]	Specialities	N	Blinded[d]
Brain MRI only continued										
Li 2015[g]	167	0 (0)	24 (-)	No	Taiwan	Other research	I	MRI operators[k]	-[k]	Yes
Illes 2004	151	82 (54)	47 (18-90)	No	USA	Other research	I	Neuroradiologist	>1	-
Wahlund 1989[g]	101	- (-)	- (16-87)[i]	-	Sweden	Other research	-	-	-	-
Menzler 2010	100	42 (42)	29 (-)	-	Germany	Other research	I	Neuroradiologist	>1	Yes
Lubman 2002[g]	98	62 (63)	27 (-)	No	Australia	Other research	R	Neuroradiologist	1	Yes
Trufyn 2014[g]	56	15 (27)	45 (-)	No	Canada	Other research	I	Neuroradiologist	1	-
Cardiac MRI only										
Vogel-Claussen 2009	254	124 (49)	61 (45-89)	No	USA	Longitudinal cohort	I	Radiologist	>1	-
Loy 2015[g]	40	40 (100)	46 (-)	No	Ireland	Other research	I	Cardiologist	1	-

MRI = magnetic resonance imaging, I = imaging, UK = United Kingdom, USA = United States of America, R = reports, - = Information not specified or not sufficiently well described

a. Sample size indicates the number of apparently asymptomatic volunteers imaged, as some studies also included patient groups (see footnote g).

b. Indicates whether or not the participants were randomly sampled from the base population.

c. Indicates whether information on incidental findings (IFs) was determined from review of images (I), or reports (R), or not specified (-).

d. Blinded to information about participants.

e. Median age; no data on mean age were available.

f. Morin 2009: All scans were reviewed by a single radiologist, and only scans with a potentially highly significant abnormality were reviewed by two radiologists.

g. Study included groups of patients and apparently asymptomatic volunteers, only data from the apparently asymptomatic volunteers were included in this review.

h. Bos 2016: All scans were reviewed by a group of trained readers (researchers or neuropsychologists) for IFs. Two neuroradiologists reviewed only scans with suspected IFs. We have assumed that the reported data on IFs pertain to those confirmed by the neuroradiologists.

i. Study was included in the review, as it was deemed to involve only a small proportion of children, as judged in consensus by two authors (LG and CLMS) based on available data on the age of the study population.

j. Kumar 2008: All scans were reviewed by a single radiologist, and only scans with suspected abnormalities were reviewed by a second reader, a neuropsychiatrist.

k. Li 2015: All scans were reviewed by an MRI operator for IFs, and confirmed by a neuroradiologist, blinded to participants' clinical status. It is not clear if the neuroradiologist reviewed all participants' scans for IFs, or just those with abnormalities detected by the MRI operator. We have assumed that the reported data on IFs pertain to those confirmed by the neuroradiologist.

Supplementary Table 2-7: Sequences evaluated for incidental findings in included studies, ordered by region imaged and descending sample size

Study	Sample size N	Magnet strength (Tesla)	Single protocol used to image participants [a]	Brain MRI						Thoracic MRI				Abdominal MRI				Other[c]
				Contrast	T1	T2	FLAIR	Haem sensitive[b]	MRA	Contrast	Non-contrast whole chest	Cardiac specific	MRA	Contrast	T1	T2 or STIR	MRA	
Brain and body MRI																		
Hegenscheid 2013	2500	1.5	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y
Baumgart 2007	1007	1.5	Y	Y	Y	Y	.	.	Y	.	Y	Y	Y	.	Y	Y[d]	Y	Y
Cieszanowski 2014	666	1.5	Y	.	.	.	Y	.	.	.	Y	.	.	.	Y	Y	.	Y
Goehde 2005	298	1.5	Y	.	Y	Y	Y	.	Y	.	Y	Y	Y	.	Y	.	Y	Y
Morin 2009	148	1.5	Y	.	Y	.	.	.	.	.	Y	.	.	.	Y	Y	.	.
Lo 2008	132	3.0	Y	N	Y	Y	.	.	.	N	Y	.	.	N	Y	Y	.	.
Saya 2017	44	1.5	Y	N	Y	Y	.	.	.	N	Y	.	.	N	Y	Y	.	.
Tarnoki 2015	22	3.0	Y	.	Y	.	.	.	Y	.	Y	.	Y	.	Y	Y	.	Y
Brain MRI only																		
Bos 2016	5800	1.5	Y	N	Y	.	Y	Y	.	.	.	.	.	.	.	.	.	Y
Yue 1997	3672	0.35/1.5[e]	Y	N	Y	Y	.	.	.	.	.	.	.	.	.	.	.	Y
Weber 2006	2536	1.0	Y	Y[d]	Y	Y	Y	.	Y[d]	.	.	.	.	.	.	.	.	.
Lee 2008	2164	1.5	Y	.	Y	Y	Y	.	Y[d]	.	.	.	.	.	.	.	.	.
Tsushima 2005	1113	1.0	Y	N	Y	Y	.	.	Y	.	.	.	.	.	.	.	.	Y
Haberg 2016	1006	1.5	Y	.	Y	Y	Y	Y	Y	.	.	.	.	.	.	.	.	.
Katzman 1999	1000	NS	N[f]	.	Y	Y	.	.	.	.	.	.	.	.	.	.	.	Y

Study (continued)	Sample size N	Magnet strength (Tesla)	Single protocol used to image participants [a]	Brain MRI						Thoracic MRI				Abdominal MRI				Other[c]
				Contrast	T1	T2	FLAIR	Haem sensitive[b]	MRA	Contrast	Non-contrast whole chest	Cardiac specific	MRA	Contrast	T1	T2 or STIR	MRA	
Sommer 2013	722	1.5	N[g]	.	.	Y	.	.	.	.	.	.	.	.	.	.	.	.
Sandeman 2013	700	1.5	Y	.	Y	Y	Y	Y	.	.	.	.	.	.	.	.	.	.
Alphs 2006	656	NS	Y	.	Y	Y	.	Y	Y[d]	.	.	.	.	.	.	.	.	.
Brugulat-Serrat 2017	575	3.0	Y	.	Y	Y	Y	Y	.	.	.	.	.	.	.	.	.	.
Hoggard 2009	525	1.5 and 3.0	N[h]	.	Y[d]	Y	.	.	.	.	.	.	.	.	.	.	.	Y
Boutet 2016	503	1.5	Y	.	Y	Y	Y			.	.	.		.	.	.	.	.
Kumar 2008	478	1.5	Y	.	Y	.	Y	.	.	.	.	.	.	.	.	.	.	.
Hartwigsen 2010	206	3.0	Y	.	Y	.	Y	.	.	.	.	.	.	.	.	.	.	.
Reneman 2012	203	1.5[i]	N[i]	Y[d]	Y	Y[d]	.	.	.	.	.	.	.	.	.	.	.	Y[d]
Li 2015	167	1.5/3.0[j]	N[j]	.	Y	.	.	.	.	.	.	.	.	.	.	.	.	Y[j]
Illes 2004	151	NS	N[k]	.	Y	Y	.	Y	.	.	.	.	.	.	.	.	.	Y
Wahlund 1989	101	NS	NS[l]	.	.	.	.	.		.	.	.	.	.	.	.	.	.
Menzler 2010	100	1.5	Y	.	Y	Y	Y	.	.	.	.	.	.	.	.	.	.	.
Lubman 2002	98	1.5	N[m]	.	Y	.	.	.	.	.	.	.	.	.	.	.	.	.
Trufyn 2014	56	3.0	Y	.	Y	Y	Y	Y	Y	.	.	.	.	.	.	.	.	Y
Cardiac MRI only																		
Vogel Claussen 2009	254	1.5	Nn	.	.	.	.	.	.	.	.	Y	Y	.	.	.	.	.
Loy 2015	40	3.0	Y	.	.	.	.	.	.	Y	.	Y	.	.	.	.	.	.

FLAIR = fluid attenuated inversion recovery, MRA = magnetic resonance angiography, STIR = short-tau inversion recovery, CE = contrast enhanced, NS = not specified

a. **Some studies report the prevalence of incidental findings (IFs) detected in participants who all attended for the same type of imaging, for example, a single research study with a single imaging protocol. These studies are indicated by 'Y.' In contrast, some studies report the prevalence of IFs detected in participants scans that have been 'pooled' from more than one research project, and therefore involve more than one imaging protocol. These studies are indicated by 'N,' with details provided in additional footnotes.**

b. **Either gradient recalled echo, or susceptibility-weighted imaging.**

c. **'Other' sequences include additional brain or chest or abdominal sequences, such as proton density, diffusion weighted imaging, colonography, mammography, in and out of phase abdominal imaging etc. Please see individual papers for details.**

d. **Sequence performed in a subset of participants.**

e. **Yue 1997: 0.35 Tesla MRI used at three imaging centres, 1.5 Tesla MRI used at one imaging centre. No data available on numbers of participants imaged using the different scanners.**

f. **Katzman 1999: Participants' scans were pooled from multiple neuroimaging studies, but all had at least T1- and T2-weighted brain imaging.**

g. **Sommer 2013: Participants' scans were pooled from multiple neuroimaging studies, but only T2 images were assessed for IFs.**

h. **Hoggard 2009: Participants' scans were pooled from multiple neuroimaging studies, but 456 had at least axial T2-weighted imaging of the whole brain, and the remaining 69 from a single neuroimaging study only had T1-weighted imaging.**

i. **Reneman 2012: Seven participants were imaged using a 3.0T MRI scanner, with only a 3D T1-weighted sequence.**

j. **Li 2015: Number of control participants scanned with each magnet was not reported. Participants imaged using the 1.5T scanner underwent 3D-fast spoiled gradient echo T1-weighted imaging. Participants imaged using the 3.0T scanner underwent 3D- magnetization-prepared rapid gradient-echo T1-weighted imaging. Additional, unspecified, sequences were performed when necessary for diagnostic purposes.**

k. **Illes 2004: Participants' scans were pooled from multiple neuroimaging studies, and each participant underwent at least one of the sequences indicated.**

l. **Wahlund 1989: No information on sequence types is available.**

m. **Lubman 2002: Participants' scans were pooled from multiple neuroimaging studies, and participants had at least a T1-weighted sequence.**

n. **Vogel-Claussen 2009: Method of imaging coronary arteries differed across the sample: 23 participants underwent steady state free precession coronary MRA images using breath-hold technique, and the remaining 231 underwent 3-dimensional steady-state free precession navigator assisted free breathing whole heart technique.**

Supplementary Table 2-8: 95% prediction intervals (and 95% confidence intervals to enable direct comparison)

IFs and region	Pooled prevalence (%)	95% prediction interval (%)	95 % confidence interval (%)
PSIFs only			
Brain	1.4	0.2 to 8.3	1.0 to 2.1
Thorax	1.3	0 to 76.8	0.2 to 8.1
Abdomen	1.9	0 to 81.1	0.3 to 12.0
Brain and body	3.9	0 to 95.4	0.4 to 27.1
PSIFs and indeterminate IFs			
Brain	1.7	0.2 to 12.3	1.1 to 2.6
Thorax	3.0	0 to 67.4	0.8 to 11.3
Abdomen	4.5	0.2 to 55.2	1.5 to 12.9
Brain and body	12.8	0.4 to 85.7	3.9 to 34.3
Suspected malignant IFs			
Brain	0.6	0.1 to 4.0	0.4 to 0.9
Thorax	0.6	0 to 29.6	0.1 to 3.1
Abdomen	1.3	0 to 76.8	0.2 to 9.3
Brain and body	2.3	0 to 86.5	0.3 to 15.4
Suspected malignant IFs and possible indicators of malignancy			
Brain	0.6	0.1 to 4.2	0.4 to 0.9
Thorax	1.0	0 to 57.4	0.2 to 5.4
Abdomen	1.6	0 to 81.0	0.2 to 10.9
Brain and body	3.0	0 to 91.8	0.4 to 20.4

IFs = incidental findings, PSIFs = potentially serious incidental findings

Supplementary Table 2-9: Types of potentially serious incidental findings (PSIFs) in descending order of frequency, as percentages of total PSIFs - brain

Types of PSIFs among 27,349 volunteers	N PSIFs	Percentage[a] of PSIFs (total N=688)
Suspected malignancy	317	46
Intracranial mass	179	26
Pituitary mass	67	9.7
Pituitary cyst	55	8.0
Extracranial mass	12	1.7
Atypical cerebellar lesion	2	0.29
Intracranial cyst[b]	1	0.15
Skull: Potentially serious lesion	1	0.15
Suspected aneurysm	214	31
Suspected vascular malformation	106	15
Suspected other:	46	6.7
Arachnoid cyst	18	2.6
Not specified: potentially serious	13	1.9
Acute infarct	4	0.58
Missing pituitary neurohypophysis signal	4	0.58
Subdural haematoma	3	0.44
Colloid cyst	2	0.29
Mesial temporal sclerosis	1	0.15
Syringomyelia	1	0.15
Possible malignancy: hydrocephalus	5	0.73

MRI = magnetic resonance imaging, PSIFs = potentially serious incidental findings

a. All percentages are rounded to two significant figures.

b. This intracranial cyst was followed up with MRI, and we presumed that there were some concerning features, and classified it as a suspected malignancy.

Supplementary Table 2-10: Types of potentially serious incidental findings (PSIFs) in descending order of frequency, as percentages of total PSIFs - thorax

Types of PSIFs among 5,111 volunteers	N PSIFs	Percentage[a] of PSIFs (total N=238)
Suspected malignancy	132	55
Lung	72	30
Nodule	56	24
Lobar pneumonia or lung consolidation	8	3.4
Lesion requiring follow-up	5	2.1
Mass	3	1.3
Other region	60	25
Neck tumor	52	22
Thyroid lesions or enlargement	4	1.7
Cardiac mass	1	0.42
Chest lesion	1	0.42
Liver lesion requiring follow-up	1	0.42
Mediastinal lesion requiring follow-up	1	0.42
Suspected other	55	23
LV hypertrophy	17	7.1
Valve defects	16	6.7
Goitre with tracheal compression	9	3.8
Heart failure	5	2.1
Reduced contractility	5	2.1
Hypertrophic cardiomyopathy	1	0.42
Suspected pulmonary hypertension	1	0.42
Thoracic aortic stenosis	1	0.42
Possible malignancy	41	17
Lymphadenopathy	22	9.2
Pleural effusion	16	6.7
Pericardial effusion	2	0.84
Urinary obstruction	1	0.42
Suspected thoracic aortic aneurysm	10	4.2

MRI = magnetic resonance imaging, PSIFs = potentially serious incidental findings

a. All percentages are rounded to two significant figures

Supplementary Table 2-11: Types of potentially serious incidental findings (PSIFs) in descending order of frequency, as percentages of total PSIFs - abdomen

Types of PSIFs among 4,817 volunteers	N PSIFs	Percentage[a] of PSIFs (total N=412)
Suspected malignancy	325	79
Renal	131	32
Ovarian	79	19
Liver	50	12
Uterine or cervical malignancy	15	3.6
Pancreas	12	2.9
Adrenal gland	8	1.9
Testicular, epididymal or seminal vesicle	7	1.7
Bladder	6	1.5
Colon or rectum	6	1.5
Spleen	5	1.2
Lumbar intraspinal neurinoma	1	0.24
Lumbar spine lesion requiring follow-up	1	0.24
Prostate	1	0.24
Psoas	1	0.24
Retroperitoneal mass	1	0.24
Stomach	1	0.24
Possible malignancy	55	13
Biliary dilatation	25	6.1
Lymphadenopathy	16	3.9
Splenomegaly	8	1.9
Chronic urinary obstruction	5	1.2
Ascites	1	0.24
Suspected other	18	4.4
Irregular/nodular liver margin	9	2.2
Haemochromatosis	5	1.2
Abdominal aortic stenosis	3	0.73
Reflux nephropathy	1	0.24
Suspected abdominal aortic aneurysm	14	3.4

MRI = magnetic resonance imaging, PSIFs = potentially serious incidental findings

a. All percentages are rounded to two significant figures.

b. Denominator is 1,921 women.

c. Denominator is 2,896 men.

Supplementary Table 2-12: Difference in estimates of prevalence of PSIFs between subgroups

Subgroup analysis and region	Prevalence of PSIFs (%) in the reference group	Prevalence of PSIFs (%) in the comparison group	Difference in estimates of prevalence of PSIFs between subgroups, % (95% CI)	p-value
Imaging setting: non-research (reference) versus research (comparison)				
Brain	0.9 (0.6 to 1.2)	1.8 (1.2 to 2.7)	0.9 (0.2 to 2.2)	0.067
Thorax	0.8 (0.01 to 31.4)	2.6 (0.6 to 10.2)	1.8 (-0.7 to 65.9)	0.335
Abdomen	1.2 (0.04 to 27.1)	3.6 (0.6 to 18.5)	2.5 (-1.1 to 63.9)	0.409
Brain and body	3.1 (0.06 to 63.3)	5.4 (0.7 to 32.0)	2.3 (-3.0 to 81.0)	0.428
Blinding of readers to participants' characteristics: Not blinded or not stated (reference) versus blinded (comparison)				
Brain	1.6 (1.0 to 2.5)	1.3 (0.7 to 2.4)	-0.4 (-1.0 to 1.1)	0.337
Thorax	2.4 (0.4 to 12.9)	0.1 (0.0 to 67.9)	-2.2 (-2.4 to 70.0)	0.659
Abdomen	3.3 (0.7 to 14.6)	0.02 (0.0 to 99.8)	-3.3 (-3.3 to 96.5)	0.659
Brain and body	8.8 (2.0 to 31.7)	0.01 (0.0 to 99.9)	-8.8 (-8.8 to 91.1)	0.633
Number of readers: One reader (reference) versus >1 (comparison)				
Brain	1.7 (1.0 to 2.9)	1.4 (0.7 to 2.6)	-0.3 (-1.1 to 1.4)	0.953

PSIF = potentially serious incidental finding, CI = confidence interval

Supplementary Table 2-13: Summary of available data on potential determinants of prevalence of potentially serious incidental findings (PSIFs, as per our definition) or IFs which required follow-up (as per each study's definition), ordered by descending sample size - age

Study	Sample size (n apparently asymptomatic people)	PSIF type	Summary of results[a]	Significant difference between age groups
Tsushima 2005	1113	IFs requiring follow-up[b,c]	Non-significantly higher prevalence in older versus younger participants (60-84 years, 4/192 [2.1%] vs 34-59 years, 11/921 [1.2%], p=0.3[d])	N
Brugulat-Serrat 2017	575	Cavernoma	Non-significantly higher prevalence in older versus younger participants (65-75 years, 4/119 [3.4%] vs 45-54 years, 2/211 [1.0%], p=0.2[d]; 55-65 years, 8/245 [3.3%] vs 45-54 years, 2/211 [1.0%], p=0.1[d])	N
		Intracranial masses	No significant difference in prevalence of pituitary, intraventricular or cerebellar masses in older versus younger participants (Pituitary masses: 45-54 years, 0/211 [0.0%] vs 55-64 years, 2/245 [0.8%, p=0.5[d]] and vs 65-75 years, 0/119 [0.0%, p=NC]; intraventricular masses: 45-54 years, 1/211 [0.5%] vs 55-64 years, 0/245 [0.0%, p=0.5[d]] vs 65-75 years, 0/119 [0.0%, p=1.0[d]]; cerebellar masses: 45-54 years, 0/211 [0.0%] vs 55-64 years, 1/245 [0.4%, p=1.0[d]] and vs 65-75 years, 0/119 [0.0%, p=NC].	N
Hartwigsen 2010	206	IFs requiring follow-up[c,e]	Participants with IFs requiring follow-up were significantly older than those with IFs which did not (p=0.04[f])	Y
Illes 2004	151	IFs requiring follow-up[c,g]	Significantly higher prevalence in older versus younger participants (≥ 60 years, 41/64 [64.1%] vs < 60 years, 30/87 [34.5%], p=0.001])	Y

PSIF = potentially serious incidental finding, N = no, Y = yes, NC = not calculated as zero frequency in both groups.

a. P-values relate to chi-square tests unless otherwise stated.

b. Tsushima 2005: IFs requiring follow-up, as judged by the study's radiologist.

c. Not all IFs requiring follow-up were classed as PSIFs but the distribution of PSIFs between age groups was not possible to calculate from the reported data; in this context, we use 'IFs requiring follow-up' as an approximate proxy for PSIFs

d. Two-tailed Fisher exact test.

e. Hartwigsen 2010: IFs requiring follow-up, the method of judging this was not reported.

f. From independent samples t-test, reported in the study paper. No numerical data on age of each group (e.g. mean) were reported.

g. Illes 2004: IFs requiring follow-up, as judged by two neuroradiologists.

Supplementary Table 2-14: Summary of available data on potential determinants of prevalence of potentially serious incidental findings (PSIFs, as per our definition) or IFs which required follow-up (as per each study's definition), ordered by descending sample size - sex

Study and body region	Sample size (N women)	PSIF type	n women with PSIF[a] (%)	n men with PSIF[a] (%)	p-value[b]
Brain					
Bos 2016	5800 (3194)	Cavernoma	18 (0.6)	19 (0.7)	0.5[c]
		Cerebral aneurysm	90 (2.8)	44 (1.7)	0.006[c]
		Pituitary cyst or mass	35 (1.1)	32 (1.2)	0.7[c]
Yue 1997	3672 (2141)	Cavernoma	4 (0.2)	1 (0.1)	0.4
		Pituitary cyst or mass	2 (0.1)	4 (0.3)	0.2
Tsushima 2005	1113 (352)	IFs requiring follow-up[d,e]	5 (1.4)	10 (1.3)	1.0
Haberg 2016	1006 (530)	Arteriovenous malformation	1 (0.2)	0 (0.0)	1.0
		Cavernoma	2 (0.4)	1 (0.2)	0.4
		Cerebral aneurysm	14 (2.6)	5 (1.1)	0.1
		Glioma	1 (0.2)	0 (0.0)	1.0
		Pituitary cyst or mass	2 (0.4)	1 (0.2)	1.0
		Vestibular schwannoma	0 (0.0)	1 (0.2)	0.5
Sandeman 2013	700 (332)	IFs requiring follow-up[f,e]	7 (2.1)	3 (0.8)	0.2
Brugulat-Serrat 2017	575 (348)	Any brain malignancy	15 (4.3)	3 (1.3)	0.1
Kumar 2008	478 (226)	Pituitary cyst or mass	3 (1.3)	1 (0.4)	0.4
Illes 2004	151 (69)	IFs requiring follow-up[g,e]	4 (5.8)	6 (7.3)	0.8
Brain and body					
Morin 2009	148 (54)	PSIFs	1 (1.9)	2 (2.1)	0.4

PSIF = potentially serious incidental finding

PSIFs, or IFs requiring follow-up, as appropriate.

a. P-values relate to two-tailed Fisher exact tests unless otherwise stated.

b. P-value relates to chi-square test.

c. Tsushima 2005: IFs requiring further evaluation, as judged by the study's radiologist.

d. Not all IFs requiring follow-up were classed as PSIFs but the distribution of PSIFs between women and men was not possible to calculate from the reported data; in this context, we use 'IFs requiring follow-up' as an approximate proxy for PSIFs.

e. Sandeman 2013: IFs requiring further referral, as judged by the study's geriatrician and radiologists.

f. Illes 2004: IFs requiring further evaluation, as judged by two neuroradiologists.

Supplementary Table 2-15: Summary of available data on potential determinants of prevalence of potentially serious incidental findings (PSIFs, as per our definition) or IFs which required follow-up (as per each study's definition), ordered by descending sample size – other factors

Study and body region	N patients[a] : N apparently asymptomatic people[b]	PSIFs	Factor	n PSIF[c]/ N with factor (%)	n PSIF[c]/ N without factor (%)	p-value[d]
Brain						
Tsushima 2005	0 : 1113	IFs requiring follow-up[e,f]	Cardiac disease[g]	0/36 (0.0)	15/1077 (1.4)	1.0
			Headache[h]	1/135 (0.7)	14/978 (1.4)	1.0
			Vertigo/dizziness[h]	4/139 (2.9)	11/974 (1.1)	0.1
			Heavy drinking[i]	0/42 (0.0)	15/1071 (1.4)	1.0
			Heavy smoking[j]	1/131 (0.8)	14/982 (1.4)	1.0
			Hypertensive[k]	4/267 (1.5)	11/846 (1.3)	0.8
			Hyperlipidaemia[l]	7/360 (1.9)	8/753 (1.1)	0.3
Sommer 2013	656 : 722	PSIFs	Psychotic episode[m]	9/656 (1.4)	13/722 (1.8)	0.7
Thorax						
Loy 2015	169 : 40	IF requiring follow-up[n]	HIV positive	12/169 (7.1)	1/40 (2.5)	0.5
Brain and body						
Saya 2017	44 : 44	Malignancy	TP53 mutation	4/44 (9.1)	0/44 (0.0)	0.1

PSIF = potentially serious incidental findings, - cannot be calculated, HIV = human immunodeficiency virus, NR = not reported

a. **Some studies included in the review involved patient groups, data from these were not included in the review, but are presented here to summarise prevalences of IFs between patient groups and apparently asymptomatic people.**

b. **Corresponds to sample sizes reported elsewhere in this review, i.e. apparently asymptomatic people only.**

c. **PSIFs, or IFs requiring follow-up, as appropriate.**

d. **P-values relate to two-tailed Fisher exact tests.**

e. **Tsushima 2005: IFs requiring further evaluation, as judged by the study's radiologist.**

f. **Not all IFs requiring follow-up were classed as PSIFs but the distribution of PSIFs between groups with and without each factor was not possible to calculate from the reported data; in this context, we use 'IFs requiring follow-up' as an approximate proxy for PSIFs.**

g. **Cardiac disease included congestive heart failure, myocardial infarction, angina pectoris, left ventricular hypertrophy atrial fibrillation; or electrocardiographic evidence of past myocardial infarction, left ventricular hypertrophy, or atrial fibrillation.**

h. **From a medical history taken by a neurologist.**

i. **≥ 60g of alcohol per day.**

j. **> 20 cigarettes per day.**

k. **Systolic blood pressure > 140 mmHg, or diastolic blood pressure > 90 mmHg, or on treatment for hypertension, or history of hypertension.**

l. **Fasting total cholesterol > 250 mg/dl, or history of hyperlipidaemia.**

m. **Sommer 2013: Not further defined.**

n. **Loy 2015: IFs requiring follow up, not further described.**

2.4 Conclusion

This chapter aimed systematically to review studies of brain, thorax, abdomen and of brain and body MRI conducted among apparently asymptomatic adults to determine the prevalence and types of PSIFs, factors associated with PSIFs, and describe what is known about the follow-up and final diagnoses of people with PSIFs.

Previous studies' estimates of the prevalence of IFs were limited by variation in the definition of IFs. By applying as consistent as possible a definition of PSIFs across studies, we have provided data on the prevalence of those IFs which may have an important impact on health. Meta-analyses of 32 published studies (n=27,643 participants) found that the pooled prevalences of PSIFs on brain, thorax, abdomen and brain and body MRI were 1.4%, 1.3%, 1.9% and 3.9% respectively. When IFs of uncertain potential seriousness were included in meta-analyses, these prevalence estimates rose to 1.7%, 3.0%, 4.5% and 12.8% respectively. Around half of PSIFs were described as suspected malignancies (brain 0.6%; thorax 0.6%; abdomen 1.3%; brain and body 2.3%). There was substantial within-study and between-study variation in prevalence estimates, but few data to reliably inform on sources of this.

Limited published follow-up data from five studies (n=234 participants with PSIFs) found that only a minority of participants (n=48, 21%) had a serious final clinical diagnosis. The available data did not allow the quantification of clinical assessments of participants with PSIFs. Informative data on the factors associated with PSIFs were also limited, but suggested some association with age, and no clear association with sex. The majority of studies employed IFs handling policies which involved systematic radiologist review of images, highlighting the lack of studies providing information on any variation in prevalence of PSIFs which may occur with other policies.

To better inform on the clinical assessments and final diagnoses of, and factors associated with PSIFs, systematic, long-term follow-up studies of unselected participants with PSIFs are needed. The next two chapters describe such a study, UK Biobank, and the evaluation of its policy to handle IFs generated during its multi-modal imaging study which is compared against a policy of systematic radiologist review.

Chapter 3 Development and evaluation of the UK Biobank incidental findings protocol

3.1 Introduction

UK Biobank is a major research resource comprising comprehensive phenotypic and genetic data on a cohort of over 500,000 British adults with data on incident diseases generated via linkages to routinely collected healthcare data (Sudlow et al., 2015). The UK Biobank dataset therefore facilitates studies of a large number of potential risk factors for a wide array of conditions with major public health impacts (Sudlow et al., 2015). To further enhance the existing dataset, the UK Biobank Imaging Study aims to collect multi-modal imaging data from a sub-cohort of 100,000 UK Biobank participants (Matthews and Sudlow, 2015). As such, potentially serious incidental findings (PSIFs) are a particularly pertinent issue for the UK Biobank Imaging Study, and it requires a protocol for handling them (Gibson et al., 2018). In this chapter, we describe the rationale behind, and process of, the UK Biobank IFs protocol: radiographer flagging of concerning images for a radiologist to review, with feedback of radiologist-confirmed PSIFs to participants and their general practitioners (GPs).

The UK Biobank IFs protocol will inevitably fail to identify all PSIFs which represent serious final diagnoses, due to the non-diagnostic nature of the research imaging and lack of systematic radiologist review of all images. Chapter 1 highlighted that members of the public may expect that all research images are reviewed by radiologists (The Royal College of Radiologists, 2011), and for research imaging to be able to generate firm clinical diagnoses (Kirschen et al., 2006). Therefore, it is imperative that UK Biobank participants have realistic expectations of how their PSIFs will be handled. This chapter will briefly describe the methods to evaluate participants' understanding of their consent to the UK Biobank IFs protocol; the results are presented in Sections 4.2 and 4.3.

Chapter 2 demonstrated a lack of studies which inform on the prevalence of PSIFs generated by protocols other than systematic radiologist review. There is no 'one-size-fits-all' incidental findings (IFs) protocol which will suit every imaging context (Gibson et al., 2018); rather, the choice of protocol should be justifiable, and based on empirical evidence. Judgements of the benefits and harms of feedback of PSIFs will be informed by data on the clinical assessments generated, and their final diagnoses. Chapter 2 demonstrated that such published data are limited, and that long-term, systematic follow-up studies of unselected participants with PSIFs are needed. To address this lack of data on the variation of the

prevalence of PSIFs under different protocols, and the methodological shortcomings of previous studies, this chapter also briefly describes the methods used to compare the UK Biobank IFs protocol with systematic radiologist review; the results are presented in Sections 4.2 and 4.3.

This chapter was originally published as a book chapter by Springer (Gibson et al., 2016b), and is included in full in Section 3.2.

3.2 Management of incidental findings on multi-modal imaging in UK Biobank

Management of Incidental Findings on Multimodal Imaging in UK Biobank

Lorna M. Gibson, Jonathan Sellors, and Cathie L.M. Sudlow

Contents

Abstract

UK Biobank is a major national health resource which aims to improve prevention, diagnosis and treatment of a wide range of serious and life-threatening illnesses. UK Biobank recruited 500,000 people aged between 40 and 69 years in 2006–2010, who underwent a range of measurements and provided detailed information about themselves, donated biological samples for future analyses and agreed to have their health followed long term. Among a range of ongoing enhancements, the UK Biobank Imaging Study aims to perform brain, cardiac and body magnetic resonance imaging, dual-energy X-ray absorptiometry and carotid Doppler ultrasound in 100,000 participants, generating the world's largest multimodal imaging dataset.

As incidental findings (IF) are an expected consequence of its imaging study, UK Biobank developed a pragmatic, scalable protocol for handling IF, in which participants and their

L.M. Gibson
University of Edinburgh, Centre for Clinical Brain Sciences, Chancellor's Building, 49 Little France Crescent, Edinburgh EH16 4SB, UK
e-mail: lorna.gibson@ed.ac.uk

J. Sellors
University of Oxford, Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health, Richard Doll Building, Old Road Campus, Oxford OX3 7LF, UK

UK Biobank Coordinating Centre, 1-2 Spectrum Way, Adswood, Stockport SK3 0SA, UK
e-mail: jonathan.sellors@ukbiobank.ac.uk

C.L.M. Sudlow (✉)
University of Edinburgh, Centre for Clinical Brain Sciences, Chancellor's Building, 49 Little France Crescent, Edinburgh EH16 4SB, UK

UK Biobank Coordinating Centre, 1-2 Spectrum Way, Adswood, Stockport SK3 0SA, UK
e-mail: cathie.sudlow@ed.ac.uk

Med Radiol Diagn Imaging (2016)
DOI 10.1007/174_2016_91, © Springer International Publishing Switzerland
Published Online: 17 Dec 2016

general practitioners receive feedback in limited circumstances: when, during image acquisition, a radiographer notices a potentially serious IF ('indicating the possibility of a condition which, if confirmed, would carry a real prospect of seriously threatening life span or of having a substantial impact on major body functions or quality of life') and a radiologist subsequently confirms a potentially serious IF.

UK Biobank has compared its IF protocol against a commonly used protocol (systematic review of all images by radiologists) and collected comprehensive data on the impact of feedback of potentially serious IF on participants and health services. The results will be published separately and will provide robust, empirical evidence to inform debates surrounding handling IF and designs of future studies' IF policies.

1 Introduction

1.1 UK Biobank

UK Biobank is a large, prospective epidemiological research resource which recruited approximately 500,000 people aged 40–69 between 2006 and 2010 (Sudlow et al. 2015). UK Biobank aims to enable studies of the prevention, diagnosis and treatment of common and serious diseases and is open to use by researchers from anywhere in the world for health-related research which is in the public interest (Collins 2012). The UK Biobank resource contains detailed baseline questionnaire and physical measurement data, genotyping and biochemical assay data, and biological samples from all participants (Sudlow et al. 2015). UK Biobank participants have agreed to have their health followed, and data on health outcomes are derived via linkages to routinely collected national healthcare datasets. Enhanced data collection is ongoing in subsets of participants, and in April 2014, UK Biobank embarked on its most ambitious enhanced data collection project to date: the UK Biobank Imaging Study.

We aim to describe the UK Biobank Imaging Study, the development of the UK Biobank incidental findings (IF) protocol and UK Biobank's programmes of evaluation of this protocol: (i) of participants' understanding of consent in relation to receiving feedback about a potentially serious IF (defined as one indicating the possibility of a condition which, if confirmed, would carry a real prospect of seriously threatening life span or of having a substantial impact on major body functions or quality of life) and (ii) the impact of the UK Biobank IF protocol on participants and health services, the results of which will be published separately.

1.2 The UK Biobank Imaging Study

Over the next seven years, UK Biobank will perform brain, cardiac and body magnetic resonance imaging (MRI), carotid Doppler ultrasound and dual-energy X-ray absorptiometry (DXA) in 100,000 of its participants and generate the world's largest multimodal imaging dataset. The data will enable researchers to investigate associations between imaging-derived phenotypes (IDP) and the wealth of exposure and outcome data from baseline and other enhanced data collections and health record linkages within the resource.

Research imaging is currently underway at the purpose-built imaging centre in Stockport, with further centres planned. On arrival at the imaging centre, participants undergo registration, prescreening and consent, followed by imaging. In order to provide contemporaneous non-imaging data, at the end of the visit, participants repeat the entire baseline assessment and an additional 12-lead electrocardiogram. Each participant's imaging visit lasts approximately four hours.

The UK Biobank Imaging Working Group collaborated with over 100 scientists to design the UK Biobank Imaging Study protocol, which aims to balance the acquisition of high-quality imaging data against feasible methods which are acceptable to participants (Matthews and Sudlow 2015; UK Biobank 2015e). These data enable UK Biobank to generate a wide range of IDP

(Table 1) and facilitate the development and testing of new image analyses methods, the results of which are being integrated into, and thus further enhancing, the UK Biobank resource (Matthews and Sudlow 2015).

Participants undergo an approximately 30-min 3.0 T brain MRI (Skyra, Siemens, Erlangen, Germany), which includes structural (T1, T2 fluid-attenuated inversion recovery, susceptibility-weighted imaging and T2*), functional and diffusion imaging (UK Biobank 2016). From these images, UK Biobank generates IDP including measures of volumes of total grey matter, cortical grey matter, total white matter, cerebrospinal fluid and structures such as the thalamus, detailed data on activation and statistical effect sizes in different regions during fMRI tasks and diffusion parameters such as fractional anisotropy in different white matter tracts (UK Biobank 2016; Miller et al. 2016).

A 20-minute non-contrast cardiac MRI is acquired using a 1.5 T Magnetom Aera scanner (Siemens Healthcare, Erlangen, Germany). Sequences include long and short axis cine, aortic distensibility cine, tagging and aortic valve flow images, from which IDP such as cardiac output, ejection fraction and end-diastolic, end-systolic and stroke volumes are calculated and from which a wide range of additional measures are being derived using novel, automated methods (UK Biobank 2015b; Petersen et al. 2013).

Participants are then repositioned within the 1.5 T scanner and undergo a 10-min body MRI. In total, these images cover tissues from the neck to the knees and include a T1 abdomen, T1 pancreas and a liver and pancreas multi-echo sequence. From these images, semiautomated measures of liver fat, fibrosis and haemosiderosis percentages can be made, in addition to body composition measurements of subcutaneous and visceral fat and thigh muscle mass (UK Biobank 2015a; West et al. 2016). Ongoing methodological developments will lead to the derivation of an increasingly wide range of measures.

Table 1 Summary of imaging modalities and imaging-derived phenotypes included in the UK Biobank Imaging Study

Imaging modality and references for further information	Scanner	Scan duration (min)	Imaging acquired	n IDP currently available	Examples of available IDP
Brain MRI (UK Biobank 2016)	3.0 T Skyra[1]	30	T1, T2 FLAIR, SWI, T2*, fMRI, DWI	749	Tissue volumes, activation during fMRI, fractional anisotropy
Cardiac MRI (UK Biobank 2015b)	1.5 T Magnetom Aera[1]	20	Cine (long axis, short axis, aorta), tagged, aortic valve flow	30	Cardiac output, ejection fraction, stroke volumes
Abdominal MRI (UK Biobank 2015a)	1.5 T Magnetom Aera[1]	10	T1 abdomen, T1 pancreas, liver and pancreas multi-echo, Dixon	5	Percentages of liver fat, fibrosis and haemosiderosis, body composition
DXA (UK Biobank 2015c)	iDXA[2]	20	Whole body, thoracolumbar spine, hips, knees	120	Bone area, mineral content and density, lean mass, fat mass
Carotid Doppler US (UK Biobank 2015d)	5–13 MHz linear array transducer and CardioHealth Station[3]	10	Video loops in longitudinal and transverse plane, CIMT measures	16	Minimum, maximum and mean CIMT

IDP imaging-derived phenotypes, *MRI* magnetic resonance imaging, *FLAIR* fluid-attenuated inversion recovery, *SWI* susceptibility-weighted imaging, *fMRI* functional MRI, *DWI* diffusion-weighted imaging, *DXA* dual-energy X-ray absorptiometry, *US* ultrasound, *CIMT* carotid intima-media thickness

[a]Siemens, Erlangen, Germany

[b]GE-Lunar, Wisconsin, USA

[c]Panasonic, Leicester, UK

Carotid Doppler ultrasound images are acquired during a 10-min examination using a 5–13 MHz linear array transducer and a CardioHealth Station (Panasonic, Leicester, UK). Two-dimensional transverse and longitudinal plane images of each carotid artery are saved as cine loops, followed by two measures of intima-media thickness per carotid artery. From these images, mean, minimum and maximum calculations of carotid intima-media thickness are generated, and additional measures of plaque characteristics will follow (UK Biobank 2015d).

DXA images of the whole body, thoracolumbar spine, hips and knees are acquired using an iDXA scanner (GE-Lunar, Wisconsin, USA). The scanner automatically generates multiple IDP of the bone area, mineral content and density and body composition measures of lean and fat mass (UK Biobank 2015c).

Descriptions of all available IDP from each modality, and non-imaging variables, are available from the UK Biobank showcase (http://www.ukbiobank.ac.uk/data-showcase).

2 UK Biobank IF Protocol

2.1 Development of the UK Biobank IF Protocol

Incidental findings (IF) are findings deemed beyond the aims of a study (Wolf et al. 2008). IF are particularly pertinent to the UK Biobank Imaging Study given the nature of IF which may be identifiable on multimodal imaging of 100,000 largely asymptomatic participants. The handling of IF in research imaging is the subject of widespread debates (Gibson et al. 2016 In Press), and while there is no 'one-size-fits-all' approach to detecting and feeding back IF, researchers should anticipate IF and design appropriate IF handling policies (Medical Research Council and Wellcome Trust 2014).

The UK Biobank IF protocol was developed following an extensive process which involved reviewing existing policies for feedback of findings to UK Biobank participants, published evidence and guidance on IF, received external legal advice on the scope of the duty of care and consultations with the independent UK Biobank Ethics and Governance Council, UK Biobank's major funders (Wellcome Trust and Medical Research Council) and with the Royal College of Radiologists and the Society and College of Radiographers. In addition, UK Biobank sought to learn from the experiences and approaches taken to handling IF used by several other large-scale research imaging projects, including the German National Cohort, the Rotterdam Scan Study, the Multi-Ethnic Study of Atherosclerosis (MESA), and the Reykjavik Heart Study. UK Biobank also consulted with relevant experts to explore the legal and ethical factors which were applicable to the development of the IF protocol.

The UK Biobank IF protocol was developed from first principles as a pragmatic protocol that could be implemented on a large scale with the objective of striking the optimum balance of most net benefit and least net harm to 100,000 largely asymptomatic participants (UK Biobank 2015e). Under this protocol, participants only receive feedback in specific, limited circumstances: when a radiographer identifies a potentially serious IF during the acquisition or quality assessment of images during the imaging visit and a radiologist subsequently confirms the presence of a potentially serious IF. UK Biobank defines a potentially serious imaging IF as 'as a finding which indicates the possibility of a condition which, if confirmed, would carry a real prospect of seriously threatening life span, or of having a substantial impact on major body functions or quality of life.'

2.2 Consent Processes

Before attending the imaging centre, UK Biobank provides participants with an information leaflet which includes a description of the IF protocol and what they should and should not reasonably expect (UK Biobank 2014b).

The information leaflet explains that the scans are not intended to diagnose an illness or identify a particular abnormality and that they will not be looked at routinely by doctors. Participants are

informed that if, during the scan, the radiographer notices something which they think may be serious, only then will the scan be reviewed by a doctor; if the doctor thinks there may be a potentially serious finding, the participant and their GP will be informed. The leaflet gives examples of IF which would be fed back to participants (a tumour) and those which would not (gallstones or a simple cyst).

UK Biobank's consent form explicitly asks for participants' consent on the basis that (a) they understand that these scans are for research purposes only and that they will not be routinely examined by medical staff and should not be regarded as part of a 'health check,' (b) that they give permission for UK Biobank to contact them and their GP in the event that a potentially serious IF is found on a scan and (c) that a lack of contact from UK Biobank does not imply that no potentially serious IF exists, but simply that no such abnormality was noticed by the staff taking the scans (UK Biobank 2014a).

2.3 Identification of IF

UK Biobank modified a list of IF developed by the German National Cohort to detail those IF which may be detected on brain, cardiac or body MR or DXA which UK Biobank would consider potentially serious and warrant feedback to participants and their GPs and those which it would consider not serious and would not be fed back. It was deemed that carotid Doppler ultrasound conducted by radiographers would not produce any IF which would be considered potentially serious.

The list is not exhaustive, and in the event that an IF is detected which is not included in the list, radiographers and radiologists are guided by the UK Biobank definition of a potentially serious IF (those which indicate the possibility of a condition which, if confirmed, would carry a real prospect of seriously threatening life span or of having a substantial impact on major body functions or quality of life) to judge whether an IF is deemed potentially serious or not (UK Biobank 2014b).

2.4 Feedback of IF

If the reviewing expert decides that the IF is not serious, then no further action is taken. If, on the other hand, the reviewing expert confirms that the IF is potentially serious, then they provide a short summary for the participant and a more comprehensive summary for the participant's GP (UK Biobank 2015e).

The GP is informed that the images have not been optimised for the purpose of identifying abnormalities and have not been reviewed in a clinical setting. Further investigations and/or referrals are left to the discretion of the GP. As required, the participant's doctors are able to review the scans collected by UK Biobank (UK Biobank 2015e).

3 Evaluation of the Impact of the UK Biobank IF Protocol

3.1 Evaluating Participants' Understanding of Consent

Given that systematic radiologist review of all acquired images is not undertaken, the UK Biobank IF Protocol will inevitably fail to identify some potentially serious IF which represent serious diseases. Public expectations relating to feedback of IF may well be unrealistic, the public associate imaging with clinical diagnoses (Kirschen et al. 2006), and expect that images will be reviewed by experts (The Royal College of Radiologists 2011). It is therefore crucial to manage participants' expectations of what will be fed back, and what will not, and specifically to ensure that they understand that the imaging does not constitute a 'health check' and that lack of feedback of a potentially serious IF does not represent an 'all clear.' The intention of the UK Biobank information materials and consent process is to provide participants with a fair, reasonable and realistic expectation of the outcome of their visit for imaging in the UK Biobank Imaging Study. UK Biobank developed a questionnaire to assess participants' understanding of this consent, which is sent to imaged participants two days after their imaging visit.

Participants are asked whether or not they thought they consented to the following: return of scans and results at the end of the imaging visit; to choose whether they and their GP would be informed; that they and their GP would automatically be contacted; that they would receive feedback of a potentially serious IF during the imaging visit; whether they would receive feedback of an IF after the imaging visit. These data are periodically reviewed so that the design of the UK Biobank Imaging Study consent materials can be improved and results will be published separately.

3.2 Comparing the UK Biobank IF Protocol with Full Review of Images by Radiologists

There is no 'best' policy for handling IF detected during research imaging of healthy populations, and existing studies vary in their approach (The Royal College of Radiologists 2011). However, there are likely to be 'better' and 'worse' policies for handling IF, which will depend on the context of the individual research study. Imaging studies should develop IF policies which are appropriate to their context (Medical Research Council and Wellcome Trust 2014), and evaluation studies which directly compare different approaches to IF will guide decisions as to which policy is more appropriate.

UK Biobank therefore designed such an evaluation study, the methods and results of which will be published in a forthcoming research article. In brief, UK Biobank assessed the prevalence of potentially serious IF and the proportions of these which were finally diagnosed as serious (i.e. true positives) and not serious (i.e. false positives) as a result of the UK Biobank IF Protocol compared with a common approach to handling IF in other imaging studies: systematic review of images by radiologists. UK Biobank also investigated the rate of serious final diagnoses which were detected by radiologists but missed by the UK Biobank IF Protocol (i.e. false negatives). The impact of feedback of potentially serious IF on participants and health services was informed by questionnaires to participants and their GPs. This evaluation was encouraged by the main funders of UK Biobank (the Medical Research Council and the Wellcome Trust) and the UK Biobank's independent Ethics and Governance Council.

Results on the rates of prevalence of potentially serious IF, false positives, false negatives and the impact of feedback of potentially serious IF were crucial in guiding judgement of the potential net benefit and net harm of each protocol.

3.3 Qualitative Work

In order to provide context for and greater exploration of the results of the quantitative evaluation study of the UK Biobank IF Protocol described above, UK Biobank commissioned the research company TNS-BMRB to conduct a parallel qualitative study of participants' experiences of the imaging visit, understanding of the consent they had given, the process and opinions of receiving feedback of a potentially serious IF and the impact of receiving feedback of a potentially serious IF (TNS-BMRB 2015). These qualitative data were collected with the aim of informing the protocol on feedback of IF for the main phase of the UK Biobank Imaging Study. The detailed methods and results of this study will be made available in a separate report.

3.4 Ongoing Evaluation

UK Biobank continues to send questionnaires to participants two days following their imaging visit in order to evaluate their understanding of consent and to send questionnaires six weeks and six months following imaging to collect data on final diagnoses, clinical follow-up and impact on participants. In addition, UK Biobank continues to send questionnaires after six months to GPs in order to collect data on final diagnoses, clinical follow-up and GPs' opinions on the net benefit and harm of providing feedback of a potentially serious IF on their patients.

This systematic follow-up of participants will provide much-needed robust, empirical data on the impact on participants and health services and data on final diagnoses and false-positive rates. Such data, along with linkages to national healthcare datasets, will enable UK Biobank to continually monitor the impact of its IF protocol and to address additional questions raised by the UK Biobank evaluation study described which warrant further research: whether or not early diagnosis of serious disease results in net benefit for asymptomatic participants and what are the health economic consequences of the UK Biobank Imaging IF Protocol. These data will contribute evidence to the debates surrounding the management of IF in research imaging and inform the practical design of appropriate and feasible IF policies for future imaging studies.

4 Summary

The UK Biobank Imaging Study aims to image 100,000 healthy participants and will generate the world's largest multimodal imaging dataset. UK Biobank has developed a pragmatic, scalable protocol for handling IF during the Imaging Study which results in feedback of IF to participants and their GPs in only limited circumstances: where a radiographer notices a potentially serious IF, images are reviewed by a radiologist, and feedback given if the radiologist confirms the presence of a potentially serious IF. This approach differs from many studies, including other large national imaging projects, in which systematic review of images by radiologists for IF is undertaken. The impact of the UK Biobank IF protocol is under continuous evaluation, and data collection is ongoing of participants' clinical follow-up and final diagnoses and the impact on participants' emotional well-being, insurance and finances and work and activities. In addition, UK Biobank has performed a head-to-head comparison of its IF protocol against systematic review by radiologists, and following initial data analyses and revision of consent materials, it continues to assess participants' understanding of consent. Such analyses will be of value not only to UK Biobank, but will provide much-needed robust, empirical data on the impact of feedback of IF which will address current gaps in knowledge and inform the design of IF policies in future imaging studies.

References

Collins R (2012) What makes UK Biobank special? Lancet 379(9822):1173–1174. http://dx.doi.org/10.1016/S0140-6736(12)60404-8

Gibson LM, Sudlow CLM, Wardlaw J (2016) Incidental findings: current ethical debates and future challenges in advanced neuroimaging. In: Illes J, Hossain S (eds) Neuroethics: anticipating the future, 2 edn. Oxford University Press, Oxford (In Press)

Kirschen MP, Jaworska A, Illes J (2006) Subjects' expectations in neuroimaging research. J Magn Reson Imaging 23(2):205–209. doi:10.1002/jmri.20499

Matthews PM, Sudlow C (2015) The UK Biobank. Brain 138(12):3463–3465. doi:10.1093/brain/awv335

Medical Research Council, Wellcome Trust (2014) Framework on the feedback of health-related findings in research. www.mrc.ac.uk/documents/pdf/mrc-wellcome-trust-framework-on-the-feedback-of-health-related-findings-in-researchpdf/

Miller KL, Alfaro-Almagro F, Bangerter NK, Thomas DL, Yacoub E, Xu J, Bartsch AJ, Jbabdi S, Sotiropoulos SN, Andersson JLR, Griffanti L, Douaud G, Okell TW, Weale P, Dragonu I, Garratt S, Hudson S, Collins R, Jenkinson M, Matthews PM, Smith SM (2016) Multimodal population brain imaging in the UK Biobank prospective epidemiological study. Nat Neurosci 19(11):1523–1536. doi:10.1038/nn.4393

Petersen SE, Matthews PM, Bamberg F, Bluemke DA, Francis JM, Friedrich MG, Leeson P, Nagel E, Plein S, Rademakers FE, Young AA, Garratt S, Peakman T, Sellors J, Collins R, Neubauer S (2013) Imaging in population science: cardiovascular magnetic resonance in 100,000 participants of UK Biobank - rationale, challenges and approaches. J Cardiovasc Magn Reson 15(1):46–46. doi:10.1186/1532-429X-15-46

Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, Liu B, Matthews P, Ong G, Pell J, Silman A, Young A, Sprosen T, Peakman T, Collins R (2015) UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 12(3):e1001779. doi:10.1371/journal.pmed.1001779

The Royal College of Radiologists (2011) Management of incidental findings detected during research imaging. https://www.rcr.ac.uk/system/files/publication/field_publication_files/BFCR%2811%298_ethics.pdf

TNS-BMRB (2015) Attitudes to and impact of the feedback of potentially serious incidental findings from UK Biobank's imaging pilot study.

UK Biobank (2014a) Consent form for the imaging assessment: UK Biobank. http://www.ukbiobank.ac.uk/wp-content/uploads/2016/06/consent_form_imaging.pdf

UK Biobank (2014b) Information leaflet: imaging assessment visit. http://www.ukbiobank.ac.uk/wp-content/uploads/2016/06/information_leaflet_imaging.pdf

UK Biobank (2015a) UK Biobank body (abdominal) MRI scan. http://biobank.ctsu.ox.ac.uk/crystal/docs/body_mri_explan.pdf

UK Biobank (2015b) UK Biobank imaging modality cardiovascular magnetic resonance (CMR). http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=349

UK Biobank (2015c) UK Biobank imaging modality DXA. http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=502

UK Biobank (2015d) UK Biobank imaging modality: carotid ultrasound. http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=511

UK Biobank (2015e) UK Biobank: report on feasibility of imaging assessments in 100,000 participants.

UK Biobank (2016) UK Biobank brain imaging documentation. http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=1977

West J, Dahlqvist Leinhard O, Romu T, Collins R, Garratt S, Bell JD, Borga M, Thomas L (2016) Feasibility of MR-based body composition analysis in large scale population studies. PLoS One 11(9):e0163332. doi:10.1371/journal.pone.0163332

Wolf SM, Lawrenz FP, Nelson CA, Kahn JP, Cho MK, Clayton W, Fletcher JG, Georgieff MK, Hammerschmidt D, Illes J, Kapur V, Keane MA, Koenig BA, Leroy BS, McFarland EG, Paradise J, Parker LS, Terry SF, Ness BV, Wilfond BS (2008) Managing incidental findings in human subjects research: analysis and recommendations. J law Med Ethics J Am Soc Law Med Ethics 36(2):211–219. doi:10.1111/j.1748-720X.2008.00266.x.Managing

3.3 Conclusion

The UK Biobank IFs protocol involves radiographer flagging of concerning images for review by radiologists. As this approach differs from other studies, which have mostly used systematic radiologist review of all images for IFs, UK Biobank aims to evaluate its protocol by 1) surveying participants to assess their understanding of consent with regards to feedback of PSIFs; 2) comparing the UK Biobank IFs protocol with systematic radiologist review of all images; 3) surveying participants with PSIFs, and their GPs, up to six months following feedback of a PSIF to collect data on clinical assessments, final diagnoses, and impacts (Gibson et al., 2016b).

This evaluation programme involves long-term, systematic follow-up of all (i.e. unselected) participants with PSIFs, a methodology which has been lacking from the majority of previous studies. Taken together with the extensive phenotypic and linked healthcare data which are available for UK Biobank participants, the UK Biobank dataset provides a unique opportunity to address several gaps in our knowledge of PSIFs (Gibson et al., 2016b). The following three chapters describe studies which utilise UK Biobank data to inform on the: differences in prevalence of PSIFs and serious final diagnoses generated by the UK Biobank IFs protocol compared to a protocol involving systematic radiologist review of all images, participants' understanding of consent with regards to feedback of PSIFs, and participants' and GPs' opinions on feedback of PSIFs (Chapter 4); factors associated with PSIFs and with serious final diagnoses (Chapters 4 and 5); economic impacts on hospital services following feedback of PSIFs (Chapter 6).

