ISER Working Paper Series

The Impact of a Personalised Blood Pressure Warning on Health Outcomes and Behaviours

Sonia Bhalotra
Department of Economics
University of Essex

Adeline Delavande
Economics Department
University of Technology Sydney

Paul Fisher
Institute for Social and Economic Research
University of Essex

Jonathan James
Department of Economics
University of Bath

No. 2020-2
February 2020

INSTITUTE FOR SOCIAL
& ECONOMIC RESEARCH

Economic and Social Research Council

University of Essex

www.iser.essex.ac.uk

Non-Technical Summary

Hypertension is often asymptomatic, so many people who have clinically serious hypertension do not know they have it. We investigate in this paper whether being told their exact blood pressure readings and their hypertension status leads to people changing their health-related behaviours. Using longitudinal individual data, we track individuals over time to test whether any change in behaviours is large enough to show up in the chances of dying from a hypertension-related cause such as cardiovascular disease.

The identification challenge is to separate underlying individual risks and behavioural tendencies from the effects of receiving clinical information. To address this we leverage a particular feature of the UK Longitudinal Household Survey (UKLHS) data that produced idiosyncratic variation in the clinical information that individuals had regarding their blood pressure status (i.e. information that was independent of their own traits and tendencies). The survey included a nurse visit during which clinical biomarkers were collected and delivered to survey participants and the variation is a result of the phasing of nurse visits across individuals over time.

Our research yields five important new findings. First, the overall prevalence of hypertension in the population is calculated to be 35.0 percent. It is highest amongst those aged 50 and over (54 percent), with above median income (42.7 percent), without a degree level education (38.3 percent), and male (36.5 percent). 9.5 percent of the population received a negative hypertension shock from the survey nurse, that is, they have high blood-pressure but did not know it. The share receiving a negative shock is highest among African and black Caribbean (13.1 percent), those aged 50 and over (12.7 percent), males (11.6 percent), those with above median income (10.6 percent) and those without a degree (9.8 percent).

Second, providing individuals with a personalised blood pressure warning led respondents to revise downwards their perceptions of their own physical health. It also triggered more formal diagnosis of hypertension by a GP. Compared to similar people who did not get the nurse feedback, the gap in clinically diagnosed hypertension persisted for a substantial four years after feedback.

Third, the sustained monitoring and advice that follows from visiting a GP in the UK led to large reductions in smoking, and some weaker evidence of improvements in diet, but no change in exercise behaviours. Future research, ideally using experimental data, may want to explore whether different types of feedback could influence diet and exercise. One possibility is personalised real time feedback from activity monitors.

Fourth, the combined effects of GP monitoring and behavioural improvements appear to have been large enough to improve respondent health. We find evidence of lower rates of coronary heart disease and congestive heart failure, although no statistical difference in the prevalence of strokes. It would be interesting to see if these findings are replicated in other studies.

Fifth, our work provides methodological insight. Some previous studies in this area have used a regression discontinuity design (RDD), comparing people just above the BP threshold that leads to a hypertension diagnosis with people just below on the premise that these people will tend to be similar. Our examination of the BP data lead us to believe that RDD designs of this sort will be vulnerable to the fact that BP measurements vary with temperature, time of day, and recent activity. In other words, measures may be too noisy to sustain comparison of people with slightly different measurements.

The Impact of a Personalised Blood Pressure Warning on Health Outcomes and Behaviours*

Sonia Bhalotra[1,2], Adeline Delavande[1], Paul Fisher[1], and Jonathan James[3]

[1]Institute of Social and Economic Research, University of Essex

[2]Department of Economics, University of Essex

[3]Department of Economics, University of Bath

February 2020

Abstract

In this paper we examine the impact of a tailored health warning on health outcomes. We exploit the design of a household panel survey that provided feedback to participants on their blood-pressure levels as a quasi-experiment. We find that many participants who were told their blood-pressure was high went on to get a formal diagnosis of hypertension from a medical practitioner. The effect of getting a formal hypertension diagnosis was to reduce the incidence of smoking and improve the quality of diets. However, we do not find changes in monthly alcohol spending. The behavioural changes (plus any prescribed medications) were large enough to reduce the incidence of cardiovascular disease.

JEL Classification: I12, I18, D8

Keywords: tailored information, hypertension, smoking

*Financial support received from the ESRC is also gratefully acknowledged, award number ES/N012038/1. Understanding Society is an initiative funded by the Economic and Social Research Council and various Government Departments, with scientific leadership by the Institute for Social and Economic Research, University of Essex, and survey delivery by NatCen Social Research and Kantar Public. The research data are distributed by the UK Data Service. We thank seminar participants at the University of York, University of Essex, the European Society of Population Economics annual conference 2018, and the European Society of Labor Economists annual conference 2019 for useful comments.

1 Introduction

There is evidence that people do not recognize the existence or severity of chronic disease conditions, many of which are asymptomatic eg. hypertension, diabetes, cancers. A key question is whether they would change their behaviours if better informed. This is important because, if they did, we may expect behavioural responses to control the condition and mitigate its impact on their lives, with direct implications for their welfare and positive externalities for health care services. In particular, even simple pieces of personalised information might lead to large behavioural changes where the initial information triggers the use of available healthcare services such as advice services or GP monitoring. On the other hand, if information was ineffective at improving behaviours, this would have important consequences for our understanding of health behaviours too and the types of information interventions that might be effective.

In general, it is difficult to identify variation in the health status of individuals or in their knowledge of their status that is independent of their characteristics. For instance, individuals from poorer socioeconomic backgrounds may be more likely to have a chronic disease condition and also less likely to know they have it. Likewise, recommended health behaviours are more commonly undertaken by individuals with higher socioeconomic status and better health (Oster (2019)). If we were designing a study, we would conduct clinical tests on a representative sample of individuals and then randomize feedback on their test results to half of them, but this would be expensive and raise ethical issues. Indeed, recent US trials that aggressively targeted blood pressure (BP) reduction were stopped early after a median follow-up of 3.26 years owing to a significantly lower rate of the composite outcome (myocardial infarction, other acute coronary syndromes, stroke, heart failure, or death from cardiovascular causes) in an intensive-treatment group than in a standard-treatment group (SPRINT Research Group (2015)). Against this backdrop, we test the extent to which individual-specific information on health status results in behaviour change and then a change in health outcomes up to five years after the original intervention using a quasi-experiment.

We focus upon a specific hypertension (HT) intervention that took place within a UK

household panel survey. The intervention warned panel participants when their BP was high and, if so, advised them to visit their GP for a formal assessment. Usually, if HT was confirmed by a GP, it would lead to sustained monitoring, that could involve the use of medications and would certainly include an assessment of lifestyle changes that could help control BP. We follow the panel participants for up to five years after the initial BP warning allowing us to study persistence of effects and medium term effects on health behaviours and outcomes of receiving a formal GP diagnosis of HT.

HT is a chronic condition and a global public health issue. In the US alone a third of the population had HT and the economic cost to the nation was estimated to be $51.2 billion per year in 2012 (Benjamin et al. (2017)). Being overweight is closely linked to high BP, as is a lack of physical activity and an unhealthy lifestyle, it can lead to diabetes, heart disease, stroke and numerous other health problems and in 2010 in the US it counted for more cardiovascular disease (CVD) deaths than any other modifiable CVD risk factor (Whelton et al. (2018)). Fortunately, HT can be controlled with medication and lifestyle choices including changes to diet, exercise, alcohol consumption, weight, and smoking (see, for example, Maryon-Davis (2005)). Psychological stress can also affect a person's risk of high BP and an inverse relationship (cross-country) between measures of pyschological well-being and BP has been documented (Blanchflower and Oswald (2008)). New US BP guidelines from the American Heart Association emphasise that BP screening can identify individuals who develop elevated BP over time and that even small improvements in population level behaviours could substantially decrease population CVD rates (Whelton et al. (2018)). In the UK, a screening program has been implemented - the NHS health check - but there is little evidence on whether it is cost effective.

Our analysis has two aims. First, we aim to understand whether information feedback can change health behaviours. We focus on two of the leading risk factors for HT which we can construct on an annual basis from our data: whether a smoker and the amount of alcohol purchased each month. As HT is asymptomatic - even at high levels - our results may be interpreted as stemming from an information effect and are unlikely to suffer confounding with direct debilitating effects of the disease that for instance may prevent

individuals working full-time. Supplementing our main results, we also study nutrition and exercise outcomes, although they are measured at a lower frequency in our data. Second, we analyse whether the information feedback triggered longer-term reductions in CVD including coronary heart disease and strokes, as HT is a leading risk factor for CVD and is the leading risk factor for coronary heart disease and strokes (WHO, 2013), the two leading causes of death in the UK and worldwide.

Smokers are known to exhibit biased assessments of their own health and overestimate lung cancer survival rates that suggest significant room for public health campaigns to educate smokers about the benefits of quiting (Ziebarth (2018)). A similar 'Dunning-Kruger' effect is seen in other health domains where overconfidence is highest among those with low levels of knowledge about the causes and high levels of misinformation (eg. Motta et al. (2018)). Information campaigns advising raised fruit and vegetable consumption or reduced salt intake and smoking are increasingly used to improve behaviours but have had mixed success (Griffith et al. (2017)) and it seems plausible that this is because individual's do not personalize risks presented in population-level terms.[1]. Our paper also contributes to the literature on the behavioural effects of medical testing and feedback e.g. HIV testing (Delavande and Kohler (2012); Thornton (2008)) and Huntington disease (Oster et al. (2013b); Oster et al. (2013a); Oster et al. (2010)).

To identify the effect of interest we exploit two features of the survey design of *Understanding Society*: the UK Household Longitudinal Study (UKHLS) that included biomarker collection including BP. First, we exploit the fact that the fieldwork period for adjacent waves in the panel overlaps by one year giving us representative samples of UK households in the same calendar year but at different waves in the panel and hence with different knowledge of their BP. Second, the high BP warning/feedback is determined by a known function of a participants BP score. This gives an additional source of variation in BP knowledge, once we account for a confounding effects related to the level of BP. We also have annual self-reports of whether a panel respondent is formally hypertensive and so use the survey feedback as an instrumental variable for the self-report.

[1]Psychologists refer to this as 'illusory superiority' - that people judge themselves superior on personality traits to an average peer. See, for example, Hoorens (1993).

We interpret our estimates as local average treatment effects (LATE), that is, we identify the effect of information for those who update their hypertensive status as a result of participating in the survey, but would not have done so otherwise.

We find that the high BP warning of the survey triggered formal diagnosis of HT but for less than 100 percent of those measured high in the survey. The Figure is less than 100 percent, at least in part, because the survey is likely to overestimate BP relative to an assessment conducted by a GP. Still, given the low up-take of GP services by at-risk of CVD groups, the new diagnosis triggered by the survey represent an important improvement in health behaviours. The increases in the formal diagnosis of HT coincided with participants revising downward subjective assessments of their physical health. We then go on to examine the health behaviours of the group formally diagnosed as HT. Our LATE estimates indicate large reductions in their smoking rates 6 months after the survey feedback which persist for 3 years. However, our LATE effects imply only a small effect on overall smoking rates of those getting a high BP warning. We find that diets marginally improved, although the estimates are imprecise and, in contrast, we find no evidence of reductions in alcohol spending. Finally, we examine whether the improvements in health behaviours were large enough to improve health outcomes over the medium term. We find important reductions in the incidence of CVD including congestive heart failure and coronary heart disease but not strokes.

The remainder of the paper is structured as follows. Section II provides some background to HT and the UK setting. Section III describes the data, feedback and empirical methods used in the analysis. Section IV presents the benchmark estimates, after which Section V shows the results from several robustness checks including different specifications of our instrument and reports evidence on the heterogeneity of the effect. Section VI concludes. Supplementary material on the data and additional results discussed throughout the paper are available in an Online Appendix.

2 Background: Hypertension, its measurement, and medical guidelines

HT occurs when BP in the arteries is persistently raised. BP is measured in millimetres of mercury (mmHg) and expressed as two readings: systolic BP measures BP when the heart beats; and diastolic BP measures BP between beats. Worldwide, an estimated 1.13 billion people have HT and it is a major cause of premature death (WHO, 2017). It is a leading risk factor for coronary heart disease (leading cause of death in the UK and worldwide), stroke (second leading cause of death in the UK and worldwide), chronic kidney disease, and aneurysm amongst others. HT is a problem in both the developed and developing world. The prevalence of HT is 34.0 percent in the US (Benjamin et al. (2018)) and 30 percent in the UK (Fat (2018)).

Central to our paper is the fact that HT is asymptomatic even at very high levels. Many therefore carry the condition but will only discover they have it at a medical check-up or if it is diagnosed during the treatment of another condition. There are exceptions as symptoms may be present at extremely raised levels (hypertensive crisis). Symptoms of a hypertensive crisis include headache (22%), epistaxis (17%), faintness, and psychomotor agitation (10%) (Papadopoulos et al. (2010)). A substantial share of the UK adult population has untreated HT although the prevalence has fallen in recent years. The prevalence of undiagnosed HT decreased from 2003 to 2017 for both men (20 to 12 percent) and women (16 to 11 percent) (Fat (2018)).

To diagnosis HT, BP is usually measured on more than one occasion or over a 24-hour period and the BP readings are then compared to clinical thresholds. If BP is measured in the clinic, as it fluctuates, multiple readings are usually taken and by the means of an inflated cuff which is placed around the upper arm and connected to an automatic device. The American Heart Association (AHA) recommends two to three of such readings on two to three separate occasions. This differs from UK practice where NICE guidelines require continuous monitoring of BP over a 24-hour period (ambulatory BP monitoring) once an individual has initially been screened as high on an office-based measure. The thresholds

of the AHA are: 'normal' systolic<120 and diastolic <80; 'elevated/pre-hypertensive' systolic 120-129 and diastolic <80; 'stage 1 hypertension' systolic 130-139 and diastolic 80-89; 'stage 2 hypertension' systolic 140 or higher and diastolic 90 or higher; 'hypertensive crisis' systolic higher than 180 or diastolic higher than 120.

There will be differences between BP as clinically assessed (e.g. by a GP) and as measured by a medical professional in a social science survey, such as UKHLS, even though both represent objective measures of BP. Differently from clinical practice, in the household survey it is typically not possible to measure BP at multiple points in time or continuously through the day to reduce the influence of environmental factors. Environmental factors affecting BP include: time of day (diurnal variation), room temperature, behaviours around the time of measurement (e.g. smoking, exercising), the presence of a medical professional (white coat effect), in addition to the random fluctuations in BP as described above. In appendix A, we adjust our survey measures of BP to account for some of the sources of difference with clinical practice. We find that on average, the survey measures tend to overestimate BP compared to what would be measured by a GP during office hours. For example, we show that survey respondents are more likely to have interviews in the evening when BP is high and that a substantial share of respondents have a home room temperature below medical guidelines that would tend to raise it. Of those measured high in the survey, only 84.6 percent would be high according to our adjusted readings they might obtain in a GP's office. Conversely, we estimate that 2.8 percent of those with normal BP according to the survey will be high according to our adjusted readings.

The benefits of lowering BP to recommended levels are well established and there is now growing evidence that lowering it further continues to bring large reductions in CVD risk (Ettehad et al. (2016)). Ettehad et al. (2016) conclude that a 10 mm Hg reduction in systolic BP reduced the risk of major cardiovascular disease events by 20%, coronary heart disease by 17%, stroke by 27%, heart failure by 28%, and all-cause mortality by 13%. Moreover, the authors find no strong evidence that proportional effects were diminished in people with lower baseline systolic levels (<130 mmHg). Results from a large randomised

control trial in the US (SPRINT trials) are consistent with this finding where targeting a systolic BP of less than 120 mm Hg, as compared with less than 140 mm Hg, resulted in lower rates of fatal and nonfatal major cardiovascular events and death from any cause (SPRINT Research Group (2015)).[2] In this direction, the American College of Cardiology and the American Heart Association lowered the guideline level for diagnosing and treating HT in 2017 substantially below the UK level (to >130 mmHg in the US whereas >140 mmHg in the UK).

HT can be treated with medications and lifestyle modifications or a combination of both. In both the US and UK, those with stage one HT and calculated to have lower CVD risk are first treated exclusively with lifestyle advice. Typical lifestyle advice includes: adopting a diet rich in fruits, vegetables, whole grains, and low-fat dairy foods; to reduce excessive sodium intake, alcohol consumption, and consumption of coffee; smokers should stop smoking; and to increase physical activity to recommend levels. In the UK, patients should receive an annual review to discuss their lifestyle and monitor BP. Both US and UK guidance emphasise the use of home BP monitors as a method to monitor BP.

3 Related Literature

There is a small but growing literature that examines the impact of disease diagnosis on changes in health and related behaviours. The closest paper to ours is by Zhao et al. (2013) who examine the impact of a HT diagnosis on diet. They use a sharp regression discontinuity design using panel data from China and find that people reduce their intake of fat in the 12 months immediately after a HT diagnosis. This effect is largest, and only statistically significant, among the richest third of the population, however there are no differences in the effect by education. Edwards (2016) examines the provision of a range of information related to several biomarkers such as BP, haemoglobin A1c, and total and HDL cholesterol. Very high BP levels (at or above 160 mmHg systolic or 110 diastolic)

[2]Indeed, the intervention was stopped early after a median follow-up of 3.26 years owing to a significantly lower rate of the composite outcome (myocardial infarction, other acute coronary syndromes, stroke, heart failure, or death from cardiovascular causes) in the intensive-treatment group than in the standard-treatment group.

were immediately fed back to the individual whereas other markers were given to the individuals by the form of a letter. He finds, using a fixed effects panel methodology, that on average there were small effects of the provision of biomarkers and indeed there is some evidence of a reduction in health care utilisation. For those who received information on their biomarkers that were in the high and dangerous categories there was an increase in pharmaceutical use, and he finds small degree of weight-loss for those with a diabetes diagnosis.

Other papers have not used a HT diagnosis but have examined other diseases or biomarkers. Carrera et al. (Forthcoming) examine whether the provision tailored health information in relation to cholesterol levels has any impact on food choices among hospital workers in the US. Hospital workers were incentivized to undertake a Health Risk Assessment (HRA), biometric health screening tests, to measure various health characteristics. The HRA covers measures of cholesterol, glucose, BP and BMI (although the authors focus solely on cholesterol levels). By combining these measures together with weekly food purchases from the hospital cafeteria both before and after the HRA. They find the impact of the information is greater for those who are medically trained. In particular, the results show a statistically significant decline in total spending on food purchases among those who were diagnosed as 'high risk', this is most pronounced among those who were unaware of their elevated cholesterol level. In addition, they find a significant increase in the proportion of healthy items purchased among high-risk participants over the age of 55 and women.

Another closely related paper is by Oster (2018) who uses scanner data from households to look at the impact of diabetes diagnosis on food purchases. In this paper, individual diagnosis of diabetes in inferred from purchases of glucose testing products (like glucose monitors) that show up on the household scanning data over time, since such items are required in order to manage their disease and track their blood sugar levels. The results show that, post-diagnosis, households record a statistically significant yet small drop in calories purchased, while in the first month healthy food purchases increase while unhealthy foods decrease, although only this latter effect persists over time.

There is a related literature that focuses on the provision of health and nutritional information to consumers and its impact on diet. Wisdom et al. (2010) find that providing calorie content information on menus at Subway restaurants reduced calorific intake by approximately 7%. Similarly, Bollinger et al. (2011) look at calorie posting at another chain restaurant, this time Starbucks, with results showing that although average calories per transaction fell by around 6%, this was solely driven by changes in food choices, with zero impact on drinks.

This paper is also related to the literature on uncertainty and updating of beliefs regarding people's health and associated behaviors. The lack of knowledge and uncertainty regarding health or disease incidence is well-documented (e.g. Crossley and Kennedy, 2002 and Barrett-Connor et al., 2011), as is the general lack of awareness regarding lifestyle risk factors (e.g. Sanderson et al., 2009). In both cases, standard economic theory would suggest that people will, when faced with new health information, update their beliefs and undertake healthier choices, as suggested by the evidence presented in studies like Carrera et al. (Forthcoming) and Zhao et al. (2013). However, there is also evidence to suggest that when it comes to certain health-related behaviors, people actually *overestimate* the risks involved in terms of falling ill or contracting a disease. For example, Viscusi and Hakes (2008) report that adults on average overestimate the lung cancer risks of smoking, as well as the mortality risks and life expectancy losses. The authors find that higher risk beliefs reduce the likelihood of starting to smoke and increase the probability of smoking cessation among smokers. It follows that new information regarding the true risks of smoking would lead to an increase in smoking and reduced efforts to quit.

Our paper has important differences when compared to these studies, both in terms of design and scope. First, as we have annual measures of whether HT has been formally diagnosed by a health care professional (rather than the survey nurse), we are able to instrument formal HT using the survey feedback as an instrumental variable. This setup allows us to account for the possibility that participants measured as high by the survey do not go on to be formally diagnosed by a healthcare professional. Second, our BP feedback comes with a clear interpretation of the measured BP score and its significance,

unlike the previous studies where only the BP reading was given to participants. Zhao et al. (2013) claim that HT cut-offs are common knowledge in China with adults getting this information from doctors, media outlets as well as the internet. However, evidence indicates this is not the case in the UK. Slark et al. (2014) find that 52% of their sample of over 1000 individuals do not know the acceptable range for BP. Unlike us, Zhao et al. (2013) have more detailed information on diet, being able to examine different macronutrients (such as fats and carbohydrates), however we examine a wider range of modifiable behaviours that contribute to HT such as alcohol consumption (spending on alcohol), regular exercise, and smoking. Even though data limitations mean we miss specific elements of diet consumption, we do have information on various food groups, as such we have a broader range compared to much of the related literature. The sample we consider is also representative of the wider population, which is not the case in the earlier studies. Edwards (2016) examines an exclusively older population of those aged 50 and above and Carrera et al. (Forthcoming)'s sample are those who work in a hospital who might be more sensitive to the provision of health information. Third, we follow participants for up to five years after feedback and so can look at the persistence of any behavioral effects. Finally, and perhaps most importantly, we can observe health outcomes allowing us to study the final impact of information on CVD.

4 Data and Methods

4.1 UKHLS Survey Data and the Personalised Blood Pressure Feedback

This subsection describes our survey data. It places emphasis on the BP measures collected by a trained nurse, variation in the subsequent (one-off) personalised feedback delivered to survey participants and the construction of our analysis sample.

We use data on adults aged at least 16 from the first eight waves of UKHLS that began in 2009. The UKHLS is a large panel survey that subsumed and expanded on the former British Household Panel Survey (BHPS) that began in 1991. It links rich

individual panel data on socio-economic variables, with objectively measured biomarkers. The socio-economic variables - including health behaviours and outcomes (smoking, alcohol spending, physical activity, diet, physical and mental health) - are collected in an annual face to face interview. The biomarker collection was a one-off and collected by a trained nurse during specially designed home visits. The collected biomarkers are BP, anthropometrics, grip strength, lung function, prescribed medications and blood samples. Collected in the annual interviews are self-reports of whether a participant has been formally diagnosed with HT by their GP or other health care professional. The data therefore allows us to distinguish between objectively measured BP at a point in time (i.e. the one-off biomarker), and participants own knowledge of whether they have been formally diagnosed with the condition as determined by their GP or similar (i.e.. the annual self-report).

Survey respondents measured to have high BP were given feedback and advised to visit their GP. Given the centrality of the feedback to our paper, we discuss the details of it below.

There are two ways that the survey generates exogenous variation in the feedback which participants receive. First, the timing of the data collection implies a feedback gap as, at a given point in time, some participants have completed the nurse visit and hence have received feedback, whilst others have not. Notionally, the survey includes two random samples drawn from the GB population $g = 1, 2$. For $g = 1$, participants enter the panel at time t, whereas group 2 enters at t+1. Participants are interviewed annually and both groups receive the nurse visit and feedback approximately five months after their second annual interview.[3] As $g = 1$ entered the panel in 2009, this implies a feedback gap opens up in 2011, as g=1 has participated in the biomarker collection and received feedback approximately 7 months ago, whereas g=2 is yet to take part. By 2012, both groups had participated in the biomarker collection and the feedback gap had closed. Figure 1 illustrates the issue graphically. Also, shown in the Figure is a third group (g=3) that corresponds to the former BHPS that enters UKHLS in 2010.[4] g=3

[3]Figure B1 shows the distribution of the months between the second interview and nurse visit.

[4]The BHPS was a representative sample of the Great Britain household population in 1991. The

likewise took part in the biomarker collection between their second and third interviews and we include it in our main analysis sample. In what follows, we present 'baseline' measures by merging information collected from the second annual interview with the data collected by the nurse.

Second, variation in feedback is generated by the feedback rules that are a function of the participants measured BP. BP was measured three times by the survey nurse and the feedback was based on the minimum of the second two readings. Specifically, participants were allocated to one of four categories and for each category the nurse read a standardised statement during their visit. The statements referred to the level of BP and guidance on how urgently a participant should visit their GP. The categories and guidance were: i.) *normal*, ii.) *mildly raised* and visit GP within 2 months, iii.) *raised* and visit GP within 2 weeks, and iv.) *considerably raised* and visit GP within five days (the threshold values were chosen based on guidelines from the British Hypertension Society and are listed in Table B2).

The oral feedback given to participants constituted a readily understandable personalised at-risk warning.[5] In addition to the verbal statements given by the nurse, participants received a written record in the form of a 'Measurement Record Card' (Figure B2). Additionally, in the cases allocated to the 'considerably raised' category the survey doctor telephoned and wrote to participants to advise them to visit their GP within five days. Participants therefore received a clear indication of their BP reading and its significance.