Chapter 4 Impact of detecting potentially serious incidental findings during multi-modal imaging

4.1 Introduction

Judgements about the benefits and harms of feeding back potentially serious incidental findings (PSIFs) should be informed by evidence on the prevalence and impact of PSIFs on participants and health services (Gibson et al., 2017b). While Chapter 2 demonstrated that the prevalence of PSIFs varies by imaged body region, there were no studies which informed on the variation in prevalence of either PSIFs or serious final diagnoses by radiologist versus non-radiologist readers, as the majority of studies included in Chapter 2 involved systematic review of all images for IFs by at least one radiologist. However, the current shortage of radiologists in the UK, and the non-clinical setting of many UK imaging research centres (The Royal College of Radiologists, 2011), means that data on the impact of alternative approaches to handling IFs are required to inform the design of pragmatic IFs policies. Although systematic radiologist review of all images may be assumed to be the most appropriate IFs handling policy (Kirschen et al., 2006; Milstein, 2008), there are limited data on the clinical assessments generated and the final diagnoses of PSIFs detected by radiologists on research imaging to confirm or refute this claim; long-term, systematic follow-up of unselected participants with PSIFs is needed (Chapter 2).

Chapter 1 described a small focus group from the Netherlands Epidemiology of Obesity (NEO) study that sought people's opinions on feedback of PSIFs (de Boer et al., 2018), and highlighted that other studies had sought opinions based on hypothetical scenarios (Brown and Knight, 2010; Kirschen et al., 2006; Opinion Leader, 2012). More data from participants who had actually received feedback of an IF would help to address several gaps in our knowledge of their experiences, such as: their opinions of receiving feedback; the impact of feedback on their emotional wellbeing, insurance and finances or on work and activities; their understanding of what they had consented to with regards to feedback. Previous studies of healthcare professionals' opinions focus on handling IFs detected on clinical imaging of patients (Booth and Boyd-Ellison, 2015; Zafar et al., 2016); staff who are managing apparently asymptomatic people with IFs generated via research imaging may have different opinions.

The previous chapter described the UK Biobank, its multi-modal imaging study of 100,000 participants, the protocol designed to handle the PSIFs which will be generated (radiographer

flagging of concerning images for a radiologist to review), and the methods for evaluating this protocol (Gibson et al., 2016b).

This chapter provides further detail on the methods of, and presents the results of, the evaluation of the UK Biobank IFs protocol. By performing long-term, systematic follow-up of all participants with PSIFs, and their general practitioners (GPs), this study addresses several of the gaps in our knowledge which are described above by providing robust, empirical data on: the prevalence and final diagnoses of PSIFs generated by two different protocols (radiographer flagging versus systematic radiologist review); clinical assessments; impacts on participants' emotional wellbeing, insurance and finances, and work and activities; participants' understanding of consent regarding the UK Biobank IFs handling policy; participants' and GPs' opinions about receiving feedback of a PSIF; factors associated with PSIFs.

The study has been published as 'Impact of detecting potentially serious incidental findings during multi-modal imaging [version 3; referees: 2 approved, 1 approved with reservations]. LM Gibson, TJ Littlejohns, L Adamska, S Garratt, N Doherty, UK Biobank Imaging Working Group, JM Wardlaw, G Maskell, M Parker, R Brownsword, PM Matthews, R Collins, NE Allen, J Sellors, CLM Sudlow. Wellcome Open Research 2018; 2:114.' The article is included in full in Section 4.2 and the supplementary materials are included in full in Section 4.3.

4.2 Impact of detecting potentially serious incidental findings during multi-modal imaging

Check for updates

RESEARCH ARTICLE

REVISED Impact of detecting potentially serious incidental findings during multi-modal imaging [version 3; referees: 2 approved, 1 approved with reservations]

Lorna M Gibson [1], Thomas J Littlejohns[2], Ligia Adamska[2], Steve Garratt [3], Nicola Doherty[3], UK Biobank Imaging Working Group, Joanna M Wardlaw [4], Giles Maskell[5], Michael Parker [6], Roger Brownsword[7], Paul M Matthews [3,8], Rory Collins[2,3], Naomi E Allen[2,3], Jonathan Sellors[2,3*], Cathie LM Sudlow[1,3*]

[1]Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK
[2]Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK
[3]UK Biobank Coordinating Centre, Stockport, UK
[4]Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK
[5]Royal Cornwall Hospitals NHS Trust, Cornwall, UK
[6]Ethox Centre, Nuffield Department of Population Health, University of Oxford, Oxford, UK
[7]The Dickson Poon School of Law, King's College London, London, UK
[8]Division of Brain Sciences, Department of Medicine, Hammersmith Hospital, Imperial College London, London, UK

[*] Equal contributors

v3 **First published:** 30 Nov 2017, **2**:114 (doi: 10.12688/wellcomeopenres.13181.1)
Second version: 03 May 2018, **2**:114 (doi: 10.12688/wellcomeopenres.13181.2)
Latest published: 02 Aug 2018, **2**:114 (doi: 10.12688/wellcomeopenres.13181.3)

Abstract

Background: There are limited data on the impact of feedback of incidental findings (IFs) from research imaging. We evaluated the impact of UK Biobank's protocol for handling potentially serious IFs in a multi-modal imaging study of 100,000 participants (radiographer 'flagging' with radiologist confirmation of potentially serious IFs) compared with systematic radiologist review of all images.

Methods: Brain, cardiac and body magnetic resonance, and dual-energy x-ray absorptiometry scans from the first 1000 imaged UK Biobank participants were independently assessed for potentially serious IFs using both protocols. We surveyed participants with potentially serious IFs and their GPs up to six months after imaging to determine subsequent clinical assessments, final diagnoses, emotional, financial and work or activity impacts.

Results: Compared to systematic radiologist review, radiographer flagging resulted in substantially fewer participants with potentially serious IFs (179/1000 [17.9%] versus 18/1000 [1.8%]) and a higher proportion with serious final diagnoses (21/179 [11.7%] versus 5/18 [27.8%]). Radiographer flagging missed 16/21 serious final diagnoses (i.e., false negatives), while systematic radiologist review generated large numbers of non-serious final diagnoses (158/179) (i.e., false positives). Almost all (90%) participants had further clinical assessment (including invasive procedures in similar numbers with serious and non-serious final diagnoses [11 and 12 respectively]), with additional impact on

Open Peer Review

Referee Status: ? ✓ ✓

	Invited Referees 1	2	3
REVISED **version 3** published 02 Aug 2018			
REVISED **version 2** published 03 May 2018		✓	✓ report
version 1 published 30 Nov 2017	? report	✓ report	

emotional wellbeing (16.9%), finances (8.9%), and work or activities (5.6%).
Conclusions: Compared with systematic radiologist review, radiographer flagging missed some serious diagnoses, but avoided adverse impacts for many participants with non-serious diagnoses. While systematic radiologist review may benefit some participants, UK Biobank's responsibility to avoid both unnecessary harm to larger numbers of participants and burdening of publicly-funded health services suggests that radiographer flagging is a justifiable approach in the UK Biobank imaging study. The potential scale of non-serious final diagnoses raises questions relating to handling IFs in other settings, such as commercial and public health screening.

Keywords
incidental findings, magnetic resonance imaging, dual-energy X-ray absorptiometry, false positives, false negatives, research ethics

1 **Bjorn Hofmann**, Norwegian University of Science and Technology (NTNU), Norway
University of Oslo, Norway

2 **Saurabh Jha**, University of Pennsylvania, USA

3 **Eline M Bunnik**, University Medical Centre Rotterdam, Netherlands

Discuss this article

Comments (0)

Corresponding author: Cathie LM Sudlow (cathie.sudlow@ed.ac.uk)

Author roles: Gibson LM: Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Software, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Littlejohns TJ**: Data Curation, Formal Analysis, Methodology, Writing – Review & Editing; **Adamska L**: Data Curation, Writing – Review & Editing; **Garratt S**: Data Curation, Writing – Review & Editing; **Doherty N**: Project Administration; **Wardlaw JM**: Supervision, Writing – Review & Editing; **Maskell G**: Writing – Review & Editing; **Parker M**: Writing – Review & Editing; **Brownsword R**: Writing – Review & Editing; **Matthews PM**: Conceptualization, Funding Acquisition, Methodology, Writing – Review & Editing; **Collins R**: Conceptualization, Funding Acquisition, Methodology, Writing – Review & Editing; **Allen NE**: Conceptualization, Funding Acquisition, Methodology, Writing – Review & Editing; **Sellors J**: Conceptualization, Funding Acquisition, Supervision, Writing – Review & Editing; **Sudlow CL**: Conceptualization, Data Curation, Funding Acquisition, Methodology, Supervision, Writing – Review & Editing

Competing interests: LMG: Member of the UK Biobank Imaging Working Group. UK Biobank Imaging Consultant, University of Edinburgh. TJL: UK Biobank Epidemiologist, University of Oxford. LA: UK Biobank Data Analyst, University of Oxford. SG: Member of the UK Biobank Imaging Working Group. Senior Project Manager of UK Biobank. ND: Senior Clinical Study Administrator of UK Biobank. JMW: Advised on imaging protocols for the UK Biobank imaging study. Currently analysing UK Biobank brain imaging data. GM: Member of the UK Biobank International Scientific Advisory Board. MP: Director of Ethox, University of Oxford. Ethox provides independent advice on ethical matters to UK Biobank. RB: None currently. Former chair of the UK Biobank Ethics and Governance Council. PMM: Chair of the UK Biobank Imaging Working Group, Member of UK Biobank Steering Committee. RC: CEO and Principle Investigator of UK Biobank. Chair of UK Biobank Steering Committee. Member of the UK Biobank Imaging, Enhancements, Follow-up and Outcomes and Infectious Diseases Working Groups. NEA: Member of UK Biobank Steering Committee, UK Biobank Imaging, Enhancements, Follow-up and Outcomes and Infectious Diseases Working Groups. UK Biobank Senior Epidemiologist. JS: UK Biobank Legal Counsel. CLMS: Member of UK Biobank Steering Committee, and UK Biobank Imaging, Enhancements, and Follow-up and Outcomes Working Groups. UK Biobank Chief Scientist.

Grant information: This work was supported by the Wellcome Trust [107190]. UK Biobank is funded by the Wellcome Trust, Medical Research Council, Department of Health, Scottish Government, Northwest Regional Development Agency, Welsh Assembly Government, British Heart Foundation, and Diabetes UK.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Gibson LM, Littlejohns TJ, Adamska L *et al.* **Impact of detecting potentially serious incidental findings during multi-modal imaging [version 3; referees: 2 approved, 1 approved with reservations]** Wellcome Open Research 2018, **2**:114 (doi: 10.12688/wellcomeopenres.13181.3)

First published: 30 Nov 2017, **2**:114 (doi: 10.12688/wellcomeopenres.13181.1)

REVISED **Amendments from Version 2**

We have added new references (de Boer *et al.*, 2018, and Bos *et al.*, 2016) and provided further detail on the UK Biobank incidental findings protocol.

See referee reports

Introduction

UK Biobank (www.ukbiobank.ac.uk) is a major resource for research into the determinants of a wide range of serious and life-threatening diseases, to improve their prevention, diagnosis and treatment[1]. It is a prospective study which recruited 500,000 men and women aged 40–69 across the UK between 2006 and 2010[1]. It includes extensive questionnaire and physical measurement data from the baseline visit, biological samples (with genotyping and biomarker assay data), longitudinal follow-up data from national health-related datasets and additional information from remote monitoring and web-based questionnaires.

The UK Biobank imaging study aims to perform brain, cardiac and body magnetic resonance imaging (MRI), dual-energy x-ray absorptiometry (DXA) and carotid Doppler ultrasound in 100,000 UK Biobank participants in dedicated imaging centres over seven years (http://imaging.ukbiobank.ac.uk/). By November 2017, over 20,000 participants had attended an imaging assessment visit (http://imaging.ukbiobank.ac.uk/), making it already the world's largest ever multi-modal imaging study[2].

Incidental findings (IFs), defined as 'findings discovered in the course of research that are beyond the aims of the study,'[3] are a predictable consequence of much research, and studies need appropriate protocols for handling them (https://wellcome.ac.uk/funding/managing-grant/wellcome-trust-policy-position-health-related-findings-research/)[4]. IFs are particularly pertinent to the UK Biobank imaging study given its large scale and the potential seriousness of IFs that may be detected. While clinical care and screening programmes aim to provide clinical benefit to patients, research studies have the primary aim of producing generalisable knowledge. Nevertheless, while research studies do not aim to benefit participants directly, they are obliged to minimise potential harms to participants and the wider public. Hence, although the UK Biobank imaging study aims to collect research data, rather than to detect or diagnose serious disease, it does require a protocol to handle IFs should they arise.

The UK Biobank imaging IFs protocol was developed as a pragmatic, scalable process, aiming to produce the best possible resource for biomedical research while minimising any potential harms for 100,000 largely asymptomatic UK Biobank participants. UK Biobank reviewed current practice, the extensive literature[3,5,6] and relevant published guidance (https://www.rcr.ac.uk/publication/management-incidental-findings-detected-during-research-imaging), sought independent legal advice, and consulted with its independent Ethics and Governance Council, the UK's Royal College of Radiologists and Society and College of Radiographers, funders, relevant experts and leading imaging research projects (including the Multi-Ethnic Study of Atherosclerosis [http://www.hopkinsmedicine.org/heart_vascular_institute/clinical_trials/preventive/mesa.html], the Reykjavik Heart Study [http://www.hjartarannsokn.is/index.aspx?GroupId=406], the Rotterdam Scan Study[7] and the German National Cohort [http://nako.de/])[2]. Key contextual factors considered were the non-clinical setting of the imaging visit, in which the scanning sequences are optimised for research use rather than clinical diagnosis, and the nature of the participants' existing consent (in particular the approach to the feedback of IFs). However, cost effectiveness was not considered relevant[2].

The UK Biobank imaging IFs protocol involves feedback to participants and their general practitioners (GPs) when a radiographer observes a potentially serious IF during image acquisition that is subsequently confirmed by a specialist radiologist. UK Biobank defines a potentially serious IF for these purposes as one indicating the possibility of a condition which, if confirmed, would carry a real prospect of seriously threatening life span, or of having a substantial impact on major body functions or quality of life.

The need for evidence to inform IFs policy

Limited data exist on the impact of feedback of IFs on participants and health services[8–11], and on how these vary by different policies for handling IFs. Most published data on opinions of receiving such feedback are based on hypothetical scenarios, rather than studies of research participants who have actually received feedback[12–14]. It is often assumed that early observation on imaging of presumed disease (prior to clinical presentation) is inevitably beneficial, but data on final clinical diagnosis and the impact of feedback of IFs are scarce[15]. Such data would inform debates about these assumptions, and the design of appropriate, acceptable protocols to handle IFs detected in research, public health screening or commercial imaging settings.

In this evaluation of the first 1000 participants in the UK Biobank imaging study, we assessed the number and types of potentially serious IFs detected and their final clinical diagnoses, comparing the UK Biobank imaging IFs protocol with systematic radiologist review of all of the images. We also assessed the impact of providing feedback about potentially serious IFs on participants, their friends, families and health services, with respect to: clinical assessments undertaken; emotional wellbeing, finances, work and daily activities; and participants' and their general practitioners' (GP) opinions about receiving feedback.

Methods

Participants

Existing participants of the UK Biobank cohort study who lived within about 100 miles of UK Biobank's first imaging centre in Stockport were invited to participate in the UK Biobank imaging study. The invitation contained a link to the UK Biobank imaging study website (http://imaging.ukbiobank.ac.uk/), and willing participants were asked to telephone the Participant Recruitment Centre where they could ask questions about the study and answer pre-screening safety questions. Participants were excluded if they had metal inside their body or an implanted

medical device which could create imaging artefacts or pose a risk during MRI, if they were likely to find it difficult to lie still, or if they were unlikely to tolerate the imaging due to known claustrophobia.

Consent

All participants received written information about the imaging study, including details about the UK Biobank imaging IFs protocol, and provided consent before taking part, including consent for UK Biobank to inform them and their GP if a potentially serious IF was identified (Supplementary File 1). We surveyed all participants with a questionnaire two days after their imaging assessment to assess their understanding of the information and consent process (http://www.ukbiobank.ac.uk/resources/).

Imaging

Participants underwent a 30 minute brain MRI (3.0 Tesla Skyra scanner, Siemens, Erlangen, Germany), a 30 minute non-contrast cardiac and body MRI (1.5 Tesla Aera scanner, Siemens, Erlangen, Germany) from neck to knees (Supplementary File 2), and a 15 minute DXA scan (iDXA, General Electric, New York, United States of America) of whole body, lumbar spine and hip, with lateral vertebral fracture assessment. Participants also underwent carotid doppler ultrasound, but this was not considered to have the potential to yield potentially serious IFs (Supplementary File 3). Imaging protocols were optimised for research purposes and did not constitute standard diagnostic examinations.

List of potentially serious IFs

UK Biobank consulted radiologists, reviewed the literature, and considered the German National Cohort's list of imaging IFs[16] to develop a list of IFs considered to be potentially serious, as well as examples of those not considered serious (Supplementary File 3). Both radiographers and reporting radiologists used this list in conjunction with UK Biobank's definition of a potentially serious IF when judging whether any observed IF was potentially serious or not.

Two protocols for handling IFs

Images from the first 1000 participants were assessed using two protocols which ran simultaneously. Under the UK Biobank IFs protocol ('radiographer flagging'), if a radiographer noticed a potentially serious IF during image acquisition and quality assessment, the relevant set of images was flagged for subsequent review by a radiologist. Under 'systematic radiologist review', all images were systematically reviewed by a radiologist. Radiographers were trained in the relevant imaging protocols but did not receive specific training in image interpretation as UK Biobank is a research resource and conducts research imaging. The radiographers were not instructed to actively look for, or to avoid looking for IFs; rather, they were instructed that should they happen to notice a concerning finding, they should flag it for review. As such, UK Biobank does not aim to provide any form of health service, including image interpretation. Radiologists and radiographers were aware of the comparison study, but were blind to each other's opinions. To aid interpretation of images assessed either during systematic radiologist review, or those flagged by radiographers, we provided reporting radiologists with data collected during the imaging visit on the participant's age, sex, body mass index, self-reported smoking status, alcohol consumption, medical history and medications.

Within a few weeks of their imaging visit, we wrote to all participants who had a potentially serious IF reported by a radiologist, whether it had been both flagged by a radiographer and confirmed by a radiologist (radiographer flagging) or detected by a radiologist during systematic review of all images (systematic radiologist review). We explained that a potentially serious abnormality (or, sometimes, abnormalities) had been observed, and advised the participant to visit his/her GP for advice about any further action required (Supplementary File 4). We also wrote to these participants' GPs, providing a copy of the radiologist's report and, if requested, copies of the relevant scans (Supplementary File 5).

Questionnaires to participants with potentially serious IFs and their GPs

We surveyed participants with potentially serious IFs approximately six weeks after writing to them and their GP and approximately six months after their imaging visit to assess the impact of this information. Both participant questionnaires collected data on clinical assessment (blood tests, imaging, specialty referral, changes in medication, invasive procedures and operations), final diagnoses, and opinions on receiving feedback and participating in the imaging study, with additional questions at six months on emotional wellbeing, insurance, finances, work and activities. We also surveyed GPs at six months about clinical assessments, final diagnoses (including copies of any relevant clinical correspondence) and their perceptions of the impact on their patients of receiving feedback (http://www.ukbiobank.ac.uk/resources/). We reconciled multiple responses on similar items from the three questionnaires by prioritising 'yes' responses and included data from coding of free text responses (http://www.ukbiobank.ac.uk/resources/).

Determining final clinical diagnoses

Because there is a paucity of empirical evidence on the natural history and final diagnoses of IFs[15], and no validated risk scores for quantitatively determining the risk to lifespan of particular IFs which are detected on research imaging, our classification of final diagnoses as 'serious' was based on clinical judgement. A consultant physician and an experienced speciality clinical radiology trainee independently classified final diagnoses for each participant who received feedback about a potentially serious IF, by reviewing all available questionnaire data together with additional relevant clinical information from further correspondence or telephone calls with the participant and/or their GP. Working from the definition of a potentially serious IF, we classified final clinical diagnoses as: serious if they were likely to significantly threaten lifespan or have a major impact on quality of life or major body functions; not serious if this was not the case, or if the available data suggested that the diagnosis was already known; and uncertain if there were insufficient data to classify as serious or not. We classified participants with more than one potentially serious IF according to their most serious final diagnosis. Given this inherent subjectivity in the classification of serious final diagnoses, we measured the repeatability of the clinical judgements of final diagnoses severity by calculating

the percentage of participants in whom both classifying doctors agreed on their initial classification. We resolved disagreements through discussion and mutual consensus.

Qualitative study

To provide additional context, UK Biobank commissioned a social research company (TNS-BMRB; www.tns-bmrb.co.uk) to conduct a parallel qualitative study. This aimed: (1) to explore participants' understanding of and opinions about the process of consent relating to feedback of potentially serious IFs through deliberative group discussions with two groups of around 10 participants each (a more and a less affluent group) prior to their imaging assessment; and (2) to assess views on the process and impact of receiving feedback through one-to-one interviews with 15–20 participants (including more and less affluent male and female participants) with IFs on different imaging modalities, and with both serious and non-serious final clinical diagnoses. Further details of the methods of recruitment, interview content and qualitative analysis methods are available at http://www.ukbiobank.ac.uk/resources/.

Statistical analyses

We summarised data from questionnaires as counts and proportions. We compared groups using chi-squared or Fisher's exact tests for proportions and Student's independent t-test for continuous variables. We considered p values of <0.05 to be statistically significant and analysed data using Microsoft Excel 2013 and SPSS Statistics version 21.

Ethics approval

UK Biobank obtained approval specifically for the imaging study, participant information and consent materials and this evaluation, including surveying participants and their GPs (North West Research Ethics Committee, Reference Number: 11/NW/0382).

Results

The first 1000 eligible participants were imaged between April and October 2014. Their mean age was 62 (range 44–77) years, and 524 (52.4%) were female. Each MRI imaging modality was conducted in >94% participants, and DXA in >99% (Figure 1).

Understanding of consent

Around 60% of the first 1000 participants (607/1000) completed the questionnaire assessing understanding of consent. The vast majority correctly understood that they would not receive their scans or results at the end of the imaging visit (540/607, 86.7%) and that they would not be told about any potentially serious IF during the visit (89.0%), but around a quarter incorrectly thought that they could choose whether or not to be informed about any potentially serious IF (158/607, 26.0%) (Supplementary File 6).

Potentially serious IFs

Radiographers flagged 66 potentially serious IFs in 66 (6.6%) participants. Of these, 18 (1.8%) were confirmed as potentially serious by radiologists. Radiologists detected potentially serious IFs in 179 (17.9%) participants (Figure 1), who included the 18 participants with potentially serious IFs flagged by radiographers. Participants with potentially serious IFs were slightly older than those without (mean age 63 versus 61 years, p=0.03), but their sex distribution did not differ significantly (55.3% vs 51.8% female, p=0.4).

Final diagnoses

Data on final diagnoses were available from one or more questionnaires, clinical correspondence and/or telephone contact in 176/179 (98.3%) participants. The two doctors agreed on the per-participant classification of final diagnoses in 172/179 (96.1%) cases. The seven cases of initial disagreement were readily resolved by discussion.

A higher proportion of participants with potentially serious IFs had serious final diagnoses (i.e. true positives) with radiographer flagging (5/18, 27.8%) than with systematic radiologist review (21/179, 11.7%, Figure 1, Table 1). A higher proportion and substantially greater absolute number had non-serious final diagnoses (i.e. false positives) with systematic radiologist review (158/179, 88.3%) than radiographer flagging (13/18, 72.2%). However, radiographer flagging missed 16 of the 21 participants with a serious final diagnosis detected by systematic radiologist review (i.e. false negatives) (Figure 1, Table 1, Supplementary File 7).

The numbers and proportions of participants with potentially serious IFs and with serious versus non-serious final diagnoses varied substantially by imaging modality. Most of the 158 false positives generated by systematic radiologist review were identified on cardiac or body MRI (54 on cardiac and 65 on body [mainly abdominal] MRI; Table 1). Participants with potentially serious IFs from brain and cardiac MRI were more likely to have a serious final diagnosis (around half under radiographer flagging, and 20% under systematic radiologist review) than those with potentially serious IFs from the other imaging modalities (Table 1).

Systematic radiologist review generated 217 potentially serious IFs in 179 participants. More than one potentially serious IF occurred in 33 participants (28 had two and five had three), although no participant had more than one serious final diagnosis. The 21 serious final diagnoses included aortic aneurysms, tumours, structural and functional cardiac disease, and osteoporotic fractures, while non-serious final diagnoses comprised benign lesions, diagnoses already known to the participant and/or their GP, and suspected lesions which were not confirmed. Radiographer flagging detected five of these 21 serious final diagnoses (one arachnoid cyst with hydrocephalus, one meningioma compressing brainstem, and three thoracic aortic aneurysms), and missed 16/21 (two pituitary tumours, two thoracic aortic aneurysms, three lung tumours, two cardiomyopathies, and one each of: atrial fibrillation, coronary heart disease, heart block with left ventricular impairment, abdominal aortic aneurysm, gastrointestinal stromal tumour, pancreatic neuroendocrine tumour, and an osteoporotic crush fracture) (See Supplementary File 7).

Follow-up questionnaires

Each of the three follow-up questionnaires was returned for ≥70% of 179 participants with a potentially serious IF; at least one questionnaire was returned for 93.3% and all three for 45.8%

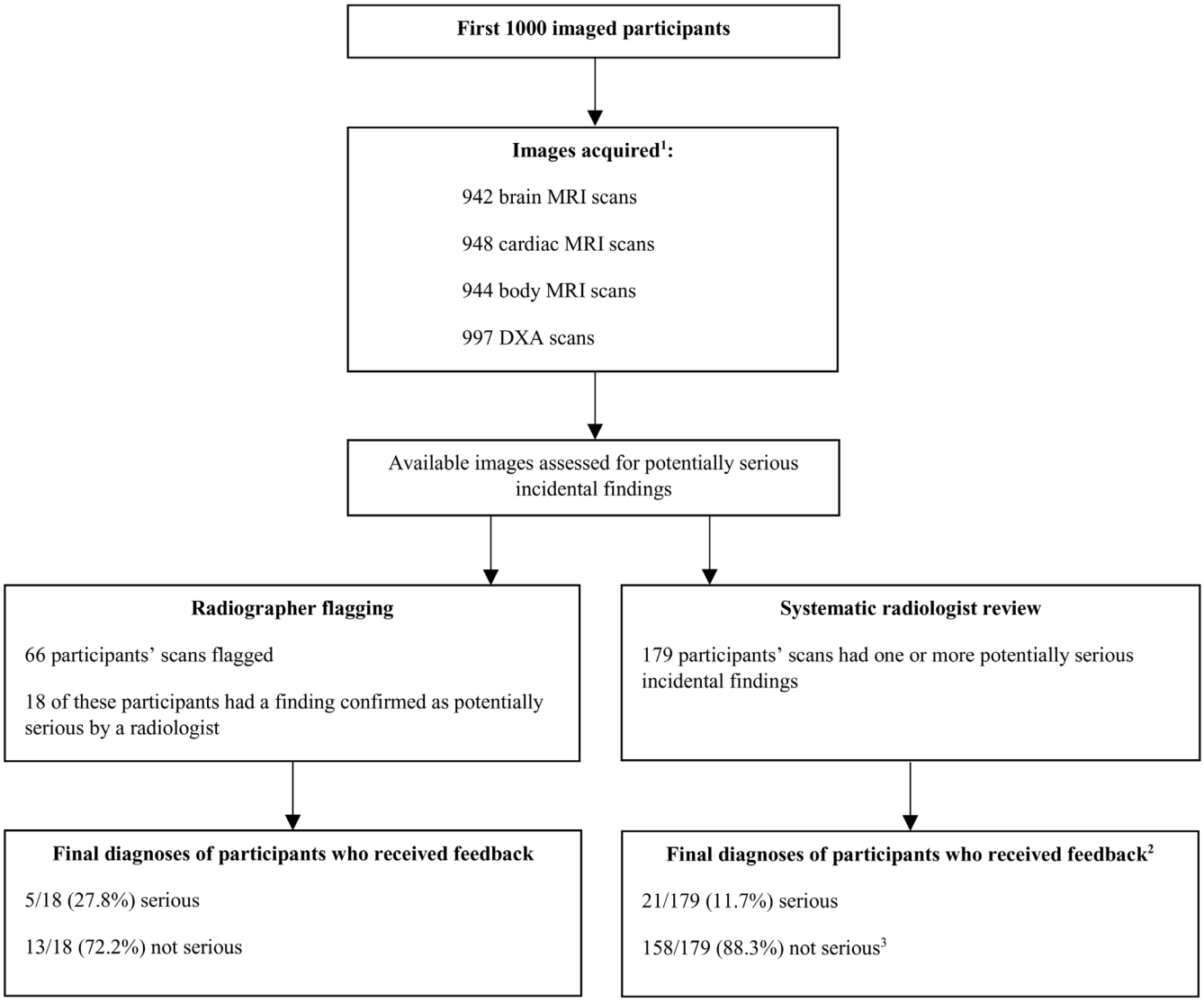


Figure 1. Participant flowchart. MRI = magnetic resonance imaging, DXA = dual energy x-ray absorptiometry. [1]68 participants had incomplete imaging: 18 underwent DXA but not MRI due to safety issues, 50 did not complete all MRI (28 due to claustrophobia, 13 due to scanner failure, nine for other reasons). [2]Final diagnosis assigned to participants with more than one potentially serious incidental finding was the most serious (serious>uncertain>not serious). [3]Three of these participants had uncertain final diagnoses, see Supplementary File 7.

(Table 2). Denominators varied for different types of clinical assessment and impact due to different proportions of completed responses to the relevant questions (Table 3).

Clinical assessment

All participants with follow-up questionnaire data had contacted their GP. Almost all had some form of clinical assessment (153/170 [90.0%]), most frequently blood tests (29.4%), further imaging (78.8%) or specialist referral (64.1%), with smaller proportions having other tests (8.8%), change of medication (10.5%) or an invasive procedure or operation (14.2%) (Table 3). The proportions having each type of clinical assessment were generally higher for those with a serious compared with non-serious final diagnosis, particularly medication changes (44.4% serious versus 6.3% non-serious) and invasive procedures (61.1% versus 8.3%). However, the absolute numbers having clinical assessment were far higher among the many more participants with non-serious final diagnoses. Of the 153 participants reporting some form of clinical assessment, 133 had a non-serious final diagnosis, suggesting that further clinical assessment might not have been necessary (Table 3).

Of particular note, similar absolute numbers of participants had invasive, potentially harmful, procedures irrespective of

Table 1. Clinical seriousness of final diagnoses of 179 participants by detection method and imaging modality.

Method of detection and imaging modality	Clinical seriousness of final diagnosis (n participants)			% of 1000 imaged participants with ≥ 1 PSIF detected	% of participants in whom a PSIF predicted a serious final diagnosis
	Serious	Non-serious[1]	Total		
Radiographer flagging					
Brain MRI	2	2	4	0.4	50.0
Cardiac MRI	3	2	5	0.5	60.0
Body MRI	0	8	8	0.8	0.0
DXA	0	1	1	0.1	0.0
> 1 modality	0	0	0	0.0	0.0
Total (any modality)	5	13	18	1.8	27.8
Systematic radiologist review					
Brain MRI	4	14	18	1.8	22.2
Cardiac MRI	13	54	67	6.7	19.4
Body MRI	3	65	68	6.8	4.4
DXA	1	10	11	1.1	9.1
> 1 modality[2]	0	15	15	1.5	0.0
Total (any modality)	21	158	179	17.9	11.7

PSIF = potentially serious incidental finding, MRI = magnetic resonance imaging, DXA = dual energy X-ray absorptiometry

[1] Includes three participants whose final diagnoses remained uncertain as of April 2016: one participant with a lung nodule was still under assessment; another participant with a lung nodule had been diagnosed with lymphoma, but it remained unclear whether the nodule was related to the lymphoma or not; and we were unable to contact one participant to determine the final diagnosis of DXA appearances suggesting a crush fracture.

[2] Fifteen participants had more than one non-serious final diagnosis arising from more than one modality.

Table 2. Available questionnaires returned by 179 participants and their GPs.

	n participants (%)
Six-week participant questionnaire	132 (74)
Six-month participant questionnaire	125 (70)
Six-month GP questionnaire	125 (70)
At least one questionnaire returned[1]	167 (93)
All three questionnaires returned	82 (46)

[1]At least one of a six-week participant, six-month participant, or six-month GP questionnaire

whether their final diagnosis was considered to be serious (n=11) or non-serious (n=12) (Supplementary File 8). The clinical management of the participants with a serious final diagnosis is summarised in Supplementary File 9.

Impact on participants

Feedback about a potentially serious IF also had an impact (presumed to be adverse) on participants' emotional wellbeing (21/124, 16.9%), insurance or finances (11/124, 8.9%), and work or activities of daily living (7/124, 5.6%). The proportion of participants reporting an impact on emotional wellbeing was higher among those with a serious final diagnosis, but the absolute numbers were higher among those with a non-serious final diagnosis, for whom these impacts could be considered to constitute net harm (Table 3). In addition to the 21 reporting an impact on emotional wellbeing in response to the relevant survey question, participants and/or their GPs spontaneously mentioned worry within questionnaire free-text responses for a further 62 participants (examples shown in Box 1).

Box 1. QUOTATIONS FROM PARTICIPANTS AND THEIR GENERAL PRACTITIONERS (GP)

Participant with a non-serious final diagnosis, six-week questionnaire: "Better to know, but I did feel anxious for a few weeks."

Participant with a serious final diagnosis, six-month questionnaire: "Life has been a physical & emotional roller-coaster since then, both for myself, family & friends. A serious risk of death on the operating table, and considering the consequences for my wife. All-in-all, I feel as if I was mugged by medical technology."

GP of a participant with a non-serious final diagnosis: "[The patient] was asymptomatic. In normal practice no investigation would be performed - this has led to unnecessary anxiety and tests."

GP of a participant with a non-serious final diagnosis: "Concerns over use of health resources regarding this. Using GP and secondary care time with potential [upper gastrointestinal endoscopy] +/- associated risks of this procedure. This for symptoms that the patient is not too concerned with at present."

Table 3. Clinical assessment, impact on participants, and opinions relating to feedback of potentially serious incidental findings[1].

	Clinical seriousness of final diagnosis[2]						p value (Fisher's exact test) for difference between those with a serious versus non-serious final diagnosis
	Serious		Non-serious[3]		Total		
Impact or opinion	**n/N[4]**	**%**	**n/N[4]**	**%**	**n/N[4]**	**%**	
Clinical assessment as reported by participant, GP or both							
Contact between participants and their GP by six months following feedback	20/20	100	146/146	100	166/166	100	-
Blood tests	6/20	30.0	44/150	29.3	50/170	29.4	1.0
Imaging	17/20	85.0	117/150	78.0	134/170	78.8	0.6
Other tests	3/20	15.0	12/150	8.0	15/170	8.8	0.4
Referral	17/20	85.0	92/150	61.3	109/170	64.1	0.05
Change of medication	8/18	44.4	9/144	6.3	17/162	10.5	<0.0001
Invasive procedure or operation	11/18	61.1	12/144	8.3	23/162	14.2	<0.0001
Any of the above clinical assessment	20/20	100	133/150	88.7	153/170	90.0	0.2
Participants' opinions							
Any impact on:							
Emotional wellbeing[5]	7/15	46.7	14/109	12.8	21/124	16.9	0.004
Insurance or finances[6]	5/15	33.3	6/109	5.5	11/124	8.9	0.004
Work or activities of daily living[7]	4/15	26.7	3/109	2.8	7/124	5.6	0.004
How their health compared to before the imaging visit							
Much better or a little better	0/15	0.0	9/109	8.3	9/124	7.3	
The same	9/15	60.0	95/109	87.2	104/124	83.9	0.0007
Much worse or a little worse	6/15	40.0	5/109	4.6	11/124	8.9	
I am glad that:							
UK Biobank told me about a potentially serious incidental finding[8]	16/16	100.0	126/129	97.7	142/145	97.9	1.0000
I took part in the UK Biobank imaging study[9]	17/17	100.0	130/131	99.2	147/148	99.3	1.0000

Wellcome Open Research 2018, 2:114 Last updated: 10 AUG 2018

	Clinical seriousness of final diagnosis[2]						p value (Fisher's exact test) for difference between those with a serious versus non-serious final diagnosis
	Serious		Non-serious[3]		Total		
Impact or opinion	**n/N[4]**	**%**	**n/N[4]**	**%**	**n/N[4]**	**%**	
GPs' opinions							
Impact on the participant's emotional wellbeing:							
Positive or very positive	6/11	54.5	10/88	11.4	16/99	16.2	
No impact	1/11	9.1	44/88	50.0	45/99	45.5	0.001
Negative or very negative	4/11	36.4	34/88	38.6	38/99	38.4	
The net impact to the participant							
Net benefit	10/11	90.9	41/75	54.7	51/86	59.3	0.02
Net harm	1/11	9.1	34/75	45.3	35/86	40.7	
Participant and GP opinions							
"Participants should always be told about potentially serious incidental findings"							
Participants who agreed with this statement[10]	6/17	35.3	49/132	37.1	55/149	36.9	1.0
GPs who agreed with this statement	6/9	66.7	55/85	64.7	61/94	64.9	1.0

-Test not performed

[1] Based on combined responses to relevant questions from the six-week and six-month participant questionnaires, and the six-month GP questionnaire

[2] 217 potentially serious incidental findings (IFs) were detected in 179 participants. For participants with more than one potentially serious IF, clinical severity of the final diagnosis per participant indicates the most severe diagnosis (serious > uncertain > not serious) of their potentially serious IFs

[3] Including three participants for whom the final diagnosis remained uncertain by April 2016 (see Supplementary File 7 for details)

[4] Denominators vary due to differences in questionnaire return rates and whether or not the relevant questions had been answered on returned questionnaires.

[5] Any impact on either the emotional wellbeing of the participant, their friends, or their family, or on family life (combined responses across several related questions)

[6] Any impact on either the cost or ability of obtaining travel, or health or life insurance or on their overall financial situation (combined responses across several related questions)

[7] Any impact on having to take time off work, change job or retire, or have help for activities of daily living (combined responses across several related questions)

[8] This question was asked on both the six-week and the six-month participant questionnaires. 145 participants answered the question at least once and formed the denominator. 98 of 100 who answered the question both times did so consistently (they were glad to have been told on both occasions) and were included in the numerator. Two of these 100 participants (both with final non-serious diagnoses) gave different answers on each questionnaire (one was glad to have been told at six weeks, but by six months would rather not have been told, while the other would rather not have been told at six weeks but was glad to have been told at six months). One further participant (who returned a single six-month participant questionnaire) reported that they would rather not have been told about their potentially serious IF, which was finally diagnosed as a non-serious condition.

[9] This question was asked on both the six-week and the six-month participant questionnaires. 148 participants answered the question at least once and formed the denominator. Answers from the 101 participants who returned both questionnaires and answered the question both times were all consistent

[10] This question was asked on both the six-week and the six-month participant questionnaires. 149 participants answered the question at least once and formed the denominator. 69 of the 105 participants who answered the question both times did so consistently and were included in the numerator. 36 gave different answers on each questionnaire: 22 changed their view from 'should always be told' to 21 'should be able to choose' and one 'no opinion'; 14 changed from 'should be able to choose,' to 11 'should always be told' and three 'other option'.

Wellcome Open Research 2018, 2:114 Last updated: 10 AUG 2018

Most participants receiving feedback reported no change in their health since the imaging visit (104/124, 83.9%). Similar absolute numbers among those with serious versus non-serious final diagnoses had worse health (6/15, 40.0% versus 5/109, 4.6%), while a few of those with a non-serious final diagnosis (but none with a serious final diagnosis) reported better health (9/109, 8.3%, Table 3).

Opinions on receiving feedback

Almost all participants reported being glad to be told about their potentially serious IF (142/145 (97.7%) (Table 3). Nonetheless, GPs who responded reported that a higher proportion of participants had experienced negative versus positive impact on emotional wellbeing (38/99, 38.4% versus 16/99, 16.2%), with most of the negative impact occurring among those with non-serious final diagnoses (Table 3). GPs also spontaneously highlighted concerns about use of health resources to manage asymptomatic people within their free-text questionnaire responses (Box 1). However, the responding GPs believed that a slightly higher proportion of participants had experienced net benefit compared to net harm (51/86, 59.3% versus 35/86, 40.7%).

A higher proportion of responding GPs (61/94, 64.9%) than participants (55/149, 36.9%) thought participants should be always told about a potentially serious IF (Table 3). Since participants were asked both at six weeks and at six months about this, we were able to assess whether the answers of 105 participants who responded on both occasions changed over time. While 69 had consistent responses, 36 changed their views (n=21, Table 3: footnote 10).

Results of the qualitative study

Deliberative group discussions about consent involved a group of 10 ‘more affluent’ participants (Townsend score <-2, four female, mean age 61, SD 9.1 years), and a group of 11 ‘less affluent’ participants (Townsend score >0, six female, mean age 66 years). One-to-one interviews involved an additional 21 participants who received feedback about a potentially serious IF (13 ‘more affluent’, 13 female, mean age 66 years). Analysis of the interview data revealed that participants were motivated to attend the imaging study by altruism, to experience MRI scanning first-hand (in case they needed to attend for investigations for a medical concern later in life), and to receive feedback about potentially serious IFs. Participants could not always recall precise details of the consent process with respect to feedback of IFs, but they were generally unconcerned about this as they trusted UK Biobank to act appropriately. One-to-one interviews further demonstrated that the implications of receiving feedback were not fully understood until after the event, that feedback resulted in short-term anxiety, and that participants tended to assume the worst on receiving feedback; indeed, some were surprised that the final diagnosis might be non-serious, having anticipated a diagnosis of cancer, an aneurysm or a serious heart condition. Further details of the qualitative study results are available at http://www.ukbiobank.ac.uk/resources/.

Discussion

Compared to systematic review of images by radiologists, the UK Biobank IFs protocol (radiographer flagging) resulted in approximately 10-fold fewer participants with non-serious diagnoses (i.e., false positives), but missed 16/21 potentially serious IFs that were diagnosed ultimately as a serious disease (i.e. false negatives).

Extrapolation of our results to the 100,000 participants who will be imaged by UK Biobank over the next few years suggests that systematic radiologist review would generate 15,800 false positives, compared with 1,300 under the UK Biobank IF protocol (radiographer flagging), and would detect serious diagnoses in 2,100 participants compared with 500 under radiographer flagging (Figure 2).

Systematic radiologist review in our study generated a prevalence of potentially serious IFs of 17.9%. The prevalence in other whole-body MRI studies of healthy populations ranged from 12.8% to 57.6%[17–20]. Since those studies used similar MRI sequences applied to similar tissue volumes, variations in prevalence are most likely to have arisen from differences in the definition of IFs, or in the age and other characteristics of the imaged populations.

Almost all participants with potentially serious IFs had subsequent clinical assessment, resulting in large numbers of investigations, referrals and procedures. Many of these were, with hindsight, unnecessary, with risk of direct harm as well as cost implications. Impact on emotional wellbeing, insurance or finances, and on work or daily activities were reported by a higher proportion of participants with serious final diagnoses, but affected a higher absolute number of participants without serious final diagnoses. In keeping with these results, over half of participants in the Study of Health in Pomerania who received feedback of an IF detected on whole-body MRI reported psychological distress[8].

Only around one-third of our participants believed that participants should always be told about potentially serious IFs. Similar proportions of participants with serious and participants with non-serious final diagnoses expressed this opinion. However, almost a quarter of participants changed their opinion over the few months between the six-week and six-month questionnaires on whether participants should or should not be able to choose to receive feedback of an IF (Table 3: footnote 10), illustrating the complexities in interpreting opinions on this issue.

The findings of this study are of practical legal and ethical importance, and can be considered with regards to the duties of care, and the ethical principles of respect for autonomy, and beneficence and non-maleficence toward participants and towards the public. The legal and ethical background to UK Biobank’s approach was developed with input from its Imaging Working Group, its independent Ethics and Governance Council, representatives of its major funders (Wellcome Trust and the Medical Research Council), UK Biobank’s legal counsel and external legal counsel and ethics advice. In brief, it was considered likely that the duty of care owed to participants by radiographers would not be of a clinical standard, but rather what a reasonably competent radiographer conducting research imaging without clinical information could reasonable observe and report. This legal duty of care informs the ethical duties of radiographers, i.e., that they must be capable of meeting

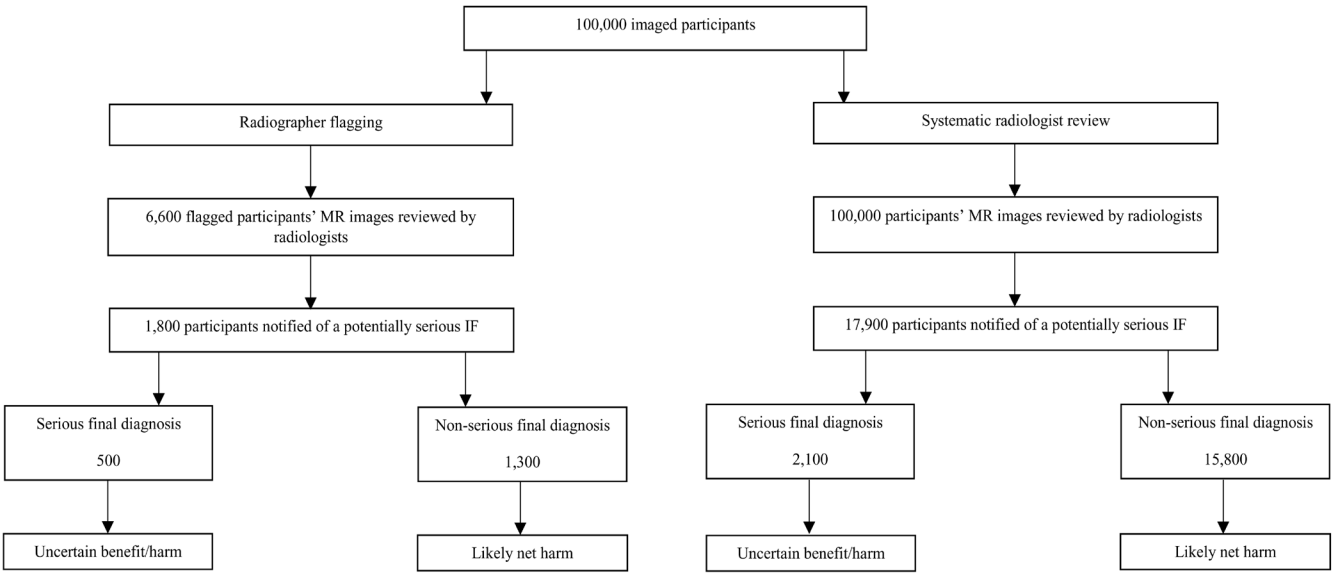


Figure 2. Extrapolation of this study's findings to the 100,000 UK Biobank imaging study participants. MR = magnetic resonance, IF = incidental finding.

the standards of care which are detailed in the consent process. Therefore, in order to respect potential participants' autonomy, it is paramount that UK Biobank have an IFs protocol in place, and that this protocol and its limitations are explained to and understood by participants. Our results reinforce the need for clarity in the information provided to participants about the feedback policy before they consent to imaging research studies. While participants' understanding of what they had consented to was generally good, a substantial minority (around a quarter) incorrectly thought that they could choose whether or not to receive feedback. The information materials for the UK Biobank imaging study now further emphasize the difference between research and clinical diagnostic imaging, that the imaging is not a 'health check,' that not all serious disease will be detected, and that some potentially serious IFs will prove to be non-serious with further investigations (http://imaging.ukbiobank.ac.uk/). Considering the ethical principles of beneficence and non-maleficence toward both participants and the public, our data suggest that feeding back potentially serious IFs which turn out not to be serious (false positives) can make some participants worse off, through exposure to the inconvenience, worry and potential harms of clinical assessments, including invasive procedures. Feedback of false positives also results in wider harm through the unnecessary use of publicly-funded health services. Missing a serious disease (false negative) does not make participants worse off compared to their status before receiving feedback of a potentially serious IF; rather, it fails to make participants better off. While the literature about IFs sometimes argues that feedback is inevitably beneficial[21], the balance of potential benefits and harms of earlier diagnosis (of IFs which are actually serious) is uncertain. It is important to reiterate that UK Biobank is a research resource which aims to facilitate research which will benefit public health, rather than provide any form of health services to individual participants. We therefore conclude that the responsibilities of researchers to avoid unnecessary harm to significant numbers of participants and disruption to publicly-funded health services mean that radiographer flagging (resulting in far fewer false positives while missing a small number of true positives with unclear benefit of earlier diagnosis) constitutes an ethically more justified approach in the UK Biobank imaging study than systematic radiologist review.

Some might argue that concerns about generating false positives suggest the case for a policy of no feedback of any IFs. However, in the light of legal advice regarding the duty of care it owed to participants as described above, UK Biobank decided not to withhold all feedback on potentially serious IFs, but to minimize the generation of false positives by only feeding back potentially serious IFs which are also confirmed by a radiologist. This approach to potentially serious IFs should be seen within the context of large-scale, population based imaging of healthy volunteers; a different approach may well be appropriate for other types of imaging studies, which may be smaller, based in clinical centres, have a different duty of care between research participants and researchers, or include participants with different characteristics (e.g., age) to those in the UK Biobank study.

While our underlying objective was to test the IFs protocol for the UK Biobank imaging study, our findings are of potential relevance in other contexts in which individuals are imaged prior to clinical presentation of disease, including public health and commercial screening. In both situations, it is important to consider the potential benefits of making a true positive diagnosis versus the potential harms to the individual and to publicly-funded health services, of a false positive diagnosis. The significant number of false positives generated by systematic radiologist reporting in

our study implies that imaging of asymptomatic people should not be undertaken without appropriate concern for ensuring that the individuals being imaged do not end up worse off than they started.

Strengths

Our study is the first to systematically follow up all participants receiving feedback about IFs and their GPs, giving the most comprehensive data on the impact of feedback of potentially serious IFs in any research imaging study to date and providing the first quantitative comparison of two different protocols for handling IFs. We have demonstrated for the first time the much lower rates of potentially serious IFs and, most importantly, false positives detected with a protocol in which radiologists report only those images which radiographers flag as having potentially serious IFs. Although the public support the principle of providing feedback of IFs[14], regardless of clinical severity[12], most previous studies did not survey people who had actually received feedback. Our findings are crucial to informing future policy surrounding feedback of IFs in research studies.

Our study was strengthened by good questionnaire response rates and near complete data on final diagnoses due to extensive efforts to gather these directly from participants and their GPs, and data collection at both early and later time periods following feedback. Results related to understanding of consent and impact of feedback on participants were confirmed and contextualised in a parallel, qualitative study.

Limitations

Radiographer flagging rates could, in principle, have been influenced by a relative lack of experience with the first 1000 imaged participants, or by knowledge that radiologists were also reviewing all images. However, ongoing collection of data on potentially serious IFs in the 7000 participants imaged subsequently showed the prevalence of IFs detected by radiographers to be broadly consistent over time with a stable prevalence of potentially serious IFs confirmed by radiologists (mean proportion of 1.7%) (Supplementary File 10).

Although questionnaire response rates by participants were generally high, only around two thirds of participants' GPs responded about participants' emotional well-being and overall net benefit/harm. The design of the questionnaires did not allow for quantification of the use of particular health services or evaluation of the associated costs. However, UK Biobank continues to collect data from participants with potentially serious IFs and their GPs through questionnaires, supplemented by linkages to national health datasets. This will enable further clinical, health economic and policy issues to be addressed using data from larger numbers of imaged participants.

Classification of final diagnoses as serious or not was based on clinical judgement of data available up to around six months following feedback of a potentially serious IF. Final diagnoses classified as serious may not actually shorten life span, or substantially impact on major body functions or quality of life in the 21 participants concerned, who were apparently healthy at the time of their imaging visits. Some potentially serious IFs may take longer than six months to diagnose, or for their full impact to become clear, potentially leading to an incomplete picture of the adverse impacts of feedback.

Conclusions

The handling of potentially serious IFs merits serious consideration by researchers undertaking imaging research studies. Our data provide evidence to inform policy for large-scale research imaging in healthy populations, and are relevant to asymptomatic populations undergoing public health screening and commercial imaging. They demonstrate that systematic radiologist review of all images leads to the diagnosis of previously unknown serious disease in some participants. However, the great majority of these findings turn out not to be serious, resulting in unnecessary anxiety for the participant and unnecessary clinical assessment, which may include invasive procedures, provided by publicly-funded health services. Further, for those participants whose IFs do turn out to be serious, it is often difficult to ascertain whether this knowledge results in clear clinical benefit.

There is no 'one-size-fits-all' approach to handling IFs, as much depends on the purpose of the imaging, be that research, screening, or clinical care. In research studies of healthy volunteers, for whom there is no direct benefit for taking part, it is particularly critical to minimise harm. Based on these results, we suggest that this is achieved in an imaging study of UK Biobank's scale and complexity with a protocol in which radiographers flag suspicious images for reporting by radiologists, rather than systematic review of all images by radiologists.

Data availability

Due to the confidential nature of questionnaire responses and clinical information on participants with potentially serious incidental findings, it is not possible to publicly share all of the data on which our analyses were based, but extensive summaries of all relevant data are included in the supplementary material and within the linked online material.

Importantly, any bona fide researcher can apply to use the UK Biobank resource, with no preferential or exclusive access, for health related research that is in the public interest. Application for access to UK Biobank data involves registration and application via the UK Biobank website, with applications considered by the UK Biobank Access Sub-Committee. Following approval, researchers and their institutions sign a Material Transfer Agreement and pay modest access charges. Further information on applying to access UK Biobank data is available at: http://www.ukbiobank.ac.uk/register-apply/.

Competing interests

LMG: Member of the UK Biobank Imaging Working Group. UK Biobank Imaging Consultant, University of Edinburgh.

TJL: UK Biobank Epidemiologist, University of Oxford.

LA: UK Biobank Data Analyst, University of Oxford.

SG: Member of the UK Biobank Imaging Working Group. Senior Project Manager of UK Biobank.

ND: Senior Clinical Study Administrator of UK Biobank.

JMW: Advised on imaging protocols for the UK Biobank imaging study. Currently analysing UK Biobank brain imaging data.

GM: Member of the UK Biobank International Scientific Advisory Board.

MP: Director of Ethox, University of Oxford. Ethox provides independent advice on ethical matters to UK Biobank.

RB: None currently. Former chair of the UK Biobank Ethics and Governance Council.

PMM: Chair of the UK Biobank Imaging Working Group, Member of UK Biobank Steering Committee.

RC: CEO and Principle Investigator of UK Biobank. Chair of UK Biobank Steering Committee. Member of the UK Biobank Imaging, Enhancements, Follow-up and Outcomes and Infectious Diseases Working Groups.

NEA: Member of UK Biobank Steering Committee, UK Biobank Imaging, Enhancements, Follow-up and Outcomes and Infectious Diseases Working Groups. UK Biobank Senior Epidemiologist.

JS: UK Biobank Legal Counsel.

CLMS: Member of UK Biobank Steering Committee, and UK Biobank Imaging, Enhancements, and Follow-up and Outcomes Working Groups. UK Biobank Chief Scientist.

Grant information

This work was supported by the Wellcome Trust [107190]. UK Biobank is funded by the Wellcome Trust, Medical Research Council, Department of Health, Scottish Government, Northwest Regional Development Agency, Welsh Assembly Government, British Heart Foundation, and Diabetes UK.