Our design suffers from non-compliance as not all respondents were willing to participate in the biomarker collection. Of the 35,501 eligible individuals[6], 20,684 agreed to take part while a further 450 did not produce 3 valid BP readings. The overall consent rate was therefore 57.0 percent. Non-consenters tended to be younger, less educated, more likely

following rules mimic the demographic processes by which the population is reproduced, including births and deaths, partnership formations and dissolutions, and emigration. The sample therefore remains representative of the GB population as it changes over time, subject to adjusting for attrition.

[5]In contrast to height, weight, percentage body fat and grip strength for which nurses were strictly prohibited from interpreting the results. For these biomarkers, participants received a record of the measured value.

[6]All UKHLS sample members who took part in the annual survey interview preceding the biomarker collection were eligible, excluding pregnant women, those ill/physically unable, those residing in Northern Ireland and, on cost grounds, a small random sample of English postal sectors in group 2.

to be employed, have better self-reported health and were less likely to have reported HT (Table B1). We drop from the consenting sample 1387 individuals (6.9 percent) for whom it was not possible to construct reported HT histories because of a refusal to answer the relevant question in at least one wave.[7] This leaves us with a baseline sample of 18,847 individuals (8410 in group one, 5843 in group two and 4594 in group three). 15,509 (82.3 percent) were measured by the survey nurse to have '*normal*' BP; 2692 (14.3 percent) to be '*mildly raised*'; 575 (3.1 percent) '*raised*' and 71 (0.4 percent) '*considerably raised*'. We focus our main analysis on the sample of individuals who are interviewed at most seven times (waves 1-7 for groups one and two and waves 2-8 for group three). This means for each individual we have up to two pre-nurse visit data points and five post nurse visit data points.

Our design suffers from attrition, although not excessively so. Of our baseline sample, 12,611 (67 percent) took part in all seven interviews, and over 90 percent took part in at least four interviews. In robustness checks we examine the sensitivity of our results to attrition.

We study a specific set of health outcomes and health behaviours which are the target of HT advice. The outcomes and behaviours we study are derived from self-reports made during the survey interviewers. Health behaviours we study are: the number of fruit and vegetable portions usually consumed per day (participants were shown pictures to help define portion size); whether a current smoker (constructed retrospectively); household monthly spending on alcohol; whether physically inactive (has not walked at least 10 minutes in the last month and reports being physically inactive); self-reported physical activity ranking (scale 0-10); and a bad diet indicator (consumes mainly whole fat milk (rather than semi-skimmed or skimmed) and white bread (rather than brown or wholemeal bread)). The health outcomes we consider are validated physical and mental health functioning scores from the short form 12-item Survey (SF-12) and indicator variables for

[7] g=3 were asked a slightly different question in their annual BHPS interviews before they entered UKHLS. The question referred to having 'Heart/blood pressure or blood circulation problems'. As this is a broader definition of HT, it may mean we over-estimate the prevalence of known HT for this group and under-estimate new cases of it. To confirm our results are not sensitive to this differing definition, we remove g=3 from our sample in sensitivity checks.

ever being diagnosed with a specific chronic condition by time t (congestive heart failure, coronary heart disease, angina, heart attack, and stroke). Not all of the variables are measured annually, rather some are collected on a rotating basis once every few waves. This means they are collected at different distances from the BP feedback. Table C1 summarises the frequency and timing of the variable collection and appendix C provides exact details of the survey questions and how we constructed our final outcome variables.

4.2 Baseline Statistics: Hypertension prevalence, the knowledge gap and health behaviours

We begin by documenting a basic fact that HT and undiagnosed HT is common. Figure 2 presents prevalence rates at baseline for different population subgroups of interest. We count as HT anyone who ever self-reported it in the survey by the time of the nurse visit or who was measured high by the nurse. The Figure decomposes the HT into 3 component parts: those measured high by the survey nurse and who previously reported HT (HT confirmed), those measured normal by the survey nurse who previously self-reported it (HT controlled), and those measured HT by the survey nurse who did not previously know they were HT (HT negative shock). The overall prevalence of HT in the population is calculated to be 35.0 percent. It is highest amongst those aged 50 and over (54 percent), with above median income (42.7 percent), without a degree level education (38.3 percent), and male (36.5 percent). 9.5 percent of the population received a negative HT shock from the survey nurse, that is, they have high BP but did not know it. The share receiving a negative shock is highest among African and black Caribbean (13.1 percent), those aged 50 and over (12.7 percent), males (11.6 percent), those with above median income (10.6 percent) and those without a degree (9.8 percent).[8]

There are compositional differences between the overall HT population and the normotensive; but also between the HT negative shock group and the HT controlled group.

[8] Our prevalence estimates line-up reasonably well with official statistics derived from the Health Survey for England. The official estimates show that HT prevalence ranged from 29-30 percent between 2005 and 2014 and that the proportion of adults with untreated HT (similar to our 'HT negative shock' group) decreased from 2003 to 2017 for both men (20 to 12 percent) and women (16 to 11 percent) (Fat (2018)).

Tables 1 and 2 show demographic comparisons. Table 1 shows that the HT compared to the normotensive are on average older (61 vs. 46), more likely to be male (48 vs. 41 percent), more likely to be white (96 vs. 93 percent), less educated (16 vs. 25 percent with a degree), and are poorer (mean monthly income of £2397 vs. £2941). Table 2 compares demographics for the HT negative shock group and the HT controlled group. Those with HT who got a negative shock are younger (58 vs 63), more likely to be male (55 vs. 45 percent), richer (mean monthly income of £2701 vs. £2284), and tend to be more educated (e.g. 19 vs. 14 percent with a degree).

There is a baseline gap in behaviours and outcomes, conditional on demographic controls, consistent with what participants know about their HT status. Table 3 shows regression estimates where we regress our health outcomes on demographic controls and dummy variables for HT status. Compared to the non-hypertensive, those who knew they were hypertensive (column 2) report worse physical and mental health, have higher measured BP, but report better diets with a mixed picture on having a sedentary lifestyle. Column 4 tests for differences between the hypertensive that did and did not know. We see that those who are HT and did not know have higher measured BP but report better physical and mental health while showing higher smoking rates and worse diets. The HT that did not know therefore show worse behaviour at baseline than their better-informed counterparts and so in principle there is a role for information feedback to close the gap in behaviours. The next Section presents our empirical strategy for determining whether information feedback could help close some of the behaviour gap between the hypertensive that knew and the negative shock group.

4.3 Estimation strategy

This subsection outlines how we use the survey feedback to identify the causal effect of (clinical) information on health behaviours and outcomes. Underpinning our empirical strategy is the notion that the survey feedback would trigger visits to a GP where a clinical diagnosis would be made. A clinical diagnosis by a GP would lead to lifestyle advice, monitoring and where appropriate the prescribing of BP medications. We implement an

IV strategy that exploits the survey feedback to instrument for a formal GP diagnosis of the condition. We identify the effect for compliers (i.e.. those for who survey feedback triggered a GP diagnosis of HT but would not have been diagnosed otherwise) in a LATE framework.

The difficulty in estimating our relationship of interest is that knowledge of one's clinical health status is likely to be correlated with other unobservable characteristics that also effect health behaviours and outcomes. For example, those with extremely high BP may show debilitating symptoms and be more likely to seek a diagnosis from their GP. This would induce a negative correlation between reported HT and good health behaviours such as exercise. On the other hand, the health-conscious may show good health behaviours and at the same time be better informed about their HT status (say if they take-up GP health checks), leading to the reverse correlation.

The ideal instrumental variable induces variation in knowledge of ones HT status but it is uncorrelated with other individual characteristics that effect health behaviours and outcomes. To instrument for whether a person has been formally diagnosed with HT, we use the variation in the timing of the survey feedback exogenously determined by the survey design and whether a participant was told to visit their GP, determined by their observable BP score (Section 4.1).

Our first stage equation is:

$$HypertensiveGP_{it} = \alpha_i + \delta_1 Instrument_{it} + X_{it}\beta + \tau_t + \epsilon_{it} \quad (1)$$

HypertensiveGP$_{it}$ indicates reporting a GP diagnosis of HT by time t and, as the interest is in the effect of the survey feedback on own health knowledge, it is fixed after feedback at the first post-feedback value[9]; X_{it} is a set of time-varying controls[10]; α_i an individual fixed effect that captures time invariant characteristics including BP if it is fixed

[9]We have also allowed HypertensiveGP$_{it}$ to vary after the information feedback, the first stage F-stat is typically smaller but broad conclusions hold.

[10]The controls are: age, age-squared, age-cubed, relationship status (married), employment status, household size, housing type. The overlapping sample design of UKHLS also allows us to also include wave fixed effects, to account for the possibility of differences in reporting behaviour (eg. questionnaire changes) across waves.

in the short-term; and $instrument_{it}$ is a dummy variable equal to one if allocated to one of the 'high' BP categories by the survey nurse and in a post feedback period. Identification comes from differences in the timing of the feedback across different subsamples of the survey and the fact that only those over the BP threshold received a high BP warning from the survey nurse.

Figure 3 shows a simplified picture of the first stage and it confirms that the BP feedback of the nurse (occurring at time zero in the Figure) triggered new diagnoses of HT as measured by the annual self-reports in the survey. Panel A of the Figure isolates new diagnosis of HT by showing, by feedback of the nurse, the share of participants not reporting HT by t-1 who newly report HT at time t (ie. HT entry rates). According to this measure, new diagnoses of HT peak six months after the survey feedback and return to their pre-feedback levels thereafter. The increases/peaks show a monotonic pattern with them being largest for those measured by the nurse as considerably raised (entry rate increases from 15% six months before feedback to 57% six months after), followed by raised (entry rate increases from 7% six months before feedback to 22% six months after), and then mildly raised (entry rate increases from 5% six months before feedback to 7% six months after); while no peak is seen for those measured as normal (which is split into normal high and normal low in the Figure). Also notable is the fact that the entry rates never reach one showing that not all people measured high in the survey subsequently received a diagnosis by their GP. Panel B shows the prevalence of reported HT in each feedback group around feedback. Those measured and assigned to the highest BP groups, prior to feedback, were more likely to report HT and show parallel but slightly increasing trends but six months after feedback, the gaps to the normal feedback groups increase and remain stable thereafter. The prevalence of reported HT for each group six months before and six months after feedback are as follows: 'considerably raised' 52% and 80%; 'raised' 51% and 62%; 'mildly raised' 44% and 48%; 'normal high' 37% and 39%; 'Normal low' 18% and 19%.

A concern might be that the post-feedback jumps in reported HT are a direct response to the feedback, rather than reflecting an increase in formal diagnosis of the condition

from a GP. Such an interpretation can be ruled out for several reasons. First, a direct response implies that participants measured high in the survey should all report being HT after feedback, yet the share doing so remains well below one. Second, under a direct response we would expect a similar jump for all those receiving a high feedback, but what is observed is a monotonic pattern with the biggest jumps occurring for participants given the most serious feedback types. This pattern is however consistent with the measurement issues discussed in Appendix A where misclassification means that a share of those measured HT in the survey turnout not to be HT when measured by their GP. Third, the survey is conducted face-to-face with trained interviewers who are there to advise respondents on the correct interpretation of the survey questions (i.e.. that the questions do not relate to the survey feedback).

In the second stage we estimate:

$$Outcome_{it} = \alpha_i + \delta_1 \widehat{Diagnosed\ with\ HBP}_{it} + X_{it}\beta + \tau_t + \epsilon_{it} \tag{2}$$

where $Diagnosed\ with\ HBP_{it}$ is the fitted value estimated from equation (1). Here we estimate a Local Average Treatment Effect (LATE), that is we estimate the impact on health behaviours of receiving a formal GP diagnosis of HT triggered by the survey nurse visit, but that would not have been diagnosed otherwise.

For our approach to be valid we need to maintain two identifying assumptions. The first, the exclusion restriction, is that conditional on X_{it} and α_i, receiving a high BP feedback is uncorrelated with characteristics that affect health behaviours and outcomes. The inclusion of the individual fixed effects, α_i, in the model accounts for the fact that the feedback, as it is a function of the BP level, will be correlated with other time-invariant characteristics that effect health outcomes. Differently, the inclusion of a range of time varying controls in X_{it} allows us to account for the possibility that there are time varying variables that may be correlated with the feedback/BP level and also effect health outcomes ie. differing trends by BP level. Finally, presented alongside the main results are estimates from models that further include trend terms for X, Y and Z, to account for the possibility of different trends in the outcome by BP level. To check the

validity of the exclusion restriction, we estimated equation (1) but replaced the dependent variable with indicators for receiving a diagnosis of one of 10 other health conditions that are unrelated to HT. If the exclusion restriction is violated, we might expect our instrument to be predictive of the other diagnoses and not just HT. The coefficients and their corresponding 95% confidence intervals are presented in Figure 4, while in Figure 5 we present the F-statistic of the excluded instruments. The results are consistent with the instrument leading to new diagnosis of HT, but not of other health conditions. Two of the coefficients presented are statistically indistinguishable from zero and of the two estimates that are statistically significant (aside from the coefficient on HT) all are very small in magnitude. The only F-statistic above 10 is when the dependent variable refers to diagnosis of HT (indeed, this statistic is very large being over 100). To check against the possibility of differential trends contaminating the main results, we perform robustness checks where we construct a sample of individuals with similar levels of BP by removing those with the lowest measured BP from the sample (Section 5.4).

The second part of the exclusion restriction requires that the instrument only effects the outcome through the first-stage, and not directly. In principle there could be a direct effect if participants did not follow the survey advice to visit their GP, and self treated instead. As the survey feedback was not a definitive HT diagnosis and this was communicated to participants, we judge that any such effect is likely to be small.

The second assumption, monotonicity, requires that the relationship between the instrument (receiving a high-BP feedback from the survey) and the endogenous variable (being diagnosed as clinically HT by their GP) is monotonic. This assumption requires that getting a high feedback from the survey can never make it less likely that an individual gets clinically diagnosed as HT by their GP.

5 Results

5.1 Health Behaviours

We begin by showing the initial effect of the HT diagnosis on two health behaviours that we have measures of each year (smoking and spending on alcohol). Specifically, the estimates shown in Table 4 report the impact 6 months after the nurse visit. The OLS estimates show a negative but imprecisely estimated association between smoking and having ever been clinically diagnosed with HT. The next column reports the reduced form (RF) estimates - this is the impact of the instrument (i.e. the nurse diagnosing HT after the visit) directly on the probability of smoking. Here we find a reduction in smoking by 0.5 percentage points. The final column shows the FE-IV estimates. Scaling the RF estimates by a factor of 20 (as the nurse visit results in someone being told they have HT led to a 5 percentage point increase in the self-reported clinical diagnosis of HT) leads to a FE-IV estimate of around a nine percentage point reduction in smoking. For context the mean shows the proportion in the sample who smoking who are "at-risk" of developing HT is 30% - the effect size we then find is then around a 30% reduction in the smoking rate for this group. While these effects appear large, and they are, it is important to re-emphasize what we estimate is a Local Average Treatment Effects (LATE), such that the impact we find are on those who update their HT status (and get a clinical diagnosis) as a result of the nurse visit and would not have done otherwise - this group are perhaps those not likely to go to the doctor and get a diagnosis in normal circumstances. Therefore, the impact of information for this group on the margin is large and larger that the average effect for the population.

Our treatment was an personalised information treatment which involved medical professionals, however the information itself was not directly related to smoking nor was a smoking cessation programme. However, it is still useful to compare our estimates to smoking cessation programmes. Two Cochrane reviews present evidence on smoking cessation interventions led by nurses (Hill Rice et al., 2017) and by physicians (Stead et al., 2013). Hill Rice et al. (2017) review 44 RCTs comparing the difference between high

intensity interventions, which includes an initial contact of more than 10 minutes with additional materials and/or strategies other than simple leaflets and additional follow-up visits and low intensity interventions which primarily inlcude providing advice (with or without a leaflet) during single consultation lasting 10 minutes or less with up to one follow-up visit. They find a relative risk effect on smoking cessation 1.29 and 1.27 for high and low intensity respectively. When physicians deliver advice, the effect increased such that high intensity interventions lead to a 1.86 relative risk increase in smoking cessation and 1.66 for less intense programmes. Therefore, while at first glance our estimates appear large they are on out of step with interventions that directly target smoking.

We next move onto how this effect evolves over time. Figure 6 shows the causal impact of the HT diagnosis (i.e. the FE-IV estimates) on the probability of smoking at 6 months and then at 12 month intervals up to 30 months[11] the nurse visit and then for data pooled together using the latest wave i.e. at 30 months. In the pooled sample we also allow the outcomes those with different HT statuses (i. mildly raised and ii. raised and considerably raised) to have a different trend. The points for the period at -6 and -18 months indicate that the participants have not received information yet so are set to zero.

The diagnosis leads to a reduction in smoking. There is initially a 9 percentage point drop in the probability of smoking 6 months after the diagnosis - repeating the analysis in Table 4. This then increases to 14 and 20 percentage points 18 and 30 months after the diagnosis respectively, the pooled sample however is close to the effect at 12 months. The pooled estimate becomes imprecisely estimated when we include the two trend variables, although the point estimate is similar to that at 6 months. In general, we cannot however distinguish between the effects at periods, i.e. the confidence intervals overlap to the extent that we cannot say that the impact at 6 months is significantly different from that at 12 months or 18 months.

In contrast, we do not find statistically significant impact on alcohol consumption. The evolution of the estimates for alcohol consumption are shown in Figure 7. We find a positive but imprecisely estimated increase in alcohol consumption up to 42 months after

[11] Due to a change in the definition of the variable we cannot examine beyond 30 months after the nurse visit for smoking but can examine up to 54 months for other outcomes.

the feedback, these effects range from around 10 to 50 pounds. The pooled estimate is of similar value, and is similarly not statistically significant, including the trends increase this value but it remains imprecise. At 54 months the estimate is a fall by pound and again not precisely estimated. Therefore, overall we find, if anything, a positive effect but this is insignificant.

We are also able to examine a wide range of health outcomes that are not measured each year but instead are only captured in waves 2 and 5. These variables are described in more detail in the data appendix. In summary, we can investigate eating patterns (these cover fruit and vegetable consumption, and the type of bread and milk) as well as measures of physical activity. We have chosen health behaviours that medical professionals advise in order to prevent or reduce the incidence of HT[12]. We combine some of the behaviours together. "Bad diet (indicator)" indicates that an individual does at least one of the following: drinks whole milk or eats white bread. "Bad diet (intensity)" takes a one if they do one of drinks whole milk or eats white bread and a two if they do both. "Sedentary lifestyle"' indicates whether someone has not walked for more than 10 minutes over the last 4 weeks and/or they play no sport at all. We also try to capture the extent to which individual engage physical activity using a variable that captures a continuous ranking (form 0 indicating very active to 10 indicating inactive).

Table 4 presents these results. We find a reduction in the probability of eating less than 5 fruits and vegetables a day. The RF point estimate is a two percentage point reduction, with the IV implying a large 36 percentage point reduction among the compliers. Both of these estimates are significant at the 10% level and are relatively imprecisely estimated. The measures of bad diet and physical activity do not show a consistent pattern, none being statistically significant.

5.2 Health Outcomes

We next turn to health outcomes. First, we report estimates from two measures of self-reported health. These are the physical and mental health scores from the short Form

[12]urlhttps://www.who.int/news-room/fact-sheets/detail/hypertension

12-item Survey (SF-12). Second, we consider whether being diagnosed with HT causally leads to subsequently being diagnosed with cardiovascular disease (CVD). In both cases we consider the longer run effects as these make take time to subsequently develop and given, we consider this longer time period we allow those with different HT statuses (either mildly raised or raised and considerably raised) to have a different trends.

In Table 4 we show that there is a negative correlation between HT status and self-reported physical health which is statistically significant. The IV estimate increases in magnitude, but the standard errors also increase, and the estimate is not precisely estimated. Similarly, the IV estimate for the mental health score is not statistically significant. Turning to CVD and the specific CVD conditions. The OLS suggests there is a positive correlation between CVD and HT, those who report having HT are 3 percentage points more likely to report having CVD. When we instrument HT, we find a negative effect, i.e we find are for those who update their HT status (and get a clinical diagnosis) as a result of the nurse visit and would not have done otherwise are less likely to report a CVD diagnosis. This is mainly due to reductions congestive heart failure and coronary heart disease which are precisely estimated. There is also a negative impact on strokes, but these are not statistically significant.

5.3 Heterogeneity

In this Section we consider various groups for whom the information on their HT status might have differing effects.

Gender — There is evidence that women are more likely to be engaged in healthy behaviours and are more likely to seek out health information (Courtenay et al. (2002)). It is also the case that HT rates are lower for women. These differences exist due to both behavioural and biological differences (Dustan (1996)). Given these differences, and that levels of awareness of HT are greater among women (Pereira et al. (2009)), it is interesting to examine whether there are any differences by gender. We present the FE-IV estimates for men and women in in Table 5 each for three different specifications. The first is on the immediate effect 6 months after the nurse visit, the next column we include all the

time periods up to 30 months after for smoking and 54 months for the other outcomes - the "pooled" sample. The final specification is also on the pooled sample but also includes trends allowing those with different levels of HT to differ. We find a statistically significant impact of the diagnosis on the reaction of smoking of men but not women. We find an immediate effect of 15 percentage points, which increases in the pooled sample. The inclusion of HT trends results in a slightly lower (ten percentage point decrease) and imprecise estimate. There also appears to be some impact on congestive heart failure (for females) and coronary heart disease (for males).

Education — One of the key assumptions behind the model of health capital (Grossman (1972)) is that those with more education are more efficient at processing information. Better educated individuals may have greater awareness and understanding of the dangers of HT and may be better placed to seek out ways in which to tackle the problem. They also might have greater access to health and medical care which makes them more likely to react to the information. On the other hand, they might be better informed about their BP to start with and so less likely to react to the information. We explore the impact of the HT diagnosis by different education levels. Specifically, we split the sample by having A-levels or not. We present the results in Table 6 We do not find a different impact of HT diagnosis by education level. There is a larger reduction in self-reported health status (SF-12 physical) for those who have more education. This might indicate that they have understood, and taken on board, the information, more than those who less education.

Age — HT is more likely as people get older. Therefore, as one ages one may become more sensitive to health information related to this condition. Conversely, being informed about having HT at a younger age may come as a greater surprise and as such my trigger a greater change in health behaviour. We examine the impact of the HT diagnosis on those who are aged above 40, 50 and 60. Table 7 presents the IV estimates for the three different specifications, as above. We find a negative and significant reduction in the probability of smoking for all three age groups (the pooled sample with trend is also negative but in each case is imprecisely estimated), as the age group increases in age the point estimate

becomes larger in absolute magnitude. This is also the case for congestive heart failure and coronary heart disease.

5.4 Robustness

We have performed a series of robustness checks on our main results. Figures 12 and appendix Figures B3–B11 present the estimates for our outcomes for the different specifications that we have considered. First, each of the Figure has the corresponding baseline FE-IV estimate. Next we consider the role that other feedback may have played. As part of the nurse visit, individuals also received feedback on their weight, percentage body fat and waist measurements but for these measures the nurses were strictly prohibited from interpreting the results and just received a record of the measured value. However, they could still influence behaviour particularly if the feedback on the other biomarkers is correlated with the HT feedback. Therefore, the next three estimates show the impact of controlling for getting these feedback after the nurse visit, and the fourth controls for all three types of feedback. Diabetes is a common co-morbidity of HT, so we examine the impact when controlling for diabetes. Next we exclude those with high and low systolic BP in order to make sure we are not including those who might be suffering from observable symptoms of very high (such as headaches or chest pain) or very low (dizziness, light-headedness or fainting) BP as these people might seek out medical attention that is unrelated to the feedback associated with the survey.

When we estimate the impact over the longer run we also include additional trends. In their discussion of difference-in-differences Kahn-Lang and Lang (*forthcoming*) suggest allowing trends to depend on group characteristics. The characteristics of these groups may be correlated with having high or low BP. Therefore, we include group specific trends for BMI, diabetes, claiming disability benefits, physical activity, sex, and being from a black Caribbean background. The inclusion of these group specific trends does not have an impact on the smoking results. The inclusion of the HT group trends does have an impact, the point estimates fall but are within a similar range to the immediate effect they do become less precisely estimated and are statistically insignificant.

6 Discussion and conclusions

There is evidence that people do not recognize the existence or severity of chronic disease conditions, many of which are asymptomatic. A key question for public policy is whether people would behave differently if they were better informed about their health. This paper has exploited features of a household survey, that included a BP data collection and feedback of results, as a quasi-experiment to present causal evidence on the effects of receiving personalised health information on health behaviours and outcomes.

There is a fairly large group in the population who are not well informed about their health but appear to respond well to receiving personalised information about it. In particular, 9.5 percent of the population had undiagnosed HT but when they learnt from the survey that their BP was high many of them went on to get a formal diagnosis of the condition from their GP. At baseline, only 45 percent of survey participants who were measured as having high BP reported the condition, but the causal effect of the feedback was to increase this Figure by 6 percentage points 6 months after feedback. Our evidence thus points to the importance of personalised information in reaching groups currently not engaging with the public health system. It is not that they are unresponsive to information or do not care about their health, but rather, they were ill informed at baseline.

Not all of those receiving a high survey feedback, go on to get a formal diagnosis of the condition from their GP and measurement error, rather than an unresponsiveness of participants to information, can explain a large part of this fact. In particular, the effects on BP of time, room temperature, behaviours around the time of measurement, the presence of a medical practitioner (white coat effects) and random noise in any group of BP measures can result in misclassification where a participant measured high in the survey will not be measured high according to a similar group of measures collected in a GP's office hours. Adjusting the survey measures for these factors, leads us to believe that only 85 percent of the survey HT will be measured high in the office by their GP. The measurement issues identified, more generally, point to the challenges of measurement for the analysis of biomarkers by economists. For example, in so far as the measurement differences may have an SES gradient future work might consider their impacts on SES

gradients in BP. Other biomarkers, and not just BP, are a function of environmental conditions and so the findings have relevance beyond BP.