The main funders of UK Biobank's imaging study (the Medical Research Council and the Wellcome Trust) and the imaging study's independent scientific review panel encouraged this evaluation.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

Members of the UK Biobank Imaging Working Group:

Imperial College London
Prof Paul Matthews (Edmond and Lily Safra Chair and Head of Brain Sciences)
Dr Tony Goldstone (Senior Clinical Research Fellow)

Newcastle University
Prof Andrew Blamire (Professor of Magnetic Resonance Physics)

Queen Mary University of London
Prof Steffen Petersen (Professor of Cardiovascular Medicine)

University of Edinburgh
Prof Cathie Sudlow (Chair of Neurology and Clinical Epidemiology)
Dr Lorna Gibson (Clinical Research Fellow)

University of Manchester
Prof Alan Jackson (Professor of Radiology)

University of Oxford
Prof Naomi Allen (Associate Professor in Epidemiology)
Prof Sir Rory Collins (BHF Professor of Medicine and Epidemiology)
Prof Paul Leeson (Professor of Cardiovascular Medicine)
Prof Karla Miller (Professor of Biomedical Engineering)
Prof Stefan Neubauer (Professor of Cardiovascular Medicine)
Prof Stephen Smith (Professor of Bioresource Systems)

University of Southampton
Prof Nicholas Harvey (Professor of Rheumatology and Clinical Epidemiology)

University of Westminster
Prof Jimmy Bell (Professor)

Other staff contributing to the UK Biobank Imaging Study:

UK Biobank Imaging Assessment Centre:
Ms Sarah Hudson (Lead Radiographer)
Mr Stephen Hill (Centre Lead Radiographer)
Ms Sam Akram (Superintendent Radiographer)
Mr Phil Bradley (Superintendent Radiographer)
Ms Hayley Daulby (Superintendent Radiographer)
Ms Nadia Damani (Senior Radiographer)
Ms Susan Jeffcock (Senior Radiographer)
Ms Anna Miners (Senior Radiographer)
Ms Philippa Scott (Senior Radiographer)
Ms Michelle Barton (Senior Radiographer)
Ms Karen Tevlin (Senior Radiographer)

Ms Abimbola Akinjiyan (Radiographer)
Mr Christo Andreou (Radiographer)
Ms Barbara Aspinall (Radiographer)
Mr Dan Birchall (Radiographer)
Ms Stacey Carter (Radiographer)
Ms Claire Collard (Radiographer)
Mr Jason Hendricks (Radiographer)
Mr Harris Khan (Radiographer)
Mr Michael Louth (Radiographer)
Ms Florence Maringanise (Radiographer)
Mr Ian Martin (Radiographer)
Ms Esnath Moyo (Radiographer)
Ms Emildah Mugoni (Radiographer)
Ms Zaynab Patel (Radiographer)
Ms Laura Serrant (Radiographer)
Ms Natasha Douglas (Radiographer)
Mr Nathan Cheetham (Radiographer)
Ms Gemma Connell (Radiographer)
Ms Elaine Sully (Radiographer)
Ms Jenny Craven (Radiographer)
Ms Angela Evans (Radiographer)

Mr Iestyn Jones (Clinic Manager)
Ms Caroline Cobb (Nurse/Study Interviewer)
Ms Mary Mitchell (Nurse/Study Interviewer)
Mr Ayaz Pasha (Nurse/Study Interviewer)
Mr Naeem Chowdhury (Senior Phlebotomist)
Mr Fidelis Fru (Healthcare Assistant)
Ms Tracey Gillon (Phlebotomist)
Ms Sarah Olma (Healthcare Assistant)
Ms Natalie Chapman (Healthcare Assistant)
Ms Charlotte Wilson (Receptionist)

UK Biobank Coordinating Centre:
Ms Vikki Arrundale (Project Support Officer)
Ms Kirsty Broadhurst (Imaging Operations Manager)
Ms Nicola Doherty (Senior Clinical Study Administrator)
Ms Brittany Devereux (Office/Project Administrator)
Mr Mark Effingham (Chief Information Officer)
Mr Steve Garratt (Senior Project Manager)
Dr Tim Peakman (Deputy CEO)
Mr Jonathan Sellors (Legal Counsel)
Dr Simon Sheard (Operations Director)

UK Biobank Imaging Consultants:

Basildon University Hospital NHS Trust
Dr Ermanno Capuano (Consultant Radiologist)

Birmingham Children's Hospital
Dr Nicola Crabtree (Principle Clinical Scientist in Bone Densitometry)

Imperial College London
Dr Neil Rane (Consultant Neuroradiologist)

Manchester Royal Infirmary
Dr Stephen Lee (Consultant Radiologist)

University of Edinburgh
Dr Lorna Gibson (Clinical Research Fellow)

University of Manchester
Dr Judith Adams (Consultant and Honorary Professor of Diagnostic Radiology)
Professor Alan Jackson (Professor of Radiology and Honorary Consultant Neuroradiologist)

William Harvey Institute
Dr Alexia Rossi (Post-doctoral Clinical Research Fellow)

Supplementary material

Supplementary File 1: Participant information leaflet. Consent for publication of the potentially identifying images was provided.

Click here to access the data.

Supplementary File 2: UK Biobank magnetic resonance imaging parameters.

Click here to access the data.

Supplementary File 3: UK Biobank lists of incidental findings.

Click here to access the data.

Supplementary File 4: Example feedback letter sent to participants.

Click here to access the data.

Supplementary File 5: Example feedback letter sent to general practitioners.

Click here to access the data.

Supplementary File 6: Understanding of consent related to the feedback of potentially serious incidental findings: Data from 607 respondents from the first 1000 imaged UK Biobank participants.

Click here to access the data.

Supplementary File 7: Final diagnoses of 217 potentially serious incidental findings.

Click here to access the data.

Supplementary File 8: Invasive procedures performed to diagnose or treat potentially serious incidental findings.

Click here to access the data.

Supplementary File 9: Clinical management (medication and procedures) of the 21 participants with potentially serious incidental findings which were finally diagnosed as serious.

Click here to access the data.

Supplementary File 10: Rates of radiographer flagging and rates of radiologist confirmation of potentially serious incidental findings in the first 7000 imaged UK Biobank participants.

Click here to access the data.

References

1. Sudlow C, Gallacher J, Allen N, *et al.*: **UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age.** *PLoS Med.* 2015; **12**(3): e1001779.
PubMed Abstract | Publisher Full Text | Free Full Text
2. Matthews PM, Sudlow C: **The UK Biobank.** *Brain.* 2015; **138**(12): 3463–5.
PubMed Abstract | Publisher Full Text
3. Wolf SM, Lawrenz FP, Nelson CA, *et al.*: **Managing incidental findings in human subjects research: analysis and recommendations.** *J Law Med Ethics.* 2008; **36**(2): 219–48, 211.
PubMed Abstract | Publisher Full Text | Free Full Text
4. Farrar J, Savill J: **UK funders' framework for health-related findings in research.** *Lancet.* 2014; **383**(9933): 1954–5.
PubMed Abstract | Publisher Full Text
5. Illes J, Kirschen MP, Edwards E, *et al.*: **Ethics. Incidental findings in brain imaging research.** *Science.* 2006; **311**(5762): 783–4.
PubMed Abstract | Publisher Full Text | Free Full Text
6. Wardlaw JM, Davies H, Booth TC, *et al.*: **Acting on incidental findings in research imaging.** *BMJ.* 2015; **351**: h5190.
PubMed Abstract | Publisher Full Text
7. Ikram MA, van der Lugt A, Niessen WJ, *et al.*: **The Rotterdam Scan Study: design update 2016 and main findings.** *Eur J Epidemiol.* 2015; **30**(12): 1299–315.
PubMed Abstract | Publisher Full Text | Free Full Text
8. Schmidt CO, Hegenscheid K, Erdmann P, *et al.*: **Psychosocial consequences and severity of disclosed incidental findings from whole-body MRI in a general population study.** *Eur Radiol.* 2013; **23**(5): 1343–51.
PubMed Abstract | Publisher Full Text
9. Sandeman EM, Valdez Hernandez M del C, Morris Z, *et al.*: **Incidental findings on brain MR imaging in older community-dwelling subjects are common but serious medical consequences are rare: a cohort study.** *PLoS One.* 2013; **8**(8): e71467.
PubMed Abstract | Publisher Full Text | Free Full Text
10. de Boer AW, Drewes YM, de Mutsert R, *et al.*: **Incidental findings in research: A focus group study about the perspective of the research participant.** *J Magn Reson Imaging.* 2018; **47**(1): 230–237.
PubMed Abstract | Publisher Full Text
11. Bos D, Poels MM, Adams HH, *et al.*: **Prevalence, Clinical Management, and Natural Course of Incidental Findings on Brain MR Images: The Population-based Rotterdam Scan Study.** *Radiology.* 2016; **281**(2): 507–515.
PubMed Abstract | Publisher Full Text
12. Kirschen MP, Jaworska A, Illes J: **Subjects' expectations in neuroimaging research.** *J Magn Reson Imaging.* 2006; **23**(2): 205–9.
PubMed Abstract | Publisher Full Text | Free Full Text
13. Brown MA, Knight H: **Attitudes towards participating in fMRI studies amongst participants in a birth cohort study.** 2010; Accessed 17/11/17.
Reference Source
14. Wellcome Trust, Medical Research Council, Opinion Leader: **Assessing public attitudes to health related findings in research.** 2012; Accessed 17/11/17.
Reference Source
15. Gibson LM, Sudlow CLM, Wardlaw JM: **Incidental findings: current ethical debates and future challenges in advanced neuroimaging.** In: Illes J, Hossain S, editors. *Neuroethics: anticipating the future.* Second ed. Oxford: Oxford University Press; 2017.
Publisher Full Text
16. Bertheau RC, von Stackelberg O, Weckbach S, *et al.*: **Management of incidental findings in the German National Cohort.** In: Weckbach S, editor. *Radiological incidental findings.* Switzerland: Springer; 2017; 57–70.
Publisher Full Text
17. Hegenscheid K, Seipel R, Schmidt CO, *et al.*: **Potentially relevant incidental findings on research whole-body MRI in the general adult population: frequencies and management.** *Eur Radiol.* 2013; **23**(3): 816–26.
PubMed Abstract | Publisher Full Text
18. Cieszanowski A, Maj E, Kulisiewicz P, *et al.*: **Non-contrast-enhanced whole-body magnetic resonance imaging in the general population: the incidence of abnormal findings in patients 50 years old and younger compared to older subjects.** *PLoS One.* 2014; **9**(9): e107840.
PubMed Abstract | Publisher Full Text | Free Full Text
19. Morin SH, Cobbold JF, Lim AK, *et al.*: **Incidental findings in healthy control research subjects using whole-body MRI.** *Eur J Radiol.* 2009; **72**(3): 529–33.
PubMed Abstract | Publisher Full Text
20. Lo GG, Ai V, Au-Yeung KM, *et al.*: **Magnetic resonance whole body imaging at 3 Tesla: feasibility and findings in a cohort of asymptomatic medical doctors.** *Hong Kong Med J.* 2008; **14**(2): 90–6.
PubMed Abstract
21. Morris Z, Whiteley WN, Longstreth WT Jr, *et al.*: **Incidental findings on brain magnetic resonance imaging: systematic review and meta-analysis.** *BMJ.* 2009; **339**: b3016.
PubMed Abstract | Publisher Full Text | Free Full Text

Open Peer Review

Current Referee Status: ? ✓ ✓

Version 2

Referee Report 27 June 2018

doi:10.21956/wellcomeopenres.14912.r33394

✓ **Eline M Bunnik**
Department of Medical Ethics and Philosophy of Medicine, Erasmus MC, University Medical Centre Rotterdam, Rotterdam, Netherlands

This much-awaited paper reports the experiences of UK Biobank, one of the largest research imaging efforts, with incidental findings. The results are of relevance to other research imaging groups around the world, and makes a valuable empirical contribution to evolving ethics and policy discussions on the management of incidental findings in research imaging contexts. A strength of this paper is that it monitors both the clinical impact and the psychosocial impact of the feedback of incidental findings on research participants. Also, the results of this study have been used to improve the informed consent process of UK Biobank (p. 12). The paper is nicely and clearly structured and comprehensive.
I have three - not too major - concerns with this paper

- The authors claim at several points in the text that e.g. "limited data exist" on the clinical and other implications of learning about incidental findings on research participants and that "data (...) are scarce" but would be much welcomed to inform the debate on appropriate protocols for handling incidental findings (page 3). Thus, the authors seem to suggest that their study is "the first" (page 12) to have looked into these implications empirically. This is not the case. The authors may have overlooked some of the available evidence[1-2] and should either discuss this evidence or rephrase sections of the paper in which they suggest that there is little evidence.

- At times, the ethical argumentation falters a little. For instance, in the introduction the authors state that in research studies, potential harms should be minimised. This is correct, but a reference might clarify the scope and nature of this assumed obligation, as there are many different conceptions and interpretations of this obligation. Also on page 11, references are missing when the authors are discussing the principles of beneficence and non-maleficence and respect for autonomy. On page 12, it is argued that (the many) false positives (associated with systematic radiologist review) will make research participants (and society) worse off through unnecessary follow-up testing, while false negatives do not make participants worse off. I do not agree. False negatives can lead to false reassurance, which may pose health risks. The authors say that the participant information materials now explain more clearly how participation in UK Biobank does not constitute a health check (page 12). However, I am concerned that a subgroup of participants will still believe or expect their images to be reviewed for abnormalities, and will thus run the risk of false reassurance. Also, there is a difficulty that the harms associated with false positives are felt on a societal level (the costs and the efforts involved in (often unnecessary) follow-up), but not on an individual level: 97,7% of participants "reported being glad to be told about their potentially serious" incidental findings (p. 10). Thus, the authors thus slightly downplay the harms associated with false

negatives and highlight the harms associated with false positives. Their conclusion that radiographer flagging is better than systematic radiologist review (with a lower rate of false positives) does not come as a surprise, but may based on a - in my view - slightly skewed weighting of benefits and harms. However, I do agree with the authors that researchers' obligations are mostly to meet the requirements detailed in the informed consent process, and also that there are good pragmatic reasons for UK Biobank to opt for a radiographer flagging policy, and that this is acceptable as long as the consent process is careful and effective in conveying that images are not being checked for abnormalities.

- And a final question: on page 4, it is explained that "radiographers were trained in the relevant imaging protocols but did not receive specific training in image interpretation". In a paper that prof.dr. Meike Vernooij and I wrote some time ago[3], we argued that whether an incidental findings is detected (in this context: whether and what kinds of findings will be flagged) will depend upon various technical, social and organisational factors, including the training, message, or instructions given to the radiographers. For this reason, I am curious to know what was said/what is being said to the radiographers by the project leads (e.g. "If you see something, you should notify X. Do try not to see things. Remember, this is a research study, not a clinical setting. Check the images for quality only, try not to look at any potential abnormalities." or something very different). May be the authors can add one sentence to the section on the two protocols to explain e.g. whether or not radiographers were discouraged from noticing findings or any other relevant variables in the instructions given to radiographers. Providing these details to research participants as part of the consent process could also be a way of conveying to participants that the research imaging does not constitute a health check.

Overall, I support the indexing of this paper.

References

1. Bos D, Poels MM, Adams HH, Akoudad S, Cremers LG, Zonneveld HI, Hoogendam YY, Verhaaren BF, Verlinden VJ, Verbruggen JG, Peymani A, Hofman A, Krestin GP, Vincent AJ, Feelders RA, Koudstaal PJ, van der Lugt A, Ikram MA, Vernooij MW: Prevalence, Clinical Management, and Natural Course of Incidental Findings on Brain MR Images: The Population-based Rotterdam Scan Study.*Radiology*. 2016; **281** (2): 507-515 PubMed Abstract | Publisher Full Text
2. de Boer AW, Drewes YM, de Mutsert R, Numans ME, den Heijer M, Dekkers OM, de Roos A, Lamb HJ, Blom JW, Reis R: Incidental findings in research: A focus group study about the perspective of the research participant.*J Magn Reson Imaging*. 2018; **47** (1): 230-237 PubMed Abstract | Publisher Full Text

3. Bunnik EM, Vernooij MW: Incidental findings in population imaging revisited.*Eur J Epidemiol*. 2016; **31** (1): 1-4 PubMed Abstract | Publisher Full Text

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
I cannot comment. A qualified statistician is required.

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Referee Expertise: Medical Ethics, Bioethics, Research Ethics

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 27 Jul 2018

Lorna Gibson, University of Edinburgh, UK

We thank Dr Bunnik for taking the time to read our manuscript and for providing helpful comments on several aspects of our work.

In particular, thank you for suggesting that we add a reference to the paper De Boer et al.[1] We became aware of this work after our manuscript was sent out for initial peer review; we appreciate the need to update our text, and we have added the reference accordingly. Similarly, we were aware of the work of Bos et al.[2] as we are conducting a systematic review of the prevalence of incidental findings on brain and body imaging.[3] We state in our introduction that 'limited data exist on the impact of feedback of IFs on participants[4] and health services[5]'[6] with references to studies of the psychological[4] and economic impacts,[5] and we agree with Dr Bunnik that a reference to Bos et al. would be suitable here, and have added this to the text. However, despite this additional reference, we do think there remains very limited robust empirical data on the impact of feedback of IFs; while we do not provide a comprehensive review of the published evidence here, we hope to describe this in forthcoming manuscripts.

We appreciate that a large body of literature exists on the obligations of researchers to research participants, and on the ethical principles of beneficence, non-maleficence and autonomy. Following an initial peer review from Professor Bjorn Hoffman, we expanded on the particular ethical and legal background from which the UK Biobank IFs protocol was developed. We agree with Dr Bunnik that a lack of feedback of IFs may be misunderstood by some participants as false reassurance of health, and UK Biobank continue to evaluate participants' understanding of consent. UK Biobank does not use questionnaires of participants and their general practitioners to follow-up participants without potentially serious IFs to determine whether or not these represent 'false negatives.' As such, we feel that we do not downplay the harms of false negatives, but simply do not have the data to comment on these at present. We also agree with Dr Bunnik that the economic impact of false positive IFs constitutes an important harm, and while we do not present data here, it is the subject of a forthcoming manuscript. We hope that the data we do present here will contribute meaningfully to the ongoing discussion of the ethics of feedback of IFs, but feel that a more extensive discussion is beyond the scope of this current work.

We agree with Dr Bunnik that the training and instructions given to the radiographers would potentially impact on the prevalence of IFs.[7] UK Biobank trains radiographers to acquire research imaging data and perform quality checks of the images at the time of the scan. If the radiographers

happen to notice something on the scan that they think could be potentially serious (either a finding listed in Supplementary File 3, or a finding that meets the UK Biobank definition of potentially serious), then they are instructed to flag the images for review by a radiologist. The radiographers are not instructed to actively look for, or to avoid looking for IFs, rather, they are instructed that if they happen to notice a concerning finding, they should flag it for review.

We thank Dr Bunnik again for her review, and for stimulating an interesting discussion of aspects of our work.

1. de Boer, A.W., et al., *Incidental findings in research: a focus group study about the perspective of the research participant.* J Magn Reson Imaging, 2018. **47**(1): p. 230-237.
2. Bos, D., et al., *Prevalence, clinical management, and natural course of incidental findings on brain MR images: the population-based Rotterdam Scan Study.* Radiology, 2016. **281**(2):507-515.
3. Gibson, L., et al., *Potentially serious incidental findings on brain and body magnetic resonance imaging conducted among apparently healthy adults: a systematic review*, in *PROSPERO*. 2016. CRD42016029472.
4. Schmidt, C.O., et al., *Psychosocial consequences and severity of disclosed incidental findings from whole-body MRI in a general population study.* Eur Radiol, 2013. **23**(5):1343-51.
5. Sandeman, E.M., et al., *Incidental findings on brain MR imaging in older community-dwelling subjects are common but serious medical consequences are rare: a cohort study.* PLOS ONE, 2013. **8**(8):e71467.
6. Gibson, L.M., et al., *Impact of detecting potentially serious incidental findings during multi-modal imaging [version 2; referees: 2 approved, 1 approved with reservations].* Wellcome Open Res, 2018. **2**:114.
7. Bunnik, E.M. and M.W. Vernooij, *Incidental findings in population imaging revisited.* Eur J Epidemiol, 2016. **31**(1): p. 1-4.

Competing Interests: This response was submitted by Dr Lorna M. Gibson on behalf of all of the authors. LMG competing interests: Member of the UK Biobank Imaging Working Group. UK Biobank Imaging Consultant, University of Edinburgh

Referee Report 08 May 2018

doi:10.21956/wellcomeopenres.14912.r33052

Saurabh Jha
Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Referee Report 08 January 2018

doi:10.21956/wellcomeopenres.14300.r28579

Saurabh Jha
Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

The paper is unique in that it quantifies the trade-offs between radiologists screening for incidental findings versus radiographers. The findings are not surprising - radiologists detect more true positives but also more false positives. The scale of the difference is surprising. The analysis is granular and the discussion is robust. The authors have anticipated many criticisms, and preemptively addressed them.

The paper would be strengthened by three additions:

1. A comparison of the operating characteristics of radiologists and radiographers graphically.
2. A tabulation of the serious incidental findings picked up by both groups. In particular, a clearer explanation of what the radiographers missed.
3. A brief explanation of how they concluded that letting radiographers screen leads to less net harms – I get it, intuitively, but many might be tempted to argue, and since this is a key point, how the authors arrived at this conclusion should be better explained. An economic model isn't needed, but expansion of some examples would help.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 26 Apr 2018

Lorna Gibson, University of Edinburgh, UK

We would like to thank Dr Jha for his comments.

Dr Jha suggested that we provide a comparison of the operating characteristics of radiologists and radiographers as a figure and we wondered if perhaps Dr Jha would like us to provide a figure showing a receiver operator characteristics curve? If so, we have deliberately chosen not to display such a figure, as it may give the misleading impression that systematic radiologist review of research images is a 'gold standard' protocol to which radiographers are being compared. Our article does not attempt to define one protocol as the 'gold standard' or 'best' protocol. Instead, we feel that there is no single 'best' protocol for handling PSIFs, rather, there will be more, and less, appropriate protocols depending on the imaging context. Our article therefore focuses on describing and weighing up the impacts, benefits and harms of each protocol in order to determine which is most appropriate to apply within the specific research context of the UK Biobank imaging study of 100,000 largely asymptomatic participants. We apologise if we have misunderstood Dr Jha's comment, and we would be more than happy to readdress this point if so.

The serious final diagnoses detected under each protocol are tabulated in Supplementary File 7. In brief, systematic radiologist review resulted in 21 serious final diagnoses. Radiographer flagging detected five of these 21 serious final diagnoses (one arachnoid cyst with hydrocephalus, one meningioma compressing brainstem, and three thoracic aortic aneurysms), and missed 16/21 (two pituitary tumours, two thoracic aortic aneurysms, three lung tumours, two cardiomyopathies, and one each of: atrial fibrillation, coronary heart disease, heart block with left ventricular impairment, abdominal aortic aneurysm, gastrointestinal stromal tumour, pancreatic neuroendocrine tumour, and an osteoporotic crush fracture). We have added this text to our results section.

Dr Jha asked for further explanation about how we concluded that radiographer flagging resulted in less net harm compared to systematic radiologist review of all images. We elaborate on this in our response to a related comment made by Professor Hofmann, and we hope that our approach addresses Dr Jha's comments.

Competing Interests: This response was submitted by Dr Lorna M. Gibson on behalf of all of the authors. LMG competing interests: Member of the UK Biobank Imaging Working Group. UK Biobank Imaging Consultant, University of Edinburgh.

Referee Report 27 December 2017

doi:10.21956/wellcomeopenres.14300.r28578

? **Bjorn Hofmann** [1,2]
[1] Norwegian University of Science and Technology (NTNU), Gjøvik, Norway
[2] University of Oslo, Oslo, Norway

Impact of detecting potentially serious incidental findings during multi-modal imaging
WellcomeOpenResearch.

This study investigates radiographer 'flagging' with radiologist confirmation of potentially serious incidental findings (IFs) compared with systematic radiologist review of images of brain, cardiac and body

magnetic resonance, and dual-energy x-ray absorptiometry scans from the first 1000 imaged UK Biobank participants. The study assessed the number and types of potentially serious IFs detected and their final clinical diagnoses. The study also includes a qualitative assessment of participants experience and understanding of participation and findings.

The study finds that radiographer flagging missed some serious diagnoses, but avoided adverse impacts for many participants with non-serious diagnoses, compared to systematic radiologist review. This makes the authors conclude that UK Biobank's responsibility to avoid both unnecessary harm to larger numbers of participants and burdening of publicly-funded health services suggests that radiographer flagging is a justifiable approach in the UK Biobank imaging study.

The study appears well conducted and is well reported. Figures and tables are informative and the manuscript is well structured. The findings are interesting and make new contributions to the field. This is a valuable study – also beyond the UK Biobank imaging study. In particular, data on final clinical diagnosis and the impact of feedback of IFs are scarce. The study is distinctive in assessing the number and types of potentially serious IFs detected and their final clinical diagnoses. It is also quite unique in investigating the impact of providing feedback about potentially serious IFs on participants, their friends, families and health services, with respect to factors such as: clinical assessments undertaken; emotional wellbeing, finances, work and daily activities; and participants' and their general practitioners' opinions about receiving feedback.

I have some detailed remarks, which hopefully can be helpful to the authors in improving the manuscript even further.

The study used a list of potentially serious Ifs (presented in a supplementary file), however, they do not discuss the inclusion criteria for this list. For instance, which criteria exist for severity, accuracy, and actionability for the various conditions? How does this relate to feedback of Ifs in other fields, e.g., ACMG's recommendations from 2013?

The reader may also want a discussion on why radiographers "did not receive specific training in image interpretation," and whether such training would alter the outcomes. Some indications are given (from the group's experience beyond the first 1000), but competency gained from formal directed training may be different from practical experience (based on volume).
From the text one may infer that radiologists in both groups had access to data collected during the imaging visit (on the participant's age, sex, body mass index, self-reported smoking status, alcohol consumption, medical history and medications), but this is not completely clear. This can easily be made explicit.

The authors classified the final clinical diagnoses as serious if the findings were likely to significantly threaten lifespan or have a major impact on quality of life or major body functions of the research participants. It is unclear how "significantly threaten" is interpreted. Is it a risk score? How does it balance the severity of the event and its probability?

The authors' claim that it is "often assumed that early observation on imaging of presumed disease (prior to clinical presentation) is inevitably beneficial" has recently been confirmed in a systematic review of the literature[1].

It is not quite clear what is meant by: "We reconciled multiple responses on similar items from the three questionnaires by prioritising 'yes' responses and included data from coding of free text responses."

Careful reading explains this, but the authors may want to help the reader here.

The ethical assessment is limited to the principle of non-maleficence. One could argue that this is a surprisingly narrow ethical analysis. Other issues, such as in professional ethics (with basis in deontological or virtue ethics), could easily be argued to be relevant as well. Moreover, the authors briefly mention the alternative of not returning information on IFs as part of this type of research (the UK Biobank imaging study), but refer to “legal advice” and “duty of care” to conclude that return of IFs are warranted. Given their findings (e.g., on lack of benefit and altruistic motivations), this conclusion may not be as obvious as the authors think and may need more elaboration to convince readers who are not part of or familiar to the project. Moreover, it also justifies some reflection on the relationship between legal and moral considerations. It is clear that this is not an article on the ethics of IFs, but when addressing ethical issues, which I strongly endorse, this should be done somewhat more elaborate.

In general this is a well-planned and well-conducted study with interesting results that will make a nice reference in the field.

References

1. Hofmann B, Skolbekken J: Surge in publications on early detection. *BMJ*. 2017. Publisher Full Text

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
I cannot comment. A qualified statistician is required.

Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 26 Apr 2018

Lorna Gibson, University of Edinburgh, UK

We would like to thank Professor Hofmann for his comments.

With regards to the UK Biobank lists of incidental findings (IFs) provided in Supplementary File 3,[1] Professor Hofmann asked us to clarify criteria used to select IFs for this list, such as severity, accuracy and actionability, and how the list relates to feedback of IFs in other fields, such as the American College of Medical Genetics (ACMG) recommendations from 2013.

The lists of IFs deemed potentially serious (i.e. for feedback), and those deemed non-serious were developed after discussion with radiologists, other relevant imaging reporting specialists, radiographers, members of UK Biobank's Imaging Working Group, and with reference to work conducted by the German National Cohort (GNC) study. The GNC lists were developed specifically for the GNC imaging study, after review of the literature and discussion of best practice by radiologists familiar with the GNC research imaging sequences, and GNC ethical framework which aimed to feedback relevant findings, and not feedback irrelevant findings.[2]

At the time of the development of the lists of IFs, there were limited empirical data available on the prevalence and types of IFs that could be expected on the types of imaging to be conducted by UK Biobank. Furthermore, the available studies differed in their definitions of IFs, some, but not all, of which included concepts such as severity and actionability within their definitions. Therefore, to further inform on the prevalence and types of IFs which may be expected on imaging conducted by UK Biobank, we conducted a systematic review of potentially serious incidental findings (PSIFs, as per the UK Biobank definition) on brain and body magnetic resonance imaging. We will report this work within a separate manuscript.

An ACMG working group generated a list of genetic mutations and recommended that these are sought out and reported when a laboratory performs any clinical exome or genome sequencing.[3] In contrast, the UK Biobank lists of IFs are certainly not used as checklists to purposefully seek out, or exclude, specific types of IFs by either the radiographers or radiologists. Rather, when a radiographer happens to see something abnormal on a scan, during image acquisition or quality assurance checks, or when a radiologist is reviewing a flagged image, they can refer to the lists in conjunction with UK Biobank's definition of a potentially serious IF when judging whether any observed IF was potentially serious (i.e. for feedback to participants and their general practitioners [GPs]) or not.

To address Professor Hofmann's comment that, 'from the text one may infer that radiologists in both groups had access to data collected during the imaging visit (on the participant's age, sex, body mass index, self-reported smoking status, alcohol consumption, medical history and medications), but this is not completely clear,' we have amended the relevant text to improve clarity.

We classified final clinical diagnoses as serious if the findings were likely to significantly threaten lifespan or have a major impact on quality of life or major body functions of the research participants. Professor Hofmann asked how we interpret the term "significantly threaten." There is a paucity of empirical evidence on the natural history and final diagnoses of IFs,[4] and to our knowledge there are no validated risk scores for quantitatively determining the risk to lifespan of particular IFs which are detected on research imaging. Our classification of final diagnoses as 'serious' is, as we mention in the limitations subsection of the discussion, a matter of clinical judgement. We also write that, as such, "'serious' final diagnoses may not actually shorten life span, or substantially impact on major body functions or quality of life in the 21 participants concerned, who were apparently healthy at the time of their imaging visits."[1] Given this inherent subjectivity in the classification of serious final diagnoses, we measured the repeatability of the

clinical judgements of final diagnoses severity, and demonstrated a very good level of agreement. Independently, a consultant physician and an experienced specialty clinical radiology trainee classified final diagnoses, and we report in our results section that these two doctors agreed in 172/179 (96.1%) cases, with the remaining seven cases easily resolved by discussion.[1]

We state that "it is often assumed that early observation on imaging of presumed disease (prior to clinical presentation) is inevitably beneficial, but data on final clinical diagnosis and the impact of feedback of IFs are scarce." Professor Hofmann kindly directed us toward articles describing a surge in publications on early detection of disease,[5] and a systematic review which demonstrates that some common screening tests are not associated with a reduction in either disease-specific or all-cause mortality.[6] However, we have chosen not to add these references to the article for three reasons. Firstly, we wish to separate PSIFs (and IFs more generally) from the concept of early detection of disease, as our data demonstrate that the vast majority of PSIFs will not be finally diagnosed as a serious conditions, i.e. the majority do not represent early detection of disease. Secondly, we wish to keep separate the concepts of screening programs from protocols for handling IFs detected during research imaging; whilst data on the benefits and harms of screening programs may be generalizable to the context of PSIFs, screening purposefully for a particular disease using a validated test is a different context to the non-optimized demonstration of an abnormality (which may or may not represent disease) on research imaging, although we accept that the populations undergoing screening and population-based imaging research (i.e. asymptomatic people) are similar. Finally, whilst a discussion of our results in the context of screening, early detection of disease, and overdiagnosis is of great interest to us, as researchers and clinicians, we wish to keep this article focused in its scope.

With regards to our methods section, Professor Hofmann commented that, 'it is not quite clear what is meant by: "we reconciled multiple responses on similar items from the three questionnaires by prioritising 'yes' responses and included data from coding of free text responses." We would like to clarify this with examples. The two questionnaires sent to participants, and the questionnaire sent to their general practitioners (available online at: http://www.ukbiobank.ac.uk/resources/) all asked whether or not the participants had been referred to a specialist. The participant questionnaires have tick-box response options of 'yes,' 'no' and 'don't know.' The GP questionnaire is different, and asks the GP to tick a box if that action has been taken (i.e. no tick may represent 'no' or 'don't know'). In addition, there are multiple free text fields available on all three questionnaires. Therefore, multiple responses may be available about specialist referrals, depending on the return of the questionnaires, and the completion of tick boxes and free text spaces. We therefore had to reconcile these multiple responses, and decided to prioritise 'yes responses,' in order to generate a maximum count. For example, if a participant responded that they did not know if they had been referred, but their GP ticked that they had been referred, we prioritized the 'yes' response of the GP, and coded the participant as being referred to a specialist. Similarly, if a participant indicated on their six-month questionnaire that 'no,' they had not been referred to a specialist, but had previously indicated 'yes,' they had been referred on their six-week questionnaire (either by ticking the box, or mentioning a specialist appointment in free text, or both), we coded the participant has having been referred to a specialist. This methodology maximizes the counts of types of follow-up and impacts and makes use of the maximum amount of data available. We have added a description of this methodology to the end of the document containing the questionnaires, hosted at http://www.ukbiobank.ac.uk/resources/, and we have added this link to the appropriate text in the methods section.

Professor Hofmann asked why radiographers did not receive specific training in image

interpretation. The UK Biobank is a research resource, and as such, is not aiming to provide any form of individual health service, including image interpretation. Accurate image interpretation, even by radiologists, is difficult in any case within the context of UK Biobank, given the lack of clinical information on current symptoms or signs, and the non-diagnostic nature of the research imaging. This is clearly evident from our results: the vast majority of PSIFs detected by radiologists are finally diagnosed as non-serious disease.[1] Within their typical roles in health services, radiographers are not trained to provide interpretation of cross-sectional imaging. Given the limitations of the research imaging, the difficulties in interpreting it (even by radiologists) and the typical role of the radiographers, rather than training radiographers to interpret multiple modalities of non-diagnostic cross-sectional imaging without any clinical information, UK Biobank opted instead to manage participants expectations of what could reasonably be expected. To this end, our consent materials state that the imaging is not a 'health check,' and lack of feedback does not constitute an 'all clear,' and we continue to evaluate participants' understanding of consent with regards to feedback of PSIFs.

Professor Hofmann stated that the ethical issues described in our article require some elaboration, including 1) some reflection on the relationship between legal and ethical considerations, 2) further explanation of our how we concluded that the return of IFs are warranted after considering "legal advice" and "duty of care," and 3) that principles other than non-maleficence, such as professional ethics, are relevant to our conclusions that radiographer flagging is the more appropriate IFs protocol in the context of the UK Biobank. Similarly, Dr Jha asked for further explanation about how we concluded that radiographer flagging resulted in less net harm compared to systematic radiologist review of all images.[7]

We thank Professor Hofmann and Dr Jha for these comments, and agree with Professor Hofmann's further statement that while this article is not focused the ethics of IFs, these issues do need to be addressed and elaborated upon. UK Biobank have carefully considered the legal and ethical background with regards to feedback of PSIFs, and with input from its Imaging Working Group, its independent Ethics and Governance Council, representatives of its major funders (Wellcome Trust and the Medical Research Council), legal input from UK Biobank's legal counsel and from external legal counsel, and ethics advice from Professor Michael Parker of Ethox (who is also a co-author on this manuscript). Following the evaluation study, UK Biobank summarised the data on PSIFs and provided a detailed and lengthy interpretation of the results in the context of both the legal and ethical backgrounds in reports to their funders. Therefore, for readers' convenience, we have summarized the key points of these reports by adding concise sentences to the discussion text to further describe the legal advice, duty of care, the relationship between the legal and ethical considerations, and justification for our conclusions. We hope that this approach addresses both Professor Hofmann's and Dr Jha's comments.

1. Gibson LM, Littlejohns TJ, Adamska L, et al. Impact of detecting potentially serious incidental findings during multi-modal imaging [version 1; referees: 1 approved, 1 approved with reservations]. *Wellcome Open Res.* 2017;2:114.
2. Bertheau RC, von Stackelberg O, Weckbach S, Kauczor HU, Schlett CL. Management of incidental findings in the German National Cohort. In: Weckbach S, ed. *Radiological incidental findings*. Switzerland: Springer; 2017:57-70.
3. Green RC, Berg JS, Grody WW, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. 2013;

http://www.acmg.net/docs/ACMG_Releases_Highly-Anticipated_Recommendations_on_Incidental_Fi
. Accessed 15th January 2018.
4. Gibson LM, Sudlow CLM, Wardlaw JM. Incidental findings: current ethical debates and future challenges in advanced neuroimaging. In: Illes J, Hossain S, eds. *Neuroethics: anticipating the future*. Second ed. Oxford: Oxford University Press; 2017.
5. Hofmann B, Skolbekken JA. Surge in publications on early detection. *BMJ.* 2017;357:j2102.
6. Saquib N, Saquib J, Ioannidis JP. Does screening for disease save lives in asymptomatic adults? Systematic review of meta-analyses and randomized trials. *Int J Epidemiol.* 2015;44(1):264-277.
7. Jha S. Referee report for: impact of detecting potentially serious incidental findings during multi-modal imaging [version 1; referees: 1 approved, 1 approved with reservations]. *Wellcome Open Res.* 2017;2:114.

Competing Interests: This response was submitted by Dr Lorna M. Gibson on behalf of all of the authors. LMG competing interests: Member of the UK Biobank Imaging Working Group. UK Biobank Imaging Consultant, University of Edinburgh.

4.3 Supplementary materials

Supplementary File 1: Participant information leaflet

This is the version of the UK Biobank imaging study Participant Information Leaflet sent to the first 1000 imaged participants, and is no longer in use. It is available here for reference only.

The current version of the UK Biobank imaging study Participant information leaflet is available at http://imaging.ukbiobank.ac.uk/ under 'Further documents.'

Improving the health of future generations

INFORMATION LEAFLET

Imaging Assessment Visit

UK Biobank is inviting you to take part in an important new study to help research. It involves taking pictures (scans) of your brain, heart, tissue and bones, so that scientists can study your internal organs in detail. This will help research into a wide range of diseases, including cancer, heart disease, dementia, diabetes, stroke and arthritis.

We aim to scan up to 100,000 people over the next few years. The scans, especially when combined with other health information you have provided to us, will create a health resource of global significance for many years to come. All of the information provided by you is stored in a confidential and secure manner. None of the data, samples and images provided to researchers will include personal identifying details.

You have been invited because you live within a reasonable travelling distance from the imaging assessment centre in Stockport. This invitation is not based on any other information that we have collected about you, either at your initial assessment or afterwards.

Taking part is entirely voluntary. Please take the time to read this leaflet carefully. It explains why we are asking you to help and what it would involve.

If anything is not clear, or if you would like more information, please telephone **0800-0-276-276 (free from most land lines) or 0292-0-765-597**, Monday-Saturday 8am-7pm, or email us at ukbiobank@ukbiobank.ac.uk

More information about UK Biobank is available at www.ukbiobank.ac.uk. There will also be further opportunities to ask questions when you arrive at the imaging assessment centre.

Thank you for your continued support of UK Biobank.

Improving the health of future generations

What is UK Biobank?

UK Biobank is a large, publicly-funded resource to help scientists from around the world to improve the prevention, diagnosis and treatment of a wide range of illnesses (such as cancer, heart disease, diabetes, dementia, and joint problems). Its goal is to improve the health of future generations.

UK Biobank allows scientists to study how health is affected by lifestyle, environment and genes. By studying the answers, measurements and samples collected from the 500,000 participants, researchers will be able to work out why some people develop particular diseases while others do not.

This research will help to find new ways to prevent premature death and disability. While taking part in UK Biobank is not intended to help participants directly, it should give future generations a much better chance of living their lives free of diseases that disable and kill.

Why does UK Biobank want to scan me?

Taking pictures of organs inside the body (such as the brain and heart) as well as the surrounding tissues and bones will allow scientists to study how the structure and function of the body's organs are related to the development of disease. The combination of these pictures with other information already collected about you will provide a substantial amount of new and important data for health research on a wide range of diseases.

Why have I been invited?

Participants who live within a reasonable distance from the imaging assessment centre are being invited to take part. The centre is at the UK Biobank's Co-ordinating Centre in Stockport. The cost of setting up such a centre means it is not possible to open one in every place in which UK Biobank undertook its original assessments. This means that some people may have to travel further to participate, although we will cover your travel expenses.

Invitations are not based on other information that has already been collected about you.

Improving the health of future generations

What scans will be performed and why?

Magnetic resonance imaging (MRI): This type of scan uses painless magnetic waves to take detailed pictures of the inside of the body (such as organs, tissues and bones). We would like to take two scans: one of the brain and the other of the heart and of the body (mainly covering the abdomen). The scanners are similar to those used in the NHS, except for being a little wider so that people are as comfortable as possible.

- ***Brain MRI scan.*** This will provide information about the structure and function of the brain. It will enable us to obtain information on, for example, which parts of the brain are important for carrying out certain tasks and how different parts of the brain are connected.

- ***Heart and body MRI scan.*** This will provide information on the size of the heart chambers and blood vessels, and changes in heart size as it beats. It will also provide detailed information on the amount and distribution of fat in the body.

Neck artery ultrasound scan. This scan uses ultrasound (high-frequency sound waves) to produce pictures of the blood vessels on either side of the neck. They will help scientists study the build-up of fatty substances (like cholesterol) in these major blood vessels.

Dual-Energy X-ray Absorptiometry (DXA) scan. This scan uses low-energy x-rays to provide a precise measure of bone density throughout the body. Detailed pictures of the spine, hips and knees will help scientists studying diseases like arthritis.

Am I eligible to take part?

All of these scans are safe and painless, and are similar to those used routinely in the NHS. However, since MRI scans involve the use of a magnet, you will not be able to take part if you have any metal or electrical implant, or if you have had an accident where metal may have entered your body.

If you have had recent surgery, you will be able to take part but your visit will not be able to occur until at least six weeks after your operation. You will also not be able to take part if you have medical problems that make it difficult to conduct the scans (e.g., severe hearing or breathing problems, tremors, etc.).

Do I have to take part in this imaging assessment?

No; attendance is entirely voluntary. We do understand that you may not have time or be able to help on this occasion.

Improving the health of future generations

What should I do if I am interested in attending?

Please let us know as soon as possible if you are willing to attend by telephoning us on:

0800-0-276-276 (free from most land lines) or 0292-0-765-597,
Monday to Saturday, 8am - 7pm

During this call, you will be asked a series of questions to find out whether you are eligible to help. Please visit our website (www.ukbiobank.ac.uk) to see the questions that will be asked, but please do not worry if there are technical terms used that you do not understand as these will be explained to you during the call. If you are willing to participate, you will also be able to ask questions about the assessment visit.

If you are eligible, you will be able to arrange an appointment during this call. Appointments are generally available from 8am to 6pm seven days a week to help find a convenient time for you.

We will then send you a letter confirming your appointment, along with detailed directions to the assessment centre. We will also send you an email or text reminder a couple of days before your planned visit.

Can I claim travel expenses for attending?

Yes, please do claim back any reasonable travel expenses (including standard train and bus fares, and mileage for car, motor cycle or bicycle journeys). A claim form will be handed to you at the end of your visit. It would help us if you attached your travel receipts.

There is ample free parking space at the centre. It is also within easy access of Stockport train station, where taxis (paid for by UK Biobank) are available. If you are registered as disabled, you can also claim travel expenses for a companion.

How do I prepare for the imaging assessment visit?

When you come for your appointment, please:

- Bring with you the confirmation letter and travel directions, so you have no difficulty in finding us.
- Bring any reading glasses that you use since you have to be able to read clearly from a computer screen (as at your original assessment visit).
- Bring along your glasses prescription (if you have it), since we will need to provide you with specially adapted glasses that match your usual prescription for the brain MRI scan.
- Bring details of the name and address of your doctor (GP).
- Remove any jewellery that can be easily removed (although we will secure with removable tape any jewellery, such as wedding rings, that cannot be removed). Please ensure that your underpants do not contain metal fastenings.

- Remove any hair grips and makeup as these may contain metallic fragments which could cause a heating sensation during the MRI scans.

- Remove any skin patches (often used for hormone-replacement therapy [HRT], nicotine replacement, pain relief, contraception, glyceryl trinitrate [GTN], etc.) as these may also contain metallic components. We recommend that you bring a spare patch with you to put on after the scans.

- Be prepared to spend about 4 hours at the assessment centre. Refreshments (such as salads, sandwiches, tea, coffee and soft drinks) will be provided during the visit. (There is no need to avoid eating before your visit.)

What happens DURING the imaging assessment visit?

The assessment will take about 4 hours. It will involve the following (in this order):

Initial steps

- A trained member of staff will ask the same questions you answered when you made the appointment. This is to double check that you are able to undergo all of the different types of scan. (If you cannot have some particular scans, then you would still be able to take part in the rest of the visit).

- You can ask any questions that you might have, and will then be asked to sign a consent form. This tells us that you agree to be scanned, and that you understand the process and the implications.

- You will be shown to a private cubicle where you will be given special, loose-fitting clothes to change in to. You will not need to remove underpants, but women will be asked to remove their bras, since they may contain metal.

- You will be asked to leave any loose metal objects (such as money, credit cards, keys, pens, mobile phone, jewellery, watches, hair pins, metal dentures, hearing aids and spectacles), as well as any skin patches (e.g., nicotine or other replacement therapy), in one of the secure lockers.

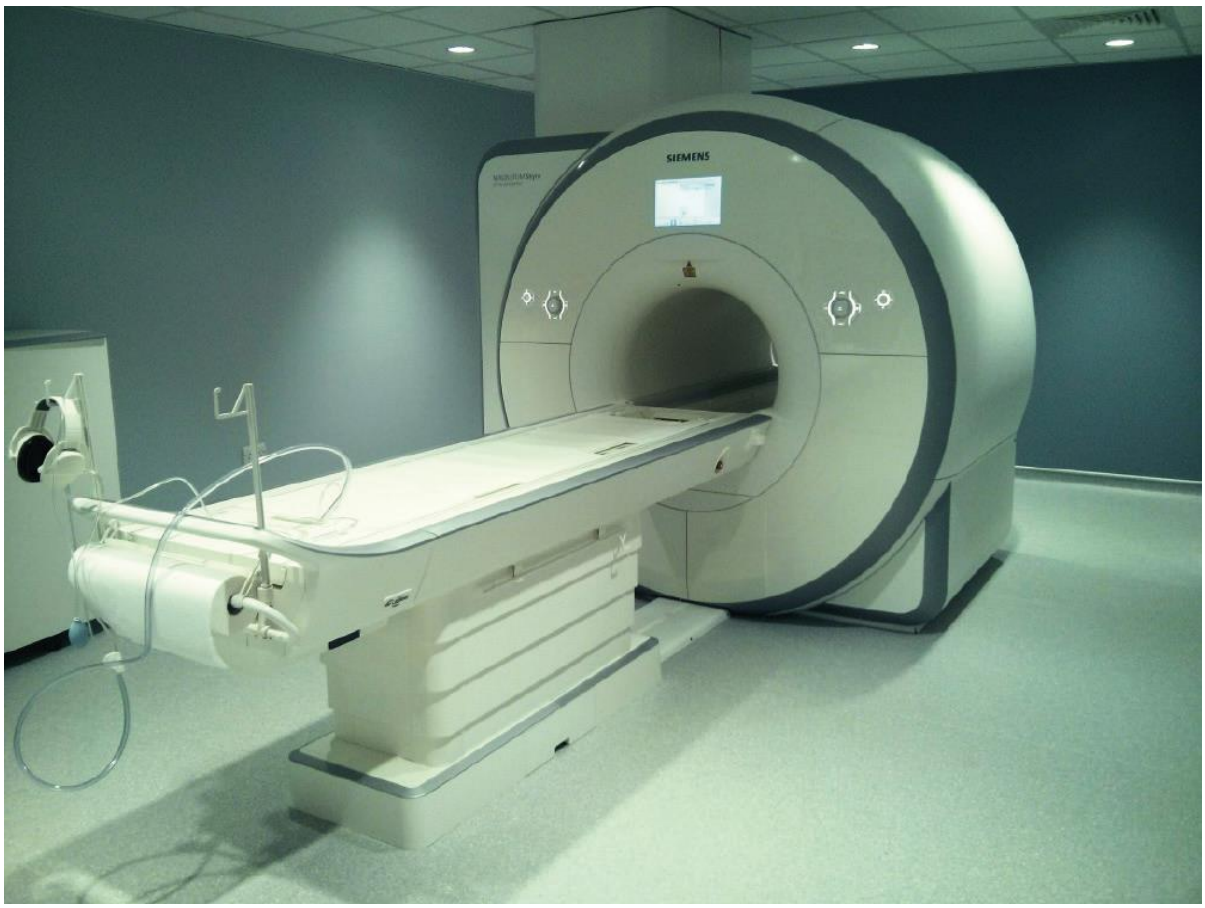
- Men may be asked if a staff member can shave a small section of their chest hair. This is so that electrical leads attached to sticky pads can be placed on the skin for the electrocardiogram (ECG; an electrical recording of the heart) and the heart MRI scan.

MRI scans

- The brain and heart/body MRI scans each take about 30 minutes.

- You will be shown into a room that houses one of the two MRI scanners. The scanner is a large cylinder with a tube running through the middle which is open at both ends (see picture below). You will be asked to lie down on a comfortably padded table that gently glides you into the scanning tube. Depending on the part of your body being scanned, you

will be moved into the scanner either head first or feet first.



A magnetic resonance imaging (MRI) scanner

- The MRI scanner is controlled by a computer which is in a different room. A specially trained technician will operate the computer. They will be able to see you through a window throughout the scan, and you will be able to talk to them through an intercom.

- MRI scanners can be noisy. We will provide you with headphones to protect your ears, but you will still be able to talk to the operator throughout the scan. You will be given a hand-held buzzer so that you can stop the process at any time if you wish.

- During the MRI scans, you will be asked to do things. For example, you will be shown something on a screen during the brain scan, and you will be asked to hold your breath for a short period of time during the heart scan.

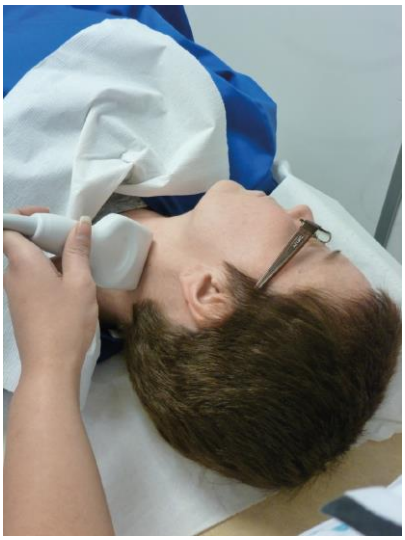
Neck ultrasound scan

- The neck scan takes about 10 minutes.

- You will be asked to lie face-up on a firm table. A clear water-based gel will be applied to your neck. A hand-held probe will then be placed against your skin and moved up and down your neck (see picture below).

- You will be asked to tilt or turn your head as the probe is swept over the entire length of both sides of your neck.

- The probe only covers your neck and does not touch your face or other parts of your body.

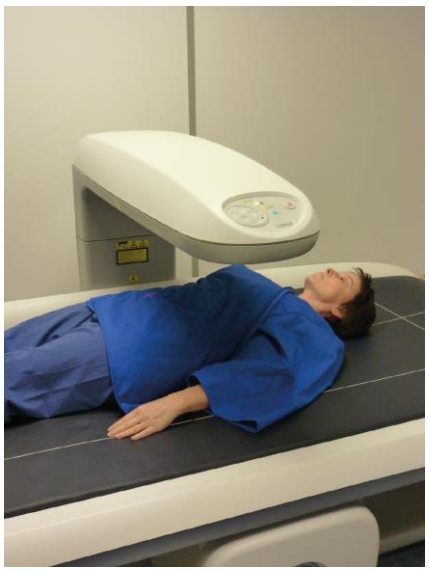


Neck ultrasound

DXA scan

- The DXA scan takes about 20 minutes.

- You will be asked to lie on a firm table while an arm passes over you (see picture below) to take X-ray pictures of your bones. You will be asked to lie in various positions so that the scanner can take pictures of different parts of your body.



DXA scanner

Other assessments

While you are with us, we would also like to take some more samples and repeat a number of measurements that we did at your first visit to UK Biobank several years ago. This information will allow scientists to take account of any changes in health and lifestyle over that time.

We will ask you to:

- Give another small sample of blood (about 3 tablespoons), saliva and urine for long-term storage and analysis.
- Answer again questions on your health, lifestyle and diet, memory, work and family history.
- Have repeat measurements of your blood pressure, pulse rate, height, weight, body fat, grip strength, heel bone density and lung function. You will also have an electrocardiogram (ECG) to measure the electrical activity of your heart.

What are the possible BENEFITS of taking part?

There are no direct benefits to you in taking part. However, the information about you from the imaging and other assessments will be valuable for helping scientists to better understand how a wide variety of diseases develop and to find new ways to prevent them.

What are the possible RISKS of taking part?

Undertaking the imaging assessment visit should not cause you any harm. We have chosen to perform scans and other physical measures that are safe, painless, relatively quick and comfortable.

The MRI scans use powerful magnets and great care is taken to prevent magnetic objects from entering the MRI room. Before you enter, we will ask you questions so that special precautions can be taken, if needed.

MRI scans involve lying flat in a slightly confined space and a small number of people may find this uncomfortable. However, the space in the scanners used in this assessment is wider (at least 70 cm, or 27 inches, in diameter) than in those typically used in hospitals, to ensure that participants are as comfortable as possible.

The low energy DXA scan involves a small dose of radiation (approximately 20 µSv units). This is the same amount as a standard chest X-ray or about one week's worth of natural background X-rays. By comparison, one trans-Atlantic flight exposes you to about four times as much radiation as that from a DXA scan.

You may feel some discomfort when you have blood taken, although our staff are specially trained to reduce this discomfort.

Improving the health of future generations

Do I need to have all the measurements?

When you call to make your appointment, we will check whether you are eligible to be scanned. This is because we do need your agreement to take part in all of the imaging scans before you make an appointment.

If it turns out when you arrive at the imaging assessment centre that you are no longer eligible for some of the scans, then you will still be able to take part in the rest of the assessment.

You do not have to have all the physical measures or to give a blood, urine or saliva sample if you don't want to. Similarly, if you feel uncomfortable about answering certain questions then you do not need to answer them.

Do I get any routine results from the visit?

As was the case at your initial assessment, you will receive information at the end of the visit about some of the physical measurements made during the assessment (blood pressure, weight, body mass index, waist circumference, percent body fat, heel bone ultrasound and lung function).

However, you will not receive any other routine results or pictures from the scans. It is important for you to be aware that this visit is not a clinical appointment or a 'health check'. You should, therefore, not rely on the absence of feedback from us as any form of reassurance regarding your health.

What will I be told if something suspicious is seen during my scans?

Imaging scans often show abnormalities, but most of these are no cause for concern.

The scans being performed in UK Biobank are not intended to diagnose an illness. They are not designed to identify any particular abnormalities and will not be routinely analysed by doctors or other specialists. Instead, they are being taken and stored for future research.

The technicians conducting the scans will be looking at the images to ensure their quality. It is important to understand that they will not be looking at them to identify particular health problems.

However, if a technician does notice something unusual that they think might be serious they will refer that scan to a specialist doctor for review. Something would be considered potentially serious if it indicated the possibility of a condition which, if confirmed, would carry a real prospect of seriously threatening life span, or of having a substantial impact on major body functions or quality of life.

If the specialist doctor agrees that the abnormality may be serious (regardless of whether or not it can be treated), then we will write to you and your GP (usually within two weeks of your visit).

For example, you and your GP would be informed if we saw an abnormality on one of your scans that looked like a tumour. However, we would not inform you if we saw an abnormality that looked like gallstones or a simple cyst, as such findings are common in healthy people

and not considered serious. Please see our website (www.ukbiobank.ac.uk) for further information about the types of potentially serious findings that we would inform you and your GP about if they are noticed during the scans and confirmed by a specialist doctor.

We estimate that about 10 to 15% of participants may have an abnormality considered to be potentially serious, but we will not know this number for sure until the project gets fully underway.

The reporting processes set in place will be carefully monitored, with ongoing training provided for the technicians doing the scanning.

What may happen if I am told about something suspicious on my scans?

The identification of an abnormality may mean that your GP refers you to specialists for further investigation and treatments. While some abnormalities may result in the diagnosis of a serious untreatable condition, others may turn out to be of little or no concern.

For some conditions, having an earlier diagnosis can be beneficial. But for other abnormalities detected on scans, knowing about them many years in advance can lead to unnecessary anxiety, investigations and treatments. The identification of some abnormalities could affect your ability to drive or work and could also make it difficult for you to obtain future travel, health or life insurance. It is important for you to understand that if we do not contact you and your GP it does not necessarily mean that you do not have any abnormalities. It simply means that no such abnormality was noticed by the technicians taking the scans.

It is also important to note that you can only take part in the imaging study if you feel able to consent to both you and your GP being informed if a potentially serious abnormality is noticed on one of your scans. If you feel that, in your case, the potential anxiety and uncertainty of being told about a possible serious abnormality, or the inconvenience and disruption to your life caused by further investigations, is likely to outweigh any benefit to you, we would quite understand if you decide not to take part in the imaging study.

How are we going to assess the impact of telling participants about potentially serious abnormalities?

For participants who are told of a potentially serious abnormality, we wish to find out how this has affected them, their family and friends, and the people involved in their care in the NHS. Very little is known about the impact of receiving information like this and this research would allow us to improve what we do and help others doing similar research.

We will, therefore, contact those participants 6 weeks and 6 months after providing such feedback to find out what, if anything, happened as a result of receiving it. We will also contact their GP about 6 months after writing to them to find out about any investigations and treatments.

Who will be able to use my information and samples?

Data and samples from your visit (scans, blood results and other data) will be stored for many years to come. The information will be used by approved researchers for medical and other

health-related research. This includes researchers who are working in other countries and in commercial companies looking for new treatments. Scientists will have to obtain scientific and, if necessary, ethics approval.

Results from all of these studies will be put back into the UK Biobank database for other researchers to use. Scientists must also publish the results of all research based on the UK Biobank resource so that everyone can benefit from it. Details of research that is being done using the resource is available on the website (www.ukbiobank.ac.uk).

We will never forward any individual's information, samples or test results to insurance companies or employers. Neither will we allow access to the police, security services, relatives or lawyers, unless forced to do so by the courts.

Published research that uses the UK Biobank resource will be made available to participants, and anyone else who might be interested, at www.ukbiobank.ac.uk.

How will information about me be kept confidential?

UK Biobank has put a number of rigorous measures in place to protect the confidentiality of participants. These include:

- No personal identifying details are included with data or samples provided to researchers.
- Information that might identify individuals (such as their names and addresses) is kept separately from other information about them in UK Biobank's databases.
- High level computer security is used to block unauthorised access (for example, by "hackers") to the computers that hold personal information.
- Access to personal information is restricted as much as possible, and all staff working on UK Biobank sign confidentiality agreements as part of their employment contracts.

Improving the health of future generations

Who do I contact if I have any questions?

If this leaflet does not answer your questions, please telephone us on 0800-0-276-276 (free from most land lines) or 0292-0-765-597, Monday-Saturday 8am-7pm for more information, or visit our website at www.ukbiobank.ac.uk.

If you would like to contact the person in charge, please send a letter or email to:

Professor Sir Rory Collins

UK Biobank
1-2 Spectrum Way
Adswood
Stockport
Cheshire
SK3 0SA

Email: ukbiobank@ukbiobank.ac.uk

We shall reply to your letter promptly in writing, unless you enclose your telephone number and wish to discuss your concerns with us.

Version January 2014

Supplementary File 2: UK Biobank magnetic resonance imaging parameters

Region and sequence	Voxel dimensions (mm)	TR (ms)	TE (ms)	N slices	Other parameters
Brain MRI[1]					
T1	1 x 1x 1	2000	-	208	TI 880 ms
FLAIR	1.05 x 1 x 1	5000	395	192	TI 1800 ms
SWI+T2*	0.8 x 0.8 x 3	27	9.4, 27[2]	48	-
Functional MRI: Rest	2.4 x 2.4 x 2.4	735	39	64	MB 8, 490 volumes
Functional MRI: Task[3]	2.4 x 2.4 x 2.4	735	39	64	MB 8, 332 volumes
Diffusion[4]	2.0 x 2.0 x 2.0	3600	92	72	MB 3
Cardiac MRI					
Long axis cine	1.8 x 1.8 x 6	32.64	1.16	-	373 ms resolution
Short axis truFISP cine	1.8 x 1.8 x 8	31.56	1.10	-	373 ms resolution
ShMOLLI	0.9 x 0.9[5] x 8	368.28	1.07	1	TI 0.1-5 s
Cine, single breath hold tagging	1.4 x 1.4 x 8	41.05	3.90	-	Grid tag, 3 short axis views
SSFP cine LV outflow tract and aorta	1.8 x 1.8 x 6	32.64	1.16	-	373 ms resolution
Flow sensitive cine aorta	1.8 x 1.8 x 6	37.12	2.47	-	TI 1.0 ms, 373 ms resolution
Body MRI					
T1 abdomen	1.9 x 1.3 x 10	450	11	12	-
3D-Dixon water fat separation neck to knees	2.2 x 1.2 x 10	3.23	1.44	-	Coronal and axial planes
T1 pancreas	1.2 x 1.2 x 1.12	2.95	1.12	80	-
Single breath hold liver and pancreas multi-echo	2.2 x 2.2 x 10	150	1.23-14.76[6]	-	-

mm = millimetres, TR = repetition time, ms = milliseconds, TE = echo time, MRI = magnetic resonance imaging, TI = inversion time, FLAIR = fluid attenuation inversion recovery, SWI = susceptibility weighted imaging, MB = multiband pulses, FISP = fast imaging with steady-state free precession, ShMOLLI = shortened modified Look-Locker inversion recovery, SSFP = steady state free precession, LV = left ventricle

- = Not applicable

[1] Participants imaged before August 18th 2014 (615/1000) also had a T2-weighted brain MRI sequence, however this was removed from the imaging protocol after this date and is no longer included in the UK Biobank imaging study. Additional modifications were made to the imaging protocol over the period during which the first 1000 participants were scanned, but these modifications were unlikely to affect the detectability or characterisation of potentially serious incidental findings and are not detailed here.

[2] Two echoes required

[3] Harari emotional faces task

[4] b=1000 and b=2000 s/mm^2, 50 directions per shell

[5] Interpolated

[6] 12 gradient recalled echoes

Supplementary File 3: UK Biobank lists of incidental findings

UK Biobank developed lists of findings which would be considered potentially serious, and findings not considered serious for use by radiographers and reporting radiologists. These lists were based on lists generated by the German National Cohort, and are subject to ongoing review.

Table i: Incidental findings on brain MRI

Potentially serious for feedback	**Not for feedback**
Acute brain infarction	Asymmetrical ventricles
Acute hydrocephalus	Chiari malformation[4]
Acute intracranial haemorrhage[1]	Chronic hydrocephalus
Arachnoid cyst[3]	Developmental anomalies (including venous anomalies)
Colloid cyst of third ventricle	Lipoma of corpus callosum
Intracranial mass lesion[2]	Non-acute brain infarction
Mastoiditis	Non-specific white matter hyperintensities
Suspected intracranial aneurysm or vascular malformation	Regional or global atrophy
	Suspected demyelination

[1] Not old bleeds, or microbleeds only detected on gradient recalled echo sequences

[2] Except meningiomata in locations considered highly unlikely to cause problems

[3] Only if large and considered likely to increase the risk of developing a subdural haematoma

[4] Descent of part of the cerebellum +/- brainstem below the foramen magnum

Table ii: Incidental findings on cardiac MRI

Potentially serious for feedback	**Not for feedback**
Aortic dissection	Atelectasis
Cardiac mass (including thrombus)	Calcified pleural plaque
Central pulmonary embolus	Calcified pulmonary nodule
Haemodynamically relevant pericardial effusion >2 cm	Emphysema
Heart valve defects[1]	Right sided descending aorta
Hilar, mediastinal, axillary or cervical lymphadenopathy[2]	
Lobar pneumonia or lung consolidation	
Lung mass > 2 cm	
Mediastinal mass > 2 cm	
Pleural effusion	
Pleural mass > 2 cm	
Pneumothorax	
Severe left or right ventricular dilation or dysfunction	
Severe left ventricular hypertrophy > 2 cm thick wall	
Thoracic aortic aneurysm > 5 cm	

[1] Severe regurgitation jet of any valve or severe turbulence (suggesting valve stenosis)

[2] >1.5 cm and >3 lymph nodes grouped in a circumscribed region

Table iii: Incidental findings on the abdominal portion of the body MRI

Potentially serious for feedback	**Not for feedback**
Abdominal aortic aneurysm > 5 cm	Abdominal wall hernia
Acute exudative pancreatitis	Bladder diverticulum
Adrenal lesion > 2 cm	Chronic cholecystitis
Ascites	Chronic pancreatitis
Cholestasis (intra- or extra-hepatic)[1]	Fatty liver
Deep vein thrombosis	Fibroids
Hepatomegaly	Gallstones
Ileus	Hiatus hernia
Intra-abdominal mass > 3 cm	Left sided inferior vena cava
Irregular/nodular liver margin	Liver cyst
Lymphadenopathy[2]	Renal calculus
Multiple small non-cystic, liver lesions (non haemangioma-like)	Simple renal cyst
Pneumoperitoneum	Single kidney
Portal vein occlusion	
Pyelonephritis	
Renal artery stenosis > 80% or bilateral	
Solid / cystic pancreatic tumour	
Solid gallbladder lesion	
Solid liver lesion	
Solid/semi-solid renal tumour > 2 cm	
Spleen infarction	
Splenomegaly > 15 cm	
Urinary obstruction	
Urinary tract mass > 2 cm	

[1] Common bile duct >15 mm (or >20 mm post-cholecystectomy)

[2] >1.5 cm and >3 lymph nodes grouped in a circumscribed region

Table iv: Incidental findings on dual energy X-ray absorptiometry

Potentially serious for feedback	Not for feedback
Major vertebral fracture	Non-skeletal findings
Primary skeletal malignancy	
Skeletal metastases	

Carotid Doppler ultrasound

Although asymptomatic carotid stenosis may be picked up by carotid ultrasound, its relevance in predicting prognosis over and above conventional vascular risk factors is not established, and so it was not considered to be a potentially serious incidental finding. Extra-carotid findings were not considered relevant for UK Biobank's imaging study as the radiographers conducting the imaging are specifically trained in the vascular component of this imaging modality only. Hence, carotid Doppler data do not form part of this manuscript.

Supplementary File 4: Example feedback letter sent to participants

Date

«participant_name»

«participant_address»

Dear «participant_name»,

Invitation to make an appointment with your GP following your visit to the UK Biobank imaging assessment visit

Thank you for your recent attendance at the UK Biobank imaging assessment visit.

We are writing to inform you that, during the scanning process, something was noticed on your [name_of] scan that your GP may want to follow up. We have informed «practice_name» of this possible abnormality, and recommend that you make an appointment to see your GP at your earliest convenience.

If these GP details are incorrect, please let us know as soon as possible by telephoning the Participant Resource Centre on 0800-0-276-276 (free from most land lines) or 0292-0-765-597, Monday-Saturday 8am to 7pm, or by emailing imaging.queries@ukbiobank.ac.uk.

Please do not be unduly alarmed by this letter. Whilst we aim to inform you only of abnormalities that might be potentially serious, it is still likely that many of these findings will turn out not to be of concern, or be something of which you are already aware. As indicated to you at the time of your assessment visit, the scans taken by UK Biobank are not specifically designed to detect clinical abnormalities, and so cannot generally be used to determine exactly what a possible abnormality is.

Your GP will be able to advise you as to what further action or investigations, if any, are needed.

Yours sincerely,

Professor Sir Rory Collins

UK Biobank Principal Investigator, and

Professor of Medicine & Epidemiology,

University of Oxford.

Supplementary File 5: Example feedback letter sent to general practitioners

Date

«GP_name»

«practice_address»

Dear «GP_name»

Cc «partipant_name»; «nhs_number»

Report of a potentially serious abnormality on an imaging scan from UK Biobank

«partipant_name»; «nhs_number» recently attended an imaging assessment visit as part of their participation in UK Biobank. UK Biobank is a population-based cohort study, established by the Medical Research Council and Wellcome Trust with the support of the Department of Health, which recruited over 500,000 middle-aged people during 2006 to 2010, some of whom are now participating in our imaging sub-study. This involved undergoing brain, heart and abdominal MRI scans, a carotid ultrasound scan and a DXA low energy X-ray scan, as well as answering questions, having non-imaging measurements and providing biological samples.

The imaging scans taken by UK Biobank are intended for research use only; they are not optimised for identifying any particular clinical abnormalities and may not provide sufficient information for diagnostic purposes. However, the radiologist/specialist who reviewed «participant_name»'s [name of] scan reported an incidental finding that may be potentially serious (i.e. indicating the possibility of a condition which, if confirmed, carries a real prospect of threatening life span, or of having a substantial impact on major body functions or quality of life).

A copy of the specialist's report is provided with this letter in case you feel further referral or clinical investigation is warranted. If required, you can request a digital copy of the relevant scans by emailing imaging.queries@ukbiobank.ac.uk or by telephoning Mr. Steve Garratt on 0161 475 5378. Alternatively you can write to us at the address shown below.

We have informed «participant_name» that a possible abnormality was found (see attached letter) on one of their scans and advised them to make an appointment to see you at their earliest convenience.

There is little consensus in the UK (or elsewhere) on the balance of benefit versus harm in telling research participants about incidental findings from imaging studies. In order to improve our procedures and those of other research projects in the future, we will contact you again in a few months with a short questionnaire asking you about the impact on the health service of providing this information.

We would be grateful if you could let us know as soon as possible if «participant_name» is not registered with your practice. Many thanks for your help.

Yours sincerely,

Professor Sir Rory Collins,

UK Biobank Principal Investigator, Prof. of Medicine & Epidemiology, University of Oxford.

Supplementary File 6: Understanding of consent related to the feedback of potentially serious incidental findings: data from 607 respondents from the first 1000 imaged UK Biobank participants

	Response[1]					
	Correct		Incorrect		Don't know	
As far as you are concerned, when you consented to participate, which of the following did you agree to?	**n/N**	**%**	**n/N**	**%**	**n/N**	**%**
My imaging scans and results would be given to me at the end of the visit[2]	526/607	86.7	49/607	8.1	32/607	5.3
In the event a potentially serious finding was identified on a scan:						
I could choose whether my GP and I would be informed[2]	381/607	62.8	158/607	26.0	68/607	11.2
Both my GP and I would automatically be contacted[3]	454/607	74.8	119/607	19.6	34/607	5.6
I would be told about this finding during the assessment visit[2]	540/607	89.0	19/607	3.1	48/607	7.9
I would be told about this finding after the assessment visit[3]	251/607	41.4	300/607	49.4	56/607	9.2

[1] Proportions of participants answering each question correctly, incorrectly or answering that they did not know were similar irrespective of whether or not participants had a potentially serious incidental finding (IF), and irrespective of whether any potentially serious IF was finally diagnosed as clinically serious or non-serious.

[2] The correct response was 'no'.

[3] The correct response was 'yes'. However, in retrospect these questions were deemed ambiguous. The participant information leaflet described the UK Biobank IF policy, including that a finding identified on a scan by a radiographer would only be fed if confirmed by a radiologist. Taking this in to account, if participants considered the case where a finding identified by a radiographer was not then confirmed by a radiologist, some participants may reasonably have concluded that they would not automatically be contacted about a finding identified on a scan, or that they might always not be told about a finding after the assessment visit.

Supplementary File 7: Final diagnoses of 217 potentially serious incidental findings

Image modality	Final diagnosis	Number of scans identified by radiographer flagging	Number of scans identified by systematic radiologist review
	Serious final diagnoses		
Brain MRI	Pituitary tumour	0	2
	Arachnoid cyst with hydrocephalus	1	1
	Meningioma compressing brainstem	1	1
Cardiac MRI	Thoracic aortic aneurysm[1]	3	5
	Lung tumour	0	3
	Cardiomyopathy	0	2
	Atrial fibrillation	0	1
	Coronary heart disease	0	1
	Heart block and LV impairment	0	1
Body MRI: Abdomen	Abdominal aortic aneurysm > 5 cm	0	1
	Gastrointestinal stromal tumour	0	1
	Pancreatic neuroendocrine tumour	0	1
DXA	Osteoporotic crush fracture	0	1
All modalities: serious final diagnoses		5	21
	Non-serious final diagnoses		
Brain MRI	Benign cyst/lesion	2	15
	Already known diagnosis	0	1
	Suspected lesion not confirmed	0	3
Cardiac MRI	Lung diagnosis – not serious	2	28
	Suspected lesion not confirmed	0	18
	Other non-serious diagnosis	0	10
	Cardiac diagnosis – not serious	0	8

	Already known cardiac diagnosis	0	7
	Already known lung diagnosis	0	2
Body MRI: Abdomen	Benign lesion (e.g. cyst)	6	57
	Suspected lesion not confirmed	0	13
	Already known diagnosis	1	4
	Other non-serious diagnosis	0	4
	Abdominal aortic aneurysm < 5 cm	1	2
Body MRI: Leg	Bone/soft tissue diagnosis – not serious	0	5
	Suspected lesion not confirmed	0	2
	Already known finding	0	1
DXA	Already known diagnosis	1	5
	Non-serious diagnosis	0	5
	Suspected lesion not confirmed	0	2
All modalities: non-serious final diagnoses		13	192
	Uncertain final diagnoses		
Cardiac MRI	Lung nodule, unclear nature	0	2
	Lung consolidation, unclear nature	0	1
DXA	Crush fracture T11, unclear relevance	0	1
All modalities: uncertain final diagnoses		0	4

[1] One participant with a thoracic aortic aneurysm was also found to have an atrial myxoma, which was resected at the time of aneurysm repair.

[2] Four findings could not be classed as serious or not serious by April 2016**:** one participant with a lung nodule was still under follow-up; another participant with a lung nodule underwent follow-up and was found to have lymphoma, but it was unclear whether the nodule was related to the lymphoma or not; one participant with lung consolidation reported that the final diagnosis may be scarring or bronchoalveolar cell carcinoma (this participant was also noted to have a meningioma compressing their brainstem on brain MRI, so that their overall final diagnosis clinical severity was classified as serious); we were unable to contact one participant with a crush fracture to determine the grade of the fracture, or whether they had been diagnosed with or treated for osteoporosis.

Supplementary File 8: Invasive procedures performed to diagnose or treat potentially serious incidental findings

Final diagnosis	n participants in diagnostic category	n/N with data available who underwent an invasive procedure[1]	Modality	Invasive procedure (one participant per procedure unless otherwise indicated)
Serious	21	11/18 (61.1%)	Brain MRI	Third-ventriculostomy and fenestration of arachnoid cyst Resection of meningioma compressing brainstem Transphenoidal resection of large pituitary macroadenoma
			Cardiac MRI	Biopsy and resection of pulmonary nodule Coronary angiogram to investigate poor LV function (n=2) Resection of lung cancer Repair of thoracic aneurysm (n=2)
			Body MRI: Abdomen	EUS FNA and Whipple procedure for pancreatic neuroendocrine tumour Resection of gastrointestinal stromal tumour
Non-serious or uncertain	158	12/144 (8.3%)	Cardiac MRI	Aspiration of breast cyst Bronchoscopy to investigate lung findings FNA of thyroid Resection of pulmonary nodule
			Body MRI: Abdomen	Colonoscopy to investigate thickened sigmoid (n=2) Hysterectomy for ovarian cysts Oopherectomy for ovarian cyst Polyps resected and uterus biopsy Transvaginal ultrasound for ovarian cyst (n=2) Upper and lower GI endoscopy for stomach mass and sigmoid diverticulosis

MRI = magnetic resonance imaging, LV = left ventricular, EUS FNA = endoscopic ultrasound fine needle aspiration, FNA = fine needle aspiration, GI = gastrointestinal

[1]Relevant questionnaire data or additional correspondence on invasive procedures were available from 18 participants with serious final diagnoses, and 144 participants with non-serious or uncertain final diagnoses

Supplementary File 9: Clinical management (medication and procedures) of the 21 participants with potentially serious incidental findings which were finally diagnosed as serious

Modality	Final diagnosis (one participant per diagnosis unless otherwise indicated)	Clinical management
Brain MRI	Arachnoid cyst with hydrocephalus	Neurosurgical drainage
	Meningioma compressing brainstem	Excision
	Pituitary tumour (n=2)	Transsphenoidal resection (n=1)
		Referred to specialist, no data on interventions (n=1)
Cardiac MRI	Atrial fibrillation	Warfarin
	Hypertrophic obstructive cardiomyopathy	Referred to specialist, no data on interventions
	Cardiomyopathy	Coronary angiogram, ramipril
	Coronary heart disease	Stress echocardiogram, coronary angiogram, aspirin, beta-blockers
	Heart block and LV impairment	Beta-blockers
	Lung tumours (n=3)	Excision and chemotherapy (n=1)
		Biopsy (n=1)
		Excision (n=1)
	Thoracic aortic aneurysm (n=5)	Repair (n=2)
		Referred to specialist, no data on interventions (n=1)
		No data (n=2)
Body MRI: Abdomen	Abdominal aortic aneurysm	No data
	Gastrointestinal stromal tumour	Excision
	Pancreatic neuroendocrine tumour	Excision and enzyme supplements
DXA	Osteoporotic crush fracture	Prophylactic medication

– = Not reported, MRI = magnetic resonance imaging, LV = left ventricular, EUS FNA = endoscopic ultrasound fine needle aspiration, DXA = dual energy X-ray absorptiometry

Supplementary File 10: Rates of radiographer flagging and rates of radiologist confirmation of potentially serious incidental findings in the first 7000 imaged UK Biobank participants

Participant blocks of 1000 (in order of attendance)	Flagged by radiographers[1]		Confirmed by radiologists[2]	
	N participants	% of total imaged	N participants	% of total imaged
1-1000	66	6.6	18	1.8
1001-2000	35	3.5	19	1.9
2001-3000	61	6.1	27	2.7
3001-4000	30	3.0	16	1.6
4001-5000	22	2.2	14	1.4
5001-6000	11	1.1	7	0.7
6001-7000	27	2.7	19	1.9
Mean per 1000 participants	36	3.6	17	1.7

[1] Chi-square 72.5, 6 degrees of freedom, $p<0.0001$

[2] Chi-square 13.0, 6 degrees of freedom, $p=0.04$

4.4 Conclusion

This chapter described the evaluation of the UK Biobank IFs protocol (radiographer flagging of concerning images for a radiologist to review) against another protocol commonly used by research imaging studies (systematic radiologist review of all imaging). This design, and the systematic long-term follow-up of participants with PSIFs and their GPs, enabled several aims of this thesis to be addressed by describing: how prevalence of PSIFs and serious final diagnoses varies by IFs protocol; clinical assessments generated by PSIFs; the impacts on participants' emotional wellbeing, insurance and finances, and work and activities; participants' understanding of consent regarding the UK Biobank IFs handling policy; participants' and GPs' opinions about receiving feedback of a PSIF; factors associated with PSIFs.

Compared to systematic radiologist review, radiographer flagging resulted in a substantially lower prevalence of participants with PSIFs (179/1,000 [17.9%] versus 18/1,000 [1.8%]) and a higher proportion of serious final diagnoses (21/179 [11.7%] versus 5/18 [27.8%]) (Gibson et al., 2018). While radiographer flagging resulted in false negatives (i.e. missed serious final diagnoses in 16/21 participants), systematic radiologist review resulted in large numbers of false positives (i.e. non-serious final diagnoses in 158 participants) (Gibson et al., 2018). While radiographer flagging was deemed to be a more appropriate approach to handling PSIFs than systematic radiologist review in a study of the scale of UK Biobank, there is no single 'best' protocol for handling IFs across all possible imaging contexts (Gibson et al., 2018).

All participants who received feedback of a PSIF had contact with their GP, and almost all (90%, 153/170 respondents) underwent some form of clinical assessment (most commonly imaging or referral to a specialist). Importantly, similar numbers of participants with serious and with non-serious final diagnoses underwent invasive procedures (n=11 and n=12 respectively) (Gibson et al., 2018). While a higher proportion of participants with serious final diagnoses reported undergoing clinical assessments compared to those with non-serious diagnoses, absolute numbers were higher in the latter group due to the higher prevalence of participants with non-serious final diagnoses (Gibson et al., 2018).

Variable proportions of participants reported that feedback of a PSIF had impacted on their emotional wellbeing (16.9%), insurance or finances (8.9%), or work or activities (5.6%). These non-medical impacts affected a higher proportion of participants with serious final diagnoses, but affected a higher absolute number of participants with non-serious final

diagnoses; in the latter group, such impacts could be regarded as unnecessary (Gibson et al., 2018).

The majority of participants correctly understood that they had consented to a protocol which would not provide feedback of a PSIF or access to their images at the end of their imaging visit (both >85%), but around a quarter of participants incorrectly thought that they could choose whether or not to receive feedback (158/607 [26.0%]); UK Biobank revised its consent materials accordingly, and evaluation of participants' understanding of consent continues (Gibson et al., 2018).

While the vast majority of participants reported being glad to have been told about a PSIF (98%, 142/145 respondents), and glad to have participated in the UK Biobank imaging study (99%, 147/148 respondents), only around one-third believed that participants should always be told about PSIFs (36.9%, 55/149 respondents [in contrast to 64.9% (61/94) of responding GPs]). Almost a quarter of participants changed their opinion on whether they should or should not be able to choose to receive feedback of a PSIF. The reasons for this are not clear, and further qualitative research is needed to understand these responses (Gibson et al., 2018). GPs also more frequently reported that participants had experienced negative impacts on emotional wellbeing rather than positive (38/99 [38.4%] versus 16/99 [16.2%] of GP respondents respectively), but that more frequently participants had experienced net benefit rather than net harm (51/86 [59.3%] versus 35/86 [40.7%] of GP respondents respectively) (Gibson et al., 2018). These data were limited by the lower response rate of GPs, and it would be useful to repeat this aspect of the evaluation study using a larger dataset which will be available in future.

This chapter also provided some preliminary data on the associations of age and sex with PSIFs; participants with PSIFs were slightly older than those without (mean age 63 versus 61 years, p=0.03), but there was no significant difference in the distribution of PSIFs between women and men (p=0.4) (Gibson et al., 2018). These results support those found during our systematic review and meta-analysis (Chapter 2). The factors associated with PSIFs, and with serious final diagnoses, are explored further in Chapter 5 using UK Biobank assessment data from a larger cohort of imaged participants.

The design of the study questionnaires did not enable quantification of use or costs of health services generated by participants with PSIFs (Gibson et al., 2018). The economic impact of feedback of PSIFs is explored further in Chapter 6 using a case-control study and linked routinely collected healthcare data.

Chapter 5 Factors associated with potentially serious incidental findings and serious final diagnoses

5.1 Introduction

Chapter 4 demonstrated that while the participants with potentially serious incidental findings (PSIFs) undergo some form of clinical assessment, the majority of PSIFs do not turn out to represent serious disease (Gibson et al., 2018). Understanding the factors associated with PSIFs and with serious final diagnoses will inform on the risks of needing some form of clinical assessment and discovering a finding which will likely impact on health respectively. Such knowledge would inform individuals' decisions to undergo imaging and researchers' estimates of the probability of generating PSIFs during imaging of a particular population, which may influence incidental findings (IFs) policy design.

Chapter 2 summarised our knowledge of the factors associated with PSIFs. There were only a small number of studies which systematically followed up unselected participants with PSIFs, and the available evidence suggested that PSIFs were associated with age (Brugulat-Serrat et al., 2017; Hartwigsen et al., 2010; Illes et al., 2004b; Tsushima et al., 2005), but not clearly associated with sex (Bos et al., 2016; Brugulat-Serrat et al., 2017; Haberg et al., 2016; Illes et al., 2004b; Kumar et al., 2008; Morin et al., 2009; Sandeman et al., 2013; Tsushima et al., 2005; Yue et al., 1997). These findings were confirmed by our study of UK Biobank participants in Chapter 4 (Gibson et al., 2018). There were not enough data to reliably comment on associations with other factors such as lifestyle or medical history, and there were no data available at all on the factors associated with serious final diagnoses (Chapter 2).

This chapter uses data from long-term, systematic follow-up of a larger cohort (N=7,334) of unselected imaged UK Biobank participants. We aim to investigate the associations of a wider range of factors with PSIFs, including sociodemographic and lifestyle factors, body mass index (BMI), morbidity and PSIFs protocol, and for the first time, the association of these factors with serious final diagnoses. This study also enables a further evaluation of the variation in prevalence of PSIFs and serious final diagnoses generated by the UK Biobank IFs protocol compared to systematic radiologist review.

This study manuscript has been edited by co-authors in preparation for submission to a journal; the current version of the draft manuscript is included in full in Section 5.2.

5.2 Factors associated with potentially serious incidental findings and with serious final diagnoses on multi-modal imaging in the UK Biobank Imaging Study: a prospective cohort study

Authors and affiliations

Lorna M. Gibson,[1] John Nolan,[1] Thomas J. Littlejohns,[2] Edouard Mathieu,[3] Steve Garratt,[4] Nicola Doherty,[4] Steffen Petersen,[5,6] Nicholas C.W. Harvey,[7] Jonathan Sellors,[2,4] Naomi E. Allen,[2] Joanna M. Wardlaw,[8] Caroline A. Jackson,[1] Cathie L.M. Sudlow.[1]

1 Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh Bioquarter, 9 Little France Road, Edinburgh, EH16 4UX

2 Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Big Data Institute, Old Road Campus, Oxford, OX3 7LF

3 Nuffield Department of Primary Care Health Sciences, University of Oxford, Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG

4 UK Biobank Co-ordinating Centre, 1-2 Spectrum Way, Adswood, Stockport, SK3 0SA

5 William Harvey Research Institute, National Institute for Health Research Barts Biomedical Research Centre, Queen Mary University of London, Charterhouse Square, London, EC1M 6BQ

6 Barts Heart Centre, St Bartholomew's Hospital, Barts Health NHS Trust, West Smithfield, London, EC1A 7BE

7 MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton General Hospital, Southampton, SO16 6YD

8 UK Dementia Research Institute at The University of Edinburgh, and Centre for Clinical Brain Sciences, University of Edinburgh, Chancellor's Building, 49 Little France Crescent, Edinburgh EH16 4SB

Correspondence

Professor Cathie L.M. Sudlow, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh Bioquarter, 9 Little France Road, Edinburgh, EH16 4UX. Cathie.Sudlow@ed.ac.uk, 0131 465 9619.

5.2.1 Abstract

Background

Feedback of potentially serious incidental findings (PSIFs) to imaging research participants generates clinical assessment in most cases. Understanding the factors associated with increased risks of PSIFs and of serious final diagnoses may influence individuals' decisions to participate in imaging research and will inform the design of PSIFs protocols for future research studies.

Methods

We included all UK Biobank participants who underwent imaging up to December 2015 (N=7,334, median age 63, 51.9% women). Brain, cardiac and body magnetic resonance, and dual-energy X-ray absorptiometry images from the first 1,000 participants were reviewed systematically by radiologists for PSIFs. Thereafter, radiographers flagged concerning images for radiologists' review. We classified final diagnoses as serious or not using data from participant surveys and clinical correspondence from general practitioners up to six months following imaging. We used binomial logistic regression models to investigate associations between age, sex, ethnicity, socio-economic deprivation, private healthcare use, alcohol intake, diet, physical activity, smoking, body mass index and morbidity, with both PSIFs and serious final diagnoses.

Results

Systematic radiologist review generated 13 times more PSIFs than radiographer flagging (179/1,000 [17.9%] versus 104/6,334 [1.6%]; age- and sex-adjusted odds ratio (OR) 13.3 [95% confidence interval (CI) 10.3–17.1] $p<0.001$) and proportionally fewer serious final diagnoses (21/179 [11.7%] versus 33/104 [31.7%]). Risks of both PSIFs and of serious final diagnoses increased with age (sex-adjusted ORs [95% CI] for oldest [67–79 years] versus youngest [44–58 years] participants for PSIFs and serious final diagnoses respectively: 1.59 [1.07–2.38] and 2.79 [0.86–9.0] for systematic radiologist review; 1.88 [1.14–3.09] and 2.99 [1.09–8.19] for radiographer flagging). No other factor was significantly associated with either PSIFs or serious final diagnoses.

Conclusion

Risks of PSIFs and serious final diagnosis are substantially influenced by PSIFs protocol and to a lesser extent by age. As most PSIFs do not represent serious disease, evidence-based

PSIFs protocols are paramount to minimise over-investigation of apparently asymptomatic research volunteers.

5.2.2 Introduction

Brain and body imaging is increasingly used for research, diagnostic and screening purposes and is accompanied by the risk of identifying abnormalities which are unrelated to the purposes of the imaging, so-called incidental findings (IFs) (Wolf et al., 2008). Since very few IFs turn out to represent serious disease (Gibson et al., 2018) it is of limited value to feedback clearly non-serious IFs. Therefore, we focus on potentially serious IFs (PSIFs), defined as those which indicate the possibility of a condition which, if confirmed, would carry a real prospect of seriously threatening life span, or of having a substantial impact on major body functions or quality of life (Gibson et al., 2018). Feedback of PSIFs detected during research imaging generates some form of clinical assessment (e.g. general practitioner [GP] appointments and specialist referrals, or further investigations including imaging and invasive procedures) in almost all cases (Gibson et al., 2018). Information on the factors associated with increased risk of detection and feedback of a PSIF (and therefore of subsequent clinical assessment), and with increased risk of eventually receiving a serious final diagnosis may influence individuals' decisions to consent to participate in imaging research and inform researchers' designs of appropriate PSIFs policies, which are required by major research funders (Farrar and Savill, 2014; Medical Research Council and Wellcome Trust, 2014).

A small number of studies (N=151 to 5,800) which followed up unselected participants with PSIFs suggest that PSIFs are associated with age, but not with sex, but none investigated the associations of PSIFs with PSIFs protocols, or any factors associated with serious final diagnoses (Bos et al., 2016; Brugulat-Serrat et al., 2017; Haberg et al., 2016; Hartwigsen et al., 2010; Illes et al., 2004b; Kumar et al., 2008; Morin et al., 2009; Sandeman et al., 2013; Tsushima et al., 2005; Yue et al., 1997).

The UK Biobank Imaging Study provides an opportunity to investigate potential risk factors for PSIFs and serious final diagnoses. In the UK Biobank Imaging Study, 100,000 of the original 500,000 participants are undergoing brain, cardiac and body magnetic resonance imaging (MRI), dual-energy X-ray absorptiometry (DXA) and carotid Doppler ultrasound (Matthews and Sudlow, 2015); over 27,000 participants have been imaged as of September 2018 (UK Biobank, 2018c). These imaging data are linked to detailed sociodemographic, lifestyle, physical measurement, genetic and routine healthcare data generating an extensive research resource (Matthews and Sudlow, 2015).

The UK Biobank Imaging Study will inevitably generate PSIFs. To inform the development of a pragmatic PSIFs protocol that aims to minimise harm to (the largely asymptomatic) 100,000 imaged participants, UK Biobank reviewed current practice, published literature and guidance, and sought advice from professional bodies and from ethical and legal experts (Gibson et al., 2018). The protocol is based on radiographers flagging images of potential concern to a radiologist for their review (Gibson et al., 2018; Gibson et al., 2016b). This approach was evaluated against a protocol involving systematic radiologist review of all images (which is more commonly used in research studies), and found to generate less harm (i.e., less unnecessary anxiety to participants) and a lower burden on the publicly-funded UK National Health Service (NHS) (Gibson et al., 2018). UK Biobank is continuing to evaluate this PSIFs protocol by systematically following up all participants identified with a PSIF.

Using data from the first 7,334 participants imaged during the first 20 months of the UK Biobank Imaging Study (including systematic follow-up data on 283 participants with PSIFs), we aimed to determine whether, and to what extent, sociodemographic, lifestyle, other health-related factors and PSIFs protocol are associated with detection of a PSIF and with a final diagnosis of serious disease.

5.2.3 Methods

We prepared this manuscript according to STROBE guidelines (von Elm et al., 2008). The statistical analysis code is available online (Gibson and Nolan, 2018). UK Biobank obtained ethics approval for the imaging study, and evaluation of the PSIFs protocol (REC Reference numbers: 11/NW/0382; 16/NW/0274). We provided all participants with written information about the imaging study and the UK Biobank imaging IFs protocol (UK Biobank, 2016). All participants provided consent to take part in the imaging study, and for UK Biobank to feed back any identified PSIFs to them and their GP.

5.2.3.1 UK Biobank Imaging Study

Of 9.2 million adults aged 40–69 invited to participate in UK Biobank, 0.5 million (5.5%) participated, providing initial baseline data between 2006 and 2010 (Fry et al., 2017). From April 2014 to December 2015, participants living within approximately 120 km of the imaging centre in Stockport were invited to take part in the UK Biobank Imaging Study (UK Biobank, 2015e). Participants were excluded if they had metal implants, penetrating metal injury, non-removable metallic items, or if they would find it difficult to complete the imaging, e.g. due to claustrophobia (Figure 5-1) (UK Biobank, 2015e).

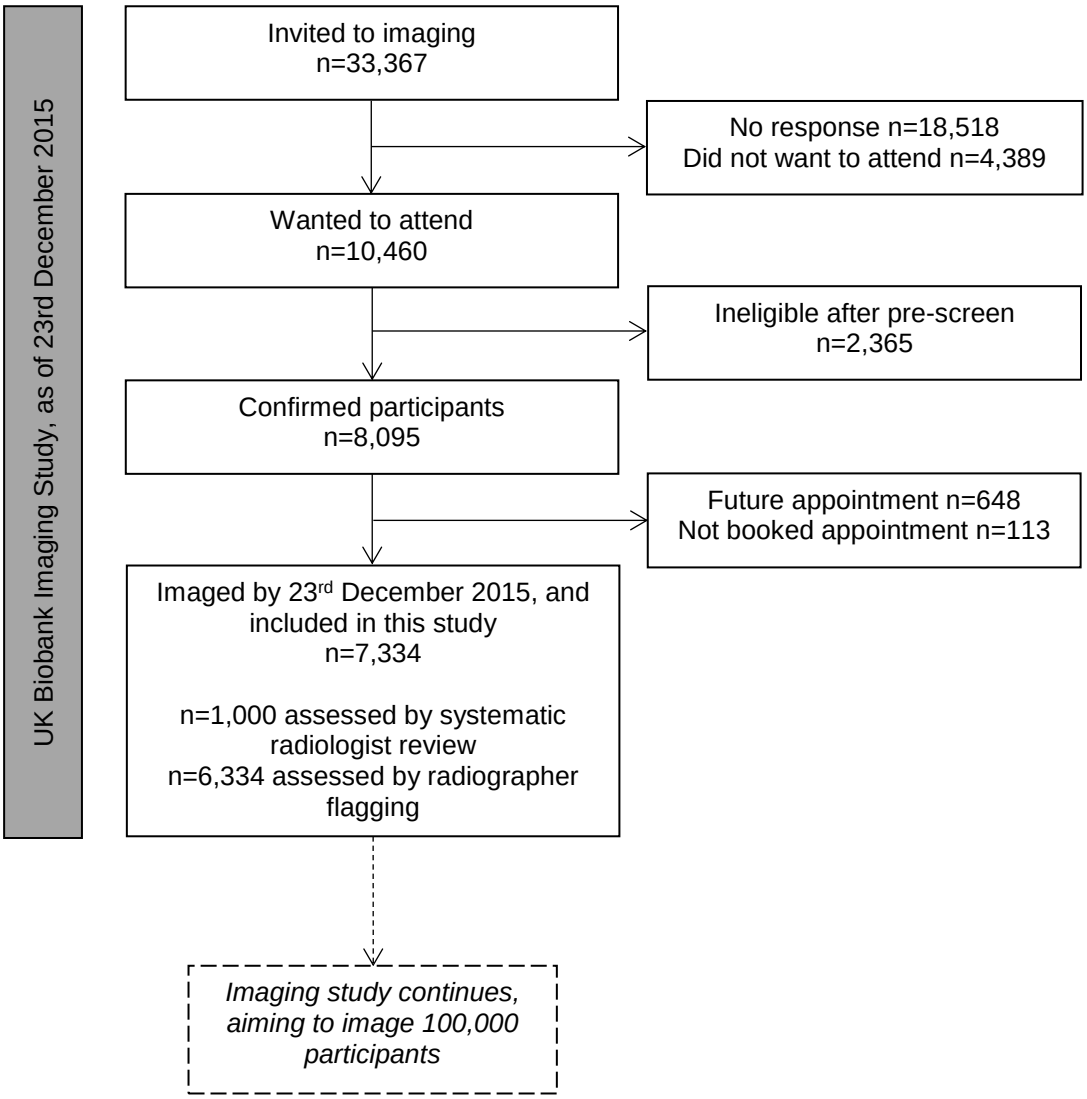


Figure 5-1. Participant flowchart

At the imaging visit, participants underwent brain, heart and body MRI, whole-body, spine and hip DXA and carotid Doppler ultrasound (UK Biobank, 2015a; b; c; d; e; 2017b). Participants also repeated the UK Biobank baseline assessment, which involved: a touchscreen questionnaire to collect data on potentially relevant risk factors for diseases, including sociodemographic, lifestyle and medical history; an interview; and physical measurements (Sudlow et al., 2015).

5.2.3.2 UK Biobank PSIFs protocol

During imaging, UK Biobank radiographers may notice PSIFs and 'flag' concerning images for radiologist review; radiologist-confirmed PSIFs are then fed back to participants and

their GP (Gibson et al., 2018). To evaluate this PSIFs protocol, all images from the first 1,000 participants were also systematically reviewed by radiologists for PSIFs (Gibson et al., 2018). Radiographers did not flag any PSIFs in addition to those detected by the radiologists within the first 1,000 imaged participants (Gibson et al., 2018). Therefore, for the purposes of this present study, we classified the first 1,000 imaged participants as undergoing the 'systematic radiologist review' PSIFs protocol, and subsequently imaged participants as undergoing the 'radiographer flagging' PSIFs protocol. For both protocols, to aid interpretation of images, radiologists received information on participants' age, sex, ethnicity, alcohol intake, smoking status, blood pressure, body mass index (BMI), employment status, and self-reported medical history.

Participants with PSIFs are surveyed six weeks and six months after receiving feedback, while their GPs are surveyed six months after feedback and asked for copies of relevant clinical correspondence; these responses include data on final diagnoses (Gibson et al., 2018).

Carotid Doppler ultrasound was deemed extremely unlikely to generate PSIFs under UK Biobank's protocol (UK Biobank, 2015e), and as such was not included in this study.

5.2.3.3 Data sources and variables

5.2.3.3.1 PSIFs and serious final diagnoses

We extracted data on the number, types and body region of each participant's PSIF(s) from radiologists' reports. A consultant physician and an experienced clinical radiology specialty trainee independently classified final diagnoses using all available survey data and clinical correspondence; we contacted participants and GPs by telephone where these data were insufficient to classify final diagnoses (Gibson et al., 2018). We classified final diagnoses as either: serious (if they were likely to threaten life span, or have a substantial impact on quality of life or major body function); not serious (if this was not the case or if available data suggested that the diagnosis was already known); or indeterminate (if there remained insufficient data to classify a final diagnosis as serious or not) (Gibson et al., 2018). We classified participants with more than one PSIF according to their most serious final diagnosis (Gibson et al., 2018).

5.2.3.3.2 Participant factors

We selected variables available from UK Biobank (UK Biobank, 2017a) which might be associated with PSIFs or would be possible confounders. These were age, sex, ethnicity, Townsend socio-economic deprivation score (which may reduce access to healthcare,

increase the risk of disease and reduce opportunities for disease detection prior to research imaging), use of private healthcare (which may be associated with reduced risk of serious final diagnoses if it increases prior knowledge of disease), alcohol intake (British Medical Association, 1995), fruit and vegetable intake (Cassidy et al., 2016), physical activity (UK Biobank, 2017c), smoking status, BMI (World Health Organisation, 2000) and morbidity. We measured the latter using the Elixhauser Index calculated using Hospital Episode Statistics data from two years before the date of imaging, and defined morbidity as ≥ 1 Elixhauser Index health conditions (Elixhauser et al., 1998; Quan et al., 2005; The National Casemix Office and Health and Social Care Information Centre, 2014).

5.2.3.4 Statistical analyses

Since our previous study showed that the 'systematic radiologist review' protocol produced approximately ten times more PSIFs compared with the 'radiographer flagging' protocol (Gibson et al., 2018) all analyses were stratified by PSIFs protocol to control for potential confounding. We compared characteristics between participants with and without PSIFs, and with and without serious final diagnoses, and calculated age- and sex-adjusted odds ratios (ORs) with 95% confidence intervals (CIs) using binomial logistic regression models.

We tested for normal distributions of continuous variables by visual inspection of graphed data and Kolmogorov-Smirnov goodness-of-fit tests. We attempted to normalise non-normally distributed data using log transformations, and if this failed, recoded variables into categories, aiming for similar numbers of participants in each category to optimise statistical efficiency. We used non-parametric tests to compare distributions of non-normally distributed variables between two groups. We considered data to be missing if participants did not respond, or if they responded 'do not know' or 'prefer not to answer'; such participants were excluded only from the relevant analyses. We present summary statistics of the characteristics of the whole UK Biobank cohort only; these cannot be compared directly to the imaged sub-cohort included in this study due to lack of independence of these two samples. The majority of variables had no, or only small proportions (< 4%) of missing data. In total, 460/7,334 (6.3%) participants had missing data for at least one variable. We performed all analyses using SPSS version 22.

5.2.4 Results

5.2.4.1 Participants

By 23rd December 2015, 7,334 of 33,367 invited participants (22.0%) had been imaged and were included in this study (Figure 5-1). Median age of the imaged participants was 63 (interquartile range 56–68) years and 3,804 (51.9%) were women (Table 5-1).

Compared to the entire UK Biobank cohort, this imaged sub-cohort included lower proportions of women, people of minority ethnicity groups, and people with less healthy lifestyles, including those with harmful alcohol intake, current smokers, low physical activity levels, or those who were overweight or obese. Conversely, a higher proportion of the imaged sub-cohort had one or more health conditions as measured using the Elixhauser Index compared to the whole cohort (Table 5-1).

5.2.4.2 PSIFs and final diagnoses

PSIFs were detected in 283/7,334 (3.9%) people: 179 of the first 1,000 (17.9%) by systematic radiologist review; 104 of the subsequent 6,334 (1.6%) by radiographer flagging (OR for systematic radiologist review versus radiographer flagging: 13.3, 95% CI 10.3–17.1, $p<0.001$, Table 5-2). The majority of PSIFs were finally diagnosed as clinically non-serious (229/283, 80.9%). Serious final diagnoses occurred in 54/7,334 (0.7%) participants: 21 of the first 1,000 (2.1%) undergoing the systematic radiologist review protocol and 33 of the 6,334 (0.5%) undergoing the radiographer flagging protocol (OR 4.2, 95% CI 2.4–7.4, $p<0.001$, Table 5-2). Radiographer flagging thus resulted in a higher proportion of PSIFs with serious final diagnoses than radiologist review (33/104 [31.7%] versus 21/179 [11.7%] respectively). The most common serious final diagnoses were tumours and vascular diseases (Table 5-3).

Table 5-1: Characteristics of the UK Biobank cohort and the imaged sub-cohort included in this study

	Entire UK Biobank cohort (of whom 100,000 will be imaged) (N=502,205)[a] n (%)	Imaged UK Biobank sub-cohort included in this study (N=7,334)[b] n (%)
Sociodemographics		
Age[c]		
Median (IQR)	63 (55 – 68)	63 (56 – 68)
Sex[d]		
Female	273,224 (54.4)	3,804 (51.9)
Male	228,981 (45.6)	3,530 (48.1)
Ethnicity[e]		
White	472,493 (94.1)	7,023 (95.8)
Minority ethnicity groups	27,012 (5.4)	225 (3.1)
TDI[e]		
Median (IQR)	-2.1 (-3.6 – 0.6)	-2.5 (-3.9 – -0.5)
Private healthcare[e]		
Never used	120,934 (70.1)	5,377 (73.3)
Ever used	49,980 (29.0)	1,850 (25.2)
Lifestyle		
Alcohol[e,f]		
None	19,942 (14.1)	848 (11.6)
Moderate	73,886 (52.3)	4,124 (56.2)
Hazardous	34,980 (24.8)	1,854 (25.3)
Harmful	9,084 (6.4)	376 (5.1)
Smoking[e]		
Never	273,400 (54.4)	4,350 (59.3)
Previous	172,980 (34.4)	2,575 (35.1)
Current	52,947 (10.5)	319 (4.3)
Fruit and vegetable portions/day[e,g]		
< 5	342,833 (68.3)	5,028 (68.6)
≥ 5	144,064 (28.7)	2,141 (29.2)
Days/week of moderate physical activity[e,h]		
0-2	169,162 (33.7)	2,149 (29.3)
3-4	118,615 (23.6)	1,967 (26.8)
5-7	187,251 (37.3)	2,983 (40.7)
Other factors		
Morbidity[i]		
None	457,301 (91.1)	6,422 (87.6)
≥1 condition	44,904 (8.9)	912 (12.4)
BMI[e,j]		
Underweight	2625 (0.5)	47 (0.6)
Normal	162,348 (32.3)	2,733 (37.3)
Overweight	212,064 (42.2)	3,061 (41.7)
Obese	122,228 (24.3)	1,454 (19.8)

IQR = interquartile range, TDI = Townsend Deprivation Index (higher score indicates greater deprivation), BMI = body mass index

a. Data collected at recruitment visit, unless otherwise indicated.

b. Data collected at the imaging visit, unless otherwise indicated.

c. Age on 30th April 2014, i.e. the start of the imaging study, for the entire cohort, and the imaged cohort.

d. Sex data were only available from the recruitment visit.

e. Data were missing for ethnicity (2,700/502,205 [0.5%], 86/7,334 [1.2%]), TDI (627/502,205 [0.1%], 0/7,334 [0.0%]), private healthcare use (1,694/172,608 [1.0%, questions on private healthcare were introduced partway through the recruitment period on 29th April 2009, thus giving a smaller denominator], 107/7,334 [1.5%]), alcohol (3,357/141,149 [2.3%, questions on subtypes of alcoholic drinks were introduced partway through the recruitment period on 29th August 2009, thus giving a smaller denominator], 132/7,334 [1.8%]), smoking (2,878/502,205 [0.6%], 90/7,334 [1.2%]), fruit and vegetable intake (15,308/502,205 [3.0%], 165/7,334 [2.2%]), physical activity (27,177/502,205 [5.4%], 235/7,334 [3.2%]), BMI (2,940/502,205 [0.6%], 39/7,334 [0.5%]), from the whole UK Biobank cohort versus the imaged sub-cohort respectively.

f. We calculated alcohol intake in units per week and categorised these using British Medical Association guidelines (women: moderate > 0 < 14, hazardous 14–35, harmful > 35; men: moderate >0 < 21, hazardous 21–50, harmful > 50) (British Medical Association, 1995).

g. We calculated portions of fruit and vegetable intake per day, and categorised these into five or more portions per day, or not (Cassidy et al., 2016).

h. Participants were asked 'in a typical week, on how many days did you do 10 minutes or more of moderate physical activities like carrying light loads, cycling at normal pace (do not include walking)?' (UK Biobank, 2017c).

i. We calculate morbidity using an Elixhauser Index score (Elixhauser et al., 1998; Quan et al., 2005) based on two-years of routinely collected Hospital Episode Statistics data, looking back from date of recruitment for the entire UK Biobank cohort, and the date of imaging for the imaged sub-cohort. Routinely collected health data are used to calculate payments for providers for services delivered for different conditions. The system for applying costs to healthcare services changed in 2012 (The National Casemix Office and Health and Social Care Information Centre, 2014), therefore the numbers of conditions coded in health records may not be directly comparable between the entire cohort, and the imaged cohort.

j. We defined BMI categories as underweight, normal, overweight and obese as BMIs of <18.5, ≥18.5 < 25.0 , ≥ 25.0 < 30.0, ≥ 30.0 respectively (World Health Organisation, 2000).

Table 5-2: Odds ratios for potentially serious incidental findings (PSIFs) and serious final diagnoses comparing two protocols

	Systematic radiologist review (N=1,000) n (%)[a]	Radiographer flagging (N=6,334) n (%)[a]	OR (95% CI)[b] systematic radiologist review versus radiographer flagging	p-value[c]
PSIFs	179 (17.9)	104 (1.6)	13.3 (10.3-17.1)	<0.001
Brain MRI	23 (2.3)	35 (0.6)	4.3 (2.5-7.3)	<0.001
Cardiac MRI	81 (8.1)	29 (0.5)	19.7 (12.8-30.2)	<0.001
Body MRI	83 (8.3)	27 (0.4)	21.3 (13.7-33.0)	<0.001
DXA	14 (1.4)	16 (0.3)	5.8 (2.8-11.9)	<0.001
Serious final diagnoses	21 (2.1)	33 (0.5)	4.2 (2.4-7.4)	<0.001
Brain MRI	4 (0.4)	13 (0.2)	2.0 (0.7-6.2)	0.221
Cardiac MRI	13 (1.3)	10 (0.2)	8.5 (3.7-19.5)	<0.001
Body MRI	3 (0.3)	5 (0.1)	4.1 (1.0-17.1)	0.056
DXA	1 (0.1)	5 (0.1)	1.3 (0.2-11.0)	0.818

OR = odds ratio, CI = confidence interval, PSIFs = potentially serious incidental findings, MRI = magnetic resonance imaging, DXA = dual-energy X-ray absorptiometry

a. Numerators are the number of participants with at least one PSIF per region. Multiple PSIFs occurred in four participants (who had two PSIFs each) under radiographer flagging, and in 33 (28 had two and five participants had three PSIFs each) under systematic radiologist review, giving a total of 325 PSIFs; therefore the sums of the body region PSIFs are greater than the 104 and 179 participants with at least one PSIF respectively. No participant had more than one serious final diagnosis.

b. Age- and sex-adjusted ORs for PSIFs and serious final diagnoses.

c. P-value from Wald test.