Perhaps our most important finding is that the BP feedback lead to a change in health behaviour by survey participants. Our preferred estimates imply that a formal HT diagnosis from a GP leads to large reductions in smoking (10 ppt 6 months after feedback compared to a smoking rate of 30 percent for an 'at-risk' of smoking group). The smoking reductions imply large improvements in cardiovascular health (associations in our data indicate that being a current smoker increases the risk of CVD incidence by 46 percent in the group with high measured BP). Differently, we do not find evidence of reductions in alcohol spending nor exercise. We do find some evidence of people adopting a healthier diet, although these effects are imprecisely estimated. Unlike generic information campaigns that have been met with mixed success (Griffith et al. (2017)), we therefore find an important role for person specific health information interventions in changing behaviour. One question is why large effects are seen for smoking but much small ones are seen for the other health behaviours. One reason, consistent with a literature in psychology, is that is easier to stop something bad than to start something good. The observed changes in health behaviours were large enough to change health outcomes. We find reductions in CVD incidence of 11 percentage points or around 30 percent of the baseline mean.

There are reasons to believe our results have external validity. They are relevant to understanding the potential of recent initiatives such as the NHS Health Check which offers those aged between 40 and 74 a diagnostic check for heart disease, stroke, diabetes, kidney disease and some forms of dementia with a view to identifying risks early in the life course. The NHS health check, like our biomarker collection, suffers from non-take up. Non-take up of the NHS health check is of a comparable magnitude to the refusals in our sample (50 and 58 percent, respectively), and both groups are of a similar composition (older and more affluent). There is therefore reason to think that the results of the paper are informative about the types of behavioural responses that might be induced by the NHS health check. Our findings are also relevant for the understanding the efficacy

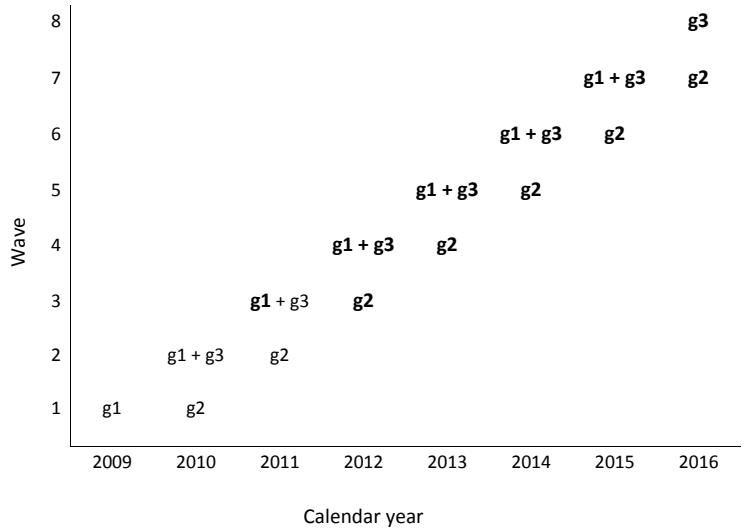
Health Risk Assessments (HRA) and online health tools which provide tailored health information based on lifestyle, family history and socio-economic characteristics.[13]

Finally, we note that there are many areas outside of health where behaviour change might be inhibited as people believe they are 'normal' in some sense e.g. belief about income, driving ability, weight, intelligence, skill level, attitudes, environmental behaviours. Our findings point to the possibility that person specific information may be effective at changing behaviours in those settings too, where general info campaigns have been ineffective.

[13]For example: The British Heart Foundation (BHF) in the UK provides an online tool that will allow an individual to calculate their heart age `www.bhf.org.uk/heart-health/risk-factors/check-your-heart-age`, a similar tool is provided by the American Heart Association and the American Stroke Association `www.heart.org/HEARTORG/Conditions/My-Life-Check---Lifes-Simple-7_UCM_471453_Article.jsp#.WYxWN1GGNPZ`, Diabetes UK (`riskscore.diabetes.org.uk/start`) and the American Diabetes Association (`www.diabetes.org/are-you-at-risk/diabetes-risk-test/` both have tools that allow visitors to their websites to calculate their individual risk of developing diabetes.

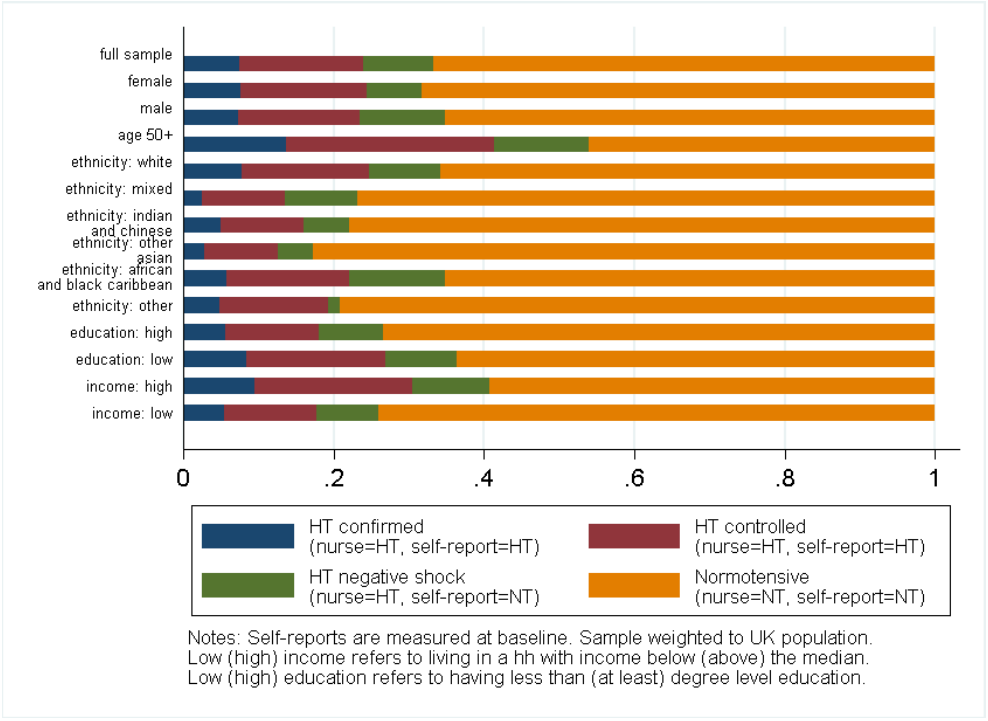
Figures and Tables

Figure 1: Timing of UKHLS data collection



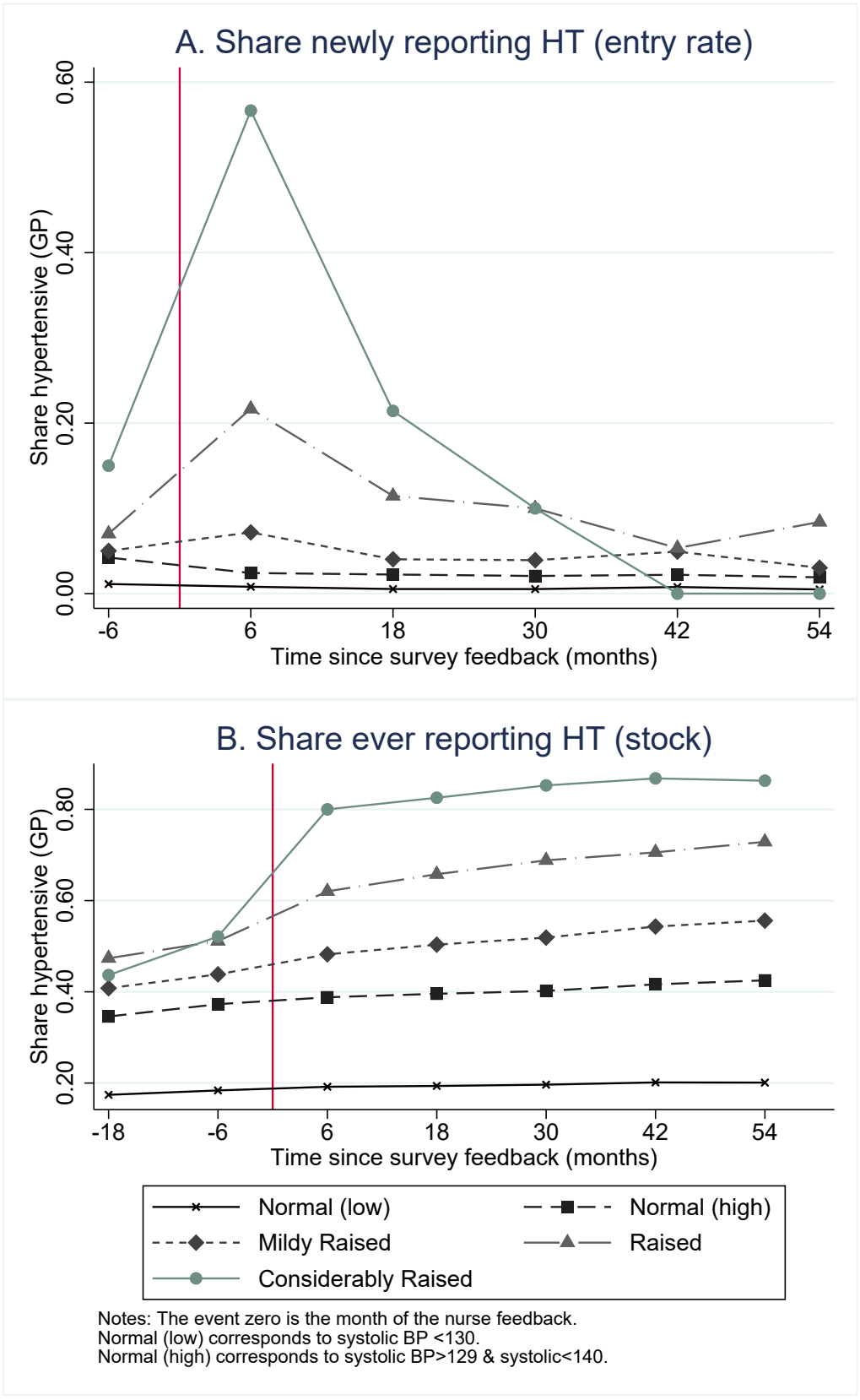
Note: Bold font indicates post feedback period. g1, g2 and g3 denote groups 1,2 and 3 respectively. g3 corresponds to the former BHPS sample.

Figure 2: Prevalence of hypertension and knowledge gap at baseline



Notes: Self-reports are measured at baseline. Sample weighted to UK population.
Low (high) income refers to living in a hh with income below (above) the median.
Low (high) education refers to having less than (at least) degree level education.

Figure 3: The effect of survey feedback on reported hypertension



Notes: The event zero is the month of the nurse feedback.
Normal (low) corresponds to systolic BP <130.
Normal (high) corresponds to systolic BP>129 & systolic<140.

Figure 4: First stage estimates of the hypertension diagnosis on self-reported hypertension and other conditions

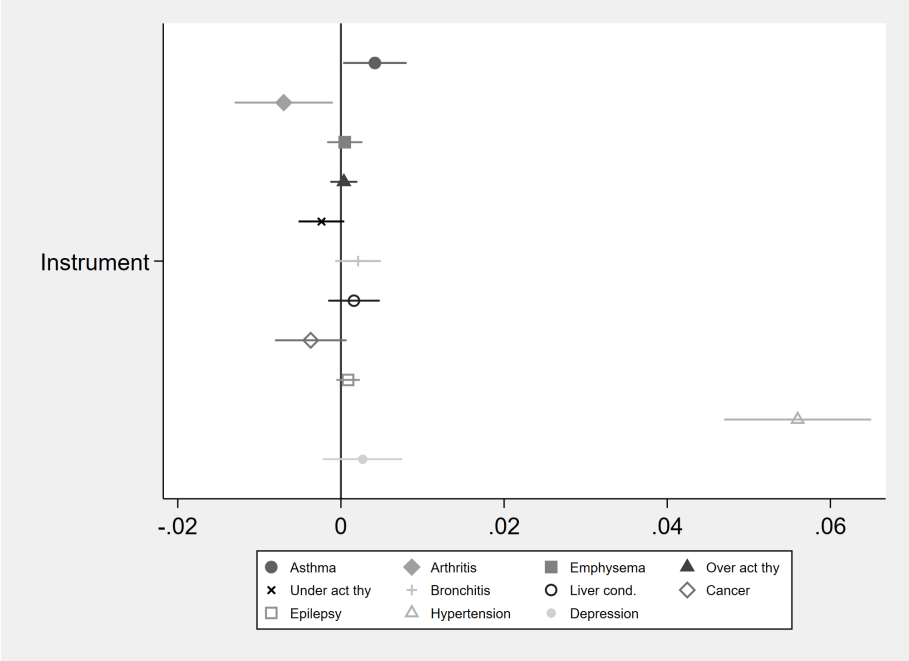


Figure 5: First stage F-statistics of the hypertension diagnosis on self-reported hypertension and other conditions