Table 5-3: Final diagnoses of 325 potentially serious incidental findings detected in 283 participants

Image modality	Final diagnoses	Systematic radiologist review (N=1,000) n participants	Radiographer flagging (N=6,334) n participants
Serious final diagnoses			
Brain MRI	Arachnoid cyst with hydrocephalus	1	-
	Arteriovenous malformation	-	1
	Cavernoma	-	1
	Meningioma requiring surgery	1	3
	Normal pressure hydrocephalus	-	1
	Pituitary tumour	2	4
	Pleomorphic adenoma requiring surgery	-	1
	Vestibular schwannoma	-	2
Cardiac MRI	Atrial fibrillation	1	1
	Cardiomyopathy	2	3
	Coronary heart disease	1	-
	Heart block and LV impairment	1	-
	Lung tumour	3	-
	Mesothelioma	-	1
	Myxoma	-	1
	Severe valve disease	-	2
	Thoracic aortic aneurysm	5	2
Body MRI: Abdomen	Abdominal aortic aneurysm > 5 cm	1	1
	Colonic tumour	-	1
	Gastrointestinal stromal tumour	1	-
	Pancreatic tumour	1	1
	Renal tumour	-	2
DXA	Osteoporotic crush fracture	1	5
All modalities: serious final diagnoses		**21**	**33**

Image modality (Continued)	Final diagnoses	Systematic radiologist review (N=1,000) n participants	Radiographer flagging (N=6,334) n participants
Non-serious final diagnoses			
Brain MRI	Already known diagnosis	1	3
	Benign cyst/lesion	15	10
	Hydrocephalus (not serious)	-	2
	Suspected lesion not confirmed	3	3
Cardiac MRI	Already known cardiac diagnosis	7	5
	Already known lung diagnosis	2	1
	Already under investigation	-	1
	Cardiac diagnosis – not serious	8	8
	Lung diagnosis – not serious	28	2
	Other non-serious diagnosis	10	1
	Suspected lesion not confirmed	18	1
Body MRI: Abdomen	Abdominal aortic aneurysm < 5 cm	2	1
	Already known diagnosis	4	3
	Benign lesion (e.g. cyst)	57	14
	Other non-serious diagnosis	4	-
	Suspected lesion not confirmed	13	2
Body MRI: Leg	Already known diagnosis	1	-
	Bone/soft tissue diagnosis – not serious	5	-
	Suspected lesion not confirmed	2	-
DXA	Already known diagnosis	5	5
	Non-serious diagnosis	5	3
	Suspected lesion not confirmed	2	2
All modalities: non-serious final diagnoses		**192**	**67**

Image modality (Continued)	Final diagnoses	Systematic radiologist review (N=1,000) n participants	Radiographer flagging (N=6,334) n participants
Uncertain final diagnoses			
Brain MRI	Lesion, unclear nature	-	4
Cardiac MRI	Lung consolidation, unclear nature	1	1
	Lung nodule, unclear nature	2	-
Body MRI: Abdomen	Cysts, unclear nature	-	2
DXA	Crush fracture T11, unclear relevance	1	-
	Fractures, unclear cause	-	1
All modalities: uncertain final diagnoses		**4**	**8**

MRI = magnetic resonance imaging, LV = left ventricular, DXA = dual-energy X-ray absorptiometry

- = zero

The two doctors agreed on the initial classification of final diagnoses in 270/283 (95.4%) of cases, and readily resolved the 13 cases of disagreement through discussion.

Systematic radiologist review generated higher proportions of PSIFs on all imaged body regions (OR range 4.3–21.3, all $p<0.001$, Table 5-2) compared to radiographer flagging. Radiologists more commonly detected PSIFs on cardiac (8.1%) and body MRI (8.3%) compared to brain MRI (2.3%) or DXA (1.4%), whereas radiographer flagging generated similar proportions of PSIFs across body regions (range 0.3–0.6%, Table 5-2). Serious final diagnoses occurred most commonly on cardiac MRI assessed by systematic radiologist review (13/1,000, 1.3%, Table 5-2).

5.2.4.3 Factors associated with PSIFs and serious final diagnoses

Across the relatively narrow age range of the included participants, older age was associated with increased odds of PSIFs and of serious final diagnoses under both protocols, albeit not statistically significant for serious final diagnoses under systematic radiologist review (sex-adjusted ORs [95% CI] for oldest [67–79 years] versus youngest [44–58 years] participants for PSIFs and serious final diagnoses respectively: 1.59 [1.07–2.38] and 2.79 [0.86–9.0] for systematic radiologist review; 1.88 [1.14–3.09] and 2.99, 95% CI [1.09–8.19] for

radiographer flagging) (Figures 5-2 and 5-3). Of the participants with PSIFs, those with serious final diagnoses were older than those with non-serious final diagnoses (median ages [range minimum-maximum] in years: 66 [50–76] versus 64 [44–76] respectively, p=0.021).

Of participants assessed by radiographer flagging, overweight participants had reduced odds of serious final diagnoses compared to those of normal or underweight BMI (age- and sex-adjusted OR 0.21, 95% CI 0.08–0.58, p=0.003, Figure 5-3), but the number of overweight participants was very small (n=5).

No significant associations were found between PSIFs or serious final diagnoses and any other investigated factor for participants assessed by either PSIFs protocol (Figures 5-2 and 5-3).

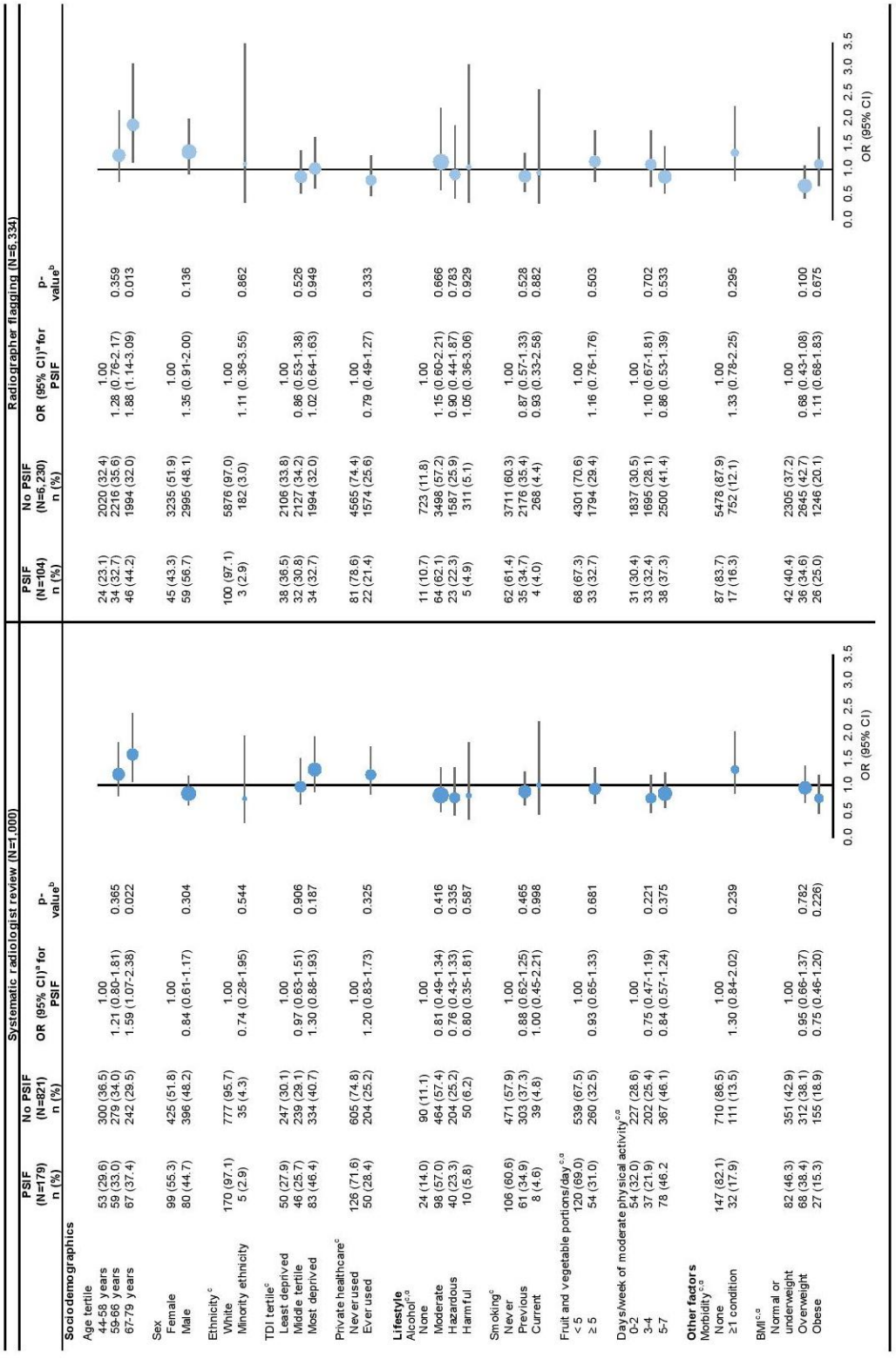


Figure 5-2. Age- and sex-adjusted odds ratios for potentially serious incidental findings (PSIFs) by PSIFs protocol.

PSIFs = potentially serious incidental findings, OR = odds ratio, CI = confidence interval, TDI = Townsend Deprivation Index, BMI = body mass index.

Circles are weighted by the proportion of participants within a category.

a. **Age- and sex-adjusted ORs, except age tertiles which are adjusted for sex only, and sex which is adjusted for age only.**

b. **P-value from Wald test.**

c. **Data were missing for ethnicity (13/1,000 [1.3%], 73/6,334 [1.2%]), TDI (1/1,000 [0.1%], 3/6,334 [<0.0%]), private healthcare use (15/1,000 [1.5%], 92/6,334 [1.5%]), alcohol (20/1,000 [2.0%], 112/6,334 [1.8%]), smoking (12/1,000 [1.2%], 78/6,334 [1.2%]), fruit and vegetable intake (27/1,000 [2.7%], 138/6,334 [2.2%]), physical activity (35/1,000 [3.5%], 200/6,334 [3.2%]) and BMI (5/1,000 [0.5%], 34/6,334 [0.5%]), for participants assessed by systematic radiologist review and by radiographer flagging respectively.**

d. **We calculated alcohol intake, fruit and vegetable intake, physical activity, morbidity and BMI as described in the footnotes to Table 5-1.**

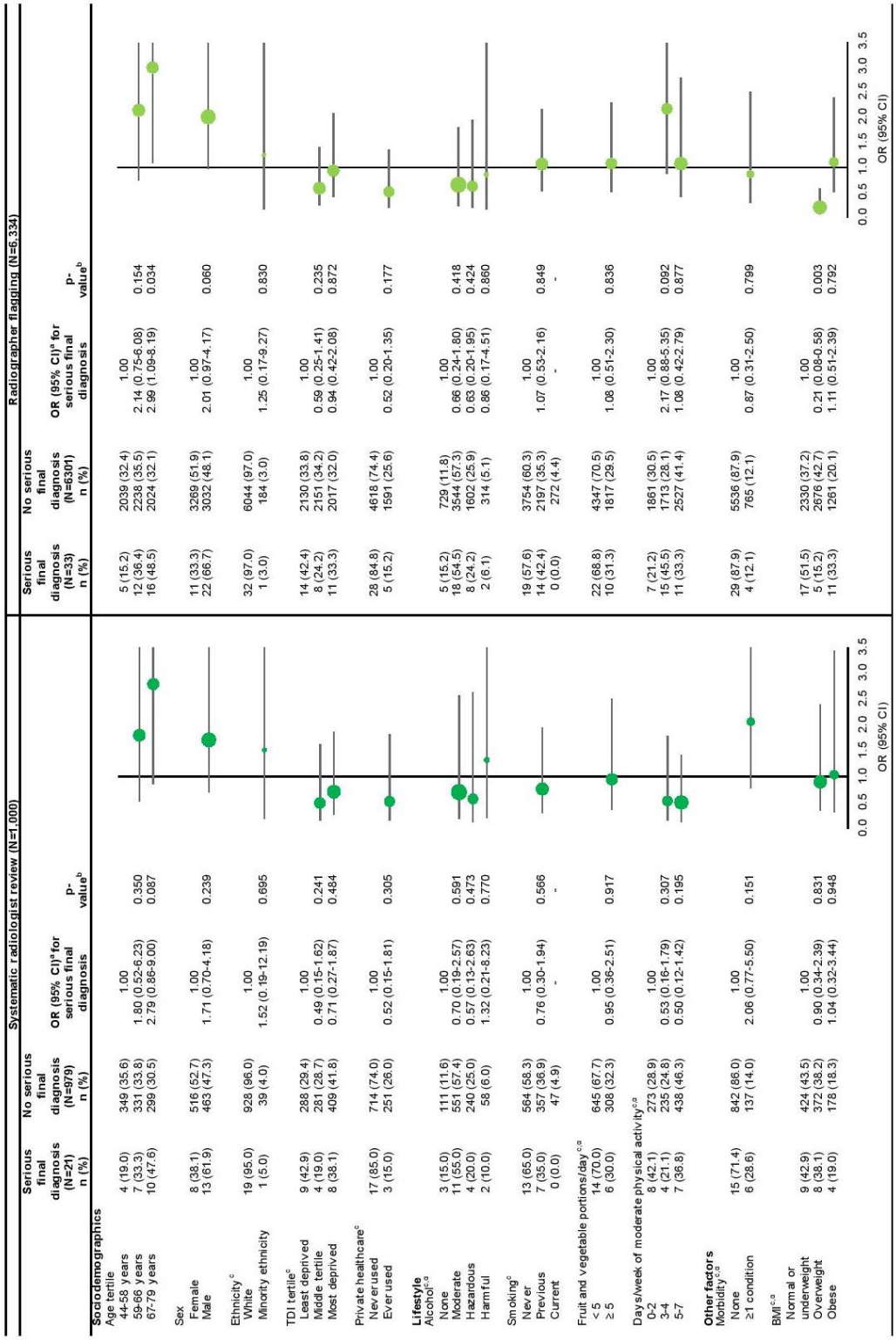


Figure 5-3. Age- and sex-adjusted odds ratios for serious final diagnoses stratified by potentially serious incidental findings (PSIFs) protocol

OR = odds ratio, CI = confidence interval, TDI = Townsend Deprivation Index, BMI = body mass index.

Circles are weighted by the proportion of participants within a category.

a. Age- and sex-adjusted ORs, except age tertiles which are adjusted for sex only, and sex which is adjusted for age only.

b. P-value from Wald test.

c. Data were missing as described in Figure 5-2, footnote c.

d. We calculated alcohol intake, fruit and vegetable intake, physical activity, morbidity and BMI as described in the footnotes to Table 5-1.

5.2.5 Discussion

Systematic radiologist review of images resulted in approximately 13 times more PSIFs, and four times more serious final diagnoses than the radiographer flagging protocol; these effect sizes are larger than those of any other risk factor assessed for association with either PSIFs or serious final diagnoses. Most (up to 80%) PSIFs did not turn out to represent serious disease. The odds of PSIFs and of serious final diagnoses increased with age, regardless of PSIFs protocol. There were no clear associations between either PSIFs or serious final diagnoses and sex, ethnicity, socio-economic deprivation, use of private healthcare, alcohol intake, diet, physical activity, smoking status, BMI or morbidity among participants assessed using either PSIFs protocol.

Our study confirms and updates our previous findings from the first 1,000 imaged UK Biobank participants (Gibson et al., 2018): compared to systematic radiologist review, radiographer flagging resulted in substantially fewer participants with PSIFs and a higher proportion of these had serious final diagnoses. We also confirm the findings of the above-mentioned smaller cohort (Gibson et al., 2018), that most PSIFs do not turn out to represent serious disease. Previous studies, mostly of brain MRI, found that PSIFs were associated with increased age (Brugulat-Serrat et al., 2017; Hartwigsen et al., 2010; Illes et al., 2004b; Tsushima et al., 2005), but not clearly associated with sex (Bos et al., 2016; Brugulat-Serrat et al., 2017; Haberg et al., 2016; Illes et al., 2004b; Kumar et al., 2008; Morin et al., 2009; Sandeman et al., 2013; Tsushima et al., 2005; Yue et al., 1997). We have further confirmed these findings in participants undergoing multi-modal imaging of multiple body regions, and shown this to be independent of the IFs protocol. Previous studies did not demonstrate any associations with PSIFs and medical history of cardiac disease (Tsushima et al., 2005), psychotic episodes (Sommer et al., 2013) or human-immunodeficiency virus (Loy et al., 2015). Given the varying nature of PSIFs (tumours, aneurysms etc.), a common biological risk factor seems unlikely. Instead, we captured morbidity using the Elixhauser Index, which

comprises 30 conditions (Elixhauser et al., 1998; Quan et al., 2005). There was no convincing association between morbidity and either PSIFs or serious final diagnoses, but sparse data on both of these outcomes and exposure data on morbidity (which may be secondary to healthy volunteer bias and a relatively short period of retrospective capture within linked hospital admissions data, chosen to limit any bias that may arise from changes in healthcare record coding practices in 2012 (The National Casemix Office and Health and Social Care Information Centre, 2014)) may have attenuated any true association. Furthermore, different definitions of morbidity may well produce different results.

Large studies are needed to investigate the factors associated with PSIFs and with serious final diagnoses, as these outcomes are relatively rare, particularly under a protocol of radiographer flagging. Our study is the largest so far to investigate the factors associated with PSIFs, and the first to investigate factors associated with serious final diagnoses, in unselected, healthy participants undergoing MRI of any body region. Our sample is approximately 25% larger than the largest previous study of factors associated with PSIFs detected on brain MRI (N=5,800) (Bos et al., 2016), and 50 times larger than the largest previous such study of multi-region MRI (N=148) (Morin et al., 2009). We systematically followed up 50% more participants for data on final diagnoses compared to the largest previous study (N=188) (Bos et al., 2016). Despite the size of our study, we still may have missed associations with PSIFs or final diagnoses due to sparsity of these outcomes within our cohort and small numbers within some exposure categories (e.g. minority ethnicity groups). Healthy volunteer selection bias likely affects the UK Biobank cohort, as participants are less deprived than non-participants and less likely to be obese, smoke, drink alcohol daily or have self-reported medical conditions compared to the general population (Fry et al., 2017). The imaged cohort are then further selected, with lower proportions of people having more ‘unhealthy’ lifestyles. As with all epidemiological studies which use self-reported data, our data on exposures may be further limited by reporting bias; participants may have inaccurately reported alcohol intake, smoking habits, physical activity and diet. The apparently reduced odds of serious final diagnoses in overweight participants may be spurious, secondary to data sparsity of both the outcome and the exposure. The direction of an association (if any) between increased BMI and PSIFs is unclear. The associations between increased BMI and certain cancers (Bhaskaran et al., 2014) may lead to increased risk of PSIFs and serious final diagnoses; alternatively, risks may be reduced if people with increased BMI tend not to complete all MRI sequences, or imaging of all body regions.

Our classifications of 'serious' final diagnoses are based on clinical judgement using data collected up to six months after feedback of a PSIF. Reaching final diagnoses of some PSIFs may take longer (Gibson et al., 2018). Feedback of PSIFs may impact on non-medical domains such as emotional wellbeing, insurance and finances and work and activities, regardless of the health-related severity of the final diagnosis (Gibson et al., 2018). 'Severity' of a final diagnosis is therefore inherently difficult to judge, though we did show good agreement between two independent physicians' classifications using a medical-based definition.

By deliberately focusing our study on participants with PSIFs and serious final diagnoses our results inform on factors associated with findings which are likely to generate clinical assessment, and those with serious health consequences, respectively. While our cohort is not representative of the general population, exposure-outcome associations can be generalised to other populations (Collins, 2012; Fry et al., 2017; Manolio and Collins, 2010), to inform the design of appropriate IFs handling policies, which are required by major funders (Farrar and Savill, 2014), and of materials to facilitate the informed consent of potential research participants.

Compared to sociodemographic, lifestyle and health-related factors, the protocol for identifying PSIFs protocol has by far the largest influence on the generation of PSIFs and serious final diagnoses. As the majority of PSIFs do not turn out to be serious, but feedback generates clinical assessments and negative impacts on emotional wellbeing, insurance and finances and work and activities (Gibson et al., 2018), our study suggests that researchers have the opportunity to greatly influence (for better or worse) the potential harms done to apparently asymptomatic research volunteers. There remain many unanswered questions on the impacts of different methodologies to feedback research results to participants (Wong et al., 2018); to inform future policy design, evaluations of the impacts of different protocols are paramount.

PSIFs are rare, and few are finally diagnosed as serious disease; hence large studies are needed to investigate the associated factors. This study represents the largest such cohort so far. Furthermore, since 100,000 participants will complete the UK Biobank imaging assessment over the next few years, it will in due course be possible to update these analyses with a substantially larger sample size, providing more comprehensive and statistically better powered estimates of the factors associated with PSIFs and with serious final diagnoses.

5.2.6 Additional information

5.2.6.1 Acknowledgements

Ms Regina Prigge (Usher Institute of Population Health Sciences and Informatics, University of Edinburgh,) and Dr Sophie Cassidy (Clinical Exercise Research Group, Institute of Cellular Medicine, Newcastle University) kindly provided advice on using UK Biobank data to calculate alcohol intake and portions of fruit and vegetables respectively. This work was conducted on behalf of the UK Biobank Imaging Working Group.*

*Members of the UK Biobank Imaging Working Group

Imperial College London

Prof Paul Matthews (Edmond and Lily Safra Chair and Head of Brain Sciences)

Dr Tony Goldstone (Senior Clinical Research Fellow)

Newcastle University

Prof Andrew Blamire (Professor of Magnetic Resonance Physics)

Queen Mary University of London

Prof Steffen Petersen (Professor of Cardiovascular Medicine)

University of Edinburgh

Prof Cathie Sudlow (Chair of Neurology and Clinical Epidemiology)

Dr Lorna Gibson (Clinical Research Fellow and Specialty Trainee in Clinical Radiology)

University of Manchester

Prof Alan Jackson (Professor of Radiology)

University of Oxford

Prof Naomi Allen (Associate Professor in Epidemiology)

Prof Sir Rory Collins (BHF Professor of Medicine and Epidemiology)

Prof Paul Leeson (Professor of Cardiovascular Medicine)

Prof Karla Miller (Professor of Biomedical Engineering)

Prof Stefan Neubauer (Professor of Cardiovascular Medicine)

Prof Stephen Smith (Professor of Biomedical Engineering)

University of Southampton

Prof Nicholas Harvey (Professor of Rheumatology and Clinical Epidemiology)

University of Westminster

Prof Jimmy Bell (Professor)

5.2.6.2 Data availability

Due to the confidential nature of questionnaire responses and clinical information on participants with potentially serious incidental findings, it is not possible to publicly share all of the data on which our analyses were based, but summaries of all relevant data are included in the manuscript, and our statistical analyses code is provided online for researchers to understand our analytical approach.

Importantly, any bona fide researcher can apply to use the UK Biobank resource, with no preferential or exclusive access, for health related research that is in the public interest. Application for access to UK Biobank data involves registration and application via the UK Biobank website, with applications considered by the UK Biobank Access Sub-Committee. Following approval, researchers and their institutions sign a Material Transfer Agreement and pay modest access charges. Further information on applying to access UK Biobank data is available at: http://www.ukbiobank.ac.uk/register-apply/.

5.2.6.3 Competing interests

L.M. Gibson: Member of the UK Biobank Imaging Working Group. UK Biobank Imaging Consultant, University of Edinburgh.

J. Nolan: UK Biobank Data Analyst, University of Edinburgh.

T.J. Littlejohns: UK Biobank Epidemiologist, University of Oxford.

E. Mathieu: Former UK Biobank Data Analyst, University of Oxford.

S. Garratt: Member of the UK Biobank Imaging Working Group. Senior Project Manager of UK Biobank Imaging Study.

N. Doherty: UK Biobank Senior Clinical Study Administrator.

S. Petersen: Member of the UK Biobank Imaging Working Group.

N.C.W. Harvey: Member of the UK Biobank Imaging Working Group.

J. Sellors: UK Biobank legal counsel.

N.E. Allen: Member of UK Biobank Steering Committee, UK Biobank Imaging, Enhancements, Follow-up and Outcomes and Infectious Diseases Working Groups. UK Biobank Senior Epidemiologist.

J.M. Wardlaw: Advised on imaging protocols for the UK Biobank imaging study. Currently analysing UK Biobank brain imaging and numeric data.

C.A. Jackson: None.

C.L.M. Sudlow: Member of UK Biobank Steering Committee, and UK Biobank Imaging, Enhancements, and Follow-up and Outcomes Working Groups. UK Biobank Chief Scientist.

5.2.6.4 Author contributions

L.M. Gibson: Conceptualisation, data curation, formal analysis, funding acquisition, investigation, methodology, software, visualisation, original draft.

J. Nolan: Data curation, formal analysis, investigation, software, review and editing.

T.J. Littlejohns: Data curation, review and editing.

E. Mathieu: Data curation, review and editing.

S. Garratt: Data curation review and editing.

N. Doherty: Project administration.

S. Petersen: Review and editing.

N.C.W. Harvey: Review and editing.

J. Sellors: Review and editing.

N.E. Allen: Conceptualisation, funding acquisition, review and editing.

J.M. Wardlaw: Conceptualisation, funding acquisition, supervision, review and editing.

C.A. Jackson: Methodology, supervision, review and editing.

C.L.M. Sudlow: Conceptualisation, data curation, funding acquisition, methodology, supervision, review and editing.

5.2.6.5 Grant information

Lorna M Gibson is funded by a Wellcome Trust Clinical Research Training Fellowship (grant reference 107190/Z/15/Z). UK Biobank is funded by the Wellcome Trust, Medical Research Council, Department of Health, Scottish Government, Northwest Regional Development Agency, Welsh Assembly Government, British Heart Foundation, and Diabetes UK.

The main funders of UK Biobank's imaging study (the Medical Research Council and the Wellcome Trust) and the imaging study's independent scientific review panel encouraged evaluation of the potentially serious incidental findings protocol, but did not contribute to the design, the collection, analyses or interpretation of data, writing of this report or the decision to submit this work for publication.

5.3 Conclusion

This chapter confirmed the findings of Chapter 4 (Gibson et al., 2018), showing that systematic radiologist review resulted in 13 times more PSIFs (179/1,000 [17.9%] versus 104/6,334 [1.6%]; age- and sex-adjusted odds ratio [OR] 13.3 [95% CI 10.3–17.1]) and four times more serious final diagnoses (OR 4.2 [95% CI 2.4–7.4]) than radiographer flagging. Furthermore, a lower proportion of PSIFs detected by radiologists resulted in serious final diagnoses compared to radiographer flagging (21/179 [11.7%] versus 33/104 [31.7%]). This chapter also confirms the findings of the systematic review presented in Chapter 2 and the results from a smaller cohort presented in Chapter 4 (Gibson et al., 2018): most PSIFs turn out not to represent serious disease.

The influence of PSIFs protocol was the largest by far of any factor investigated for an association with either PSIFs or serious final diagnoses. We confirmed our findings from our systematic review (Chapter 2), that age is (and sex is not) associated with PSIFs. We further demonstrated that age is (and sex is not) associated with serious final diagnoses. In addition, we found no clear evidence of associations between either PSIFs or serious final diagnoses with ethnicity, deprivation, use of private healthcare, alcohol intake, diet, physical activity, smoking status, BMI or morbidity.

While this study was the largest so far to investigate the factors associated with PSIFs and the first to investigate the factors associated with serious final diagnoses, we still may have missed significant associations due to lack of power, healthy volunteer bias and misclassification of exposure variables. The outcomes of PSIFs and serious final diagnoses are rare, and even more so under a PSIFs protocol of radiographer flagging, and as such, even larger studies are needed in order to have the statistical power to detect significant associations. UK Biobank participants are largely healthier than the general population (Fry et al., 2017), and the imaged sub-cohort are further selected and healthier than the broader UK Biobank cohort. In addition, participants may inaccurately report their alcohol intake, smoking habits, physical activity and diet. Limited data on exposures, due to the combination of healthy volunteer and reporting bias, therefore may have attenuated associations toward the null.

Ongoing long-term systematic follow up of imaged UK Biobank participants will enable further investigation of the factors associated with PSIFs and serious final diagnoses in future. As PSIFs protocol had by far the largest effect on the prevalence of PSIFs and serious final diagnoses, evaluations of different IFs policies are paramount to inform the design of

IFs policies which minimise the unnecessary impacts resulting from feedback of PSIFs which turn out not to be serious. We explored the impact of feedback of PSIFs in Chapter 4 in terms of the clinical assessments generated and impacts on participants' emotional wellbeing, insurance and finances and work and activities. The following chapter will build further on this work by evaluating the economic impact of feedback of PSIFs on hospital services.

Chapter 6 Economic impact of potentially serious incidental findings on hospital services

6.1 Introduction

Chapter 1 outlined the challenges of studying the economic impact of feedback of incidental findings (IFs). Due to the range of different types of IFs that can occur across body regions, care will be required from many different health services, and a comprehensive assessment of costs will therefore be complex. While a number of studies have been conducted (Bendix et al., 2011; Flicker et al., 2008; Gluecker et al., 2003; Goehler et al., 2014; Lee et al., 2010; Machaalany et al., 2009; Maizlin et al., 2007; Mutneja et al., 2017; Pickhardt et al., 2008; Schramm et al., 2016; Veerappan et al., 2010; Wagner et al., 2002), the majority have focused on patient populations undergoing single-region clinical computed tomography (CT), and most were conducted in non-UK health systems, which limits their generalisability to apparently asymptomatic people undergoing multi-region research magnetic resonance imaging (MRI) within the UK.

Our systematic review (Chapter 2) demonstrated that there are limited published data on the long-term systematic follow-up of unselected apparently asymptomatic participants with IFs, and that which is available does not enable the use and costs of health services during follow-up to be quantified. Similarly, while UK Biobank perform long-term systematic follow-up of their participants with potentially serious IFs (PSIFs), the follow-up surveys do not enable quantification of clinical assessments (i.e., binary data are collected on whether or not a participant saw a specialist doctor, but not on the number of outpatient appointments (Chapter 4) (Gibson et al., 2018)). Without quantification of the types of follow-up, costs cannot be attached. The impact of feedback of PSIFs detected on research MRI of apparently asymptomatic volunteers on use or costs of publicly-funded UK health services is still not clearly understood.

However, simply quantifying the health service uses and costs associated with the clinical assessment of PSIFs, or with serious final diagnoses, will be of limited value. UK Biobank participants are older adults, and as such will already use health services for any current health concerns, symptoms or diagnoses. It is not known if PSIFs, or serious final diagnoses, result in any significant increase in use or cost of health services (or indeed which services bear the majority of these burdens), either compared to such participants' uses and costs before they received feedback of a PSIF, or to control participants without PSIFs.

The UK Biobank cohort offers a unique opportunity to study the economic impact of PSIFs given the availability of linked healthcare data for participants with and without PSIFs, which covers the time periods both before and after feedback of PSIFs. In addition, the UK Biobank participants with PSIFs are systematically followed-up through participant and GP surveys to collect data on final diagnoses. The majority of PSIFs turn out not to be serious, and as such any follow-up may be deemed unnecessary (Chapter 2, Chapter 4) (Gibson et al., 2018). Understanding the relative impacts on health services of subgroups of participants with serious and with non-serious final diagnoses will inform judgements on the benefits and harms of feedback, and in turn influence the design of policies to handle IFs.

This chapter describes a study which used data from the UK Biobank Imaging Study to assess the NHS hospital costs generated by feedback of PSIFs. We used data from cases with PSIFs and controls without, both before and after feedback of PSIFs, to conduct four-way comparisons of the hospital contacts and costs generated by: cases after versus controls after feedback; cases after versus cases before feedback; cases before versus controls before feedback; controls after versus controls before feedback. We explored potential explanations for any observed differences in costs. As hospital data linkages were complete, and primary care data linkages were incomplete for UK Biobank participants at the time of this study, the comparisons are of hospital use and costs.

This study manuscript has been edited by co-authors in preparation for submission to a journal; the current version of the draft manuscript is included in full in Section 6.2, and supplementary materials are included in full in Section 6.3.

6.2 Use and cost of hospital services by UK Biobank participants with potentially serious incidental findings: a case-control and before-after study utilising linked English Hospital Episode Statistics data from 2013-2016

Authors and affiliations

Lorna M. Gibson,[1] John Nolan,[1] Edouard Mathieu,[2] Thomas J. Littlejohns,[3] Steve Garratt,[4] Simon Sheard,[4] Nicola Doherty,[4] Catriona Keerie,[1] Naomi E. Allen,[3] Joanna M. Wardlaw,[5] José Leal,[6] Alastair M. Gray,[6] Cathie L.M. Sudlow.[1]

1 Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh Bioquarter, 9 Little France Road, Edinburgh, EH16 4UX

2 Nuffield Department of Primary Care Health Sciences, University of Oxford, Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG

3 Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Big Data Institute, Old Road Campus, Oxford, OX3 7LF

4 UK Biobank Co-ordinating Centre, 1-2 Spectrum Way, Adswood, Stockport, SK3 0SA

5 Centre for Clinical Brain Sciences and UK Dementia Research Institute, University of Edinburgh, Chancellor's Building, 49 Little France Crescent, Edinburgh EH16 4SB

6 University of Oxford, Health Economics Research Centre, Nuffield Department of Population Health, Richard Doll Building, Old Road Campus, Oxford, OX3 7LF

Correspondence

Professor Cathie L.M. Sudlow, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh Bioquarter, 9 Little France Road, Edinburgh, EH16 4UX. Cathie.Sudlow@ed.ac.uk, 0131 465 9619.

6.2.1 Abstract

Background

Potentially serious incidental findings (PSIFs) on research imaging lead to clinical assessment in almost all cases. Previous studies have been small, or focused on patient populations, or non-UK health services. We assessed the economic impact on UK hospital services of feeding back PSIFs to population-based apparently asymptomatic participants of the UK Biobank Imaging Study.

Methods

We matched (by age, sex, imaging date and morbidity score) 179 cases with PSIFs on either magnetic resonance imaging or dual-energy X-ray absorptiometry to controls without. We attached National Reference Costs to linked Hospital Episode Statistics inpatient, outpatient, accident and emergency and critical care datasets from 2013-2016. Using data from cases and controls during the year before and after feedback of a PSIF, we conducted four-way comparisons of hospital contacts and costs using Wilcoxon signed rank tests, and proportions with ≥ 1 hospital contact using McNemar's tests, and plotted cumulative costs.

Results

There were no differences in hospital contacts or costs between cases and controls before feedback, or between controls before and after feedback of a PSIF. Following feedback, 144 (80.5%) cases, and 94 (52.5%) controls used hospital services; cases' median numbers of hospital contacts and median costs were significantly higher compared to controls, and to the year before (hospital contacts: three versus one and versus one; costs: £522 versus £114 and versus £128 [all $p<0.001$]). Rates of cases' cumulative costs began to increase approximately 30-60 days following feedback of a PSIF. A year after feedback of a PSIF, total cases' hospital costs (£431,114) were higher than controls' (£147,817, 2.9-fold) and cases' costs the year before (£167,434, 2.6-fold); these increases were greater in serious than non-serious cases (10.5-fold, 1.9-fold respectively). Most PSIFs were non-serious (158/179 [88%]) and non-serious cases generated greater total absolute costs than serious cases (£239,021 versus £192,093). These patterns of costs persisted over longer follow-up. After feedback of a PSIF, the majority of cases' cost and service use impacts were borne by inpatient (68.3%) and outpatient services (82.3%) respectively.

Conclusions

After feedback of a PSIF, research volunteers use substantially more hospital services than controls, and compared to the year before; the majority of cost and service impacts are borne by inpatient and outpatient services respectively. Absolute cost and service impacts are higher in cases with non-serious, rather than serious final diagnoses, as most PSIFs represent non-serious disease; avoidance of unnecessary feedback through the design of appropriate PSIFs policies would enable researchers and policymakers to minimise unwarranted impacts on publicly-funded healthcare services.

6.2.2 Introduction

While imaging is increasingly used to collect data for research, it may generate incidental findings (IFs) unrelated to the aim of the study (Wolf et al., 2008). IFs vary widely in the clinical severity of their associated final diagnosis, and it is of limited value to feed back clearly non-serious IFs to individuals and their healthcare team. We therefore focus on potentially serious IFs (PSIFs), i.e. findings which indicate the possibility of a condition which, if confirmed, would carry a real prospect of seriously threatening life span, or of having a substantial impact on major body functions or quality of life (Gibson et al., 2018).

Unlike clinical imaging, research imaging is usually not optimised to accurately diagnose an abnormality. Hence, feedback of PSIFs generates some form of clinical assessment in almost all cases in order to resolve diagnostic uncertainty and determine the appropriate clinical management (Gibson et al., 2018; Gibson et al., 2017b). Any structural disease affecting any body region (such as a tumour or an aneurysm) may be detected as a PSIF. As such, people with PSIFs may require care from a broad range of health services, including primary care, outpatient specialty appointments, imaging and other diagnostic services, invasive procedures and/or hospital admission (Gibson et al., 2018; Gibson et al., 2017b). This makes assessment of the economic impact of feedback of PSIFs on health services challenging.

Our knowledge of the economic impact of IFs is mostly informed by studies performed in the USA, the results of which may not be generalisable to the UK or other socialised health systems (Bendix et al., 2011; Flicker et al., 2008; Gluecker et al., 2003; Goehler et al., 2014; Lee et al., 2010; Machaalany et al., 2009; Maizlin et al., 2007; Morgan et al., 2015; Pickhardt et al., 2008; Veerappan et al., 2010; Wagner et al., 2002). Furthermore, isolated data on the costs of clinical assessment of people with PSIFs may be of limited usefulness to policymakers without information on the usual costs of healthcare for comparison; such context is paramount to enable judgments of the significance of any financial impact on health services. Most PSIFs turn out not to represent serious disease (Gibson et al., 2018), so the resulting costs generated during investigation to determine their non-serious nature may be deemed in retrospect to have been unnecessary. Empirical data on the economic impact of feedback of PSIFs on publicly-funded healthcare services (and the relative impacts generated by cases with serious or with non-serious final diagnoses) will inform judgments of the benefits and harms of feedback of PSIFs, and pragmatic approaches to the design of policies for handling imaging IFs.

The UK Biobank Imaging Study presents a key opportunity to investigate the economic impact of feedback of PSIFs in research volunteers. Brain, cardiac and body MRI, carotid Doppler ultrasound and dual-energy X-ray absorptiometry imaging of 100,000 largely asymptomatic adults already participating in UK Biobank is ongoing (over 27,000 have been imaged by September 2018 (UK Biobank, 2018c)). Linked routinely collected UK National Health Service (NHS) hospital data are available for all UK Biobank participants (Matthews and Sudlow, 2015) as well as systematic follow-up data from participant and general practitioner (GP) surveys to determine the final diagnoses of participants with PSIFs (Gibson et al., 2018).

Using data from participants in the UK Biobank Imaging Study, we aimed to assess the impact on NHS hospital services in terms of hospital contacts (which we define as inpatient admissions, outpatient appointments, accident and emergency care attendances and critical care admissions) and associated costs generated by feedback of PSIFs. To do this, we used data from cases with PSIFs (and subgroups with serious and non-serious final diagnoses) and controls without PSIFs, both before and after feedback of PSIFs, to conduct four-way comparisons of the hospital contacts and costs generated by: cases after versus controls after feedback; cases after versus cases before feedback; cases before versus controls before feedback; and controls after versus controls before feedback. We aimed to explore potential explanations for any observed differences in costs between groups. As hospital data linkages were complete, and primary care data linkages were incomplete for UK Biobank participants at the time of this study, this study focuses on hospital use and costs.

6.2.3 Methods

We prepared this manuscript according to RECORD guidelines (Benchimol et al., 2015). UK Biobank provided all imaged participants with written information about the imaging study and the UK Biobank imaging IFs protocol (UK Biobank, 2016). All participants provided written consent to take part in the imaging study and for UK Biobank to feed back any PSIFs to them and their GP. UK Biobank obtained specific ethics approval for the imaging study, and for evaluation of the PSIFs protocol (Research Ethics Committee reference numbers: 11/NW/0382; 16/NW/0274).

6.2.3.1 UK Biobank Imaging Study

UK Biobank is a population-based prospective cohort study that recruited half a million participants between 2006 and 2010 from England, Scotland and Wales. Between April and October 2014, UK Biobank invited participants living within approximately 120 km of the imaging centre in Stockport to the first phase of the UK Biobank Imaging Study (Gibson et al., 2018; UK Biobank, 2015e). Participants with metal implants, penetrating metal injury and non-removable metallic items were excluded, as were those likely to have difficulties completing the imaging, e.g. people with claustrophobia (Figure 6-1) (Gibson et al., 2018; UK Biobank, 2015e).

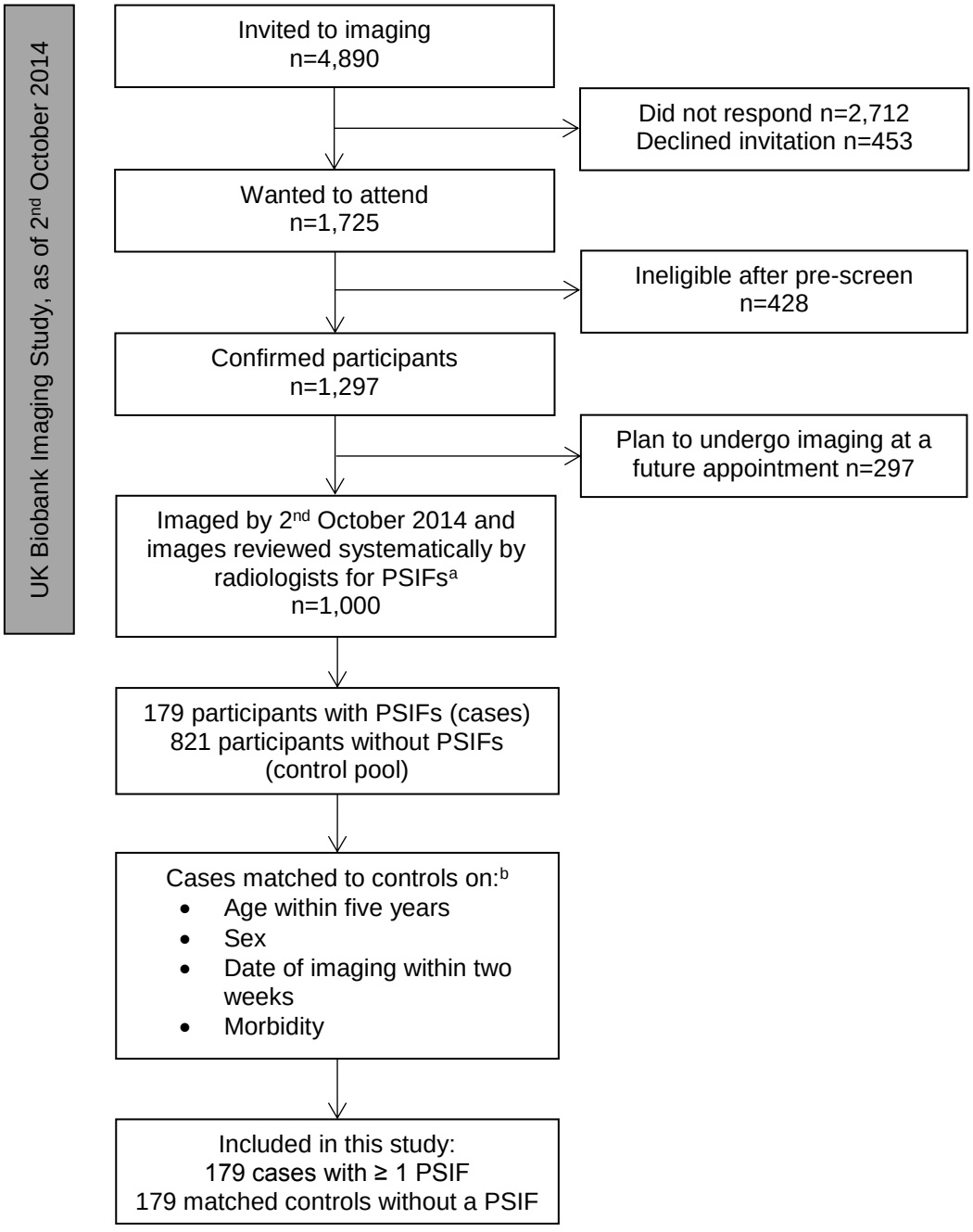


Figure 6-1: Participant flowchart

PSIFs = potentially serious incidental findings

a. **Radiologists systematically reviewed 942 brain, 948 cardiac and 944 body magnetic resonance imaging (MRI) scans and 997 dual-energy X-ray absorptiometry (DXA) scans from the first 1,000 imaged UK Biobank participants for PSIFs. PSIFs were defined as those 'indicating the possibility of a condition which, if confirmed, would carry a real prospect of seriously threatening life span, or of having a substantial impact on major body functions or quality of life (Gibson et al., 2018).'**

b. **Cases were matched to controls on: age and sex (which may influence health and access to health services); date of imaging within two weeks (to reduce differences in lengths of follow-up within Hospital Episode Statistics (HES) data which may bias calculations of hospital service use and cost); and morbidity (calculated using a binary Elixhauser Index score (Elixhauser et al., 1998; Quan et al., 2005) of 0 or ≥ 1 using HES data from the two years before the date of imaging, to reduce the risk of bias which may occur following changes in health resource group coding structures in 2012 (The National Casemix Office and Health and Social Care Information Centre, 2014).**

At the imaging visit, participants underwent brain, heart and body MRI, whole-body dual-energy X-ray absorptiometry (DXA) and carotid Doppler ultrasound (UK Biobank, 2015a; b; c; d; 2017b). Carotid Doppler ultrasound was deemed unlikely to generate PSIFs as the clinical relevance of asymptomatic carotid stenosis is not well established, and extra-carotid abnormalities were not likely to be relevant as UK Biobank sonographers are trained in vascular Doppler US only (Gibson et al., 2018); therefore carotid Doppler ultrasound was not included in this study.

6.2.3.2 UK Biobank PSIFs protocol and classifications of final diagnoses

Images from the first 1,000 participants were reviewed systematically by radiologists, who detected PSIFs (most commonly tumours and aneurysms) in 179 participants (Gibson et al., 2018). UK Biobank sent two surveys to participants with PSIFs, six weeks and six months after giving feedback, and sent surveys and requests for copies of relevant clinical correspondence to their GPs six months after feedback. The surveys collect data on follow-up and final diagnoses of PSIFs (UK Biobank, 2018b); the classification of these and summary diagnoses for these 179 participants have been published previously (Gibson et al., 2018). In brief, a consultant physician and an experienced specialty trainee in clinical radiology independently classified final diagnoses as either: serious (if they were likely to threaten life span, or have a substantial impact on quality of life or major body function); not serious (if this was not the case or if available data suggested that the diagnosis was already known); or uncertain (if there remained insufficient data to classify them as serious or not). We used the most serious final diagnosis to classify participants with more than one PSIF (Gibson et al., 2018).

6.2.3.3 Matching of cases and controls

We matched each of the 179 cases with PSIFs to one of the available 821 controls without a PSIF on age (within five years), sex, date of imaging (within two weeks, to reduce bias generated by differing lengths of follow-up period) and morbidity score (Figure 6-1). We defined the latter by calculating Elixhauser Index scores (Elixhauser et al., 1998; Quan et al., 2005) using linked Hospital Episode Statistics (HES) inpatient data covering the two years prior to the date of imaging (The National Casemix Office and Health and Social Care Information Centre, 2014). We matched participants on a binary score (0 or ≥ 1), as the majority of participants did not have a health condition listed in the Elixhauser Index.

6.2.3.4 Obtaining linked HES data and attaching costs

HES data were linked at individual-level to UK Biobank participants' data by NHS Digital matching algorithms; the quality of matching was assessed using matched rank scores (UK Biobank, 2014b). NHS Digital de-duplicated records, and both NHS Digital and UK Biobank checked and cleaned HES data to optimise the availability of valid, correctly formatted codes (HES Data Quality Team, 2016a; b; UK Biobank, 2013b). However, some invalid codes and duplicate records could still exist. We obtained linked HES admitted patient care (i.e. inpatient), outpatient, accident and emergency and critical care (together termed 'hospital') data from financial years 2013-2016 from the UK Biobank linked healthcare datasets for all cases and controls. We double-checked for duplicate records using NHS Digital methods (HES Data Quality Team, 2016b), and based on the assumption that no patient had multiple outpatient appointments on the same date with the same consultant within the same specialty after being referred on the same date by the same GP; we amalgamated such duplicate records and retained the maximum number of available treatment and diagnosis codes.

To attach costs to each record, we first generated currency codes (known as healthcare resource groups [HRGs]), using HRG4+ 2016-2017 Reference Costs Grouper software (NHS Digital, 2017b) (NHS Digital, UK) and NHS Digital methodology (NHS Digital, 2017a). Errors generated by the HRG4+ 2016-2017 Reference Costs Grouper software due to ICD-10 codes which were subsequently deleted following the introduction of the 5th edition of ICD-10 (n=34 inpatient records, 21 cases, 13 controls) were checked by a doctor (6-years qualified), who chose an appropriate updated 5th edition ICD-10 code from lists provided by NHS Digital blind to cost and PSIF final diagnosis (The National Casemix Office and NHS Digital, 2017). There was insufficient diagnostic information available for one control inpatient record, which was excluded from analyses as a HRG could not be

generated. Second, we attached national average unit costs to HRGs using the National Schedule of Reference Costs 2016-2017 (NHS Improvement, 2017) to enable comparisons across the different years of data (i.e. 2013-2016).

None of the included cases or controls had died, moved outside of England or withdrawn from UK Biobank prior to March 31st 2016, so we censored HES data follow up for all cases and controls at this date.

6.2.3.5 Statistical analyses

We calculated the frequency distributions of baseline characteristics between cases with PSIFs and controls without. We defined the date of feedback of PSIFs as one day after UK Biobank sent feedback letters to cases via first class postal mail. For these analyses, we defined each control's 'date of feedback' as the same as their matched case. We tested for normal distributions of continuous cost variables by visual inspection of graphed data and Kolmogorov-Smirnov goodness-of-fit tests; substantial proportions of both cases and controls had no costs, and given this skewed data we conducted comparisons using non-parametric statistical analyses (Min and Agresti, 2002).

Using data from cases and controls from the year before and the year after feedback of a PSIF, we performed four-way comparisons of: 1) cases after versus controls after feedback; 2) cases after versus cases before feedback; 3) controls after versus controls before feedback; and 4) cases before versus controls before feedback. In addition, we compared hospital costs between cases and controls from the entire available follow-up period after feedback of a PSIF. We used McNemar's tests to compare proportions of discordant case-control and before-and-after feedback pairs with 0 versus ≥ 1 hospital contacts. We used related-samples Wilcoxon signed rank tests to compare the numbers of hospital contacts, and hospital costs, between groups. We repeated all analyses for subgroups of cases with serious and non-serious final diagnoses. We considered p-values of <0.05 to be significant.

We used SPSS version 22 for analyses.

6.2.4 Results

6.2.4.1 Cases and controls

Of the 179 included cases, 21 (11.7%) had serious final diagnoses (mostly suspected tumours and aneurysms) and 158 (88.3%) had non-serious final diagnoses (Gibson et al., 2018).

Cases and controls were sufficiently matched on age (median age of 64, interquartile range [IQR] 57–69 years), sex (99/179 [55.3%] women), duration of available follow-up (median number of days 575 [IQR 551–623]) and on a binary measure of morbidity (32/179 [17.9%] had at least one Elixhauser Index health condition) (Supplementary Tables 6-1 and 6-2). Both groups were predominantly white, had not used private healthcare, drank alcohol in moderation, were non-smokers, and were either of normal body mass index (BMI) or overweight. A higher proportion of controls ate five or more portions of fruit and vegetables per day compared to cases (40.8% versus 30.2%) and undertook moderate exercise on at least three days per week (72.1% versus 64.3%). Cases had a higher median Townsend Deprivation Score (i.e. more deprived) (-1.87 [IQR -3.70–0.31]) compared to controls (-2.47 [IQR -3.95–0.13]) (Supplementary Table 6-1).

6.2.4.2 Background hospital contacts and costs

6.2.4.2.1 Cases versus controls before feedback of a PSIF

Before feedback, there were no significant differences between cases and controls with respect to: the proportions with ≥ 1 hospital contact (Table 6-1); their median numbers of total hospital (or inpatient, outpatient, emergency or critical care) contacts (Figure 6-2, Table 6-2) or costs (Figure 6-3, Table 6-3).

6.2.4.2.2 Controls' hospital contacts and costs following feedback of a PSIF to their matched case

Comparing the year before and the year after feedback of a PSIF (to their matched cases), there were no significant differences between: the proportions of controls with ≥ 1 hospital contact (Table 6-1); controls' median numbers of hospital contacts (Figure 6-2, Table 6-2) or costs (Figure 6-3, Table 6-3).

Table 6-1: Numbers (%) of 179 cases (21 with serious and 158 with non-serious final diagnoses) and their 1:1 matched controls with ≥ 1 hospital contact during the year before and the year after feedback of a potentially serious incidental finding

n (%) of participants (179 cases [21 with serious and 158 with non-serious final diagnoses] and matched controls) with ≥ 1 hospital contact					Case-control and before-after comparisons							
	After feedback of a PSIF		Before feedback of a PSIF		Cases after versus controls after		Cases after versus cases before		Controls after versus controls before		Cases before versus controls before	
	Cases n (%)	Controls n (%)	Cases n (%)	Controls n (%)	N discordant pairs (%)[a]	p-value[b]	N discordant pairs (%)[a]	p-value[b]	N discordant pairs (%)[a]	p-value[b]	N discordant pairs (%)[a]	p-value[b]
Total hospital	144 (80.4)	94 (52.5)	98 (54.7)	92 (51.4)	86 (48.0)	<0.001	58 (32.4)	<0.001	50 (27.9)	0.888	80 (44.7)	0.576
Serious	21 (100)	14 (66.7)	14 (66.7)	10 (47.6)	7 (33.3)	0.016[c]	7 (33.3)	0.016[c]	4 (19.0)	0.125[c]	10 (47.6)	0.344[c]
Non-serious	123 (77.8)	80 (50.6)	84 (53.2)	82 (51.9)	79 (50.0)	<0.001	51 (32.3)	<0.001	46 (29.1)	0.883	70 (44.3)	0.905
Inpatient	58 (32.4)	32 (17.9)	33 (18.4)	34 (19.0)	66 (36.0)	0.002	67 (37.4)	0.003	36 (20.1)	0.868	35 (19.6)	1.000
Serious	13 (61.9)	8 (38.1)	4 (19.0)	6 (28.6)	15 (71.4)	0.302[c]	13 (61.9)	0.022[c]	4 (19.0)	0.625[c]	6 (28.6)	0.687[c]
Non-serious	45 (25.1)	24 (13.4)	29 (16.2)	28 (15.6)	51 (32.3)	0.005	54 (34.2)	0.040	32 (20.3)	0.597	29 (18.4)	1.000
Outpatient	140 (78.2)	88 (49.2)	93 (52.0)	88 (49.2)	90 (50.3)	<0.001	59 (33.0)	<0.001	48 (26.8)	1.000	75 (41.9)	0.644
Serious	21 (100)	13 (61.9)	13 (61.9)	10 (47.6)	8 (38.1)	0.008[c]	8 (38.1)	0.008[c]	3 (14.3)	0.250[c]	9 (42.9)	0.508[c]
Non-serious	119 (66.5)	75 (41.9)	80 (44.7)	78 (43.6)	82 (51.9)	<0.001	51 (32.3)	<0.001	45 (28.5)	0.766	66 (41.8)	0.902

	n (%) of participants (179 cases [21 with serious and 158 with non-serious final diagnoses] and matched controls) with ≥ 1 hospital contact				Case-control and before-after comparisons							
	After feedback of a PSIF		Before feedback of a PSIF		Cases after versus controls after		Cases after versus cases before		Controls after versus controls before		Cases before versus controls before	
	Cases n (%)	Controls n (%)	Cases n (%)	Controls n (%)	N discordant pairs (%)[a]	p-value[b]	N discordant pairs (%)[a]	p-value[b]	N discordant pairs (%)[a]	p-value[b]	N discordant pairs (%)[a]	p-value[b]
Continued												
Emergency	36 (20.1)	30 (16.8)	23 (12.8)	29 (16.2)	38 (21.2)	0.471	47 (26.3)	0.079	45 (25.1)	1.000	46 (25.7)	0.461
Serious	7 (33.3)	5 (23.8)	5 (23.8)	1 (4.8)	8 (38.1)	0.727[c]	10 (47.6)	0.754[c]	6 (28.6)	0.291[c]	4 (19.0)	0.125[c]
Non-serious	29 (16.2)	25 (14.0)	18 (10.1)	28 (15.6)	30 (19.0)	0.636	37 (23.4)	0.099	39 (24.7)	0.749	42 (26.6)	0.164
Critical care	6 (3.4)	0 (0)	2 (1.1)	0 (0)	6 (3.4)	0.031[c]	8 (4.5)	0.289[c]	0 (0)	1.000	2 (1.1)	0.500[c]
Serious	6 (28.6)	0 (0)	0 (0)	0 (0)	6 (28.6)	0.031[c]	6 (28.6)	0.031[c]	0 (0)	1.000	0 (0)	1.000
Non-serious	0 (0)	0 (0)	2 (1.1)	0 (0)	0 (0)	1.000	2 (1.3)	0.500[c]	0 (0)	1.000	2 (1.3)	0.500[c]

a. Only pairs with differences in binary (0 or ≥ 1) numbers of contacts (i.e. discordant pairs) are included in the McNemar test.

b. Exact two-tailed p-value from McNemar test.

c. Binomial distribution used.

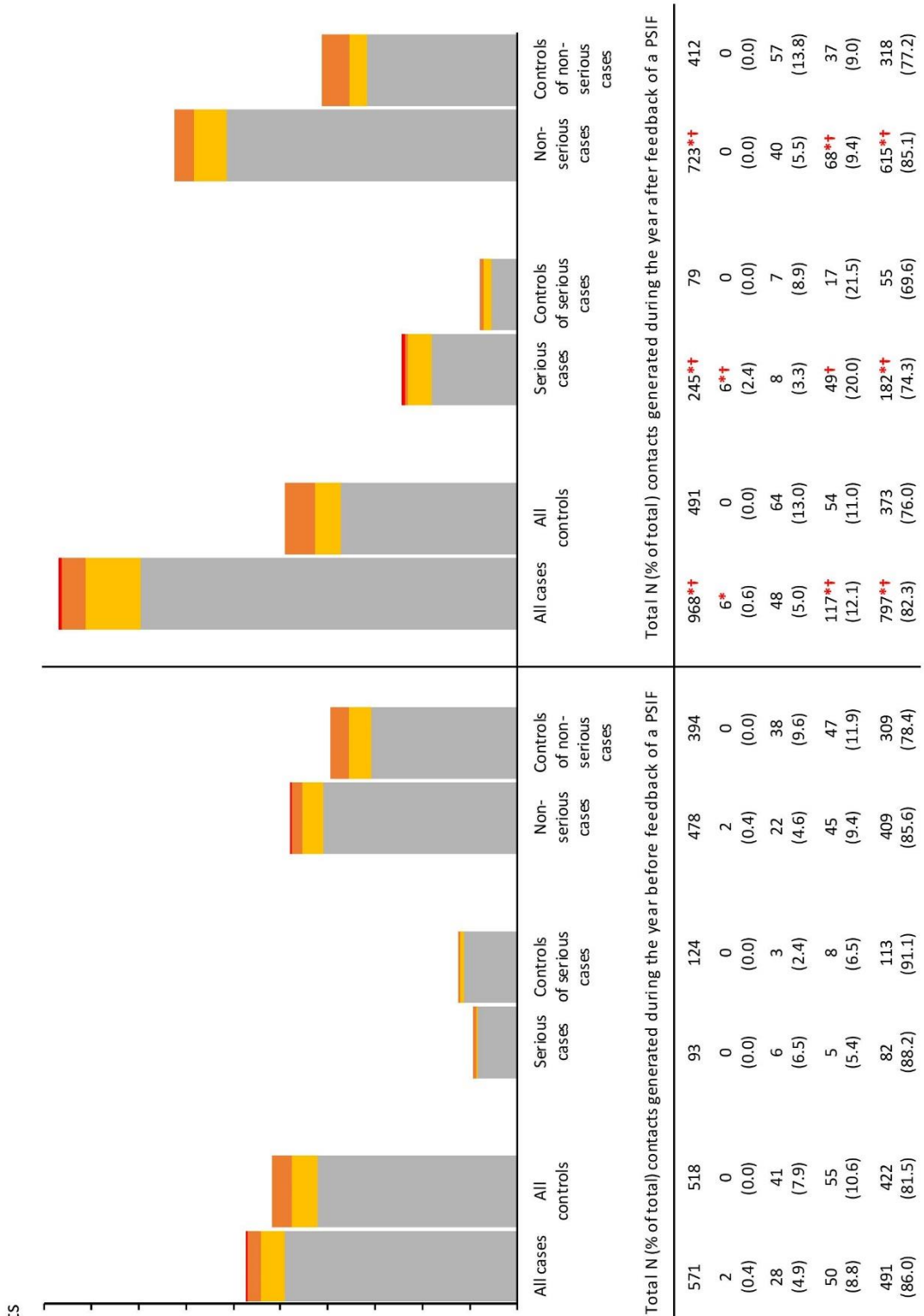


Figure 6-2: Total numbers (% of total) of hospital contacts generated by 179 cases (21 with serious and 158 with non-serious final diagnoses) and their 1:1 matched controls during the year before and after feedback of a potentially serious incidental finding

PSIF = potentially serious incidental finding.

*** Statistically significantly higher median numbers of contacts generated by cases compared to controls during the year after feedback of a PSIF. See Table 6-2 for details.**

† Statistically significantly higher median numbers of contacts generated by cases during the year after feedback of a PSIF, compared to the year before. See Table 6-2 for details.

There were no statistically significant differences in median numbers of contacts generated: by controls during the year after feedback of a PSIF, compared to the year before; or by cases compared to controls during the year before feedback of a PSIF. See Table 6-2 for details.

Table 6-2: Median numbers (interquartile and full range) of hospital contacts generated by 179 cases (21 with serious and 158 with non-serious final diagnoses) and their 1:1 matched controls during the year before and after feedback of a potentially serious incidental finding

Median (IQR, range) numbers of hospital contacts by 179 cases (21 with serious and 158 with non-serious final diagnoses) and their matched controls					Case-control and before-after comparisons							
	After feedback of a PSIF		Before feedback of a PSIF		Cases after versus controls after		Cases after versus cases before		Controls after versus controls before		Cases before versus controls before	
	Cases	Controls	Cases	Controls	N discordant pairs (%)[b]	p-value[c]	N discordant pairs (%)[b]	p-value[c]	N discordant pairs (%)[b]	p-value[c]	N discordant pairs (%)[b]	p-value[c]
	Median (IQR, range)[a]	Median (IQR, range)[a]	Median (IQR, range)[a]	Median (IQR, range)[a]								
Total hospital	3 (1-8, 0-34)	1 (0-4, 0-27)	1 (0-4, 0-35)	1 (0-4, 0-38)	156 (87.2)	<0.001	139 (77.7)	<0.001	106 (59.2)	0.595	126 (70.4)	0.390
Serious	12 (5-15, 2-29)	2 (0-5, 0-16)	2 (0-6, 0-35)	0 (0-5, 0-38)	21 (100)	0.002	19 (90.5)	0.002	13 (61.9)	0.531	16 (76.2)	0.949
Non-serious	3 (1-6, 0-34)	1 (0-3, 0-27)	1 (0-4, 0-33)	1 (0-3, 0-28)	135 (85.4)	<0.001	120 (75.9)	<0.001	93 (58.9)	0.746	110 (69.6)	0.328
Inpatient	0 (0-1, 0-11)	0 (0-0, 0-6)	0 (0-0, 0-5)	0 (0-0, 0-8)	73 (40.8)	0.001	74 (41.3)	0.001	42 (23.5)	0.617	45 (25.1)	0.826
Serious	1 (0-3, 0-11)	0 (0-1, 0-6)	0 (0-0, 0-2)	0 (0-1, 0-2)	17 (81.0)	0.054	14 (66.7)	0.002	7 (33.3)	0.172	7 (33.3)	0.438
Non-serious	0 (0-1, 0-4)	0 (0-0, 0-6)	0 (0-0, 0-5)	0 (0-0, 0-8)	56 (35.4)	0.004	60 (38.0)	0.036	35 (22.2)	0.196	38 (24.1)	0.929
Outpatient	3 (1-7, 0-31)	0 (0-3, 0-19)	1 (0-3, 0-32)	0 (0-3, 0-36)	150 (83.8)	<0.001	132 (73.7)	<0.001	100 (55.9)	0.596	121 (67.6)	0.274
Serious	9 (4-13, 2-17)	1 (0-4, 0-13)	2 (0-6, 0-32)	0 (0-4, 0-36)	21 (100)	0.001	19 (90.5)	0.002	12 (57.1)	0.198	15 (71.4)	0.903
Non-serious	2 (1-5, 0-31)	0 (0-3, 0-19)	1 (0-3, 0-32)	0 (0-2, 0-26)	129 (81.6)	<0.001	113 (71.5)	<0.001	88 (55.7)	0.973	106 (67.1)	0.219

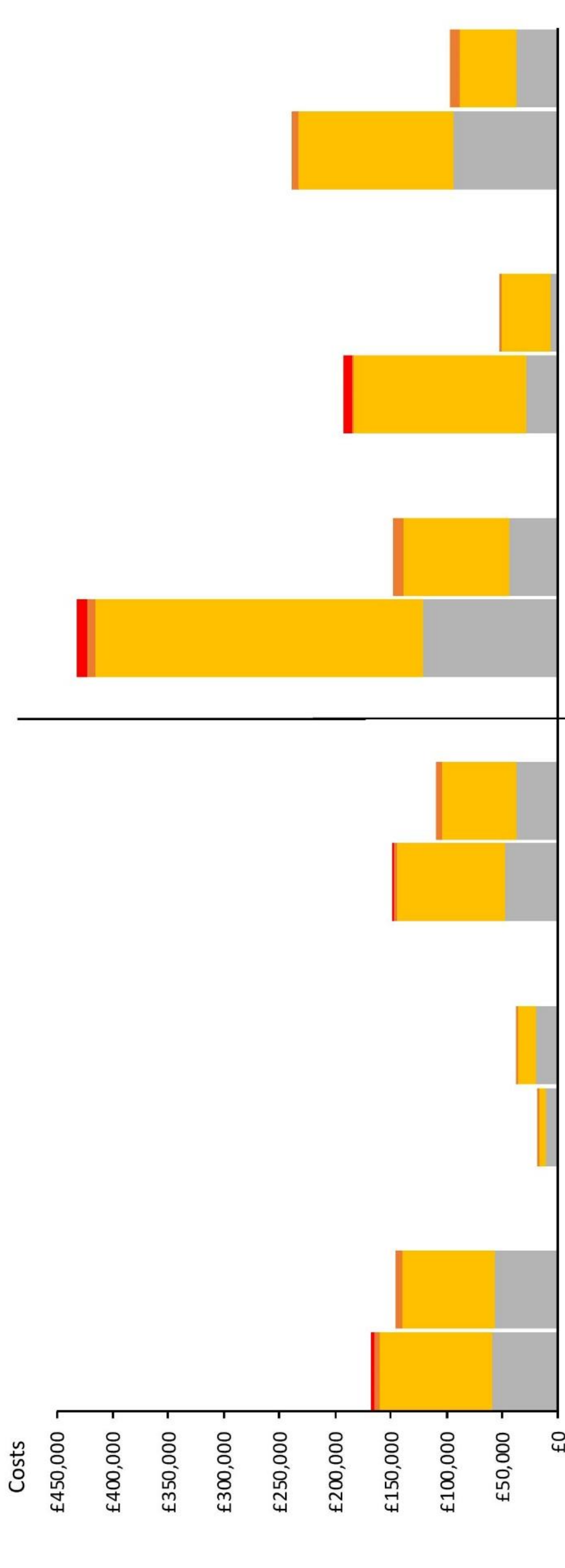
	Median (IQR, range) numbers of hospital contacts by 179 cases (21 with serious and 158 with non-serious final diagnoses) and their matched controls				Case-control and before-after comparisons							
	After feedback of a PSIF		Before feedback of a PSIF		Cases after versus controls after		Cases after versus cases before		Controls after versus controls before		Cases before versus controls before	
	Cases Median (IQR, range)[a]	Controls Median (IQR, range)[a]	Cases Median (IQR, range)[a]	Controls Median (IQR, range)[a]	N discordant pairs (%)[b]	p-value[c]	N discordant pairs (%)[b]	p-value[c]	N discordant pairs (%)[b]	p-value[c]	N discordant pairs (%)[b]	p-value[c]
Continued												
Emergency	0 (0-0, 0-3)	0 (0-0, 0-24)	0 (0-0, 0-2)	0 (0-0, 0-4)	52 (29.1)	0.467	48 (26.8)	0.050	48 (26.8)	0.683	48 (26.8)	0.269
Serious	0 (0-1, 0-2)	0 (0-0, 0-2)	0 (0-0, 0-2)	0 (0-0, 0-3)	9 (42.9)	0.973	10 (47.6)	0.793	6 (28.6)	0.344	5 (23.8)	0.563
Non-serious	0 (0-0, 0-3)	0 (0-0, 0-24)	0 (0-0, 0-2)	0 (0-0, 0-4))	43 (27.2)	0.503	38 (24.1)	0.053	42 (26.6)	0.949	43 (27.2)	0.131
Critical care	0 (0-0, 0-1)	0 (0-0, 0-0)	0 (0-0, 0-1)	0 (0-0, 0-0)	6 (3.4)	0.031	8 (4.5)	0.289	0 (0)	1.000	2 (1.1)	0.500
Serious	0 (0-1, 0-1)	0 (0-0, 0-0)	0 (0-0, 0-0)	0 (0-0, 0-0)	6 (28.6)	0.031	6 (28.6)	0.031	0 (0)	1.000	0 (0)	1.000
Non-serious	0 (0-0, 0-0)	0 (0-0, 0-0)	0 (0-0, 0-1)	0 (0-0, 0-0)	0 (0)	1.000	2 (1.3)	0.500	0 (0)	1.000	2 (1.3)	0.500

IQR = interquartile range, PSIF = potentially serious incidental finding

a. 25th and 75th percentile values calculated using Tukey's Hinges.

b. Only pairs with differences in numbers of contacts (i.e. discordant pairs) are included in the related-samples Wilcoxon signed rank test.

c. Exact two-tailed p-value from related-samples Wilcoxon signed rank test for differences in numbers of contacts between groups.



	Year before feedback of a PSIF: All cases	All controls	Serious cases	Controls of serious cases	Non-serious cases	Controls of non-serious cases	Year after feedback of a PSIF: All cases	All controls	Serious cases	Controls of serious cases	Non-serious cases	Controls of non-serious cases
	Total costs in £ (% of total) generated during the year before feedback of a PSIF						Total costs in £ (% of total) generated during the year after feedback of a PSIF					
Total hospital	167,434	145,515	18,219	36,175	149,215	109,341	431,114*†	147,817	192,093*†	51,385	239,021*†	96,432
Critical care	2,332 (1.4)	0 (0.0)	0 (0.0)	0 (0.0)	2,332 (1.6)	0 (0.0)	7,642* (1.8)	0 (0.0)	7,642* (4.0)	0 (0.0)	0 (0.0)	0 (0.0)
Emergency	4,417 (2.5)	5,774 (4.0)	1,036 (5.7)	421 (1.2)	3,111 (2.1)	5,353 (4.9)	7,188 (1.7)	8,866 (6.0)	1,236 (0.6)	902 (1.8)	5,952 (2.5)	7,965 (8.3)
Inpatient	102,620 (61.3)	82,894 (57.0)	5,991 (32.9)	16,139 (44.6)	96,629 (64.8)	66,755 (61.1)	294,469*† (68.3)	95,326 (64.5)	155,040*† (80.7)	43,909 (85.5)	139,429*† (58.3)	51,417 (53.3)
Outpatient	58,334 (34.8)	56,847 (39.1)	11,191 (61.4)	19,615 (54.2)	47,143 (31.6)	37,233 (34.1)	121,815*† (28.3)	43,625 (29.5)	28,175*† (14.7)	6,575 (12.8)	93,639*† (39.2)	37,051 (38.4)

Figure 6-3: Total hospital costs in £ (% of total) generated by 179 cases (21 with serious and 158 with non-serious final diagnoses) and their 1:1 matched controls during the year before and the year after feedback of a potentially serious incidental finding

PSIF = potentially serious incidental finding.

*** Statistically significantly higher median costs in cases compared to controls during the year after feedback of a PSIF. See Table 6-3 for details.**

† Statistically significantly higher median costs in cases during the year after feedback of a PSIF, compared to the year before. See Table 6-3 for details.

There were no statistically significant differences in median costs generated: by controls during the year after feedback of a PSIF, compared to the year before; or by cases compared to controls during the year before feedback of a PSIF. See Table 6-3 for details.

Table 6-3: Median (interquartile and full range) hospital costs (£) generated by 179 cases (21 with serious and 158 with non-serious final diagnoses) and their 1:1 matched controls during the year before and after feedback of a potentially serious incidental finding

Median (IQR, range) hospital costs (£) generated by 179 cases (21 with serious and 158 with non-serious final diagnoses) and their matched controls					Case-control and before-after comparisons									
	After feedback of a PSIF		Before feedback of a PSIF		Cases after versus controls after		Cases after versus cases before		Controls after versus controls before		Cases before versus controls before			
	Cases	Controls	Cases	Controls	N discordant pairs (%)[b]	p-value[c]	N discordant pairs (%)[b]	p-value[c]	N discordant pairs (%)[b]	p-value[c]	N discordant pairs (%)[b]	p-value[c]		
	Median (IQR, range)[a]	Median (IQR, range)[a]	Median (IQR, range)[a]	Median (IQR, range)[a]										
Total hospital	522 (137-2441, 0-38995)	114 (0-635, 0-18917)	128 (0-864, 0-22798)	69 (0-456, 0-12943)	162 (90.5)	<0.001	148 (82.7)	<0.001	116 (64.8)	0.329	135 (75.4)	0.457		
Serious	6050 (906-16281, 262-38995)	349 (0-2783, 0-18917)	206 (0-581, 0-8612)	0 (0-2699, 0-12943)	21 (100)	0.019	21 (100)	0.001	14 (66.7)	0.583	17 (81.0)	0.712		
Non-serious	464 (115-1749, 0-13137)	59 (0-543, 0-10525)	109 (0-912, 0-22798)	71 (0-444, 0-8647)	141 (89.2)	<0.001	127 (80.4)	<0.001	102 (64.6)	0.208	118 (74.7)	0.337		
Inpatient	0 (0-1485, 0-36465)	0 (0-0, 0-18158)	0 (0-0, 0-20539)	0 (0-0, 0-8029)	78 (43.6)	<0.001	79 (44.1)	<0.001	50 (27.9)	0.644	51 (28.5)	0.987		
Serious	4672 (0-13843, 0-36465)	0 (0-2004, 0-18158)	0 (0-0, 0-2598)	0 (0-0, 0-4196)	18 (85.7)	0.043	15 (71.4)	0.001	9 (42.9)	0.359	8 (38.1)	0.109		
Non-serious	0 (0-1078, 0-9660)	0 (0-0, 0-9360)	0 (0-0, 0-20539)	0 (0-0, 0-8029)	60 (38.0)	<0.001	64 (40.5)	0.018	41 (26.0)	0.245	43 (27.2)	0.536		

	Median (IQR, range) hospital costs (£) generated by 179 cases (21 with serious and 158 with non-serious final diagnoses)and their matched controls				Case-control and before-after comparisons							
	After feedback of a PSIF		Before feedback of a PSIF		Cases after versus controls after		Cases after versus cases before		Controls after versus controls before		Cases before versus controls before	
	Cases Median (IQR, range)[a]	Controls Median (IQR, range)[a]	Cases Median (IQR, range)[a]	Controls Median (IQR, range)[a]	N discordant pairs (%)[b]	p-value[c]	N discordant pairs (%)[b]	p-value[c]	N discordant pairs (%)[b]	p-value[c]	N discordant pairs (%)[b]	p-value[c]
Continued												
Outpatient	363 (130-937, 0-9913)	0 (0-307, 0-2604)	103 (0-409, 0-5615)	0 (0-310, 0-10035)	159 (88.8)	<0.001	143 (79.9)	<0.001	110 (61.5)	0.569	128 (71.5)	0.165
Serious	1334 (787-1873, 262-2960)	141 (0-600, 0-1281)	202 (0-581, 0-5615)	0 (0-438, 0-10035)	21 (100)	<0.001	21 (100)	0.001	13 (61.9)	0.305	16 (76.2)	0.860
Non-serious	296 (58-673, 0-9913)	0 (0-293, 0-2604)	86 (0-403, 0-3034)	0 (0-294, 0-4227)	138 (87.3)	<0.001	122 (77.2)	<0.001	97 (61.4)	0.891	112 (70.9)	0.113
Emergency	0 (0-0, 0-654)	0 (0-0, 0-3778)	0 (0-0, 0-399)	0 (0-0, 0-467)	57 (31.8)	0.301	52 (29.1)	0.135	52 (29.1)	0.944	49 (27.4)	0.282
Serious	0 (0-119, 0-307)	0 (0-0, 0-268)	0 (0-0, 0-399)	0 (0-0, 0-421)	10 (47.6)	0.625	11 (52.4)	0.880	6 (28.6)	0.438	5 (23.8)	0.438
Non-serious	0 (0-0, 0-654)	0 (0-0, 0-3778)	0 (0-0, 0-386)	0 (0-0, 0-467)	47 (29.7)	0.359	41 (26.0)	0.115	46 (29.1)	0.531	44 (27.8)	0.110

	Median (IQR, range) hospital costs (£) generated by 179 cases [21 with serious and 158 with non-serious final diagnoses] and their matched controls				Case-control and before-after comparisons							
	After feedback of a PSIF		Before feedback of a PSIF		Cases after versus controls after		Cases after versus cases before		Controls after versus controls before		Cases before versus controls before	
	Cases Median (IQR, range)[a]	Controls Median (IQR, range)[a]	Cases Median (IQR, range)[a]	Controls Median (IQR, range)[a]	N discordant pairs (%)[b]	p-value[c]	N discordant pairs (%)[b]	p-value[c]	N discordant pairs (%)[b]	p-value[c]	N discordant pairs (%)[b]	p-value[c]
Continued												
Critical care	0 (0-0, 0-1633)	0 (0-0, 0-0)	0 (0-0, 0-1633)	0 (0-0, 0-0)	6 (3.4)	0.031	8 (4.5)	0.188	0 (0)	1.000	2 (1.1)	0.500
Serious	0 (0-905, 0-1633)	0 (0-0, 0-0)	0 (0-0, 0-0)	0 (0-0, 0-0)	6 (28.6)	0.031	6 (28.6)	0.500	0 (0)	1.000	0 (0)	1.000
Non-serious	0 (0-0, 0-0)	0 (0-0, 0-0)	0 (0-0, 0-1633)	0 (0-0, 0-0)	0 (0)	1.000	2 (1.3)	0.500	0 (0)	1.000	2 (1.3)	0.500

IQR = interquartile range, PSIF = potentially serious incidental finding

a. 25th and 75th percentile values calculated using Tukey's Hinges.

b. Only pairs with differences in costs (i.e. discordant pairs) are included in the related-samples Wilcoxon signed rank test.

c. Exact two-tailed p-value from related-samples Wilcoxon signed rank test for differences in costs between groups.

6.2.4.3 Cases' hospital contacts and costs following feedback of a PSIF

After feedback, 144 (80.5%) cases, and 94 (52.5%) controls had contact with hospital services and generated some hospital costs (i.e. substantial portions had no such costs) (Table 6-1). By one year after feedback of a PSIF, a significantly higher proportion of cases had had ≥ 1 hospital contact compared to controls, and compared to the year before (Table 6-1). After feedback of a PSIF, cases' median numbers of hospital contacts were significantly higher than controls, and compared to cases the year before (three versus one and versus one, both comparisons $p<0.001$ [Figure 6-2, Table 6-2]).

Rates of cases' hospital (and inpatient and outpatient) costs began to increase around 30–60 days following feedback of a PSIF, regardless of the severity of the final diagnosis, and continued at an increased rate for the remainder of the year (Figure 6-4). After feedback of a PSIF, cases generated higher median hospital costs compared to controls, and to the year before (£522 versus £114 and versus £128, both $p<0.001$); the same pattern of significantly higher median hospital costs was seen in subgroups of serious and non-serious cases [Figure 6-3, Table 6-3]). The total costs generated by the group of serious cases by one year after feedback was 10.5-fold higher than their total costs the year before, compared to a 1.9-fold increase for non-serious cases (Figure 6-3). However, the group of non-serious cases generated higher absolute total hospital costs than serious cases by one year after feedback (£239,021 versus £192,093, Figure 6-3) because of their greater number (n=158 versus n=21).

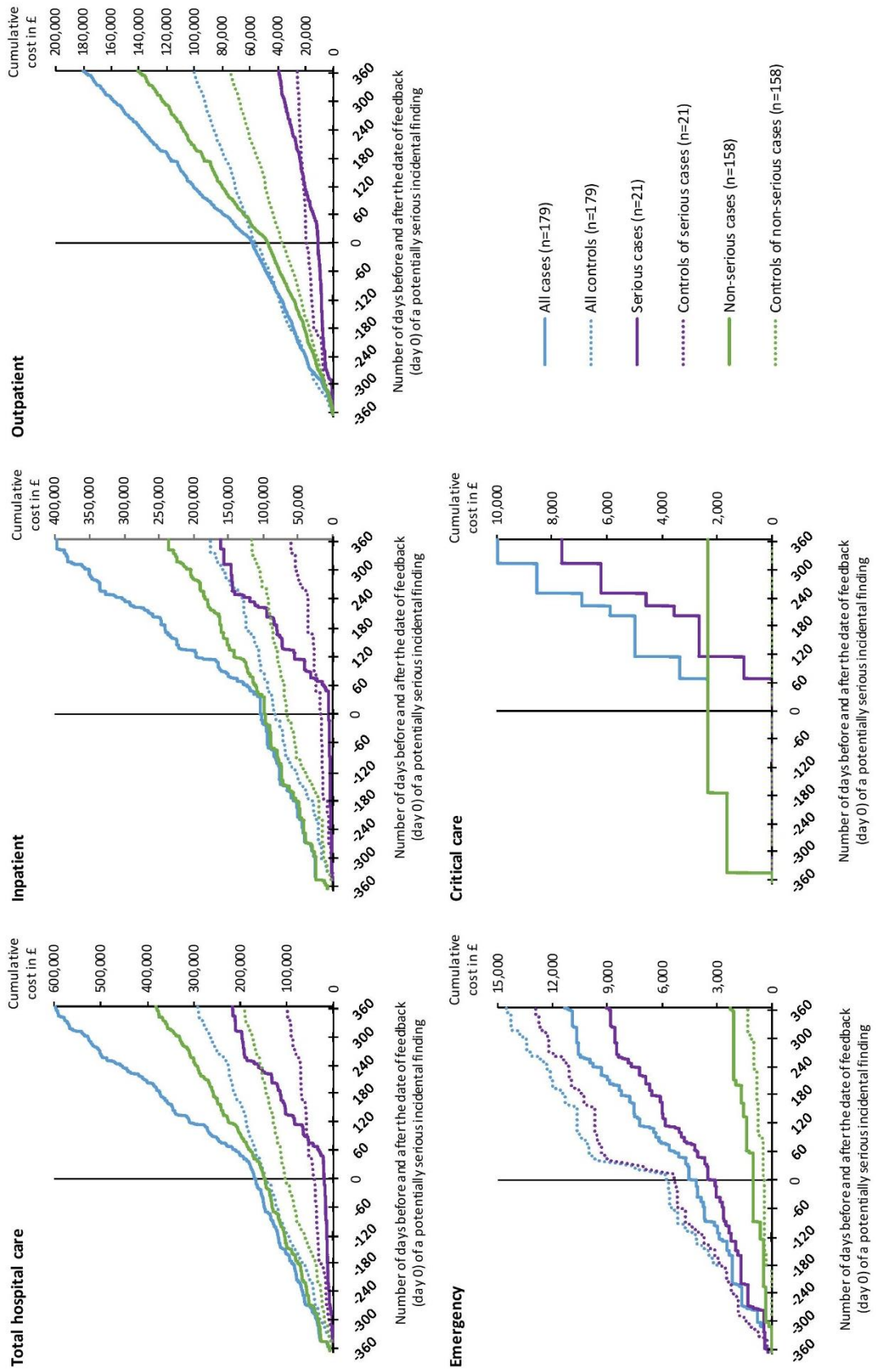


Figure 6-4: Cumulative hospital costs (£) generated by cases and controls over the year before and after feedback of a potentially serious incidental finding (time zero)

6.2.4.3.1 Inpatient care

Following feedback of a PSIF, higher proportions of cases (and serious and non-serious subgroups) had ≥ 1 hospital admission compared to the year before, and compared to controls (although this was not statistically significant for serious cases compared to their controls) (Table 6-1). This same pattern was evident for cases' median numbers of inpatient admissions (Table 6-2).

Inpatient costs accounted for the majority of the total hospital costs generated by cases during the year after feedback (68.3%), and for a higher proportion of serious cases' total hospital costs than non-serious cases' (80.7% versus 58.3% respectively) (Figure 6-3).

During the year after feedback of a PSIF, cases generated higher total inpatient costs compared to controls, and compared to the year before (3.1 and 2.9-fold higher respectively, Figure 6-3). The magnitude of total inpatient cost increase following feedback of a PSIF was higher for the group of serious cases compared to the non-serious cases (25.9-fold, 1.4-fold respectively) (Figure 6-3). After feedback of a PSIF, cases (and serious and non-serious subgroups) had significantly higher median inpatient costs compared to controls, and compared to the year before (Figure 6-3, Table 6-3).

Elective admissions contributed to the majority of cases' inpatient costs after feedback (80.1%), with smaller proportions contributed by non-elective admissions (17.6%), and other types of costs (1.3% [including chemotherapy, excess bed days, high cost drugs or rehabilitation]) (Supplementary Figure 6-1). After feedback of a PSIF, cases' median elective inpatient costs were significantly higher compared to controls', and to cases' costs the year before (both p=0.001) (Supplementary Figure 6-1, Supplementary Table 6-3).

6.2.4.3.2 Outpatient care

While inpatient care accounted for the majority of cases' hospital costs after feedback of a PSIF, outpatient care accounted for the majority of cases' hospital contacts (82.3%, Figure 6-2). Following feedback of a PSIF, a significantly higher proportion of cases had ≥ 1 outpatient appointment compared to controls and compared to cases the year before (Table 6-1), and the median number of outpatient appointments was significantly higher amongst cases compared to these two groups (Figure 6-2, Table 6-2).

One year after feedback of a PSIF, cases' total outpatient costs were higher than controls' and than cases' costs the year before (2.8 and 2.1-fold respectively), and cases' median outpatient costs were significantly higher than these two groups (both p<0.001 [Figure 6-3,

Table 6-3]). Compared to the year before feedback, both serious and non-serious cases' total outpatient costs had increased by a similar magnitude (2.5 and 2.0-fold respectively) and median outpatient costs were significantly higher (Figure 6-3, Table 6-3).

Consultant-led appointments accounted for 69.8% of cases' total outpatient costs after feedback of a PSIF; non-consultant-led appointments, procedures and other costs (including imaging, chemotherapy, high cost drugs, nuclear medicine and radiotherapy costs [the latter of which were generated by nine cases and three controls]) accounted for the remainder of the total outpatient costs (Supplementary Figure 6-2). Only a minority (2.0%) of cases' outpatient costs appeared to be attributable to imaging (n=21, Supplementary Figure 6-2). After feedback of a PSIF, median consultant-led appointment costs were higher for cases compared to controls, and to cases costs the year before (both $p<0.001$), but small numbers of case-control and before-and-after case pairs precluded any firm conclusion on any differences of other outpatient costs (Supplementary Table 6-4).

6.2.4.3.3 Emergency and critical care

After feedback of a PSIF, there were no significant differences between cases (or subgroups of serious and non-serious cases) and controls, or compared to the year before, regarding: proportions with ≥ 1 contact with emergency care services (Table 6-1); median number of emergency care contacts (Figure 6-2, Table 6-2); median emergency care costs (Figure 6-3, Table 6-3).

While the proportions of admitted cases, median number of contacts, and median costs of critical care were significantly higher for serious cases compared to their controls during the year after feedback of a PSIF (Tables 6-1, 6-2 and 6-3, Figures 6-2 and Figure 6-3), the absolute numbers of cases admitted to critical care, and their associated costs, were small (6/21 [28.6%] serious cases generated £7,642 of costs); no controls were admitted to critical care (Table 6-1, Figure 6-3).

6.2.4.4 Costs generated by cases and controls over a longer follow-up period

As HES data were available until the end of the 2015-2016 financial year, in total 282.5 person-years of follow-up was available for 179 cases and their controls. Magnitudes of total cost differences, and median cost differences during this longer period of follow-up were consistent with the analyses reported above, which were based on one year (i.e. 179 person-years) of follow-up. Cases generated 2.3-fold higher total costs compared to controls, with a greater difference in costs for serious cases than non-serious cases compared to their respective controls (3.9-fold higher, 1.9-fold higher respectively) (Supplementary Figure 6-

3). Cases (and serious and non-serious subgroups) generated significantly higher median costs of total hospital, inpatient and outpatient care compared to controls (Supplementary Table 6-5).

There were no significant differences in median costs of emergency care between cases (or serious or non-serious subgroups) and controls (Supplementary Table 6-5).

Again, serious cases had significantly higher median costs of critical care compared to controls (p=0.016) (Supplementary Table 6-5), but the absolute numbers of cases admitted to critical care, and their associated costs, remained small (7/21 [33.3%] serious cases generated £8,664); no controls were admitted to critical care (Supplementary Figure 6-3 and Supplementary Table 6-5).

6.2.5 Discussion

6.2.5.1 Main findings

Before feedback of a PSIF, there were no differences between numbers of hospital contacts or costs between cases and controls, and controls' numbers of hospital contacts and costs did not significantly differ during the year after compared to the year before feedback of a PSIF to their matched case. However, after feedback of a PSIF, 144 (80.5%) cases and 94 (52.5%) controls used hospital services, and cases' median numbers of hospital contacts and median hospital costs were significantly higher compared to controls, and compared to the year before (hospital contacts: three versus one and versus one; hospital costs: £522 versus £114 and versus £128 [all $p<0.001$]). Rates of cases' cumulative costs began to increase approximately 30–60 days following feedback of a PSIF. A year after feedback of a PSIF, total hospital costs were higher amongst cases compared to controls', and to cases' costs the year before (2.9- and 2.6-fold respectively); serious cases' total hospital costs had increased by a far greater relative magnitude than non-serious cases (10.5- and 1.9-fold respectively). However, the majority of PSIFs (88%) were finally diagnosed as non-serious, and non-serious cases generated greater absolute costs than serious cases (£239,021 versus £192,093). These patterns of cost differences between cases and controls persisted over a longer follow-up period. After feedback of a PSIF, the majority of the cost burden generated by cases was borne by inpatient services (68%), but the majority of their service burden fell to outpatient services (82%).

6.2.5.2 Strengths of the study

Images from the first 1,000 imaged participants were all reviewed systematically by radiologists for PSIFs under an established definition and protocol (Gibson et al., 2018). We

selected all the cases with PSIFs, and matched them to controls from this imaged group. Our selection of controls therefore satisfies two key requirements: they were selected from the same imaged population from which the cases are drawn and would have been selected as cases had they had a PSIF; all images from the controls were also reviewed by radiologists for PSIFs, and therefore their exposure status has been measured with the same accuracy as cases. These methods minimise selection and information bias respectively (Wacholder et al., 1992).

Our study is approximately ten-fold larger than the largest previous study to assess the economic impact of IFs in healthy volunteers undergoing MRI (N=18) (Pinato et al., 2012), over twice as large as similar studies of patient groups (N=65) (Hayes et al., 2016) (N=83) (Wagner et al., 2002), and larger than other previous studies of the economic impact of IFs studied within a UK health service (N=18-114) (Bromage et al., 2012; Pinato et al., 2012). Unlike previous studies, we did not limit our assessment of healthcare costs to a single type of clinical assessment (such as imaging or other diagnostic tests) (Flicker et al., 2008; Gluecker et al., 2003; Lee et al., 2010; Maizlin et al., 2007; Priola et al., 2013; Schramm et al., 2016; Veerappan et al., 2010; Wagner et al., 2002) but used HES datasets and well-established methods to calculate service use and costs generated across inpatient, outpatient, emergency and critical care services. To our knowledge, no previous study has attempted to provide a context for the costs generated by people with PSIFs by comparing these to either a control group without PSIFs, or cases' costs before feedback of a PSIF. By checking for censored records, and matching cases to controls who were imaged within a fortnight, we have minimised any bias which may have been introduced by variability of length of follow-up periods within the HES datasets. Two doctors rigorously applied a definition of PSIFs after going to considerable lengths to obtain data on final diagnoses from participants and their GPs in order to classify cases into subgroups of those with serious versus non-serious final diagnoses.

6.2.5.3 Limitations

The routinely collected data that our study is based on does not permit accurate classification of hospital contacts into those directly due to the clinical assessment of PSIFs and those which are unrelated. However, the magnitudes of differences in hospital contacts and costs between cases and controls after feedback of a PSIF, the lack of difference between these groups before feedback of a PSIF, and the increase in costs after compared to before feedback which occurred in cases, but not in controls, strongly implies that feeding back

PSIFs to apparently asymptomatic research volunteers has an impact on hospital use and costs.

We matched cases and controls using a binary Elixhauser Index score, rather than on exact conditions, as the latter method would have reduced the number of case-control pairs available for analyses. It is unlikely that the difference in conditions, or differences in other baseline characteristics, between small numbers of case-control pairs accounts entirely for the observed differences in hospital contacts and costs, but this could be explored further using adjusted modelling techniques and/or sensitivity analyses.

We followed up all participants within HES data for at least one year (i.e. 179 person-years), and for a maximum of 282.5 person-years; however, this may not be long enough for the full impact of diagnosis and treatment of all PSIFs to become manifest. A similar proportion of cases and controls reported having used private healthcare services, although the dates of these are not known. As stated previously, linked primary care data were not available during the time of this study, hence our focus on the impact on hospital services only. As a result of these limitations, our study likely underestimates the economic impact of feedback of PSIFs.

All the cases and controls underwent research imaging and attended healthcare services in England, and we used English National Reference Costs, which are averages that may result in under- or over-estimates of costs in individual cases, but enable comparisons across groups. The costs of healthcare will vary between countries, healthcare systems and over time, and so our cost results may not be directly generalisable beyond the English NHS. However, the magnitude of increased costs in cases and controls is worthy of further study in different healthcare systems.

Substantial proportions of participants (both cases and controls) did not generate any hospital cost, which is not surprising given our cohort is comprised of community-based research volunteers, who are healthier than the general population (Fry et al., 2017). This skewed cost data necessitated the use of non-parametric tests, which have lower statistical power than their parametric counterparts as they discard concordant pairs of data from analyses (Kirkwood and Stern, 2003). As such, we may have missed some significant differences. Conversely, due to the number of comparisons performed, some results may be statistically significant through chance. We considered performing a correction for multiple comparisons, but these may be too conservative and may potentially place too much

emphasis on the statistical p-value; instead, we report all of the p-values and interpret these in the context of the available data and the clinical plausibility (Perneger, 1998).

6.2.5.4 Comparisons with other studies

Compared with previously published self-reported data from the cases included in this current study and their GPs, data from HES records showed that fewer cases had undergone imaging after feedback of a PSIF (n=134 versus n=21), and more cases had attended outpatient appointments (n=109 versus n=140) (Gibson et al., 2018). These differences may be due to incomplete coding of imaging attendances within HES datasets, and reflect the longer follow-up period in this current study (1 year versus 6 months) respectively. The Diagnostic Imaging Dataset, which is compiled from hospital radiology information systems (NHS England, 2018), may capture cases' and controls' use of imaging services more completely than HES datasets, and should be considered for use in future studies.

We found that cases with non-serious final diagnoses generated higher total costs than those with serious final diagnoses, due to the majority of PSIFs turning out not to represent serious disease. Similarly, a study of follow-up of IFs detected on CT of 114 patients with haematuria found that costs were higher for patients with IFs who did not need intervention, compared to those who did need intervention (£34,734, £12,622 respectively), as there was a higher prevalence of the former (Bromage et al., 2012).

Follow-up of 116 patients with IFs on CT colonography over 12–24 months generated £34,329 in total costs (including inpatient, outpatient, diagnostic tests and surgical procedures (Xiong et al., 2006). This is far lower than the results of our study even accounting for differences in sample size and use of HRGs and National Reference Costs from different financial years, and is likely due to the inclusion of selected hospital records only: Xiong et al. reviewed patient records and attached costs only to those hospital contacts which were directly related to follow-up of IFs (Xiong et al., 2006), whereas we attached costs to all hospital contacts which occurred during specified time periods in order to enable four-way comparisons between cases and controls before-and-after feedback and thus provide a context for our cases' costs after feedback of a PSIF.

Using cost data from UK hospital imaging departments in 2012, imaging performed for 19 participants with IFs on research MRI of the torso generated £7,775. This is relatively high compared to our estimates for outpatient imaging, which were based on average 2016-2017 National Reference Costs, rather than costs from a single department (Pinato et al., 2012),

and as noted above, our use of HES records may have underestimated the volume (and therefore the cost) of imaging performed.

6.2.5.5 Implications for policy, practice and research

The findings of this study are of practical importance to current debates on the design of ethical policies for handling PSIFs detected in apparently asymptomatic research volunteers. The vast majority of PSIFs turned out not to be serious; such findings may be considered as 'false-positives', and as such our results may also be informative to public health and commercial screening, and clinical imaging contexts. It may be assumed that feedback of a PSIF will inevitably benefit a participant, however our previous study demonstrated that feedback results in clinical assessments such as invasive procedures (with the inconvenience and harms associated with these) and impacts on individuals' emotional wellbeing, insurance and finances and work and activities, regardless of the severity of the final diagnosis (Gibson et al., 2018). Our current work focuses on the impacts of feedback of PSIFs on a societal level, namely, on publicly-funded hospital services. Feeding back findings which turn out not to be serious (i.e. the majority) impacts on already overstretched healthcare services by shifting publicly-funded services away from patients in need while providing little benefit (and certainly some harm (Gibson et al., 2018)) to research volunteers with PSIFs, ultimately limiting the value of health services overall. There is considerable scope for better handling of PSIFs in order to reduce the unnecessary impacts on research volunteers, patients and health services. Robust, empirical data on the natural history of IFs, head-to-head comparisons of different IFs handling protocols, and randomised controlled trials of active treatment versus no treatment or surveillance of different types of IFs are needed (Gibson et al., 2017b).

Our study involves participants whose images were assessed by radiologists for PSIFs, and followed up within the UK health care system. Different protocols for detection of PSIFs may result in different proportions of serious and non-serious final diagnoses, and in turn different economic impacts. In future, data from UK Biobank will provide the opportunity to study the impact of feedback of PSIFs from a radiographer flagging protocol on both primary care and hospital services, using a larger cohort of participants, with a longer follow-up period. The economic impacts on health services, and societal costs such as lost working days or wages as a result of feedback of PSIFs from public health, commercial screening and clinical imaging contexts would further inform the debate on the value of feedback of such findings.

Our study of UK Biobank participants is the first to use a four-way comparison approach to enable data on hospital use and costs generated by cases with PSIFs to be set in context, and is the largest study of the health economic impact of feedback of PSIFs to apparently asymptomatic research volunteers. Over the next few years, UK Biobank will complete imaging of 100,000 participants and add linkages to other health-related datasets including primary care; this will enable us to complete a more comprehensive assessment of the economic impact of the feedback of PSIFs, and of those which result in serious and non-serious final diagnoses.

6.2.6 Additional information

6.2.6.1 Data availability

Due to the confidential nature of questionnaire responses and clinical information on participants with potentially serious incidental findings, it is not possible to publicly share all of the data on which our analyses were based, but summaries of all relevant data are included within the manuscript and supplementary material.

Importantly, any bona fide researcher can apply to use the UK Biobank resource, with no preferential or exclusive access, for health related research that is in the public interest. Application for access to UK Biobank data involves registration and application via the UK Biobank website, with applications considered by the UK Biobank Access Sub-Committee. Following approval, researchers and their institutions sign a Material Transfer Agreement and pay modest access charges. Further information on applying to access UK Biobank data is available at: http://www.ukbiobank.ac.uk/register-apply/.

6.2.6.2 Competing interests

L.M. Gibson: Member of the UK Biobank Imaging Working Group. UK Biobank Imaging Consultant, University of Edinburgh.

J. Nolan: UK Biobank Data Analyst, University of Edinburgh.

E. Mathieu: Former UK Biobank Data Analyst, University of Edinburgh.

T.J. Littlejohns: UK Biobank Epidemiologist, University of Oxford.

S. Garratt: Member of the UK Biobank Imaging Working Group. Senior Project Manager of UK Biobank Imaging Study.

S. Sheard: UK Biobank Director of Operations.

N. Doherty: UK Biobank Senior Clinical Study Administrator.

C. Keerie: None.

N.E. Allen: Member of UK Biobank Steering Committee, UK Biobank Imaging, Enhancements, Follow-up and Outcomes and Infectious Diseases Working Groups. UK Biobank Senior Epidemiologist.

J.M. Wardlaw: Advised on imaging protocols for the UK Biobank imaging study. Author of reports on IFs and guidance on their management in the UK; Currently analysing UK Biobank brain imaging and numeric data.

J. Leal: None.

A.M. Gray: None.

C.L.M. Sudlow: Member of UK Biobank Steering Committee, and UK Biobank Imaging, Enhancements, and Follow-up and Outcomes Working Groups. UK Biobank Chief Scientist.

6.2.6.3 Author contributions

L.M. Gibson: Conceptualisation, data curation, formal analysis, funding acquisition, investigation, methodology, software, visualisation, original draft, review and editing

J. Nolan: Data curation, formal analysis, investigation, software, review and editing

E. Mathieu: Data curation, review and editing

T.J. Littlejohns: Data curation, review and editing

S. Garratt: Data curation, review and editing

S. Sheard: Data curation, review and editing

N. Doherty: Project administration

C. Keerie: Methodology, review and editing

N.E. Allen: Review and editing

J.M. Wardlaw: Conceptualisation, funding acquisition, supervision, review and editing

J. Leal: Methodology, supervision, review and editing

A.M. Gray: Conceptualisation, funding acquisition, methodology, supervision, review and editing

C.L.M. Sudlow: Conceptualisation, data curation, funding acquisition, methodology, supervision, review and editing

6.2.6.4 Grant information

Lorna M Gibson is funded by a Wellcome Trust Clinical Research Training Fellowship (grant reference 107190/Z/15/Z). UK Biobank is funded by the Wellcome Trust, Medical Research Council, Department of Health, Scottish Government, Northwest Regional Development Agency, Welsh Assembly Government, British Heart Foundation, and Diabetes UK.

The main funders of UK Biobank's imaging study (the Medical Research Council and the Wellcome Trust) and the imaging study's independent scientific review panel encouraged evaluation of the potentially serious incidental findings protocol, but did not contribute to the design, the collection, analyses or interpretation of data, writing of this report or the decision to submit this work for publication.

6.3 Supplementary materials

Supplementary Table 6-1: Baseline characteristics of cases and controls

	Cases with potentially serious incidental findings (N=179) n (%)	Controls without potentially serious incidental findings[a] (N=179) n (%)
Matched variables		
Age		
Median (IQR)	64 (57-69)	64 (57-69)
Sex		
Female	99 (55.3)	99 (55.3)
Male	80 (44.7)	80 (44.7)
Duration of available follow-up[b]		
Median N days (IQR)	575 (551-623)	575 (551-623)
Morbidity[c]		
None	147 (82.1)	147 (82.1)
≥1 condition	32 (17.9)	32 (17.9)
Other variables		
Ethnicity[d]		
White	170 (95.0)	175 (97.8)
Minority ethnic groups	5 (2.8)	3 (1.7)
Townsend Deprivation Index		
Median (IQR)	-1.87 (-3.70-0.31)	-2.47 (-3.95-0.13)
Private healthcare[d]		
Never used	126 (70.4)	127 (70.9)
Ever used	50 (27.9)	51 (28.5)
Alcohol[d,e]		
None	24 (13.4)	15 (8.4)
Moderate	98 (54.7)	96 (53.6)
Hazardous	40 (22.3)	53 (29.6)
Harmful	10 (5.6)	14 (7.8)
Smoking[d]		
Never	106 (59.2)	100 (55.9)
Previous	61 (34.1)	71 (39.7)
Current	8 (4.5)	7 (3.9)
Fruit and vegetable portions/day[d,e]		
< 5	120 (67.0)	101 (56.4)
≥ 5	54 (30.2)	73 (40.8)
Days/week of moderate physical activity[d,e]		
0-2	54 (30.2)	44 (24.6)
3-4	37 (20.7)	55 (30.7)
5-7	78 (43.6)	74 (41.3)
BMI[d,e]		
Underweight	3 (1.7)	2 (1.1)
Normal	79 (44.1)	78 (43.6)
Overweight	68 (38.0)	65 (36.3)
Obese	27 (15.1)	34 (19.0)

IQR = interquartile range, BMI = body mass index

a. One control was matched to each case based on age (within 5 years), sex, date of imaging (within two weeks), and morbidity.

b. We calculated the duration of follow-up as the number of days between the date of feedback of a potentially serious incidental finding to a case and March 31st 2016, the last date of the available Hospital Episode Statistics (HES) dataset. We assigned the same date of feedback to controls to match their cases.

c. Morbidity was calculated using an Elixhauser Index score based on two years of HES data prior to the date of imaging (Elixhauser et al., 1998; Quan et al., 2005; The National Casemix Office and Health and Social Care Information Centre, 2014).

d. Data were missing for ethnicity (4 [2.2%], 1 [0.6%]), private healthcare (3 [1.7%], 1 [0.6%]), alcohol (7 [3.9%], 1 [0.6%]), smoking (4 [2.2%], 1 (0.6%]), fruit and vegetable intake (5 [2.8%], 5 [2.8%]), physical activity (10 [5.6%], 6 [3,4%]) and BMI (2 [1.1%], 0 [0.0%]) for 179 cases and 179 controls respectively.

e. Methodology for classifying alcohol intake, fruit and vegetable intake, physical activity and BMI was the same as that used in Chapter 5.

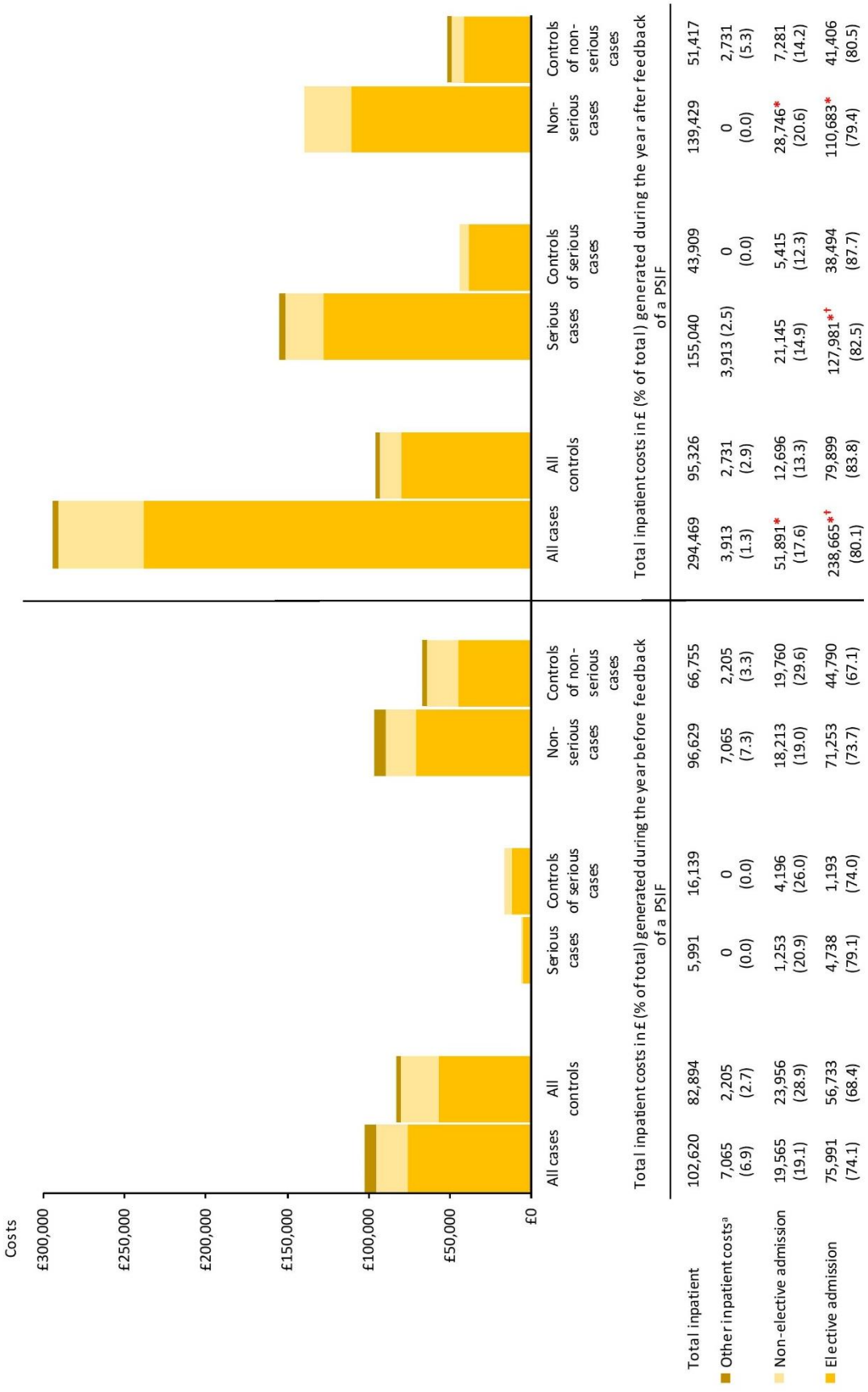
Supplementary Table 6-2: Details of 32 cases with at least one Elixhauser Index condition and their matched controls

Case-control pair number	Case's Elixhauser Index condition(s)[a]	Control's Elixhauser Index condition(s)[a]
1	Cardiac arrhythmias	Alcohol abuse Diabetes, uncomplicated Hypertension, uncomplicated Solid tumour without metastases
2	Cardiac arrhythmias	Hypertension, uncomplicated
3	Cardiac arrhythmias	Hypertension, uncomplicated Hypothyroidism
4	Depression	Diabetes, uncomplicated Hypertension, uncomplicated Solid tumour without metastases
5	Depression	Hypertension, uncomplicated
6	Hypertension, uncomplicated	Chronic pulmonary disease Hypertension, uncomplicated
7	Hypertension, uncomplicated	Hypertension, uncomplicated
8	Hypertension, uncomplicated	Hypertension, uncomplicated Hypothyroidism
9	Hypertension, uncomplicated	Chronic pulmonary disease
10	Hypertension, uncomplicated	Alcohol abuse Cardiac arrhythmias
11	Hypothyroidism	Diabetes, uncomplicated
12	Solid tumour without metastases	Alcohol abuse Hypertension, uncomplicated Renal failure Solid tumour without metastases
13	Weight loss	Hypothyroidism
14	Alcohol abuse Depression	Chronic pulmonary disease
15	Chronic pulmonary disease Hypothyroidism	Cardiac arrhythmias Valvular disease
16	Congestive heart failure Hypertension, uncomplicated	Alcohol abuse Chronic pulmonary disease Hypertension, uncomplicated
17	Depression Obesity	Hypertension, uncomplicated
18	Diabetes, uncomplicated Hypertension, uncomplicated	Hypertension, uncomplicated Weight loss
19	Fluid and electrolyte disorders Hypertension, uncomplicated	Hypertension, uncomplicated
20	Hypertension, uncomplicated Hypothyroidism	Hypertension, uncomplicated
21	Hypertension, uncomplicated Rheumatoid arthritis/collagen vascular diseases	Diabetes, uncomplicated
22	Hypothyroidism Weight loss	Chronic pulmonary disease Fluid and electrolyte disorders

Case-control pair number (continued)	Case's Elixhauser Index condition(s)[a]	Control's Elixhauser Index condition(s)[a]
23	Metastatic cancer Solid tumour without metastases[b]	Alcohol abuse Cardiac arrhythmias Chronic pulmonary disease
24	Metastatic cancer Solid tumour without metastases[b]	Other neurological disorders
25	Cardiac arrhythmias Hypertension, uncomplicated Obesity	Hypertension, uncomplicated
26	Cardiac arrhythmias Hypertension, uncomplicated Hypothyroidism	Hypertension, uncomplicated Liver disease
27	Cardiac arrhythmias Rheumatoid arthritis/collagen vascular diseases Solid tumour without metastases	Hypothyroidism
28	Chronic pulmonary disease Diabetes, uncomplicated Hypertension, uncomplicated	Alcohol abuse
29	Coagulopathy Hypertension, uncomplicated Renal failure	Diabetes, uncomplicated
30	Chronic pulmonary disease Diabetes, uncomplicated Hypertension, uncomplicated Valvular disease	Diabetes, uncomplicated
31	Depression Diabetes, uncomplicated Hypothyroidism Other neurological disorders	Alcohol abuse
32	Alcohol abuse Chronic pulmonary disease Congestive heart failure Fluid and electrolyte disorders Hypertension, uncomplicated	Chronic pulmonary disease Hypertension, uncomplicated Rheumatoid arthritis/collagen vascular diseases

a. Elixhauser Index conditions were identified using Hospital Episode Statistics (HES) inpatient data covering the two years prior to the date of imaging (Elixhauser et al., 1998; Quan et al., 2005; The National Casemix Office and Health and Social Care Information Centre, 2014)

b. These two cases had multiple HES records, at least one of which contained codes relating to both 'metastatic cancer' and 'solid tumour without metastases'; such coding may represent either a diagnostic coding error, or simultaneous primary tumours.