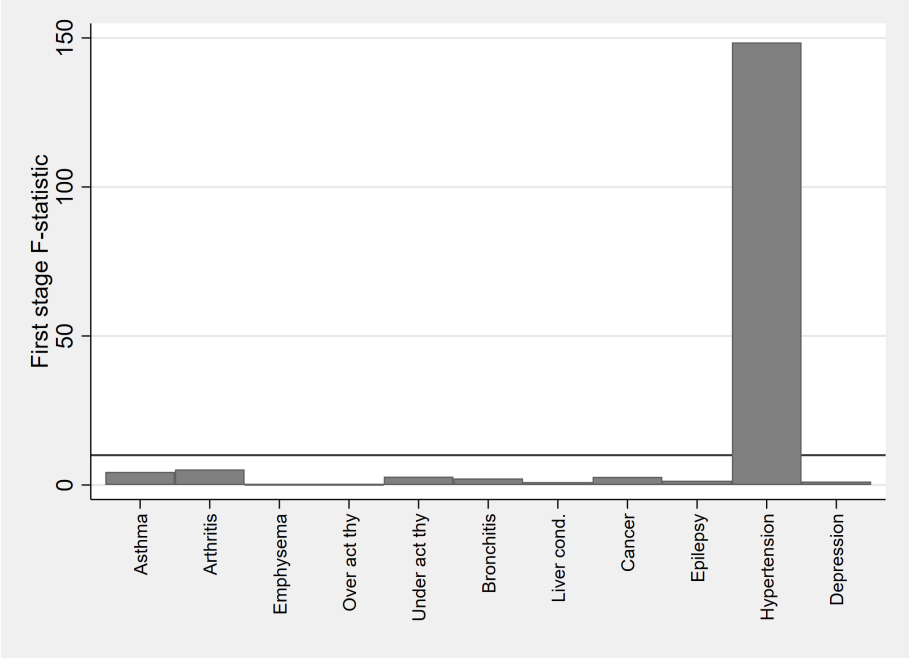
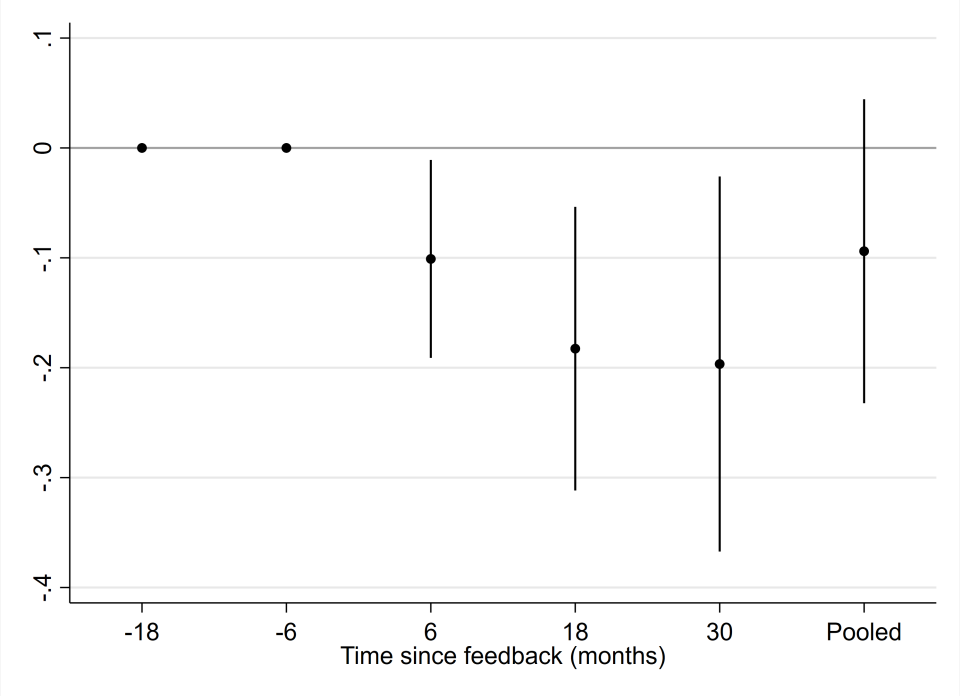
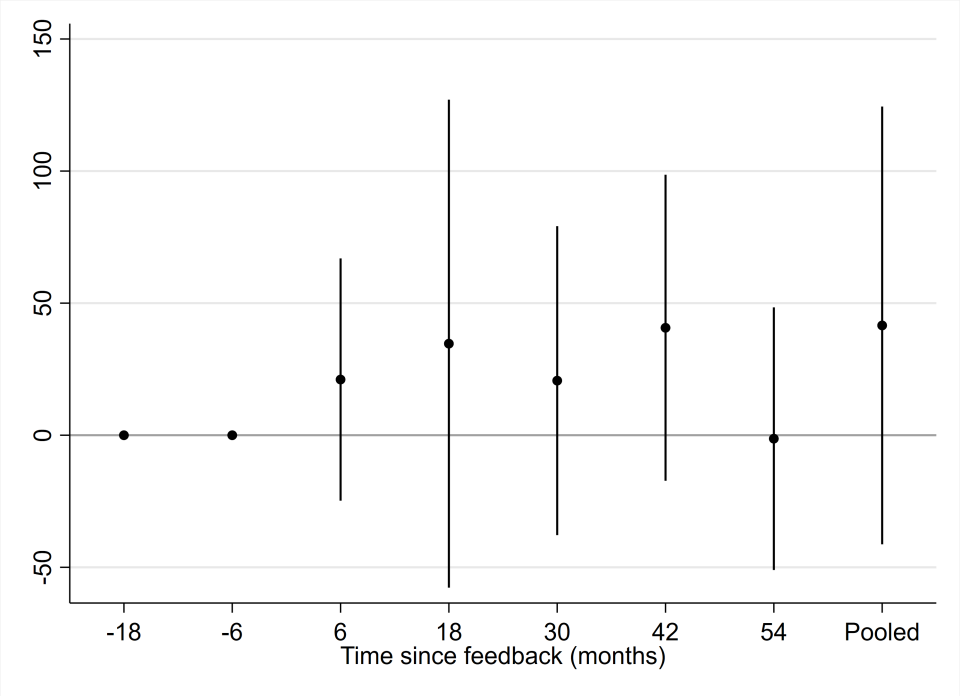


Figure 6: The impact of hypertension diagnosis on smoking



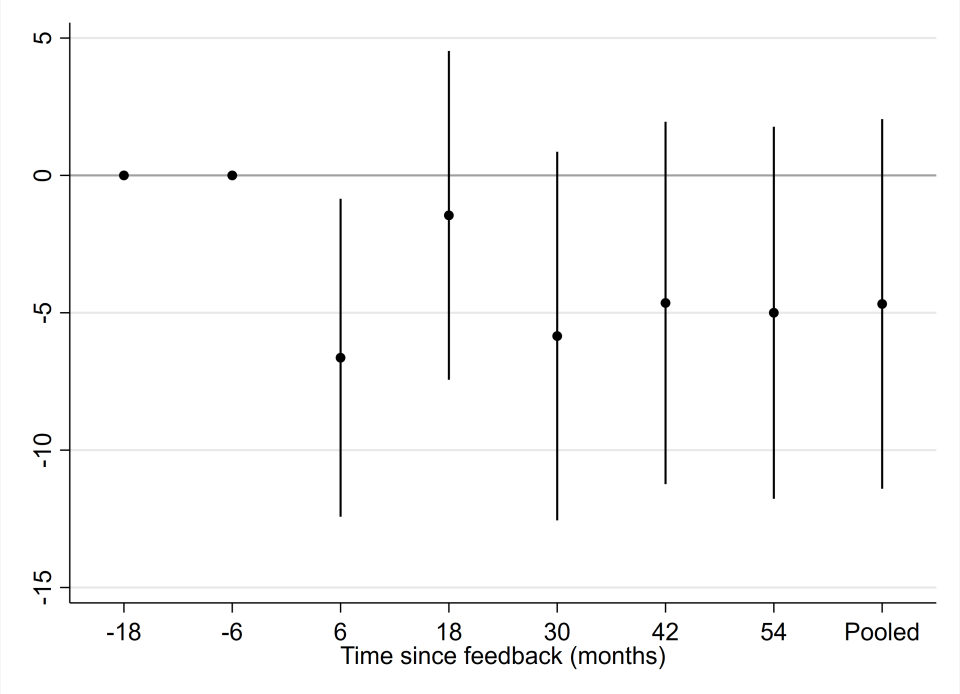
Note: Each dot is from a separate instrumental variable estimates set out in equation (2). The whiskers represent the 95% confidence interval calculated from standard errors clustered at the individual level. All the regressions include individual fixed effects, and the following controls: age, age-squared, age-cubed, household size, and dummies indicating being married, being employed or self-employed, and having qualifications of A-level or about, plus a set of time dummies. The estimates for pooled also includes separate trends for those who have i. mildly raised hypertension and ii. raised and considerably raised hypertension.