Supplementary Figure 6-1: Total costs in £ (% of total) of inpatient care generated by 179 cases (21 with serious and 158 with non-serious final diagnoses) and their 1:1 matched controls during the year before and the year after feedback of a potentially serious incidental finding

PSIF = potentially serious incidental finding.

*** Statistically significantly higher median inpatient costs in cases compared to controls during the year after feedback of a PSIF. See Supplementary Table 6-3 for details.**

† Statistically significantly higher median inpatient costs in cases during the year after feedback of a PSIF, compared to the year before. See Supplementary Table 6-3 for details.

There were no statistically significant differences in median inpatient costs generated: by controls during the year after feedback of a PSIF, compared to the year before; or by cases compared to controls during the year before feedback of a PSIF. See Supplementary Table 6-3 for details.

a. Including chemotherapy, excess bed days, high cost drugs and rehabilitation costs.

Supplementary Table 6-3: Median (interquartile and full range) inpatient costs (£) generated by 179 cases (21 with serious and 158 with non-serious final diagnoses) and their 1:1 matched controls during the year before and after feedback of a potentially serious incidental finding

Median (IQR, range) inpatient costs (£) generated by 179 cases (21 with serious and 158 with non-serious final diagnoses) and their matched controls					Case-control and before-after comparisons							
	After feedback of a PSIF		Before feedback of a PSIF		Cases after versus controls after		Cases after versus cases before		Controls after versus controls before		Cases before versus controls before	
	Cases	Controls	Cases	Controls	N discordant pairs (%)[b]	p-value[c]	N discordant pairs (%)[b]	p-value[c]	N discordant pairs (%)[b]	p-value[c]	N discordant pairs (%)[b]	p-value[c]
	Median (IQR, range)[a]	Median (IQR, range)[a]	Median (IQR, range)[a]	Median (IQR, range)[a]								
Elective	0 (0-983, 0-18661)	0 (0-0, 0-18158)	0 (0-0, 0-10064)	0 (0-0, 0-7008)	68 (38.0)	<0.001	68 (38.0)	0.001	40 (22.3)	0.913	44 (24.6)	0.338
Serious	3451 (0-11157, 0-18661)	0 (0-912, 0-18158)	0 (0-0, 0-1468)	0 (0-0, 0-2928)	17 (81.0)	0.040	15 (71.4)	0.001	7 (33.3)	0.469	7 (33.3)	0.156
Non-serious	0 (0-0, 0-9660)	0 (0-0, 0-6630)	0 (0-0, 0-10064)	0 (0-0, 0-7008)	51 (32.3)	0.001	53 (33.5)	0.089	33 (20.9)	0.714	37 (23.4)	0.117
Non-elective	0 (0-0, 0-17804)	0 (0-0, 0-3411)	0 (0-0, 0-6399)	0 (0-0, 0-4196)	20 (11.2)	0.008	24 (13.4)	0.070	16 (8.9)	0.292	17 (9.5)	0.359
Serious	0 (0-0, 0-17804)	0 (0-0, 0-3411)	0 (0-0, 0-1253)	0 (0-0, 0-4196)	5 (23.8)	0.625	5 (23.8)	0.313	3 (14.3)	1.000	2 (9.5)	1.000
Non-serious	0 (0-0, 0-4628)	0 (0-0, 0-3258)	0 (0-0, 0-6399)	0 (0-0, 0-3619)	15 (9.5)	0.004	19 (12.0)	0.199	13 (8.2)	0.142	15 (9.5)	0.462

	Median (IQR, range) inpatient costs (£) generated by 179 cases (21 with serious and 158 with non-serious final diagnoses) and their matched controls				Case-control and before-after comparisons							
	After feedback of a PSIF		Before feedback of a PSIF		Cases after versus controls after		Cases after versus cases before		Controls after versus controls before		Cases before versus controls before	
	Cases Median (IQR, range)[a]	Controls Median (IQR, range)[a]	Cases Median (IQR, range)[a]	Controls Median (IQR, range)[a]	N discordant pairs (%)[b]	p-value[c]	N discordant pairs (%)[b]	p-value[c]	N discordant pairs (%)[b]	p-value[c]	N discordant pairs (%)[b]	p-value[c]
Continued												
Other[d]	0 (0-0, 0-3786)	0 (0-0, 0-2731)	0 (0-0, 0-4076)	0 (0-0, 0-2205)	8 (4.5)	0.523	12 (6.7)	0.906	3 (1.7)	0.750	7 (3.9)	0.281
Serious	0 (0-0, 0-3786)	0 (0-0, 0-0)	0 (0-0, 0-0)	0 (0-0, 0-0)	5 (23.8)	0.438	6 (28.6)	0.531	0 (0)	1.000	0 (0)	1.000
Non-serious	0 (0-0, 0-0)	0 (0-0, 0-2731)	0 (0-0, 0-4076)	0 (0-0, 0-2205)	3 (1.9)	1.000	6 (3.8)	0.406	3 (1.9)	0.750	6 (3.8)	0.344

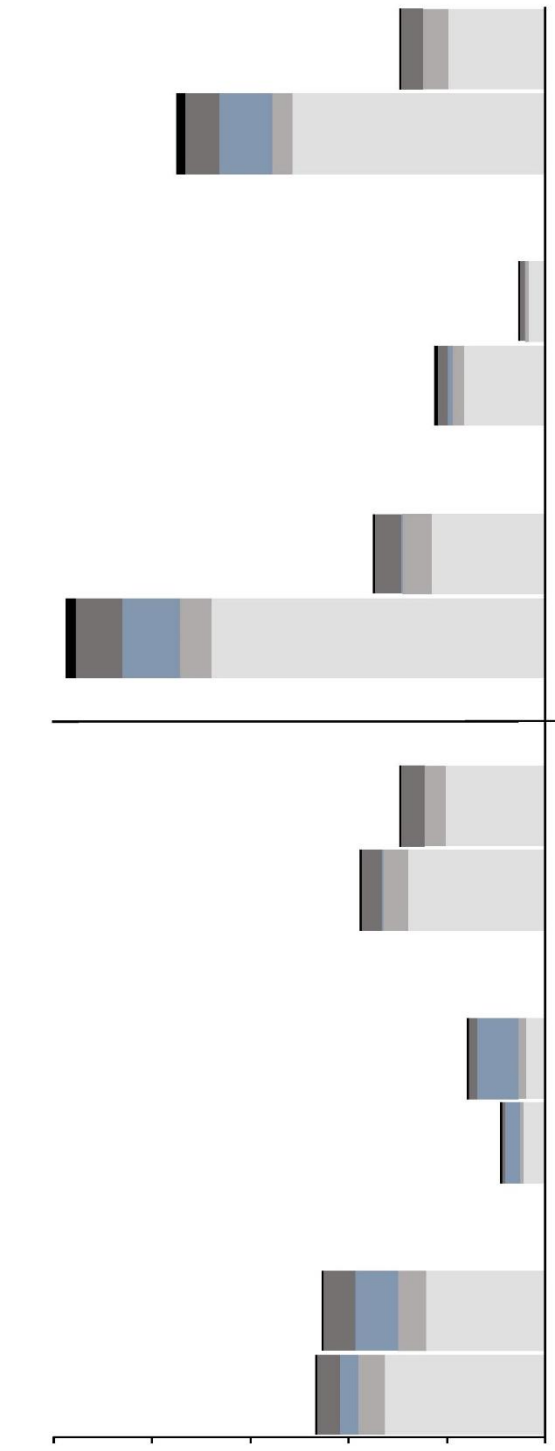
IQR = interquartile range, PSIF = potentially serious incidental finding

a. **25th and 75th percentile values calculated using Tukey's Hinges.**

b. **Only pairs with differences in costs (i.e. discordant pairs) are included in the related-samples Wilcoxon signed rank test.**

c. **Exact two-tailed p-value from related-samples Wilcoxon signed rank test for differences in costs between groups.**

d. **Including chemotherapy, excess bed days, high cost drugs and rehabilitation costs.**



	Total outpatient costs in £ (% of total) generated during the year before feedback of a PSIF							Total outpatient costs in £ (% of total) generated during the year after feedback of a PSIF						
	All cases	All controls	Serious cases	Controls of serious cases	Non-serious cases	Controls of non-serious cases		All cases	All controls	Serious cases	Controls of serious cases	Non-serious cases	Controls of non-serious cases	
Total outpatient	58,334	56,847	11,191	19,615	47,143	37,233		121,815	43,625	28,175	6,575	93,639	37,051	
Imaging	314 (0.5)	501 (0.9)	74 (0.7)	74 (0.4)	241 (0.5)	428 (1.1)		2,385*† (2.0)	451 (1.0)	559* (2.0)	52 (0.8)	1,826*† (1.9)	399 (1.1)	
Procedures	5,992 (10.3)	8,003 (14.1)	725 (6.5)	2,049 (10.4)	5,267 (11.1)	5,954 (16.0)		11,810*† (9.7)	6,662 (15.3)	2,879 (10.2)	1,409 (21.4)	8,931*† (9.5)	5,253 (14.2)	
Other outpatient costs[a]	4,287 (7.3)	10,861 (19.1)	4,104 (36.7)	10,861 (55.4)	183 (0.4)	0 (0.0)		14,807* (12.2)	251 (0.6)	1,223 (4.3)	0 (0.0)	13,584 (14.5)	251 (0.7)	
Non-consultant-led	6,977 (12.0)	7,086 (12.5)	536 (4.8)	1,630 (8.3)	6,441 (13.7)	5,455 (14.7)		7,732 (6.3)	7,342 (16.8)	2,731*† (9.7)	872 (13.3)	5,000 (5.3)	6,470 (17.5)	
Consultant-led	40,764 (69.9)	30,396 (53.5)	5,735 (51.4)	5,000 (25.5)	35,011 (74.3)	25,396 (68.2)		85,081*† (69.8)	28,920 (66.3)	20,784*† (73.8)	4,242 (64.5)	64,298*† (68.7)	24,679 (66.6)	

Supplementary Figure 6-2: Total costs in £ (% of total) of outpatient care generated 179 cases (21 with serious and 158 with non-serious final diagnoses) and their 1:1 matched controls during the year before and the year after feedback of a potentially serious incidental finding

PSIF = potentially serious incidental finding.

*** Statistically significantly higher median outpatient costs in cases compared to controls during the year after feedback of a PSIF. See Supplementary Table 6-4 for details.**

† Statistically significantly higher median outpatient costs in cases during the year after feedback of a PSIF, compared to the year before. See Supplementary Table 6-4 for details.

There were no statistically significant differences in median outpatient costs generated: by controls during the year after feedback of a PSIF, compared to the year before; or by cases compared to controls during the year before feedback of a PSIF. See Supplementary Table 6-4 for details.

a. Including chemotherapy, high cost drugs, nuclear medicine and radiotherapy costs.

Supplementary Table 6-4: Median (interquartile and full range) outpatient costs (£) generated by 179 cases (21 with serious and 158 with non-serious final diagnoses) and their 1:1 matched controls during the year before and after feedback of a potentially serious incidental finding

Median (IQR, range) outpatient costs (£) generated by 179 cases (21 with serious and 158 with non-serious final diagnoses) and their matched controls					Case-control and before-after comparisons							
	After feedback of a PSIF		Before feedback of a PSIF		Cases after versus controls after		Cases after versus cases before		Controls after versus controls before		Cases before versus controls before	
	Cases Median (IQR, range)[a]	Controls Median (IQR, range)[a]	Cases Median (IQR, range)[a]	Controls Median (IQR, range)[a]	N discordant pairs (%)[b]	p-value[c]	N discordant pairs (%)[b]	p-value[c]	N discordant pairs (%)[b]	p-value[c]	N discordant pairs (%)[b]	p-value[c]
Consultant-led	284 (0-767, 0-4077)	0 (0-187, 0-2309)	0 (0-238, 0-2805)	0 (0-290, 0-2563)	150 (83.8)	<0.001	134 (74.9)	<0.001	95 (53.1)	0.959	120 (67.0)	0.091
Serious	956 (598-1345, 0-2790)	116 (0-336, 0-958)	155 (0-362, 0-1281)	0 (0-288, 0-1517)	21 (100)	<0.001	21 (100)	<0.001	11 (52.4)	0.831	15 (71.4)	0.679
Non-serious	223 (0-549, 0-4077)	0 (0-172, 0-2309)	0 (0-215, 0-2805)	0 (0-202, 0-2563)	129 (81.6)	<0.001	113 (71.5)	<0.001	84 (53.2)	0.855	105 (66.5)	0.113
Non-consultant-led	0 (0-0, 0-767)	0 (0-0, 0-629)	0 (0-0, 0-1062)	0 (0-0, 0-823)	66 (36.9)	0.776	52 (29.1)	0.100	56 (31.3)	0.769	54 (30.2)	0.109
Serious	62 (0-234, 0-767)	0 (0-0, 0-281)	0 (0-0, 0-451)	0 (0-0, 0-823)	13 (61.9)	0.046	12 (57.1)	0.032	8 (38.1)	0.742	7 (33.3)	0.234
Non-serious	0 (0-0, 0-674)	0 (0-0, 0-629)	0 (0-0, 0-1062)	0 (0-0, 0-376)	53 (33.5)	0.483	40 (25.3)	0.606	48 (30.4)	0.627	47 (29.7)	0.180

	Median (IQR, range) outpatient costs (£) generated by 179 cases (21 with serious and 158 with non-serious final diagnoses) and their matched controls				Case-control and before-after comparisons							
	After feedback of a PSIF		Before feedback of a PSIF		Cases after versus controls after		Cases after versus cases before		Controls after versus controls before		Cases before versus controls before	
	Cases Median (IQR, range)[a]	Controls Median (IQR, range)[a]	Cases Median (IQR, range)[a]	Controls Median (IQR, range)[a]	N discordant pairs (%)[b]	p-value[c]	N discordant pairs (%)[b]	p-value[c]	N discordant pairs (%)[b]	p-value[c]	N discordant pairs (%)[b]	p-value[c]
Continued												
Procedures	0 (0-111, 0-1140)	0 (0-0, 0-701)	0 (0-0, 0-525)	0 (0-0, 0-1288)	65 (36.3)	0.009	60 (33.5)	0.008	41 (22.9)	0.387	52 (29.1)	0.462
Serious	0 (0-148, 0-700)	0 (0-0, 0-701)	0 (0-0, 0-350)	0 (0-129, 0-818)	12 (57.1)	0.225	10 (47.6)	0.080	7 (33.3)	0.406	7 (33.3)	0.219
Non-serious	0 (0-99, 0-1140)	0 (0-0, 0-689)	0 (0-0, 0-525)	0 (0-0, 0-1288)	53 (33.5)	0.021	50 (31.6)	0.037	34 (21.5)	0.540	45 (28.5)	0.913
Imaging	0 (0-0, 0-373)	0 (0-0, 0-117)	0 (0-0, 0-139)	0 (0-0, 0-177)	24 (13.4)	0.001	23 (12.8)	<0.001	11 (6.1)	0.752	9 (5.0)	0.773
Serious	0 (0-52, 0-147)	0 (0-0, 0-52)	0 (0-0, 0-74)	0 (0-0, 0-74)	7 (33.3)	0.047	7 (33.3)	0.109	2 (9.5)	1.000	2 (9.5)	1.000
Non-serious	0 (0-0, 0-373)	0 (0-0, 0-117)	0 (0-0, 0-139)	0 (0-0, 0-177)	17 (10.8)	0.012	16 (10.1)	0.001	9 (5.7)	0.895	7 (4.4)	0.672

	Median (IQR, range) outpatient costs (£) generated by 179 cases (21 with serious and 158 with non-serious final diagnoses) and their matched controls				Case-control and before-after comparisons							
	After feedback of a PSIF		Before feedback of a PSIF		Cases after versus controls after		Cases after versus cases before		Controls after versus controls before		Cases before versus controls before	
	Cases Median (IQR, range)[a]	Controls Median (IQR, range)[a]	Cases Median (IQR, range)[a]	Controls Median (IQR, range)[a]	N discordant pairs (%)[b]	p-value[c]	N discordant pairs (%)[b]	p-value[c]	N discordant pairs (%)[b]	p-value[c]	N discordant pairs (%)[b]	p-value[c]
Continued												
Other[d]	0 (0-0, 0-5457)	0 (0-0, 0-251)	0 (0-0, 0-4104)	0 (0-0, 0-8643)	8 (4.5)	0.023	9 (5.0)	0.094	3 (1.7)	0.500	3 (1.7)	1.000
Serious	0 (0-0, 0-611)	0 (0-0, 0-0)	0 (0-0, 0-4104)	0 (0-0, 0-8643)	2 (9.5)	0.500	3 (14.3)	1.000	2 (9.5)	0.500	2 (9.5)	1.000
Non-serious	0 (0-0, 0-5457)	0 (0-0, 0-251)	0 (0-0, 0-183)	0 (0-0, 0-0)	6 (3.8)	0.094	6 (3.8)	0.063	1 (0.6)	1.000	1 (0.6)	1.000

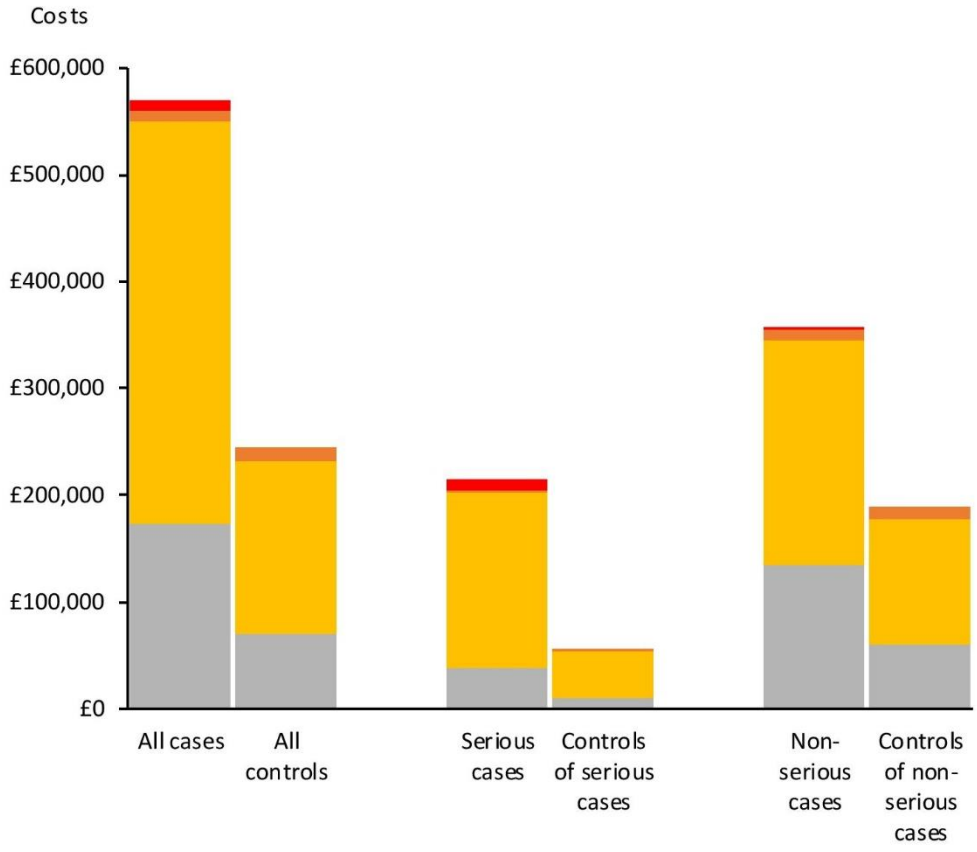
IQR = interquartile range, PSIF = potentially serious incidental finding

a. **25th and 75th percentile values calculated using Tukey's Hinges.**

b. **Only pairs with differences in costs (i.e. discordant pairs) are included in the related-samples Wilcoxon signed rank test.**

c. **Exact two-tailed p-value from related-samples Wilcoxon signed rank test for differences in costs between groups.**

d. **Including chemotherapy, high cost drugs, nuclear medicine and radiotherapy costs.**



	Total costs in £ (% of total) generated during the 282.5 person-years follow-up after feedback of a PSIF					
Total hospital	570,470*	244,721	213,894*	55,476	356,576*	189,245
Critical care	9,686 (1.7)	0 (0.0)	8,664* (4.1)	0 (0.0)	1,022 (0.3)	0 (0.0)
Emergency	11,084 (1.9)	12,266 (5.0)	1,508 (0.7)	1,188 (2.1)	9,576 (2.7)	11,078 (5.9)
Inpatient	376,991* (66.1)	161,688 (66.1)	164,939* (77.1)	43,909 (79.1)	212,052* (59.5)	117,780 (62.2)
Outpatient	172,708* (30.3)	70,766 (28.9)	38,783* (18.1)	10,379 (18.7)	133,925* (37.6)	60,387 (31.9)

Supplementary Figure 6-3: Total hospital costs in £ (% of total) generated by 179 cases (21 with serious and 158 with non-serious final diagnoses) and their 1:1 matched controls during the 282.5 person-years of follow-up available after feedback of a potentially serious incidental finding

PSIF = potentially serious incidental finding.

*** Statistically significantly higher median costs in cases compared to controls during the 282.5 person-years of follow-up after feedback of a PSIF. See Supplementary Table 6-5 for details.**

Supplementary Table 6-5: Median (interquartile and full range) hospital costs (£) generated by 179 cases (21 with serious and 158 with non-serious final diagnoses) and their 1:1 matched controls after feedback of a potentially serious incidental finding over 282.5 person-years of follow-up

	Median (IQR, range) hospital costs (£) generated by 179 cases (21 with serious and 158 with non-serious final diagnoses) and their matched controls over 282.5 person-years of follow-up		Case-control comparison	
	Cases Median (IQR, range)[a]	Controls Median (IQR, range)[a]	N discordant pairs (%)[b]	p-value[c]
Total hospital	846 (202-3257, 0-43100)	252 (0-1229, 0-19434)	171 (95.5)	<0.001
Serious	6050 (1947-16621, 0-43100)	593 (0-2783, 0-19434)	21 (100)	0.011
Non-serious	668 (171-2507, 0-32703)	240 (0-1159, 0-17574)	150 (94.9)	0.001
Inpatient	0 (0-1797, 0-40033)	0 (0-440, 0-18158)	89 (49.7)	0.002
Serious	4672 (0-13843, 0-40033)	0 (0-2004, 0-18158)	18 (85.7)	0.043
Non-serious	0 (0-1443, 0-30128)	0 (0-0, 0-15164)	71 (44.9)	0.030
Outpatient	548 (180-1279, 0-16489)	146 (0-590, 0-2926)	169 (94.4)	<0.001
Serious	1947 (1234-2497, 0-3996)	410 (0-849, 0-1592)	21 (100)	<0.001
Non-serious	468 (148-1019, 0-16489)	141 (0-537, 0-2926)	148 (93.7)	<0.001
Emergency	0 (0-16, 0-1276)	0 (0-0, 0-3778)	71 (39.7)	0.631
Serious	0 (0-119, 0-570)	0 (0-120, 0-300)	10 (47.6)	1.000
Non-serious	0 (0-0, 0-1276)	0 (0-0, 0-3778)	61 (38.6)	0.624
Critical care	0 (0-0, 0-1633)	0 (0-0, 0-0)	8 (4.5)	0.008
Serious	0 (0-1022, 0-1633)	0 (0-0, 0-0)	7 (33.3)	0.016
Non-serious	0 (0-0, 0-1022)	0 (0-0, 0-0)	1 (0.6)	1.000

IQR = interquartile range, PSIF = potentially serious incidental finding

a. 25th and 75th percentile values calculated using Tukey's Hinges.

b. Only pairs with differences in costs (i.e. discordant pairs) are included in the related-samples Wilcoxon signed rank test.

C. Exact two-tailed p-value from related-samples Wilcoxon signed rank test for differences in costs between groups.

6.4 Conclusion

This chapter explored the impact on hospital services of feedback of PSIFs to apparently asymptomatic volunteers. There were no differences in hospital contacts or costs between cases and controls before feedback (or between controls before and after feedback of a PSIF to their matched case). However, after feedback of a PSIF, 144 (80.5%) cases with PSIFs had contact with hospital services (either inpatient, outpatient, emergency or critical care) at least once, compared to 94 (52.5%) of controls without a PSIF, and cases' numbers of hospital contacts and costs were significantly higher compared to controls, and compared to cases' hospital contacts and costs the year before (hospital contacts: three versus one and versus one; hospital costs: £522 versus £114 and versus £128 [all $p<0.001$]). The rates of cases' cumulative costs began to increase approximately 30–60 days following feedback of a PSIF. A year after feedback of a PSIF, cases' total hospital costs were higher than controls' and higher than cases' costs the year before (£431,114 versus £147,817 [2.9-fold] and versus £167,434 [2.6-fold] respectively). Serious cases' total costs had increased much more than non-serious cases' total costs one year after feedback of a PSIF (10.5-fold, 1.9-fold respectively). However, as the vast majority (158/179 [88%]) of PSIFs were finally diagnosed as non-serious, non-serious cases generated greater absolute costs than serious cases (£239,021 versus £192,093). These patterns of costs were similar over a longer follow-up period which included an additional 103.5 person-years of follow-up. After feedback, the majority of cases' cost burden impacted on inpatient services (68%), but the majority of cases' service use burden impacted on outpatient services (82%).

This chapter did not address the economic impact of feedback of PSIFs on primary healthcare services (as such data were not available), the length of follow-up (median number of days 575 [interquartile range 551-623]) may not have been long enough for the economic impact of some PSIFs to fully manifest, and HES datasets did not appear to capture all of the imaging performed in cases with PSIFs when compared to cases' self-reported data presented in Chapter 4 (Gibson et al., 2018); as such our results on the magnitudes of the economic impact of feedback of PSIFs are likely underestimates.

This chapter includes cases with PSIFs identified using a protocol of systematic radiologist review; subsequent UK Biobank participants' images are now handled using a protocol of radiographer flagging (Gibson et al., 2018), which may result in different proportions of serious and non-serious final diagnoses, and a different economic impact. The UK Biobank Imaging Study will generate more cases with PSIFs over the forthcoming years, a greater length of follow-up time will pass, and linkages other healthcare datasets (including primary

care) will become available, which will all enable further exploration of the economic impact of feedback of PSIFs in future.

The next chapter summarises the key results from the studies included in Chapters 2, 4, 5 and 6, and compares these results with other studies, describes the strengths, limitations and implications of this thesis, and suggests directions for future work.

Chapter 7 Discussion

7.1 Summaries of key findings relating to potentially serious incidental findings

7.1.1 Prevalence and nature

7.1.1.1 Key findings

Chapter 2 reported that the pooled prevalences of potentially serious incidental findings (PSIFs) on brain, thorax, abdomen and brain and body magnetic resonance imaging (MRI) were 1.4%, 1.3%, 1.9% and 3.9% respectively. These prevalence estimates rose when IFs of uncertain potential seriousness were included, to 1.7%, 3.0%, 4.5% and 12.8% respectively. Suspected malignancies accounted for around half of PSIFs on each imaged body region (brain 0.6%; thorax 0.6%; abdomen 1.3%; brain and body 2.3%). There was substantial between-study and within-study heterogeneity, but few data to reliably inform on sources of this.

7.1.1.2 Comparison with other studies

Pooled point prevalence estimates of PSIFs appeared to increase from brain and thorax, to abdomen, to brain and body MRI. Although this observation was not statistically robust, this finding is biologically plausible, given the range of pathologies possible across regions, and is supported by similar patterns found in primary studies of brain and body MRI (Gibson et al., 2018; Hegenscheid et al., 2013) and recent summary reports (The Royal College of Radiologists, 2011).

Individual studies of whole-body MRI of apparently asymptomatic people found prevalences of incidental findings (IFs) ranging from 12.8–57.6% (Baumgart and Egelhof, 2007; Cieszanowski et al., 2014; Goehde et al., 2005; Hegenscheid et al., 2013; Lo et al., 2008; Morin et al., 2009; Saya et al., 2017; Tarnoki et al., 2015). These differences in prevalence are likely due to different definitions of IFs and PSIFs across these, and our, studies, and may be influenced by other characteristics, such as participants' ages (see Section 7.1.3).

An umbrella review which repeated meta-analyses of studies of patient populations undergoing cardiac MRI, and of mixed patient and apparently asymptomatic populations undergoing brain MRI, in existing systematic reviews found prevalences of IFs that were much higher than our estimates (around 34% and 22% respectively (Dunet et al., 2016; Morris et al., 2009; O'Sullivan et al., 2018; Takashima et al., 2017)). However these differences are likely due to differences between our reviews' definitions (IFs versus PSIFs),

as well as populations (patients versus apparently asymptomatic people) (Dunet et al., 2016; Morris et al., 2009; O'Sullivan et al., 2018; Takashima et al., 2017).

7.1.2 Follow-up and final diagnoses

7.1.2.1 Key findings

Our systematic review did not identify any study which enabled the types of follow-up of PSIFs to be quantified (Chapter 2). We provided new evidence on the follow-up of PSIFs in Chapter 4, and found that every participant with survey data reported that they had contacted their general practitioner (GP) (Gibson et al., 2018). The vast majority (90%) of participants with PSIFs had some form of clinical assessment, most commonly imaging or referral to a specialist (Gibson et al., 2018). Similar numbers of participants had invasive procedures, regardless of the seriousness of their final diagnoses (Gibson et al., 2018).

Regarding final diagnoses, three of our studies consistently found that the majority (around 80%) of PSIFs do not turn out to represent serious disease (Chapters 2, 4 and 5) (Gibson et al., 2018).

7.1.2.2 Comparison with other studies

Authors of a recent umbrella review aimed to summarise the final diagnoses of IFs across a range of modalities, but inadvertently presented suspected final diagnoses as firm final diagnoses in some cases, notably for studies of brain MRI (O'Sullivan et al., 2018). This highlights an issue with reporting and interpreting studies of IFs: descriptions of IFs detected on research imaging should not be taken as firm diagnoses unless the report specifically states that participants have had systematic clinical follow-up. As we show in the UK Biobank cohort, the majority of PSIFs turn out not to be serious (Gibson et al., 2018), that is to say, final diagnoses may differ from the IF description in a large number of cases.

7.1.3 Factors associated with PSIFs and with serious final diagnoses

7.1.3.1 Key findings

Of all the factors we investigated in Chapter 5, PSIFs protocol had by far the greatest effect on prevalence of PSIFs and serious final diagnoses. Systematic radiologist review of images resulted in 13 times more PSIFs (179/1,000 [17.9%]; 104/6,334 [1.6%]; age- and sex-adjusted odds ratio [OR] 13.3 [95% confidence interval (CI) 10.3–17.1]) and four times more serious final diagnoses (OR 4.2 [95% CI 2.4–7.4]) compared to radiographer flagging. A lower proportion of PSIFs detected by radiologists resulted in serious final diagnoses compared to radiographer flagging (21/179 [11.7%]; 33/104 [31.7%], Chapters 4 and 5)

(Gibson et al., 2018). Chapters 2, 4 and 5 also found that PSIFs were more common with increasing age, and Chapter 5 found that increasing age was also associated with serious final diagnoses.

With regards to other participant factors, the studies presented in Chapters 2, 4 and 5 found no significant associations of PSIFs or serious final diagnoses with sex (Gibson et al., 2018). Neither were there any significant associations between PSIFs or serious final diagnoses and participants' ethnicity, Townsend Deprivation Index score, use of private healthcare services, alcohol intake, fruit and vegetable intake, physical activity, smoking status, body mass index (BMI) or morbidity (Chapter 5).

Of several imaging factors investigated in our systematic review (including imaging setting [research versus non-research], blinding versus non-blinding of image readers to information on participants, and numbers of image readers [one versus more than one]) there was no evidence of any clinically meaningful differences in the prevalence of PSIFs between these subgroups (Chapter 2). There was not sufficient data on any other investigated factor to reliably inform on any other associations with PSIFs.

7.1.3.2 Comparison with other studies

No other studies of the effect of different PSIFs protocols on the prevalence of either PSIFs or serious final diagnoses were identified by the time of writing. The results in this thesis support previously published data that demonstrate trends toward increased IFs (of all clinical severity) with age, and not with sex (Bos et al., 2016; Brugulat-Serrat et al., 2017; Haberg et al., 2016; Hartwigsen et al., 2010; Illes et al., 2004a; Kumar et al., 2008; Morin et al., 2009; Sandeman et al., 2013; Tsushima et al., 2005; Yue et al., 1997). This thesis extends upon this knowledge by demonstrating that these associations hold when selecting only IFs of potential seriousness, and with serious final diagnoses, on multiple body regions, and in the case of age, independent of the great effect of PSIFs protocol (Chapter 5). There was no association of PSIFs on brain and body imaging with morbidity, extending on the findings of a systematic review of IFs on brain MRI only, which found no evidence of a difference in prevalence among participants either with comorbidities, without, or where their medical history was not known (Morris et al., 2009).

7.1.4 Participants' understanding of consent to feedback

7.1.4.1 Key findings

Chapter 4 reported the results of a survey of UK Biobank participants to assess their understanding of consent to feedback of PSIFs. Around one quarter of respondents (158/607, 26%) incorrectly thought that they could choose whether or not to be informed about a PSIF.

7.1.4.2 Comparison with other studies

Participants of the Netherlands Epidemiology of Obesity (NEO) study also misunderstood consent materials with regards to feedback of IFs. Despite consent materials stating that only IFs with 'serious health consequences' would be fed back, focus groups found that participants incorrectly thought that all IFs would be fed back (de Boer et al., 2018).

7.1.5 Non-medical impacts of feedback on participants

7.1.5.1 Key findings

Participants who received feedback of PSIFs reported impacts on their emotional wellbeing, insurance and finances and work and activities (17%, 9%, and 6% respectively). These non-medical impacts affected a higher proportion of participants with serious final diagnoses, but affected a higher absolute number of participants with non-serious final diagnoses (Gibson et al., 2018).

7.1.5.2 Comparison with other studies

These findings are in keeping with data from the Study of Health in Pomerania (SHIP) and NEO studies which found that around half the participants who received feedback of an IF reported that they experienced some psychological distress or worry (de Boer et al., 2018; Schmidt et al., 2013). Furthermore, a similar magnitude of participants with IFs in the SHIP reported impairments to work life (4%) and leisure (5%) compared to the UK Biobank participants (6%) (Schmidt et al., 2013). No informative published data on the impacts of feedback of PSIFs on insurance and finances had been identified by the time of writing to allow comparison.

7.1.6 Opinions of receiving feedback

7.1.6.1 Key findings

Despite participants reporting that feedback of PSIFs results in non-medical impacts, and GPs reporting that a higher proportion of participants experienced negative, compared to positive, impacts on emotional wellbeing, the vast majority of participants were glad to have been told about a PSIF (142/145, 98%) (Gibson et al., 2018).

When asked if participants should always receive feedback of a PSIF, a higher proportion of GPs agreed with this statement than participants (61/94 [65%] vs 55/149 [37%]), and around a quarter of participants changed their mind on this between the six-week and six-month questionnaires (Gibson et al., 2018).

7.1.6.2 Comparison with other studies

Similarly, participants of the NEO study experienced impacts on their wellbeing following feedback of an IF, but were glad to have been informed; this apparent contradiction may be due to participants' misunderstandings of research imaging as a screening service (de Boer et al., 2018). Regarding the results of screening, finding disease is seen as an opportunity to benefit from early treatment, and lack of disease is seen as conferring health; both scenarios are seen as beneficial by patients (Ransohoff et al., 2002). This does however highlight the possibility that research participants do not understand the limitations of research imaging. UK Biobank has since updated its consent materials to emphasise more strongly that the research imaging does not detect all disease, that some findings will turn out not to be serious, and that lack of feedback does not constitute an 'all clear' (Gibson et al., 2018; UK Biobank, 2018c). As such, it is imperative to continue to monitor participants' understanding of consent, and if possible, with regards to these concepts in particular.

No other study was identified that asked GPs and research participants their opinions on whether or not feedback of PSIFs should always be given, but this binary question may not do justice to the potential answers to this question, as respondents may consider the different types of findings, their clinical severity (Opinion Leader, 2012), or the circumstances and wishes that may differ between individuals.

7.1.7 Economic impact of feedback

7.1.7.1 Key findings

Before feedback of a PSIF, there were no differences in hospital contacts or costs between cases and controls, and controls' hospital contacts and costs did not differ between the years before and after feedback of a PSIF to their matched case. However, after feedback of a PSIF, 144 (81%) cases generated some hospital costs, compared to 94 (53%) controls, and cases' median numbers of hospital contacts and median costs were significantly higher than controls, and compared to cases' hospital contacts and costs the year before feedback (hospital contacts: three versus one and versus one; hospital costs: £522 versus £114 and versus £128 [all $p<0.001$]). Cases' cumulative hospital costs began to increase approximately 30–60 days after feedback of a PSIF. A year after feedback of a PSIF, cases' total hospital

costs were higher compared to controls', and to cases' costs the year before (£431,114 versus £147,817 and versus £167,434 respectively). Compared to the year before feedback, after feedback of a PSIF serious cases' total hospital costs had increased by a far greater magnitude than non-serious cases (10.5- and 1.9-fold respectively). However, the majority of PSIFs (158/179 [88%]) turned out not to be serious, and these non-serious cases generated greater absolute costs than serious cases (£239,021 versus £192,093).

After feedback of a PSIF, the majority of cases' cost burden fell to inpatient services (68%) while the vast majority of cases' service use burden fell to outpatient services (82%).

7.1.7.2 Comparison with other studies

Compared to self-reported data from the cases with PSIFs (Gibson et al., 2018), Hospital Episode Statistics (HES) data found that they had apparently attended fewer imaging appointments and more outpatient appointments; this is likely due to incomplete coding of imaging appointments in HES data and a longer length follow-up respectively (six months of participant surveys versus one year within HES datasets). Incomplete coding of imaging appointments also likely explains the lower imaging costs in our study of 179 cases (total £2,385) compared to follow-up imaging of IFs detected on research MRI in a much smaller cohort (n=19) of elderly volunteers (£7,775) (Pinato et al., 2012).

Cases with non-serious final diagnoses generated higher total costs than those with serious final diagnoses (due to the higher number of the former). Similarly, patients who did not need intervention for IFs detected on computed tomography (CT) generated higher costs than those who did need intervention, due to their higher absolute prevalence (Bromage et al., 2012).

Total costs of follow-up of 116 patients with IFs on CT colonography were far lower than our study (£34,329) (Xiong et al., 2006). This is likely due to attachment of costs only to those hospital contacts which were directly related to follow-up of IFs (as identified by review of patient records) (Xiong et al., 2006), rather than to all hospital contacts over a specified time period as in our study, which allowed us to make comparisons between groups.

7.2 Strengths and limitations

The strengths and limitations of each study are discussed in detail within each chapter. This section serves to summarise the broader strengths and limitations of this thesis.

The studies in this thesis have focused on PSIFs (defined as those 'indicating the possibility of a condition which, if confirmed, would carry a real prospect of seriously threatening life span, or of having a substantial impact on major body functions or quality of life' (Gibson et al., 2018)) rather than IFs more generally. By specifically excluding non-serious IFs, this thesis has generated data on findings which are potentially clinically significant, and have implications for research participants, study staff and healthcare services. Furthermore, we made extensive efforts to gather data on and rigorously apply a methodology to classify final diagnoses. There were good questionnaire response rates in Chapter 4, and minimal missing data in Chapters 5 and 6, which minimises any information bias which may have affected results. The majority of previous studies of IFs on MRI had focused only on participants undergoing brain MRI, which may not be generalisable to those undergoing MRI of other body regions. This thesis presents data on both brain and body MRI, and the results are therefore generalisable to participants undergoing either whole-body MRI, or a component body region.

This thesis is strengthened by the originality and size of the studies presented. To my knowledge, it presents the first studies to: systematically review and meta-analyse the prevalence and types of PSIFs on both brain and body MRI and to attempt to systematically summarise the literature on follow-up and final diagnoses of unselected participants, and factors associated with PSIFs (Chapter 2); compare two protocols for PSIFs (Chapter 4); quantitatively assess the understanding of consent to feedback of PSIFs of an unselected group of imaged participants (Chapter 4); assess the impact of feedback of PSIFs on unselected participants' insurance and finances (Chapter 4); and to use routinely collected healthcare data to investigate the economic impact of feedback of PSIFs on hospital services (Chapter 6). Chapter 5 also represents the largest study so far of factors associated with serious final diagnoses, and Chapter 6 is the largest study so far to assess the economic impact on any health service of feedback of IFs detected in a cohort undergoing MRI, and is the first study to set the economic impact of PSIFs in the context of usual healthcare costs in order to aid judgements of the magnitude of this impact.

Limited data precluded analyses in some studies which may have affected our ability to reliably detect or exclude associations between exposures and outcomes. It was possible to perform only a small number of subgroup analyses to investigate the heterogeneity in prevalence estimates of PSIFs in the systematic review (Chapter 2) due to either incomplete data or imbalanced subgroups. The study of the factors associated with PSIFs and serious final diagnoses was in part limited by the lack of data on these outcomes, due to their relative

rarity (Chapter 5). In addition, healthy volunteer bias may explain the lack of data on particular exposures in Chapter 5 (e.g. harmful levels of alcohol intake), and also the substantial proportion of participants who had no hospital costs in Chapter 6, although most people do not generate hospital costs during any given time period. Conversely, given the numbers of different statistical tests performed in Chapters 5 and 6, some results may be significant due to chance; in line with published advice, each result was interpreted within the context of the available data and the clinical plausibility, rather than analysed further with statistical corrections which may potentially put too much emphasis on a statistical p-value (Perneger, 1998). This thesis was not able to investigate the economic impact of feedback of PSIFs on primary care services, as linkages to routinely collected primary care datasets were not available at the time of the study.

Classification of serious final diagnoses was subjective, although there was good reliability of this process between two doctors. The definition of 'serious final diagnosis' focuses on the medical impact of PSIFs rather than non-medical impacts. While the study presented in Chapter 4 demonstrated impacts of non-serious final diagnoses on some non-medical domains (Gibson et al., 2018), this thesis was not able to capture impacts on other non-medical domains such as socio-economic impacts on participants related to costs of travel to appointments, or numbers of workdays lost when attending for healthcare. Final diagnoses classified as serious may have not in fact shortened life span, or had a substantial impact on major organ function or quality of life. Furthermore, some PSIFs may take longer than six months to diagnose, or for their impact to fully manifest, and as such, our data on impacts are likely underestimates.

7.3 Impact on participants, research and policy

In 2014, the Medical Research Council and the Wellcome Trust published a framework to guide researchers on the points that they should consider when designing IFs policies (Medical Research Council and Wellcome Trust, 2014), and the results of this thesis inform on a number of these. Firstly, researchers are advised to consider the probability of identifying health-related findings, their potential severity and the certainty of this knowledge, and the risks of false positives, and that 'factors relating to the study population, such as age, may also be relevant' (Medical Research Council and Wellcome Trust, 2014). The results in this thesis from Chapters 2, 4, and 5 inform on the prevalence of PSIFs, including those which were finally diagnosed as non-serious (i.e., false positives), and the factors associated with these, which included the PSIFs protocol and age (Gibson et al., 2018). Secondly, researchers are advised to anticipate the types of IFs which may arise

(Medical Research Council and Wellcome Trust, 2014), and Chapters 2 and 4 demonstrated that on brain, thorax and abdominal MRI the most common types of PSIFs are suspected tumours and aneurysms (Gibson et al., 2018). Thirdly, researchers should follow good practice guidelines where available, and tailor their protocols to the context of the study (including the local practice and legal structures) (Medical Research Council and Wellcome Trust, 2014); the detailed information presented in this thesis on the UK Biobank protocol and its evaluation provides one such approach from which researchers could draw if appropriate (Medical Research Council and Wellcome Trust, 2014). Finally, the best interests of the participants, and the potential benefits and harms of feedback should be considered (Medical Research Council and Wellcome Trust, 2014). Participants within research studies do not generally stand to benefit directly from their involvement; rather the benefit of research is for the public good, and as such, the risks of harms to individual participants must be minimised. This thesis provides data to show the impact of feedback of PSIFs on participants' emotional wellbeing, insurance and finances, and work and activities, as well as the economic impacts on hospital services, and that these impacts occur even in participants with non-serious final diagnoses. As this thesis also shows that the majority of PSIFs turned out not to be serious, feeding back such findings unnecessarily burdens overstretched healthcare services and shifts these publicly-funded services away from the patients who need them, while providing little benefit (but certainly some harm) to research volunteers, and ultimately limiting the value of health care services overall. Such harms can never be completely avoided, as the specificity for serious disease of non-targeted, non-diagnostic imaging is not optimal. However, researchers and policymakers should aim to minimise these unnecessary impacts to research volunteers, patients and health services, and (as this thesis shows in Chapter 5) they would be able to substantially influence these impacts through the design of appropriate IFs protocols.

UK Biobank took a pragmatic approach, and aimed to design an IFs protocol that was scalable and resulted in minimal harm to the 100,000 apparently asymptomatic volunteers. We investigated two possible options to minimise the harms associated with feedback of PSIFs, and in the context of a study the scale of UK Biobank, deemed a protocol of radiographer flagging to be more appropriate than one of radiologist review based on the empirical evidence presented in this thesis. While this approach may not be suitable for other studies, the principle of minimising harm will still be important (Gibson et al., 2018).

The prevalence of PSIFs, and of those which are finally diagnosed as serious versus non-serious diseases, will vary by PSIFs protocol. While systematic radiologist review of images

for PSIFs may be assumed to be the most appropriate protocol (Bunnik et al., 2017) and indeed, expected by some participants (Kirschen et al., 2006), the results presented in this thesis challenge this assumption within the context of large-scale imaging of apparently asymptomatic populations. As such, the results may be informative to other research studies, particularly those of large-scale, or of multi-region imaging, or those including healthy controls or population-based participants, and also to public health screening and direct-to-consumer commercial imaging contexts.

Informed consent for imaging is necessary across all imaging contexts to promote the autonomy of individuals, and facilitating potential research participants' informed decision making to take part in a study is essential to maintaining public trust in medical research (Farrar and Savill, 2014; Jefford and Moore, 2008). This thesis shows, however, that despite informing participants about the PSIFs protocol, a substantial proportion did misunderstand elements of it. This highlights the need for researchers and policymakers to make consent materials as clear as possible; the data from this thesis may inform the content included in such materials, and inform the assessment of these materials by research ethics committees.

7.4 Unanswered questions and future research

The studies presented in this thesis include cohorts of up to 7,334 imaged UK Biobank participants. The UK Biobank Imaging Study continues, and aims to image up to 100,000 participants. Data from this larger cohort will enable several questions on PSIFs to be addressed in future. In response to the study presented in Chapter 4, UK Biobank amended their consent materials with a view to improving participants' understanding of consent, the effect of which could be evaluated using an interrupted time series study. A larger cohort may also enable an exploration of the effects of factors such as age, education, and opinions on the participant information leaflet on participants' understanding of consent to potentially identify further opportunities to optimise informed consent processes. Relatively few participants reported impacts of feedback of PSIFs on their emotional wellbeing, insurance and finances and work and activities in Chapter 4; a larger cohort would provide more robust data on the prevalence of these impacts, and may enable exploration of how these vary by other factors, such as severity of final diagnoses, or participant age, sex, or employment status. Chapter 4 also demonstrated that around one third of participants thought feedback of PSIFs should always be given, rather than that they should be able to choose whether they receive feedback or not, and around one quarter changed their minds on this issue between the six-week and six-month questionnaires, although the reasons for this remain unclear. Data from a larger cohort could inform how opinions or changes in opinion vary by factors

such as follow-up, final diagnoses or non-medical impacts. In addition, participants' reasons for changing their minds could be further explored by analysing the available free-text responses from the participants' questionnaires, or using other qualitative methods such as interviews. Only around two-thirds of GPs provided opinions on the impacts of feedback of PSIFs on their patients' emotional wellbeing and the net benefit versus net harm of feedback; the response rate and opinions from the larger cohort may provide more robust results. Furthermore, robust qualitative analyses of the vast quantity of free text data provided by participants and their GPs on their follow-up questionnaires may generate insights in to the experiences of receiving feedback of a PSIF, and inform the direction of future studies. A larger cohort may overcome the limitations of the study presented in Chapter 5, in which sparse data on exposures and limited power due to relatively rare outcomes may have led to associations between several factors and either PSIFs or serious final diagnoses being missed.

Lack of linkages to the relevant datasets precluded investigations of the health economic impact of feedback of PSIFs on primary care services, and underestimated the burden on imaging services but these impacts could be investigated when these linkages become available in future. In addition, Chapter 6 investigates the economic impact of PSIFs detected using a protocol of systematic radiologist review. UK Biobank now use a protocol of radiographer flagging of concerning images for a radiologist to review, which results in lower overall numbers of PSIFs and different proportions of those which are finally diagnosed as serious versus non-serious. UK Biobank continues to collect data on participants with PSIFs, and to update and add linkages to routinely collected datasets; this will provide the opportunity in future to study the impact of feedback of PSIFs from a radiographer flagging protocol on both primary and secondary care services, using a larger cohort of participants, with a longer follow-up period. Although it would unlikely change the overall pattern of observed differences in hospital contacts and costs between groups, adjusted statistical models, such as generalised estimating equations, and/or sensitivity analyses would enable further exploration of the economic impact of feedback of PSIFs on health services. Our novel methodology of a four-way comparison using data from cases and controls before and after feedback of a PSIF may inform the design of future studies of the impacts of PSIFs in other healthcare systems.

Further questions could be addressed using other quantitative and qualitative studies. Firstly, our knowledge of the effect of PSIFs protocol on the prevalence of PSIFs and serious final diagnoses is informed by studies of systematic radiologist review or radiographer flagging.

The influence on prevalence of other PSIFs protocols, which may use radiology trainee or non-medical image readers, multiple versus single readers, readers blind to clinical information or not, is well worthy of further study, as systematic radiologist review of images may not be feasible in some contexts, and radiographer flagging may not be deemed appropriate in others. Different PSIFs protocols could be investigated either in head-to-head comparisons within individual imaging studies, or by creating an online image set and inviting different groups of readers to interpret research images. To inform the design of such studies, quantitative and qualitative work with different stakeholders, including research participants, patients, research staff and primary and secondary care physicians could identify these groups' preferred outcomes for PSIFs protocols, such as high specificity for serious final diagnoses. Qualitative studies would be useful to shed light on the reasons behind the participants' gratitude for feedback of a PSIF despite experiencing negative impacts, and their changes in opinions on whether or not feedback of PSIFs should always be provided. Further work is needed to assess participants' understanding of consent, in particular, how this may vary by the provision of different information materials and how factors such as age, education level, and cognitive function influence understanding. Such work would facilitate evidence-based design of consent materials and processes by researchers and their evaluation by research ethics committees; feasibly UK Biobank data would enable such investigations.

While UK Biobank survey participants up to six months following feedback of a PSIF, studies with longer follow-up periods are likely needed to elucidate final diagnoses in some cases, and to more fully capture the medical and non-medical impacts of feedback. Furthermore, while the UK Biobank surveys GPs, and qualitative studies have interviewed small groups of physicians who manage patients with IFs (Booth and Boyd-Ellison, 2015; Zafar et al., 2016), the views of secondary healthcare professionals such as radiologists or speciality physicians and surgeons on managing apparently asymptomatic research volunteers with PSIFs within publicly-funded healthcare systems would also be useful to inform the debates on the benefits and harms of feedback.

Finally, while this thesis did not aim to evaluate the management of particular types of PSIFs, such as conservative versus active treatments, our data on the prevalence of PSIFs would inform the power calculations for trials of treatments, which are needed in order to ensure that individuals receive good clinical care, and that publicly-funded healthcare services are used appropriately.

7.5 Final summary

PSIFs, i.e. findings with the potential to impact on health, occur in around 1.5% of apparently asymptomatic adults undergoing either brain or thorax MRI, and in around 2% undergoing abdomen MRI and 4 to 13% on brain and body MRI. Around half of PSIFs are suspected to be malignant on research imaging, but most (around 80%) PSIFs turn out not to represent serious disease.

Feedback of PSIFs results in substantial impacts in terms of clinical assessments, non-medical impact on participants (including on emotional wellbeing, insurance and finances and work and activities), and economic impacts on hospital services. Importantly, as the majority of PSIFs turned out not to be serious, many of these impacts may well be deemed unnecessary in retrospect.

Researchers and policymakers may mitigate against these potential harms through careful design of their PSIFs protocols, as this has the greatest effect on the prevalence of PSIFs and on serious final diagnoses, far more than any participant-level characteristic. We show that systematic radiologist review of images generated around 13 times more PSIFs and a lower proportion of serious final diagnoses than a protocol of radiographer flagging of concerning images for radiologists to review. There is no 'one-size-fits-all' best PSIFs protocol, and neither of these protocols may be feasible or appropriate for other studies conducted in other contexts; data on the impacts of other PSIFs protocols would further facilitate evidence-based policy design, and provide further empirical data to inform the ongoing debate on the benefits and harms of feedback.

Researchers must take care to clearly explain to potential participants how their images and PSIFs will be handled in order to obtain truly informed consent. Ideally, researchers would evaluate participants' understanding of consent, and factors associated with correct understanding, which in turn would inform on methods to design consent materials and processes for future studies. Furthermore, qualitative work is needed to understand participants' and GPs' opinions on whether or not PSIFs should always be provided, participants' gratitude for feedback of PSIFs despite the measured negative impacts, and different stakeholders' preferred outcomes of a PSIFs protocol. Finally, further work is needed to confirm the results in this thesis using data from a larger cohort, which will be available in future as the UK Biobank Imaging Study continues to work toward imaging 100,000 of its participants.

Abbreviations

ACMG	American College of Medical Genetics and Genomics
BMI	Body mass index
CI	Confidence interval
CT	Computed tomography
DXA	Dual-energy X-ray absorptiometry
GNC	German National Cohort
GP	General practitioner
HES	Hospital episode statistics
HRG	Healthcare resource group
IFs	Incidental findings
IQR	Interquartile range
MRI	Magnetic resonance imaging
NEO	Netherlands Epidemiology of Obesity
NHS	National Health Service
OR	Odds ratio
PSIFs	Potentially serious incidental findings
RECORD	Reporting of studies Conducted using Observational Routinely-collected health Data
SHIP	Study of Health in Pomerania
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
VOMIT	Victims of medical imaging technology

Glossary

Apparently asymptomatic people	People who were not selected on the basis of any symptoms, risk factors, or disease, and who attended for population-based research imaging studies, commercial or occupational screening, or as research controls
Brain and body MRI	MRI of the brain, thorax and abdomen
IFs	Findings unrelated to the purpose of imaging
Indeterminate final diagnoses	See ***uncertain final diagnoses***
Indeterminate IFs	IFs which could not be classified as either ***PSIFs*** or ***non-serious IFs***
Non-malignant PSIFs	PSIFs which were neither ***suspected malignancies*** or ***possible indicators of malignancy***
Non-serious final diagnoses	Final diagnoses which were not likely to significantly threaten lifespan or have a major impact on quality of life or major body functions, or diagnoses which were already known
Non-serious IFs	IFs which were not likely to indicate a condition which would seriously threaten life span, or of have a substantial impact on major body functions or quality of life
Patients	People selected for a study based on symptoms, risk factors or disease, or those admitted to or attending a health care facility for clinical assessment
Possible indicator of malignancy	PSIFs which were not masses, but could be related to malignancy, e.g. pleural effusions
PSIFs	Findings indicating the possibility of a condition which, if confirmed, would carry a real prospect of

	seriously threatening life span, or of having a substantial impact on major body functions or quality of life
Serious final diagnoses	Final diagnoses which were likely to significantly threaten lifespan or have a major impact on quality of life or major body functions
Suspected malignancy	PSIFs which were described as tumours, masses, complex cysts or lesions
Uncertain final diagnoses	Final diagnoses which could not be classified as either ***serious final diagnoses*** or ***non-serious final diagnoses***

References

ACMG Board of Directors. 2015. ACMG policy statement: updated recommendations regarding analysis and reporting of secondary findings in clinical genome-scale sequencing. Genet Med 17(1):68-9.

Alphs HH, Schwartz BS, Stewart WF, Yousem DM. 2006. Findings on brain MRI from research studies of occupational exposure to known neurotoxicants. AJR Am J Roentgenol 187(4):1043-7.

Anand SS, Tu JV, Awadalla P, Black S, Boileau C, Busseuil D, Desai D, Despres JP, de Souza RJ, Dummer T, Jacquemont S, Knoppers B, Larose E, Lear SA, Marcotte F, Moody AR, Parker L, Poirier P, Robson PJ, Smith EE, Spinelli JJ, Tardif JC, Teo KK, Tusevljak N, Friedrich MG. 2016. Rationale, design, and methods for Canadian alliance for healthy hearts and minds cohort study (CAHHM) - a Pan Canadian cohort study. BMC Public Health 16:650.

Anon. 2005. How volunteering for an MRI scan changed my life. Nature 434(7029):17.

Apold VS, Downie J. 2011. Bad news about bad news: the disclosure of risks to insurability in research consent processes. Account Res 18(1):31-44.

Baumgart D, Egelhof T. 2007. [Preventive whole-body screening encompassing modern imaging using magnetic resonance tomography]. Herz 32(5):387-94.

Bazzocchi A, Ferrari F, Diano D, Albisinni U, Battista G, Rossi C, Guglielmi G. 2012. Incidental findings with dual-energy X-ray absorptiometry: spectrum of possible diagnoses. Calcif Tissue Int 91(2):149-56.

Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sorensen HT, von Elm E, Langan SM. 2015. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS Med 12(10):e1001885.

Bendix K, Jensen JM, Poulsen S, Mygind N, Norgaard BL. 2011. Coronary dual source multi detector computed tomography in patients suspected of coronary artery disease: prevalence of incidental extra-cardiac findings. Eur J Radiol 80(1):109-14.

Bertheau RC, von Stackelberg O, Weckbach S, Kauczor H-U, Schlett CL. 2016. Management of incidental findings in the German National Cohort. In: Weckbach S, editor. Incidental radiological findings. First ed. Cham, Switzerland: Springer.

Bhaskaran K, Douglas I, Forbes H, dos-Santos-Silva I, Leon DA, Smeeth L. 2014. Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5.24 million UK adults. Lancet 384(9945):755-65.

Bland M. 2006. Meta-analysis: publication bias. Accessed: 26/09/18. https://www-users.york.ac.uk/~mb55/msc/systrev/week7/pub_text.pdf.

Booth TC, Boyd-Ellison JM. 2015. The current impact of incidental findings found during neuroimaging on neurologists' workloads. PLoS One 10(2):e0118155.

Booth TC, Jackson A, Wardlaw JM, Taylor SA, Waldman AD. 2010. Incidental findings found in "healthy" volunteers during imaging performed for research: current legal and ethical implications. Br J Radiol 83(990):456-465.

Borenstein M, Hedges LV, Higgins JP, Rothstein HR. 2010. A basic introduction to fixed-effect and random-effects models for meta-analysis. Res Synth Methods 1(2):97-111.

Bos D, Poels MM, Adams HH, Akoudad S, Cremers LG, Zonneveld HI, Hoogendam YY, Verhaaren BF, Verlinden VJ, Verbruggen JG, Peymani A, Hofman A, Krestin GP, Vincent AJ, Feelders RA, Koudstaal PJ, van der Lugt A, Ikram

MA, Vernooij MW. 2016. Prevalence, clinical management, and natural course of incidental findings on brain MR images: the population-based Rotterdam Scan Study. Radiology 281(2):507-515.

Bos D, Vernooij MW. 2016. Management of incidental findings on neuroimaging in the Rotterdam Study. In: Weckbach S, editor. Incidental radiological findings. First ed. Cham, Switzerland: Springer.

Boutet C, Vassal F, Celle S, Schneider FC, Barthelemy JC, Laurent B, Barral FG, Roche F. 2017. Incidental findings on brain magnetic resonance imaging in the elderly:the PROOF study. Brain Imaging Behav 11(1):293-299.

British Medical Association. 1995. Alcohol: guidelines on sensible drinking. London, UK.

Bromage SJ, Liew MP, Moore KC, Raju B, Shackley DC. 2012. The economic implications of unsuspected findings from CT urography performed for haematuria. Br J Radiol 85(1017):1303-6.

Brown DA, Hasso AN. 2008. Toward a uniform policy for handling incidental findings in neuroimaging research. Am J Neuroradiol 29(8):1425-7.

Brown M, Knight HM. 2010. Attitudes towards participating in fMRI studies amongst participants in a birth cohort study. Accessed: 26/09/18. http://www.cls.ioe.ac.uk/library-media/documents/CLS_WP_2010_8_.pdf.

Brugulat-Serrat A, Rojas S, Bargallo N, Conesa G, Minguillon C, Fauria K, Gramunt N, Molinuevo JL, Gispert JD. 2017. Incidental findings on brain MRI of cognitively normal first-degree descendants of patients with Alzheimer's disease: a cross-sectional analysis from the ALFA (Alzheimer and Families) project. BMJ Open 7(3):e013215.

Bui TH, Raymond FL, Van den Veyver IB. 2014. Current controversies in prenatal diagnosis 2: should incidental findings arising from prenatal testing always be reported to patients? Prenat Diagn 34(1):12-7.

Bunnik EM, van Bodegom L, Pinxten W, de Beaufort ID, Vernooij MW. 2017. Ethical framework for the detection, management and communication of incidental findings in imaging studies, building on an interview study of researchers' practices and perspectives. BMC Med Ethics 18(1):10.

Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, Social Sciences and Humanities Research Council of Canada. 2014. Tri-council policy statement: ethical conduct for research involving humans. Accessed: 26/09/18. http://www.pre.ethics.gc.ca/pdf/eng/tcps2-2014/TCPS_2_FINAL_Web.pdf.

Cassidy S, Chau JY, Catt M, Bauman A, Trenell MI. 2016. Cross-sectional study of diet, physical activity, television viewing and sleep duration in 233,110 adults from the UK Biobank; the behavioural phenotype of cardiovascular disease and type 2 diabetes. BMJ Open 6(3):e010038.

Chen Z, Chen J, Collins R, Guo Y, Peto R, Wu F, Li L. 2011. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. Int J Epidemiol 40(6):1652-66.

Chief Medical Officer for Scotland. 2016. Annual Report 2014-15: realistic medicine. Accessed: 26/09/18. http://www.gov.scot/Resource/0049/00492520.pdf.

China Kadoorie Biobank. 2015. Genotyping data. Accessed: 26/09/18. http://www.ckbiobank.org/site/Study+Resources/Genotyping+Data.

Choi SH, Kwon HJ, Lee SY, Park HJ, Kim MS, Sohn JH, Chung EC, Park HW. 2016. Focal hepatic solid lesions incidentally detected on initial ultrasonography in 542 asymptomatic patients. Abdom Radiol (NY) 41(2):265-72.

Cieszanowski A, Maj E, Kulisiewicz P, Grudzinski IP, Jakoniuk-Glodala K, Chlipala-Nitek I, Kaczynski B, Rowinski O. 2014. Non-contrast-enhanced whole-body

magnetic resonance imaging in the general population: the incidence of abnormal findings in patients 50 years old and younger compared to older subjects. PLoS One 9(9):e107840.
Collins R. 2012. What makes UK Biobank special? Lancet 379(9822):1173-1174.
Council of Europe. 1997. Convention for the protection of human rights and dignity of the human being with regard to the application of biology and medicine: convention on human rights and biomedicine. Accessed: 20/06/19. https://www.coe.int/en/web/conventions/full-list/-/conventions/rms/090000168007cf98.
CTSU. 2018. Mexico City Prospective Study. Accessed: 26/09/18. https://www.ctsu.ox.ac.uk/research/prospective-blood-based-study-of-150-000-individuals-in-mexico.
Danaei G, Vander Hoorn S, Lopez AD, Murray CJ, Ezzati M. 2005. Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors. Lancet 366(9499):1784-93.
de Boer AW, Drewes YM, de Mutsert R, Numans ME, den Heijer M, Dekkers OM, de Roos A, Lamb HJ, Blom JW, Reis R. 2018. Incidental findings in research: a focus group study about the perspective of the research participant. J Magn Reson Imaging 47(1):230-237.
de Mutsert R, den Heijer M, Rabelink TJ, Smit JW, Romijn JA, Jukema JW, de Roos A, Cobbaert CM, Kloppenburg M, le Cessie S, Middeldorp S, Rosendaal FR. 2013. The Netherlands Epidemiology of Obesity (NEO) study: study design and data collection. Eur J Epidemiol 28(6):513-23.
Debette S, Markus HS. 2010. The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis. BMJ 341:c3666.
Delaney SK, Hultner ML, Jacob HJ, Ledbetter DH, McCarthy JJ, Ball M, Beckman KB, Belmont JW, Bloss CS, Christman MF, Cosgrove A, Damiani SA, Danis T, Delledonne M, Dougherty MJ, Dudley JT, Faucett WA, Friedman JR, Haase DH, Hays TS, Heilsberg S, Huber J, Kaminsky L, Ledbetter N, Lee WH, Levin E, Libiger O, Linderman M, Love RL, Magnus DC, Martland A, McClure SL, Megill SE, Messier H, Nussbaum RL, Palaniappan L, Patay BA, Popovich BW, Quackenbush J, Savant MJ, Su MM, Terry SF, Tucker S, Wong WT, Green RC. 2016. Toward clinical genomics in everyday medicine: perspectives and recommendations. Expert Rev Mol Diagn 16(5):521-32.
Dunet V, Schwitter J, Meuli R, Beigelman-Aubry C. 2016. Incidental extracardiac findings on cardiac MR: systematic review and meta-analysis. J Magn Reson Imaging 43(4):929-39.
Elixhauser A, Steiner C, Harris DR, Coffey RM. 1998. Comorbidity measures for use with administrative data. Med Care 36(1):8-27.
Evans BJ. 2013. Minimizing liability risks under the ACMG recommendations for reporting incidental findings in clinical exome and genome sequencing. Genet Med 15(12):915-20.
Farrar J, Savill J. 2014. UK funders' framework for health-related findings in research. Lancet 383(9933):1954-5.
Flicker MS, Tsoukas AT, Hazra A, Dachman AH. 2008. Economic impact of extracolonic findings at computed tomographic colonography. J Comput Assist Tomogr 32(4):497-503.
Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, Collins R, Allen NE. 2017. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. Am J Epidemiol 186(9):1026-1034.

Furtado CD, Aguirre DA, Sirlin CB, Dang D, Stamato SK, Lee P, Sani F, Brown MA, Levin DL, Casola G. 2005. Whole-body CT screening: spectrum of findings and recommendations in 1192 patients. Radiology 237(2):385-94.

German National Cohort (GNC) Consortium. 2014. The German National Cohort: aims, study design and organization. Eur J Epidemiol 29(5):371-382.

Gibson L, Paul L, Wardlaw J, Sudlow C. 2016a. Potentially serious incidental findings on brain and body magnetic resonance imaging conducted among apparently healthy adults: a systematic review. CRD42016029472. Accessed: 27/09/18. http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42016029472.

Gibson LM, Littlejohns TJ, Adamska L, Garratt S, Doherty N, UK Biobank Imaging Working Group, Wardlaw JM, Maskell G, Parker M, R. B, Matthews PM, Collins R, Allen NE, Sellors J, Sudlow CLM. 2018. Impact of detecting potentially serious incidental findings during multi-modal imaging [version 3; referees: 2 approved, 1 approved with reservations]. Wellcome Open Res 2:114.

Gibson LM, Nolan J. 2018. Factors associated with potentially serious incidental findings and with serious final diagnoses on multimodal imaging in the UK Biobank Imaging Study: a prospective cohort study, [software]. Accessed: 26/09/18. https://datashare.is.ed.ac.uk/handle/10283/3112.

Gibson LM, Paul L, Sudlow CLM. 2017a. Potentially serious incidental findings on brain and body magnetic resonance imaging conducted among apparently healthy adults: a systematic review and meta-analysis [dataset] Accessed: 26/09/18. https://datashare.is.ed.ac.uk/handle/10283/2773.

Gibson LM, Sellors J, Sudlow CLM. 2016b. Management of incidental findings on multi-modal imaging in UK Biobank. In: Weckbach S, editor. Incidental radiological findings. First ed. Cham, Switzerland: Springer.

Gibson LM, Sudlow CLM, Wardlaw JM. 2017b. Incidental findings: current ethical debates and future challenges in advanced neuroimaging. In: Illes J, Hossain S, editors. Neuroethics: anticipating the future. Second ed. Oxford, UK: Oxford University Press.

Gluecker TM, Johnson CD, Wilson LA, Maccarty RL, Welch TJ, Vanness DJ, Ahlquist DA. 2003. Extracolonic findings at CT colonography: evaluation of prevalence and cost in a screening population. Gastroenterology 124(4):911-6.

Goehde SC, Hunold P, Vogt FM, Ajaj W, Goyen M, Herborn CU, Forsting M, Debatin JF, Ruehm SG. 2005. Full-body cardiovascular and tumor MRI for early detection of disease: feasibility and initial experience in 298 subjects. AJR Am J Roentgenol 184(2):598-611.

Goehler A, McMahon PM, Lumish HS, Wu CC, Munshi V, Gilmore M, Chung JH, Ghoshhajra BB, Mark D, Truong QA, Gazelle GS, Hoffmann U. 2014. Cost-effectiveness of follow-up of pulmonary nodules incidentally detected on cardiac computed tomographic angiography in patients with suspected coronary artery disease. Circulation 130(8):668-75.

Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, McGuire AL, Nussbaum RL, O'Daniel JM, Ormond KE, Rehm HL, Watson MS, Williams MS, Biesecker LG. 2013. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med 15(7):565-74.

Haberg AK, Hammer TA, Kvistad KA, Rydland J, Muller TB, Eikenes L, Garseth M, Stovner LJ. 2016. Incidental intracranial findings and their clinical impact; the HUNT MRI study in a general population of 1006 participants between 50-66 years. PLoS One 11(3):e0151080.

Hamza TH, van Houwelingen HC, Stijnen T. 2008. The binomial distribution of meta-analysis was preferred to model within-study variability. J Clin Epidemiol 61(1):41-51.

Harris R, Sawaya GF, Moyer VA, Calonge N. 2011. Reconsidering the criteria for evaluating proposed screening programs: reflections from 4 current and former members of the U.S. Preventive Services Task Force. Epidemiol Rev 33:20-35.

Harris RP, Sheridan SL, Lewis CL, Barclay C, Vu MB, Kistler CE, Golin CE, DeFrank JT, Brewer NT. 2014. The harms of screening: A proposed taxonomy and application to lung cancer screening. JAMA Intern Med 174(2):281-286.

Hartwigsen G, Siebner HR, Deuschl G, Jansen O, Ulmer S. 2010. Incidental findings are frequent in young healthy individuals undergoing magnetic resonance imaging in brain research imaging studies: a prospective single-center study. J Comput Assist Tomogr 34(4):596-600.

Hayes LM, Frebault JS, Landercasper J, Borgert AJ, Vang CA, Johnson JM, Linebarger JH. 2016. Extramammary findings in diagnostic breast magnetic resonance imaging among patients with known breast cancer: incidence and cost analysis. Am J Surg 212(6):1194-1200.

Hayward R. 2003. VOMIT (victims of modern imaging technology) — an acronym for our times. BMJ 326(7401):1273.

Hegenscheid K, Kuhn JP, Volzke H, Biffar R, Hosten N, Puls R. 2009. Whole-body magnetic resonance imaging of healthy volunteers: pilot study results from the population-based SHIP study. ROFO 181(8):748-759.

Hegenscheid K, Seipel R, Schmidt CO, Volzke H, Kuhn JP, Biffar R, Kroemer HK, Hosten N, Puls R. 2013. Potentially relevant incidental findings on research whole-body MRI in the general adult population: frequencies and management. Eur Radiol 23(3):816-826.

Heller MT, Harisinghani M, Neitlich JD, Yeghiayan P, Berland LL. 2013. Managing incidental findings on abdominal and pelvic CT and MRI, part 3: white paper of the ACR Incidental Findings Committee II on splenic and nodal findings. J Am Coll Radiol 10(11):833-9.

HES Data Quality Team. 2016a. Data quality checks performed on SUS and HES data. Accessed: 26/09/18. https://digital.nhs.uk/binaries/content/assets/legacy/pdf/d/r/data_quality_checks_performed_on_sus_and_hes_data.pdf.

HES Data Quality Team. 2016b. Methodology for identifying and removing duplicate records from the HES dataset. Accessed: 26/09/18. https://digital.nhs.uk/binaries/content/assets/legacy/pdf/m/o/hes_duplicate_identification_and_removal_methodology.pdf.

Hoang JK, Langer JE, Middleton WD, Wu CC, Hammers LW, Cronan JJ, Tessler FN, Grant EG, Berland LL. 2015. Managing incidental thyroid nodules detected on imaging: white paper of the ACR Incidental Thyroid Findings Committee. J Am Coll Radiol 12(2):143-50.

Hoggard N, Darwent G, Capener D, Wilkinson ID, Griffiths PD. 2009. The high incidence and bioethics of findings on magnetic resonance brain imaging of normal volunteers for neuroscience research. J Med Ethics 35(3):194-9.

Hunter JP, Saratzis A, Sutton AJ, Boucher RH, Sayers RD, Bown MJ. 2014. In meta-analyses of proportion studies, funnel plots were found to be an inaccurate method of assessing publication bias. J Clin Epidemiol 67(8):897-903.

Icelandic Heart Association. 1999. About the Icelandic Heart Association. Accessed: 26/09/18. http://www.hjartarannsokn.is/index.aspx?GroupId=406.

Ikram MA, van der Lugt A, Niessen WJ, Koudstaal PJ, Krestin GP, Hofman A, Bos D, Vernooij MW. 2015. The Rotterdam Scan Study: design update 2016 and main findings. Eur J Epidemiol 30(12):1299-1315.

Ikram MA, van der Lugt A, Niessen WJ, Krestin GP, Koudstaal PJ, Hofman A, Breteler MM, Vernooij MW. 2011. The Rotterdam Scan Study: design and update up to 2012. Eur J Epidemiol 26(10):811-824.

Illes J. 2006. 'Pandora's box' of incidental findings in brain imaging research. Nat Clin Pract Neurol 2(2):60-1.

Illes J, Chin VN. 2008. Bridging philosophical and practical implications of incidental findings in brain research. J Law Med Ethics 36(2):298-304.

Illes J, Kann D, Karetsky K, Letourneau P, Raffin TA, Schraedley-Desmond P, Koenig BA, Atlas SW. 2004a. Advertising, patient decision making, and self-referral for computed tomographic and magnetic resonance imaging. Arch Intern Med 164(22):2415-9.

Illes J, Kirschen MP, Edwards E, Stanford LR, Bandettini P, Cho MK, Ford PJ, Glover GH, Kulynych J, Macklin R, Michael DB, Wolf SM. 2006. Ethics. Incidental findings in brain imaging research. Science 311(5762):783-4.

Illes J, Rosen AC, Huang L, Goldstein RA, Raffin TA, Swan G, Atlas SW. 2004b. Ethical consideration of incidental findings on adult brain MRI in research. Neurology 62(6):888-890.

Jacobs PC, Mali WP, Grobbee DE, van der Graaf Y. 2008. Prevalence of incidental findings in computed tomographic screening of the chest: a systematic review. J Comput Assist Tomogr 32(2):214-21.

James MK, Francois MP, Yoeli G, Doughlin GK, Lee SW. 2017. Incidental findings in blunt trauma patients: prevalence, follow-up documentation, and risk factors. Emerg Radiol 24(4):347-353.

Jefford M, Moore R. 2008. Improvement of informed consent and the quality of consent documents. Lancet Oncol 9(5):485-93.

Kang SK, Spector-Bagdady K, Caplan AL, Braithwaite RS. 2016. Exome and Genome Sequencing and Parallels in Radiology: Searching for Patient-Centered Management of Incidental and Secondary Findings. J Am Coll Radiol 13(12 Pt A):1467-1472.

Katzman GL, Dagher AP, Patronas NJ. 1999. Incidental findings on brain magnetic resonance imaging from 1000 asymptomatic volunteers. JAMA 282(1):36-9.

Khosa F, Krinsky G, Macari M, Yucel EK, Berland LL. 2013. Managing incidental findings on abdominal and pelvic CT and MRI, Part 2: white paper of the ACR Incidental Findings Committee II on vascular findings. J Am Coll Radiol 10(10):789-94.

Kim JW, Kang EY, Yong HS, Kim YK, Woo OH, Oh YW, Lee KY, Han H. 2009. Incidental extracardiac findings at cardiac CT angiography: comparison of prevalence and clinical significance between precontrast low-dose whole thoracic scan and postcontrast retrospective ECG-gated cardiac scan. Int J Cardiovasc Imaging 25 Suppl 1:75-81.

Kingsley-Loso JL, Grey KR, Hanson JL, Raju SI, Parks PR, Bershow AL, Warshaw EM. 2015. Incidental lesions found in veterans referred to dermatology: the value of a dermatologic examination. J Am Acad Dermatol 72(4):651-5.

Kirkwood BR, Stern JAC. 2003. Essential medical statistics. Oxford, UK: Blackwell Publishing Ltd.

Kirschen MP, Jaworska A, Illes J. 2006. Subjects' expectations in neuroimaging research. J Magn Reson Imaging 23(2):205-9.

Kumar R, Sachdev PS, Price JL, Rosenman S, Christensen H. 2008. Incidental brain MRI abnormalities in 60- to 64-year-old community-dwelling individuals:

data from the Personality and Total Health Through Life study. Acta Neuropsychiatr 20(2):87-90.
Lee CI, Tsai EB, Sigal BM, Plevritis SK, Garber AM, Rubin GD. 2010. Incidental extracardiac findings at coronary CT: clinical and economic impact. AJR Am J Roentgenol 194(6):1531-8.
Lee WJ, Chang LB, Lee YC. 2008. Incidental findings on brain MRI. N Engl J Med 358(8):853-854.
Li WC, Tu CH, Chao HT, Yeh TC, Chen LF, Hsieh JC. 2015. High prevalence of incidental brain findings in primary dysmenorrhoea. Eur J Pain 19(8):1071-4.
Lo GG, Ai V, Au-Yeung KM, Chan JK, Li KW, Chien D. 2008. Magnetic resonance whole body imaging at 3 Tesla: feasibility and findings in a cohort of asymptomatic medical doctors. Hong Kong Med J 14(2):90-6.
Loy A, Morgan R, O'Dea S, Daly C, Mulcahy F. 2015. Clinically significant extra-cardiac findings in asymptomatic HIV-positive men undergoing cardiac magnetic resonance imaging. Int J STD AIDS 26(5):346-51.
Lubman DI, Velakoulis D, McGorry PD, Smith DJ, Brewer W, Stuart G, Desmond P, Tress B, Pantelis C. 2002. Incidental radiological findings on brain magnetic resonance imaging in first-episode psychosis and chronic schizophrenia. Acta Psychiatr Scand 106(5):331-6.
Lumbreras B, Donat L, Hernandez-Aguado I. 2010. Incidental findings in imaging diagnostic tests: a systematic review. Br J Radiol 83(988):276-89.
Machaalany J, Yam Y, Ruddy TD, Abraham A, Chen L, Beanlands RS, Chow BJ. 2009. Potential clinical and economic consequences of noncardiac incidental findings on cardiac computed tomography. J Am Coll Cardiol 54(16):1533-41.
MacMahon H, Naidich DP, Goo JM, Lee KS, Leung ANC, Mayo JR, Mehta AC, Ohno Y, Powell CA, Prokop M, Rubin GD, Schaefer-Prokop CM, Travis WD, Van Schil PE, Bankier AA. 2017. Guidelines for management of incidental pulmonary nodules detected on CT images: from the Fleischner Society 2017. Radiology 284(1):228-243.
Magrini N, Font M. 2007. Direct to consumer advertising of drugs in Europe. BMJ 335(7619):526.
Maizlin ZV, Barnard SA, Gourlay WA, Brown JA. 2007. Economic and ethical impact of extrarenal findings on potential living kidney donor assessment with computed tomography angiography. Transpl Int 20(4):338-42.
Manolio TA, Collins R. 2010. Enhancing the feasibility of large cohort studies. JAMA 304(20):2290-1.
Marwick TH, Neubauer S, Petersen SE. 2013. Use of cardiac magnetic resonance and echocardiography in population-based studies: why, where, and when? Circ Cardiovasc Imaging 6(4):590-6.
Matthews PM, Sudlow C. 2015. The UK Biobank. Brain 138(12):3463-3465.
McCaffery KJ, Jansen J, Scherer LD, Thornton H, Hersch J, Carter SM, Barratt A, Sheridan S, Moynihan R, Waller J, Brodersen J, Pickles K, Edwards A. 2016. Walking the tightrope: communicating overdiagnosis in modern healthcare. BMJ 352:i348.
Medical Research Council. 2014. Maximising the value of population based cohorts: MRC strategic review of the largest UK population cohort studies. Accessed: 27/09/18. https://mrc.ukri.org/publications/browse/maximising-the-value-of-uk-population-cohorts/.
Medical Research Council, Wellcome Trust. 2014. Framework on the feedback of health-related findings in research. Accessed: 25/09/18. https://mrc.ukri.org/documents/pdf/mrc-wellcome-trust-framework-on-the-feedback-of-health-related-findings-in-researchpdf/.

Meldrum H, Douglas N. 2010. Doctors call for an end to misleading advertising claims on health tests: a joint statement from the Academy of Medical Royal Colleges and the BMA. Accessed: 19/09/18. http://www.aomrc.org.uk/wp-content/uploads/2016/05/AOMRC_Statement_2010-06-24_Academy_BMA_misleading_advertising_health_tests.pdf.

Menzler K, Iwinska-Zelder J, Shiratori K, Jaeger RK, Oertel WH, Hamer HM, Rosenow F, Knake S. 2010. Evaluation of MRI criteria (1.5 T) for the diagnosis of hippocampal sclerosis in healthy subjects. Epilepsy Res 89(2-3):349-54.

Milstein AC. 2008. Research malpractice and the issue of incidental findings. J Law Med Ethics 36(2):356-60.

Min Y, Agresti A. 2002. Modeling nonnegative data with clumping at zero: a survey. Journal of the Iranian Statistical Society 1(1):7-33.

Mohr JP, Parides MK, Stapf C, Moquete E, Moy CS, Overbey JR, Al-Shahi Salman R, Vicaut E, Young WL, Houdart E, Cordonnier C, Stefani MA, Hartmann A, von Kummer R, Biondi A, Berkefeld J, Klijn CJ, Harkness K, Libman R, Barreau X, Moskowitz AJ. 2014. Medical management with or without interventional therapy for unruptured brain arteriovenous malformations (ARUBA): a multicentre, non-blinded, randomised trial. Lancet 383(9917):614-21.

Morgan AE, Berland LL, Ananyev SS, Lockhart ME, Kolettis PN. 2015. Extraurinary incidental findings on CT for hematuria: the radiologist's role and downstream cost analysis. AJR Am J Roentgenol 204(6):1160-7.

Morin SH, Cobbold JF, Lim AK, Eliahoo J, Thomas EL, Mehta SR, Durighel G, Fitzpatrick J, Bell JD, Taylor-Robinson SD. 2009. Incidental findings in healthy control research subjects using whole-body MRI. Eur J Radiol 72(3):529-33.

Morris Z, Whiteley WN, Longstreth WT, Jr., Weber F, Lee YC, Tsushima Y, Alphs H, Ladd SC, Warlow C, Wardlaw JM, Al-Shahi Salman R. 2009. Incidental findings on brain magnetic resonance imaging: systematic review and meta-analysis. BMJ 339:b3016.

Murphy E, Thompson A. 2009. An exploration of attitudes among black Americans towards psychiatric genetic research. Psychiatry 72(2):177-94.

Mutneja A, Saling L, Yano M, Vijayan A. 2017. Financial impact of incidental radiologic findings during living kidney donor evaluation. Transplant Proc 49(9):2007-2010.

National Human Genome Research Institute. 2012. Overview of the Human Genome Project. Accessed: 26/09/19. https://www.genome.gov/12011241/the-completion-of-the-sequence-and-remaining-goals/.

National Human Genome Research Institute. 2016. The cost of sequencing a human genome. Accessed: 26/09/18. https://www.genome.gov/27565109/the-cost-of-sequencing-a-human-genome/.

Nationale Kohorte. 2018. NAKO. Accessed: 26/09/18. http://nako.de/.

Natori S, Lai S, Finn JP, Gomes AS, Hundley WG, Jerosch-Herold M, Pearson G, Sinha S, Arai A, Lima JA, Bluemke DA. 2006. Cardiovascular function in multi-ethnic study of atherosclerosis: normal values by age, sex, and ethnicity. AJR Am J Roentgenol 186(6 Suppl 2):S357-65.

Nguyen XV, Davies L, Eastwood JD, Hoang JK. 2017. Extrapulmonary findings and malignancies in participants screened with chest CT in the National Lung Screening Trial. J Am Coll Radiol 14(3):324-330.

NHS Digital. 2017a. Grouper user manual: HRG4+ 2016/17 reference costs grouper. Accessed: 26/09/18. https://digital.nhs.uk/binaries/content/assets/legacy/pdf/b/8/hrg4__201617_reference_costs_grouper_user_manual_v1.0.pdf.

NHS Digital. 2017b. HRG4+ 2016/17 reference costs grouper. Accessed: 26/09/18. https://files.digital.nhs.uk/assets/HRG4__201617_Reference_Costs_Grouper.zip.

NHS England. 2018. Diagnostic imaging dataset. Accessed: 26/09/18. https://www.england.nhs.uk/statistics/statistical-work-areas/diagnostic-imaging-dataset/.

NHS Improvement. 2017. National schedule of reference costs 1. Accessed: 26/09/18. https://improvement.nhs.uk/resources/reference-costs/.

O'Sullivan JW, Muntinga T, Grigg S, Ioannidis JPA. 2018. Prevalence and outcomes of incidental imaging findings: umbrella review. BMJ 361:k2387.

Opinion Leader. 2012. Assessing public attitudes to health related findings in research. Accessed: 26/09/18. https://wellcome.ac.uk/sites/default/files/wtvm055196_0.pdf.

Orme NM, Fletcher JG, Siddiki HA, Harmsen WS, O'Byrne MM, Port JD, Tremaine WJ, Pitot HC, McFarland EG, Robinson ME, Koenig BA, King BF, Wolf SM. 2010. Incidental findings in imaging research: evaluating incidence, benefit, and burden. Arch Intern Med 170(17):1525-32.

Parker LS, London AJ, Aronson JD. 2013. Incidental findings in the use of DNA to identify human remains: an ethical assessment. Forensic Sci Int Genet 7(2):221-9.

Patel MD, Ascher SM, Paspulati RM, Shanbhogue AK, Siegelman ES, Stein MW, Berland LL. 2013. Managing incidental findings on abdominal and pelvic CT and MRI, part 1: white paper of the ACR Incidental Findings Committee II on adnexal findings. J Am Coll Radiol 10(9):675-81.

Perneger TV. 1998. What's wrong with Bonferroni adjustments. BMJ 316(7139):1236-8.

Petersen SE, Matthews PM, Bamberg F, Bluemke DA, Francis JM, Friedrich MG, Leeson P, Nagel E, Plein S, Rademakers FE, Young AA, Garratt S, Peakman T, Sellors J, Collins R, Neubauer S. 2013. Imaging in population science: cardiovascular magnetic resonance in 100,000 participants of UK Biobank - rationale, challenges and approaches. J Cardiovasc Magn Reson 15:46.

Petersen SE, Sanghvi MM, Aung N, Cooper JA, Paiva JM, Zemrak F, Fung K, Lukaschuk E, Lee AM, Carapella V, Kim YJ, Piechnik SK, Neubauer S. 2017. The impact of cardiovascular risk factors on cardiac structure and function: Insights from the UK Biobank imaging enhancement study. PLoS One 12(10):e0185114.

Petitti DB. 2001. Approaches to heterogeneity in meta-analysis. Stat Med 20(23):3625-33.

Pickhardt PJ, Hanson ME, Vanness DJ, Lo JY, Kim DH, Taylor AJ, Winter TC, Hinshaw JL. 2008. Unsuspected extracolonic findings at screening CT colonography: clinical and economic impact. Radiology 249(1):151-9.

Pinato DJ, Stavraka C, Tanner M, Esson A, Jacobson EW, Wilkins MR, Libri V. 2012. Clinical, ethical and financial implications of incidental imaging findings: experience from a phase I trial in healthy elderly volunteers. PLoS One 7(11):e49814.

Post W. 2014. The multi-ethnic study of atherosclerosis. Accessed: 26/09/18. http://www.hopkinsmedicine.org/heart_vascular_institute/clinical_trials/preventive/mesa.html.

Prabhakar AM, Le TQ, Abujudeh HH, Raja AS. 2015. Incidental findings and recommendations are common on ED CT angiography to evaluate for aortic dissection. Am J Emerg Med 33(11):1639-41.

Presidential Commission for the Study of Bioethical Issues. 2013. Anticipate and communicate: ethical management of incidental and secondary findings in the clinical, research and direct-to-consumer contexts. Accessed: 26/09/18. https://bioethicsarchive.georgetown.edu/pcsbi/sites/default/files/FINALAnticipateCommunicate_PCSBI_0.pdf.

Prince MR, Zhang H, Zou Z, Staron RB, Brill PW. 2011. Incidence of immediate gadolinium contrast media reactions. AJR Am J Roentgenol 196(2):W138-43.

Priola AM, Priola SM, Giaj-Levra M, Basso E, Veltri A, Fava C, Cardinale L. 2013. Clinical implications and added costs of incidental findings in an early detection study of lung cancer by using low-dose spiral computed tomography. Clin Lung Cancer 14(2):139-48.

Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi JC, Saunders LD, Beck CA, Feasby TE, Ghali WA. 2005. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care 43(11):1130-9.

Rang M. 1972. The Ulysses syndrome. Can Med Assoc J 106(2):122-3.

Ransohoff DF, McNaughton Collins M, Fowler FJ. 2002. Why is prostate cancer screening so common when the evidence is so uncertain? A system without negative feedback. Am J Med 113(8):663-7.

Reneman L, de Win MM, Booij J, van den Brink W, den Heeten GJ, Freling N, Majoie CB. 2012. Incidental head and neck findings on MRI in young healthy volunteers: prevalence and clinical implications. AJNR Am J Neuroradiol 33(10):1971-4.

Ridyard CH, Hughes DA. 2010. Methods for the collection of resource use data within clinical trials: a systematic review of studies funded by the UK Health Technology Assessment program. Value Health 13(8):867-72.

Riley RD, Higgins JP, Deeks JJ. 2011. Interpretation of random effects meta-analyses. BMJ 342:d549.

Roache R. 2014. Why is informed consent important? J Med Ethics 40(7):435-6.

Roberts JS, Ostergren J. 2013. Direct-to-consumer genetic testing and personal genomics services: a review of recent empirical studies. Curr Genet Med Rep 1(3):182-200.

Samuel GN, Farsides B. 2017. The UK's 100,000 Genomes Project: manifesting policymakers' expectations. New Genet Soc 36(4):336-353.

Sandeman EM, Hernandez M, Morris Z, Bastin ME, Murray C, Gow AJ, Corley J, Henderson R, Deary IJ, Starr JM, Wardlaw JM. 2013. Incidental findings on brain MR imaging in older community-dwelling subjects are common but serious medical consequences are rare: a cohort study. PLOS ONE 8(8):e71467.

Saya S, Killick E, Thomas S, Taylor N, Bancroft EK, Rothwell J, Benafif S, Dias A, Mikropoulos C, Pope J, Chamberlain A, Gunapala R, Izatt L, Side L, Walker L, Tomkins S, Cook J, Barwell J, Wiles V, Limb L, Eccles D, Leach MO, Shanley S, Gilbert FJ, Hanson H, Gallagher D, Rajashanker B, Whitehouse RW, Koh DM, Sohaib SA, Evans DG, Eeles RA. 2017. Baseline results from the UK SIGNIFY study: a whole-body MRI screening study in TP53 mutation carriers and matched controls. Fam Cancer 16(3):433-440.

Schelbert EB, Cao JJ, Sigurdsson S, Aspelund T, Kellman P, Aletras AH, Dyke CK, Thorgeirsson G, Eiriksdottir G, Launer LJ, Gudnason V, Harris TB, Arai AE. 2012. Prevalence and prognosis of unrecognized myocardial infarction

determined by cardiac magnetic resonance in older adults. JAMA 308(9):890-6.

Schmidt CO, Hegenscheid K, Erdmann P, Kohlmann T, Langanke M, Volzke H, Puls R, Assel H, Biffar R, Grabe HJ. 2013. Psychosocial consequences and severity of disclosed incidental findings from whole-body MRI in a general population study. Eur Radiol 23(5):1343-51.

Schmidt CO, Sierocinski E, Hegenscheid K, Baumeister SE, Grabe HJ, Volzke H. 2016. Impact of whole-body MRI in a general population study. Eur J Epidemiol 31(1):31-9.

Schramm D, Jasaabuu C, Bach AG, Tennstedt O, Spielmann RP, Surov A. 2016. Costs associated with evaluation of incidental breast lesions identified on computed tomography. Br J Radiol 89(1059):20140847.

Scottish Radiology Transformation Programme, NHS National Services Scotland. 2017. Bulletin. Accessed: 25/09/19. http://www.sharedservices.scot.nhs.uk/media/1475/srtp-bulletin-january-2018.pdf.

Sebastian S, Araujo C, Neitlich JD, Berland LL. 2013. Managing incidental findings on abdominal and pelvic CT and MRI, Part 4: white paper of the ACR Incidental Findings Committee II on gallbladder and biliary findings. J Am Coll Radiol 10(12):953-6.

See MS, Pacifico MD, Harley OJ, Francis I, Smith RW, Jones ME. 2010. Incidence of 'incidentalomas' in over 100 consecutive CT angiograms for preoperative DIEP flap planning. J Plast Reconstr Aesthet Surg 63(1):106-10.

Sherrer RL, Lai WS, Thomas JV, Nix JW, Rais-Bahrami S. 2018. Incidental findings on multiparametric MRI performed for evaluation of prostate cancer. Abdom Radiol (NY) 43(3):696-701.

Shie P, Cardarelli R, Sprawls K, Fulda KG, Taur A. 2009. Systematic review: prevalence of malignant incidental thyroid nodules identified on fluorine-18 fluorodeoxyglucose positron emission tomography. Nucl Med Commun 30(9):742-8.

Siddiki H, Fletcher JG, McFarland B, Dajani N, Orme N, Koenig B, Strobel M, Wolf SM. 2008. Incidental findings in CT colonography: literature review and survey of current research practice. J Law Med Ethics 36(2):320-31.

Sierink JC, Saltzherr TP, Russchen MJ, de Castro SM, Beenen LF, Schep NW, Goslings JC. 2014. Incidental findings on total-body CT scans in trauma patients. Injury 45(5):840-4.

Smith EE, O'Donnell M, Dagenais G, Lear SA, Wielgosz A, Sharma M, Poirier P, Stotts G, Black SE, Strother S, Noseworthy MD, Benavente O, Modi J, Goyal M, Batool S, Sanchez K, Hill V, McCreary CR, Frayne R, Islam S, DeJesus J, Rangarajan S, Teo K, Yusuf S. 2015. Early cerebral small vessel disease and brain volume, cognition, and gait. Ann Neurol 77(2):251-261.

Smith EE, Saposnik G, Biessels GJ, Doubal FN, Fornage M, Gorelick PB, Greenberg SM, Higashida RT, Kasner SE, Seshadri S. 2017. Prevention of stroke in patients with silent cerebrovascular disease: a scientific statement for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 48(2):e44-e71.

Sommer IE, de Kort GA, Meijering AL, Dazzan P, Hulshoff Pol HE, Kahn RS, van Haren NE. 2013. How frequent are radiological abnormalities in patients with psychosis? A review of 1379 MRI scans. Schizophr Bull 39(4):815-9.

Stephan BC, Tzourio C, Auriacombe S, Amieva H, Dufouil C, Alperovitch A, Kurth T. 2015. Usefulness of data from magnetic resonance imaging to improve prediction of dementia: population based cohort study. BMJ 350:h2863.

Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, Liu B, Matthews P, Ong G, Pell J, Silman A, Young A, Sprosen T, Peakman T, Collins R. 2015. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 12(3):e1001779.

Sui Y. 2017. Antiplatelet therapy for silent brain infarction (ANTISBI). Accessed: 26/09/18. https://clinicaltrials.gov/ct2/show/NCT03318744?term=silent+infarct&rank=10.

Sullivan DR, Golden SE, Ganzini L, Hansen L, Slatore CG. 2015. 'I still don't know diddly': a longitudinal qualitative study of patients' knowledge and distress while undergoing evaluation of incidental pulmonary nodules. NPJ Prim Care Respir Med 25:15028.

Suzuki H, Gao H, Bai W, Evangelou E, Glocker B, O'Regan DP, Elliott P, Matthews PM. 2017. Abnormal brain white matter microstructure is associated with both pre-hypertension and hypertension. PLoS One 12(11):e0187600.

Takashima K, Takimoto Y, Nakazawa E, Hayashi Y, Tsuchiya A, Fujita M, Akabayashi A. 2017. Discovery and informing research participants of incidental findings detected in brain magnetic resonance imaging studies: Review and multi-institutional study. Brain Behav 7(5):e00676.

Tarnoki DL, Tarnoki AD, Richter A, Karlinger K, Berczi V, Pickuth D. 2015. Clinical value of whole-body magnetic resonance imaging in health screening of general adult population. Radiol Oncol 49(1):10-6.

Taylor HA, Jr. 2005. The Jackson Heart Study: an overview. Ethn Dis 15(4 Suppl 6):S6-1-3.

The National Casemix Office, Health and Social Care Information Centre. 2014. Casemix companion. Accessed: 26/09/18. http://webarchive.nationalarchives.gov.uk/20171013001217/http://content.digital.nhs.uk/media/13634/Casemix-Companion/pdf/Casemix_Companion_v1.0.pdf.

The National Casemix Office, NHS Digital. 2017. ICD-10 5th edition update: HRG4+ 2016/17 reference costs grouper. Accessed: 26/09/18. https://digital.nhs.uk/binaries/content/assets/legacy/pdf/b/b/hrg4__201617_reference_costs_grouper_icd-10_5th_edition_update_v1.0.pdf.

The Royal College of Radiologists. 2011. Management of incidental findings detected during research imaging. Accessed: 26/09/18. https://www.rcr.ac.uk/system/files/publication/field_publication_files/BFCR%2811%298_ethics.pdf.

The Royal College of Radiologists Standing Scottish Committee. 2016. Unreported X-rays, computed tomography (CT) and magnetic resonance imaging (MRI) studies: results of the September 2016 survey of Scottish NHS health boards Accessed: 19/09/18. https://www.rcr.ac.uk/sites/default/files/unreported_studies_report_-_scotland_-_september_2016.pdf.

Thomas ET, Del Mar C, Glasziou P, Wright G, Barratt A, Bell KJL. 2017. Prevalence of incidental breast cancer and precursor lesions in autopsy studies: a systematic review and meta-analysis. BMC Cancer 17(1):808.

Trufyn J, Hill MD, Scott JN, Modi J, Ciura V, Frayne R, Goyal M, Lautner D, Bhayana D, Davenport WJ, Mah JK, Burton JM, Costello F. 2014. The prevalence of incidental findings in multiple sclerosis patients. Can J Neurol Sci 41(1):49-52.

Tsao CW, Gona P, Salton C, Danias PG, Blease S, Hoffmann U, Fox CS, Albert M, Levy D, O'Donnell CJ, Manning WJ, Yeon SB. 2011. Subclinical and clinical

correlates of left ventricular wall motion abnormalities in the community. Am J Cardiol 107(6):949-55.
Tsushima Y, Taketomi-Takahashi A, Endo K. 2005. Prevalence of abnormal findings on brain magnetic resonance (MR) examinations in adult participants of brain docking. BMC Neurol 5:18.
UK Biobank. 2013a. Feedback in the UK Biobank imaging study.
UK Biobank. 2013b. Hospital episode statistics data in showcase. Accessed: 26/09/18. http://biobank.ctsu.ox.ac.uk/showcase/docs/HospitalEpisodeStatistics.pdf.
UK Biobank. 2014a. Consent form for the imaging assessment: UK Biobank. Accessed: 3rd February 2019. https://imaging.ukbiobank.ac.uk/assets/Uploads/consent-form-copy.pdf.
UK Biobank. 2014b. External data providers: matching algorithms. Accessed: 26/09/18. https://biobank.ctsu.ox.ac.uk/~bbdatan/matching_algorithms_documentation_v1.pdf.
UK Biobank. 2015a. UK Biobank body (abdominal) MRI scan. Accessed: 26/09/18. http://biobank.ctsu.ox.ac.uk/crystal/docs/body_mri_explan.pdf.
UK Biobank. 2015b. UK Biobank imaging modality cardiovascular magnetic resonance (CMR). Accessed: 26/09/18. https://biobank.ctsu.ox.ac.uk/crystal/docs/cardiac_mri_explan.pdf.
UK Biobank. 2015c. UK Biobank imaging modality DXA. Accessed: 26/09/18. http://biobank.ctsu.ox.ac.uk/crystal/docs/DXA_explan_doc.pdf.
UK Biobank. 2015d. UK Biobank imaging modality: carotid ultrasound. Accessed: 26/09/18. http://biobank.ctsu.ox.ac.uk/crystal/docs/carult_explan_doc.pdf.
UK Biobank. 2015e. UK Biobank: report on feasibility of imaging assessments in 100,000 participants.
UK Biobank. 2016. UK Biobank imaging assessment visit: incidental findings. Accessed: 26/09/18. http://www.ukbiobank.ac.uk/wp-content/uploads/2017/11/incidental-findings-2.pdf.
UK Biobank. 2017a. List of available fields. Accessed: 26/09/18. http://biobank.ctsu.ox.ac.uk/crystal/util/ukb_field.txt.
UK Biobank. 2017b. UK Biobank brain imaging documentation. Accessed: 26/09/18. https://biobank.ctsu.ox.ac.uk/crystal/docs/brain_mri.pdf.
UK Biobank. 2017c. UK Biobank touch-screen questionnaire: final version. Accessed: 26/09/18. http://biobank.ctsu.ox.ac.uk/crystal/docs/TouchscreenQuestionsMainFinal.pdf.
UK Biobank. 2018a. Genetic data. Accessed: 26/09/18. http://www.ukbiobank.ac.uk/scientists-3/genetic-data/.
UK Biobank. 2018b. Resources. Accessed: 26/09/18. http://www.ukbiobank.ac.uk/resources/.
UK Biobank. 2018c. UK Biobank Imaging Study. Accessed: 26/09/18. https://imaging.ukbiobank.ac.uk/.
UK Biobank Ethics and Governance Council. 2015. Feedback of health related findings: foreground principles and background perspectives.
UK National Screening Committee, National Health Service. 2014. Thinking of having a private screening test? Accessed: 26/09/18. https://www.nhs.uk/Conditions/breast-cancer-screening/Documents/Public_leaflet_-_Thinking_of_having_a_private_screening_test_FINAL_Nov_2014.pdf.
United Nations Educational Scientific and Cultural Organization. 1997. Universal declaration on the human genome and human rights. Accessed: 20/06/19.

http://portal.unesco.org/en/ev.php-URL_ID=13177&URL_DO=DO_TOPIC&URL_SECTION=201.html.
Veerappan GR, Ally MR, Choi JH, Pak JS, Maydonovitch C, Wong RK. 2010. Extracolonic findings on CT colonography increases yield of colorectal cancer screening. AJR Am J Roentgenol 195(3):677-86.
Vernooij MW, Ikram MA, Tanghe HL, Vincent AJ, Hofman A, Krestin GP, Niessen WJ, Breteler MM, van der Lugt A. 2007. Incidental findings on brain MRI in the general population. N Engl J Med 357(18):1821-8.
Victor RG, Haley RW, Willett DL, Peshock RM, Vaeth PC, Leonard D, Basit M, Cooper RS, Iannacchione VG, Visscher WA, Staab JM, Hobbs HH. 2004. The Dallas Heart Study: a population-based probability sample for the multidisciplinary study of ethnic differences in cardiovascular health. Am J Cardiol 93(12):1473-80.
Vogel-Claussen J, Li D, Carr J, Liu K, Szklo M, Lima JA, Bluemke DA. 2009. Extracoronary abnormalities on coronary magnetic resonance angiography in the multiethnic study of atherosclerosis study: frequency and clinical significance. J Comput Assist Tomogr 33(5):752-4.
Volzke H, Alte D, Schmidt CO, Radke D, Lorbeer R, Friedrich N, Aumann N, Lau K, Piontek M, Born G, Havemann C, Ittermann T, Schipf S, Haring R, Baumeister SE, Wallaschofski H, Nauck M, Frick S, Arnold A, Junger M, Mayerle J, Kraft M, Lerch MM, Dorr M, Reffelmann T, Empen K, Felix SB, Obst A, Koch B, Glaser S, Ewert R, Fietze I, Penzel T, Doren M, Rathmann W, Haerting J, Hannemann M, Ropcke J, Schminke U, Jurgens C, Tost F, Rettig R, Kors JA, Ungerer S, Hegenscheid K, Kuhn JP, Kuhn J, Hosten N, Puls R, Henke J, Gloger O, Teumer A, Homuth G, Volker U, Schwahn C, Holtfreter B, Polzer I, Kohlmann T, Grabe HJ, Rosskopf D, Kroemer HK, Kocher T, Biffar R, John U, Hoffmann W. 2011. Cohort profile: the study of health in Pomerania. Int J Epidemiol 40(2):294-307.
von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. 2008. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol 61(4):344-9.
Wacholder S, McLaughlin JK, Silverman DT, Mandel JS. 1992. Selection of controls in case-control studies. I. Principles. Am J Epidemiol 135(9):1019-28.
Wagner SC, Morrison WB, Carrino JA, Schweitzer ME, Nothnagel H. 2002. Picture archiving and communication system: effect on reporting of incidental findings. Radiology 225(2):500-5.
Wahlund LO, Agartz I, Saaf J, Wetterberg L, Marions O. 1989. [Magnetic resonance tomography in psychiatry--clear benefits for health care services]. Lakartidningen 86(46):3991-4.
Wardlaw JM, Bastin ME, Valdes Hernandez MC, Maniega SM, Royle NA, Morris Z, Clayden JD, Sandeman EM, Eadie E, Murray C, Starr JM, Deary IJ. 2011. Brain aging, cognition in youth and old age and vascular disease in the Lothian Birth Cohort 1936: rationale, design and methodology of the imaging protocol. Int J Stroke 6(6):547-59.
Wardlaw JM, Davies H, Booth TC, Laurie G, Compston A, Freeman C, Leach MO, Waldman AD, Lomas DJ, Kessler K, Crabbe F, Jackson A. 2015. Acting on incidental findings in research imaging. BMJ 351:h5190.
Wardlaw JM, Dorman PJ, Lewis SC, Sandercock PA. 1999. Can stroke physicians and neuroradiologists identify signs of early cerebral infarction on CT? J Neurol Neurosurg Psychiatry 67(5):651-3.
Weber F, Knopf H. 2006. Incidental findings in magnetic resonance imaging of the brains of healthy young men. J Neurol Sci 240(1-2):81-4.

Wilson JMG, Jungner G, World Health Organisation. 1968. Principles and practice of screening for disease. Public Health Papers. Geneva: World Health Organisation.

Wolf SM, Lawrenz FP, Nelson CA, Kahn JP, Cho MK, Clayton EW, Fletcher JG, Georgieff MK, Hammerschmidt D, Hudson K, Illes J, Kapur V, Keane MA, Koenig BA, Leroy BS, McFarland EG, Paradise J, Parker LS, Terry SF, Van Ness B, Wilfond BS. 2008. Managing incidental findings in human subjects research: analysis and recommendations. J Law Med Ethics 36(2):219-248.

Wong CA, Hernandez AF, Califf RM. 2018. Return of research results to study participants: uncharted and untested. JAMA 320(5):435-436.

World Health Organisation. 2000. Obesity: preventing and managing the global epidemic. Accessed: 26/09/18. http://apps.who.int/iris/bitstream/10665/42330/1/WHO_TRS_894.pdf.

World Medical Organisation. 2013. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 310(20):2191-4.

Xiong T, McEvoy K, Morton DG, Halligan S, Lilford RJ. 2006. Resources and costs associated with incidental extracolonic findings from CT colonogaphy: a study in a symptomatic population. Br J Radiol 79(948):948-61.

Yue NC, Longstreth WT, Jr., Elster AD, Jungreis CA, O'Leary DH, Poirier VC. 1997. Clinically serious abnormalities found incidentally at MR imaging of the brain: data from the Cardiovascular Health Study. Radiology 202(1):41-46.

Zafar HM, Bugos EK, Langlotz CP, Frasso R. 2016. "Chasing a ghost": factors that influence primary care physicians to follow up on incidental imaging findings. Radiology 281(2):567-573.

Zins M, Goldberg M. 2015a. The CONSTANCES cohort. Accessed: 26/09/18. http://www.constances.fr/assets/pdf/Scientific-protocol-01-2015.pdf.

Zins M, Goldberg M. 2015b. The French CONSTANCES population-based cohort: design, inclusion and follow-up. Eur J Epidemiol 30(12):1317-28.