Figure 7: The impact of hypertension diagnosis on alcohol expenditure



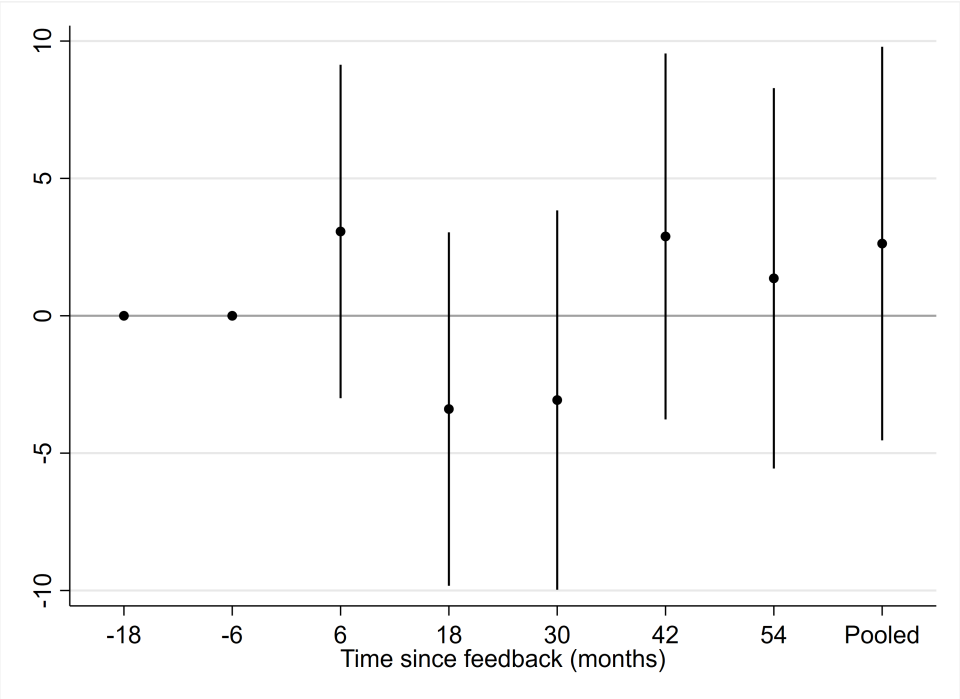
Note: see note to figure 6

Figure 8: The impact of hypertension diagnosis on self-reported physical health



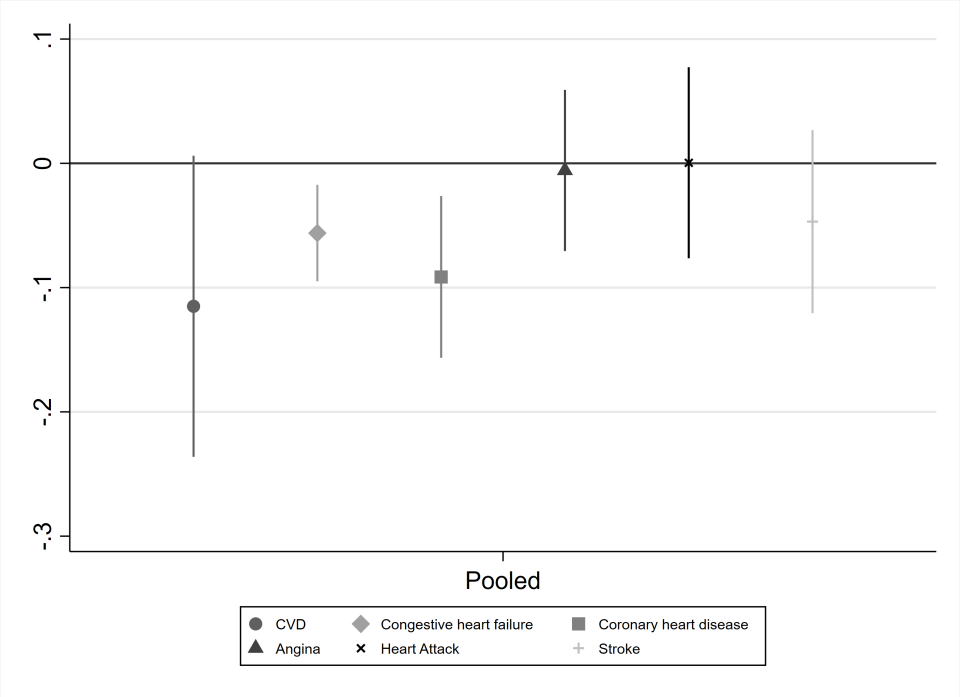
Note: see note to figure 6

Figure 9: The impact of hypertension diagnosis on self-reported mental health



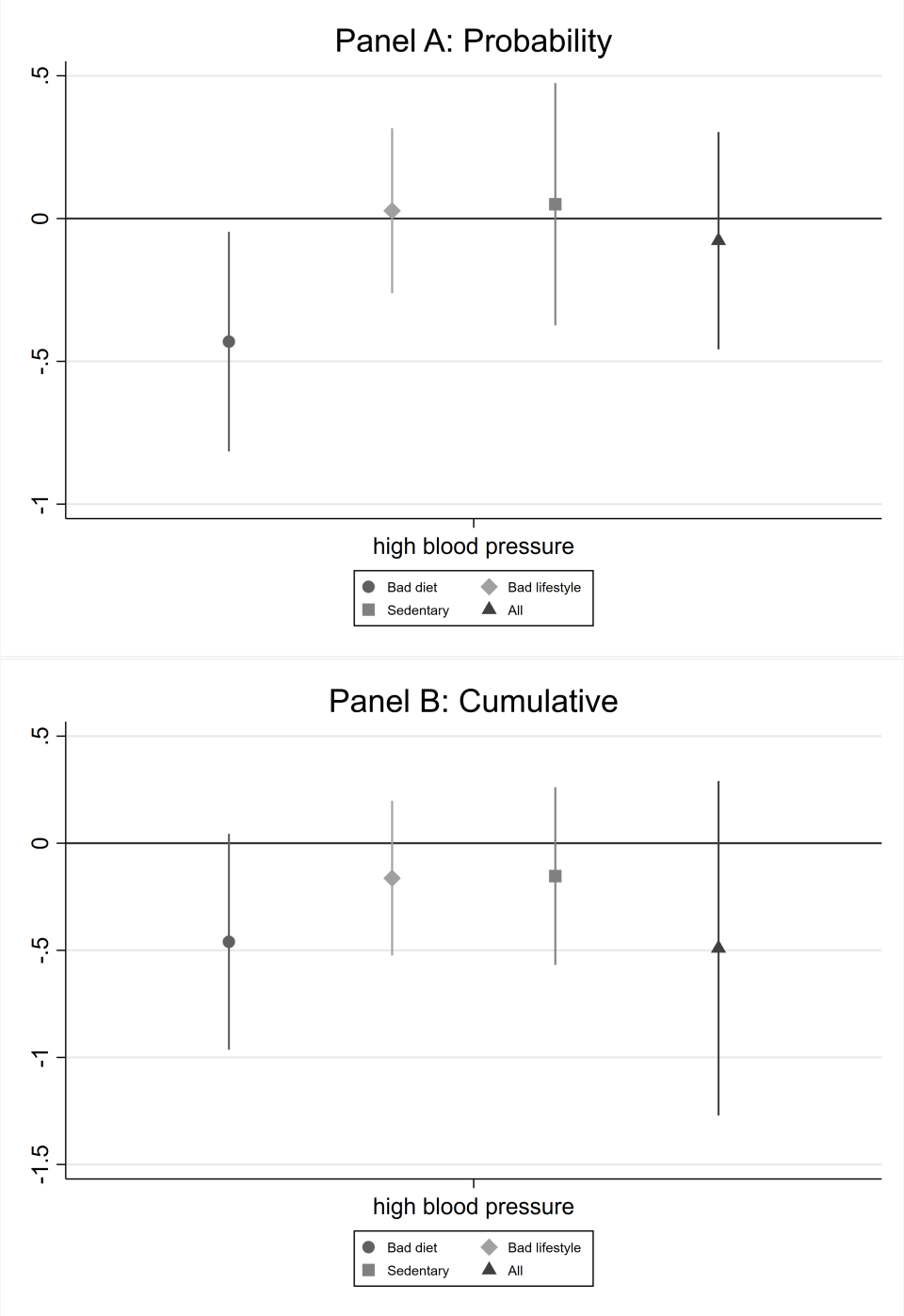
Note: see note to figure 6

Figure 10: The impact of hypertension diagnosis on heart disease related conditions



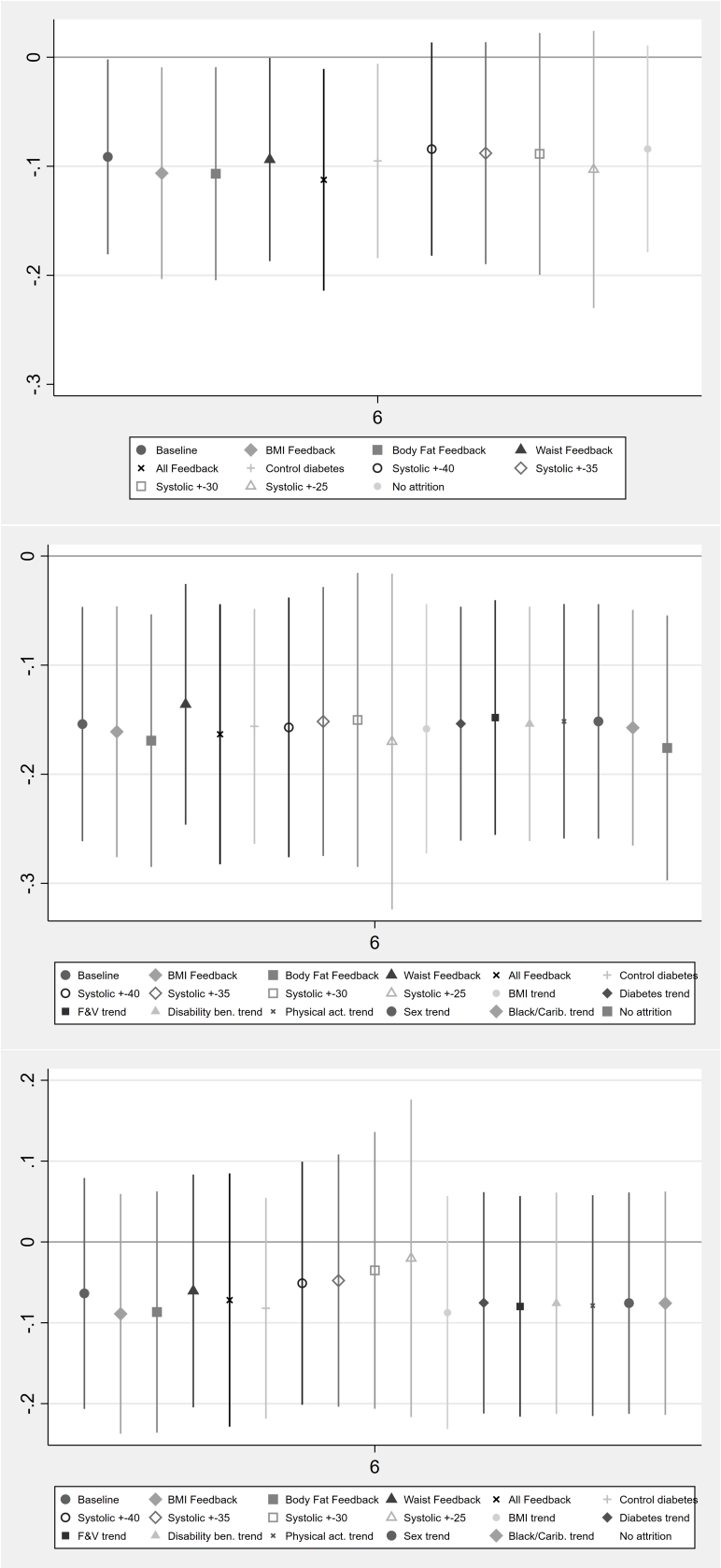
Note: Each marker is from a separate instrumental variable estimates set out in equation (2). The whiskers represent the 95% confidence interval calculated from standard errors clustered at the individual level. All are estimated based on the "pooled" sample (i.e. the sample that goes up to and includes 54 months after the feedback) that have separate trends for those who have i. mildly raised hypertension and ii. raised and considerably raised hypertension.

Figure 11: The impact of hypertension diagnosis on health behaviours



Note: Each dot is from a separate regression (FE-IV) set out in equation (2). The whiskers represent the 95% confidence interval calculated from standard errors clustered at the individual level. Data is from waves 2 and 5 of the UKHLS. "Bad diet" indicates that an individual either never eats fruit, never eats vegetables, drinks wholemilk or eats white bread. "Bad lifstyle" indicates whether an individual either smokes or drinks alcohol 5 or more days a week. Sedentary indicates whether someone has not walked for more than 10 minutes over the last 4 weeks or they play no sport at all. "All" is a combination of all of these.

Figure 12: The impact of hypertension diagnosis on smoking - alternative specifications



Note: Each dot is from a separate regression (FE-IV) set out in equation (2). The whiskers represent the 95% confidence interval calculated from standard errors clustered at the individual level. The top panel shows the immediate impact, the middle panel shows the long run impact not including trends, and the bottom panel is also for the long run and does include separate trends for those who have i. mildly raised hypertension and ii. raised and considerably raised hypertension.

Table 1: Demographic profile by whether hypertensive

	Normotensive	Hypertensive	Mean Diff	SE
Male	0.41	0.48	0.07***	0.01
Age	45.77	61.45	15.67***	0.24
Total household net income	2940.68	2396.95	-543.73***	75.13
Ethnicity:				
White	0.93	0.96	0.02***	0.00
Mixed	0.01	0.01	-0.00*	0.00
Indian and chinese	0.02	0.01	-0.01***	0.00
Other asian	0.02	0.01	-0.01***	0.00
African or black caribean	0.01	0.01	0.00	0.00
Other	0.01	0.00	-0.00**	0.00
Highest qualification:				
Degree	0.25	0.16	-0.10***	0.01
Other higher degree	0.13	0.12	-0.01	0.01
A-level	0.21	0.17	-0.04***	0.01
GCSE	0.22	0.18	-0.04***	0.01
Other	0.09	0.14	0.05***	0.00
No qualification	0.10	0.24	0.14***	0.01
N	12104	6743		

Notes: Unweighted sample

Table 2: Demographic profile of hypertensive by knowledge status

	Know	-ve shock	Mean Diff	SE
Male	0.45	0.55	0.10***	0.01
Age	62.78	57.86	-4.92***	0.40
Total household net income	2283.79	2701.44	417.65**	130.55
Ethnicity:				
White	0.96	0.95	-0.01	0.01
Mixed	0.01	0.01	0.00	0.00
Indian and chinese	0.01	0.01	0.00	0.00
Other asian	0.01	0.01	0.00	0.00
African or black caribean	0.01	0.02	0.01	0.00
Other	0.00	0.00	-0.00	0.00
Highest qualification:				
Degree	0.14	0.19	0.05***	0.01
Other higher degree	0.12	0.14	0.02*	0.01
A-level	0.16	0.17	0.01	0.01
GCSE	0.17	0.19	0.02	0.01
Other	0.15	0.13	-0.01	0.01
No qualification	0.26	0.18	-0.08***	0.01
N	4916	1827		

Notes: Unweighted sample

Table 3: Pre-feedback (mean) gaps by knowledge status and conditional on observables

	(1)	(2)	(3)	(4)	(5)
	Constant	HT	-ve shock	(3)=(2)	N
	(non-HT)	and knew		p-value	
A. Annual measures					
Currently smokes	0.251***	-0.008	0.013	**0.041**	18508
	(15.67)	(-1.14)	(1.33)		
Monthly alcohol spending	61.369***	2.673	6.478**	0.070	18503
	(18.08)	(1.92)	(3.13)		
SF-12 Physical	50.429***	-2.066***	-0.543*	**0.000**	16123
	(144.28)	(-12.46)	(-2.54)		
SF-12 Mental	50.925***	-0.809***	-0.029	**0.004**	16123
	(127.23)	(-4.32)	(-0.12)		
B. Blood pressure					
Systolic	120.311***	10.944***	28.178***	**0.000**	18698
	(223.67)	(39.59)	(96.83)		
Diastolic	69.062***	5.912***	15.425***	**0.000**	18698
	(175.69)	(30.35)	(60.98)		
C. Other behaviours					
Fruit and vegetable consumption:					
Consumes less than WHO guidelines	0.737***	0.007	0.014	0.513	18491
	(47.27)	(0.80)	(1.31)		
# fruit and veg per day	3.521***	-0.021	-0.048	0.528	18491
	(60.07)	(-0.69)	(-1.20)		
Diet:					
Bad diet (indicator)	0.090***	-0.017***	0.014	**0.000**	18506
	(8.23)	(-3.67)	(1.95)		
Bad diet (intensity)	0.482***	-0.009	0.050**	**0.000**	18506
	(19.46)	(-0.85)	(3.13)		
Sedentary lifestyle:					
Sedentary (indicator)	0.060***	0.010	0.002	0.275	18498
	(6.20)	(1.87)	(0.24)		
Sedentary (intensity)	7.441***	0.445***	0.338***	0.135	18499
(0-10 scale 0=very active, 10=inactive)	(77.76)	(9.12)	(4.94)		

Notes: Panel A. lists variables recorded annually; panel B. those collected once at the nurse visit and panel C those recorded on a rotating basis. Estimates are from linear regression models that include a constant (column 1); a dummy for self-reporting HT or being HT according to the survey nurse (column 2); a dummy for being HT according to the survey nurse but self-reporting as not HT (column 3) and demographic controls. Controls are: age up to the third power; education dummies (4); continuous household income; whether children in the household; ethnicity dummies (6); dummies for long-standing health conditions (12) and region dummies (10). t-statistics in parenthesis.

Table 4: Hypertension diagnosis effects on health and health behaviours at six month follow-up

Outcome	Mean at-risk	OLS	RF	IV	First stage *F-stat*	*N*
Smoking	0.30 †	-0.00224	-0.00547**	-0.101**	124.3	57,053
		(0.00442)	(0.00245)	(0.0460)		
Alcohol Consumption (£per month)	50.68 ‡	1.452	1.159	21.08	143.1	67,397
		(2.062)	(1.283)	(23.40)		
SF-12 Physical	49 .22‡	-0.905***	-0.366**	-6.636**	128.4	61,373
		(0.313)	(0.161)	(2.953)		
SF-12 Mental	49.99‡	-0.0807	0.169	3.068	128.4	61,373
		(0.377)	(0.170)	(3.096)		
CVD	0.28$^{\lambda}$	0.0373***	0.00177	0.0305	116	55,154
		(0.00838)	(0.00253)	(0.0437)		
Congestive heart failure	0.02$^{\lambda}$	0.00448	-0.00213***	-0.0368***	116	55,154
		(0.00303)	(0.000567)	(0.0104)		
Coronary heart disease	0.09$^{\lambda}$	0.00911**	-0.00204*	-0.0352*	116	55,154
		(0.00431)	(0.00110)	(0.0193)		
Angina	0.12$^{\lambda}$	0.0148***	0.000144	0.00248	116	55,154
		(0.00538)	(0.00149)	(0.0257)		
Heart Attack	0.12$^{\lambda}$	0.00781*	0.000658	0.0113	116	55,154
		(0.00436)	(0.00172)	(0.0297)		
Stroke	0.07$^{\lambda}$	0.00825*	0.000938	0.0162	116	55,154
		(0.00435)	(0.00162)	(0.0279)		

Note: These estimates are based on models that use data from the UKHLS up to and including the wave after the nurse visit took place. All the models include individual fixed effects. The column labeled OLS presents the ordinary least squares estimates. The OLS models regress the outcome on the self-reported hypertension diagnosis. The column labeled "RF" presents reduced-form estimates of the outcome on the instrument (the nurse diagnosing hypertension interacted with a post-nurse visit dummy). All models include the following controls: age, age-squared, age-cubed, household size, and dummies indicating being married, being employed or self-employed, and having qualifications of A-level or about, plus a set of time dummies. Standard errors are clustered at the individual level. *, ** and *** respectively denote significance at the 10, 5 or 1 percent level.

† at-risk is those with no qualifications and not self-reporting hypertensive.

‡ at-risk is full population.

$^{\lambda}$ at-risk is males aged over 65.

Table 5: Hypertension diagnosis effects on health and health behaviours by gender

	Male			Female		
Outcome	Immediate	Pooled w/o trend	Pooled w trend	Immediate	Pooled w/o trend	Pooled w trend
Smoking	-0.155**	-0.187***	-0.108	-0.0198	-0.145*	-0.0602
	(0.0612)	(0.0725)	(0.0845)	(0.0715)	(0.0878)	(0.128)
N	24,593	38,795	31,885	32,460	51,339	41,738
First stage *F-stat*	78.62	79.06	71.99	45.75	48.98	32.56
Alcohol Consumption (£)	36.25	32.76	75.18	2.415	27.99	14.20
	(31.28)	(29.28)	(55.51)	(36.14)	(24.97)	(42.47)
N	29,138	59,558	48,672	38,259	78,360	63,493
First stage *F-stat*	85.07	88.98	90.92	57.20	60.66	61.51
SF-12 Physical	-6.848*	-6.034**	-3.802	-7.496	-4.651	-6.116
	(3.557)	(2.740)	(4.169)	(5.139)	(4.023)	(5.824)
N	26,599	55,510	45,508	34,774	73,194	59,608
First stage *F-stat*	77.20	86.68	84.17	50.85	58.77	59.70
SF-12 Mental	1.991	3.447	-0.0822	4.337	-1.958	6.419
	(3.623)	(2.785)	(4.282)	(5.494)	(4.085)	(6.431)
N	26,599	55,510	45,508	34,774	73,194	59,608
First stage *F-stat*	77.20	86.68	84.17	50.85	58.77	59.70
CVD	-0.00694	0.129*	-0.149*	0.0740	0.194**	-0.0633
	(0.0528)	(0.0738)	(0.0773)	(0.0761)	(0.0968)	(0.103)
N	23,753	47,557	36,652	31,401	63,436	48,464
Congestive heart failure	-0.0271**	-0.0211	-0.0347	-0.0544***	-0.0128	-0.0848**
	(0.0123)	(0.0231)	(0.0231)	(0.0188)	(0.0307)	(0.0370)
N	23,753	47,557	36,652	31,401	63,436	48,464
Coronary heart disease	-0.0463**	-0.0804**	-0.100**	-0.0252	0.0186	-0.0762
	(0.0227)	(0.0367)	(0.0438)	(0.0348)	(0.0497)	(0.0522)
N	23,753	47,557	36,652	31,401	63,436	48,464
Angina	-0.00536	-0.0174	-0.0160	0.00687	0.0577	0.00750
	(0.0342)	(0.0423)	(0.0430)	(0.0399)	(0.0511)	(0.0532)
N	23,753	47,557	36,652	31,401	63,436	48,464
Heart Attack	0.0646	0.0960*	0.0382	-0.0720**	-0.0607	-0.0604
	(0.0462)	(0.0548)	(0.0584)	(0.0346)	(0.0442)	(0.0471)
N	23,753	47,557	36,652	31,401	63,436	48,464
Stroke	-0.0287	0.0971**	-0.0995**	0.0820	0.110*	0.0291
	(0.0291)	(0.0461)	(0.0461)	(0.0556)	(0.0649)	(0.0653)
N	23,753	47,557	36,652	31,401	63,436	48,464
First stage *F-stat*	70.09	72.39	72.69	45.37	49.03	48.65

Note: All these estimates are from fixed-effect instrumental variable models. The column "Immediate" presents estimates based on models that use data from the UKHLS up to and including the wave after the nurse visit took place. The columns "Pooled w/o trend" and "Pooled w trend" both use data from the UKHLS up to and including five waves after the nurse visit took place. The trends include are for two groups i) mildly raised and ii) raised & considerably raised. All models include the following controls: age, age-squared, age-cubed, household size, and dummies indicating being married, being employed or self-employed, and having qualifications of A-level or about, plus a set of time dummies. Standard errors are clustered at the individual level. *, ** and *** respectively denote significance at the 10, 5 or 1 percent level. CVD, congestive heart failure, coronary heart disease, angina, heart attack and stroke are all based on the same sample and share the same first stage results reported under stroke.

Table 6: Hypertension diagnosis effects on health and health behaviours by qualification level

	A-level & above			Below A-level		
Outcome	Immediate	Pooled w/o trend	Pooled w trend	Immediate	Pooled w/o trend	Pooled w trend
Smoking	-0.105	-0.197**	-0.0815	-0.101	-0.146**	-0.0914
	(0.0682)	(0.0853)	(0.114)	(0.0622)	(0.0729)	(0.0888)
N	31,427	50,248	40,602	25,331	39,659	32,841
First stage *F-stat*	58.33	58.89	45.42	66.61	68.69	58.09
Alcohol Consumption (£)	7.831	64.34*	88.28	30.08	3.319	11.83
	(35.46)	(36.82)	(69.03)	(30.99)	(21.52)	(38.52)
N	36,093	76,409	61,629	30,912	61,203	50,290
First stage *F-stat*	64.72	68.28	67.73	79.54	80.56	83.09
SF-12 Physical	-12.35***	-6.587*	-13.27**	-2.081	-3.403	3.244
	(4.482)	(3.460)	(5.342)	(4.012)	(3.118)	(4.753)
N	33,810	72,981	59,062	27,132	55,389	45,783
First stage *F-stat*	61.50	64.88	64.03	67.55	79.64	77.90
SF-12 Mental	7.317*	6.573*	9.418*	-1.185	-3.445	-5.278
	(4.417)	(3.487)	(5.312)	(4.409)	(3.280)	(5.324)
N	33,810	72,981	59,062	27,132	55,389	45,783
First stage *F-stat*	61.50	64.88	64.03	67.55	79.64	77.90
CVD	0.0667	0.213**	-0.0583	0.00246	0.130	-0.165*
	(0.0611)	(0.0844)	(0.0882)	(0.0614)	(0.0831)	(0.0887)
N	29,611	61,751	46,904	25,185	48,971	38,006
Congestive heart failure	-0.00845	0.0194	-0.0371	-0.0597***	-0.0450*	-0.0707**
	(0.0135)	(0.0313)	(0.0268)	(0.0158)	(0.0234)	(0.0300)
N	29,611	61,751	46,904	25,185	48,971	38,006
Coronary heart disease	-0.0282	-0.0629*	-0.0949**	-0.0393	-0.0114	-0.0908*
	(0.0225)	(0.0342)	(0.0472)	(0.0297)	(0.0460)	(0.0484)
N	29,611	61,751	46,904	25,185	48,971	38,006
Angina	0.0363	0.0246	0.00830	-0.0249	0.0121	-0.0241
	(0.0363)	(0.0439)	(0.0480)	(0.0358)	(0.0473)	(0.0473)
N	29,611	61,751	46,904	25,185	48,971	38,006
Heart Attack	0.00880	0.0856	-0.0120	0.0138	-0.000801	0.0111
	(0.0380)	(0.0544)	(0.0604)	(0.0437)	(0.0495)	(0.0522)
N	29,611	61,751	46,904	25,185	48,971	38,006
Stroke	0.0451	0.115**	0.0123	-0.00782	0.0957*	-0.102*
	(0.0395)	(0.0517)	(0.0502)	(0.0390)	(0.0555)	(0.0571)
N	29,611	61,751	46,904	25,185	48,971	38,006
First stage *F-stat*	52.65	53.58	52.87	64.33	66.74	66.73

Note: for estimation details see note to table 5

Table 7: Hypertension diagnosis effects on health and health behaviours by age

	Aged 40 and above			Aged 50 and above			Aged 60 and above		
Outcome	Immediate	Pooled w/o trend	Pooled w trend	Immediate	Pooled w/o trend	Pooled w trend	Immediate	Pooled w/o trend	Pooled w trend
Smoking	-0.122**	-0.160***	-0.104	-0.157**	-0.177**	-0.130	-0.215**	-0.235**	-0.190
	(0.0485)	(0.0559)	(0.0739)	(0.0624)	(0.0704)	(0.0930)	(0.0875)	(0.0990)	(0.119)
N	41,423	67,228	55,442	29,779	49,049	40,627	19,133	32,056	26,837
First stage *F-stat*	111.8	113.2	87.77	72.45	72.26	59.24	40.82	39.63	36.36
Alcohol Consumption (£)	26.43	31.67	51.78	26.32	40.53	67.32	50.04	80.70**	119.2
	(24.44)	(21.10)	(40.79)	(29.74)	(27.18)	(54.05)	(39.01)	(38.63)	(76.32)
N	48,106	104,274	85,861	34,660	77,098	63,762	22,705	51,377	42,944
First stage *F-stat*	130.2	131.9	129.1	87.73	86.48	84.50	51.42	49.46	47.98
SF-12 Physical	-5.764*	-3.264	-3.623	-6.862*	-5.382	-5.041	-6.327	-3.020	-2.808
	(3.108)	(2.447)	(3.761)	(4.088)	(3.290)	(4.999)	(5.468)	(4.509)	(6.772)
N	43,358	96,816	80,029	30,873	70,968	58,946	19,744	46,419	38,993
First stage *F-stat*	114.8	127.5	121	75.97	82.21	77.73	45.46	47.71	44.73
SF-12 Mental	0.575	-0.571	-3.030	1.123	1.454	-2.293	5.258	1.769	-2.544
	(3.235)	(2.482)	(3.998)	(4.170)	(3.205)	(5.269)	(5.460)	(4.233)	(6.911)
N	43,358	96,816	80,029	30,873	70,968	58,946	19,744	46,419	38,993
First stage *F-stat*	114.8	127.5	121	75.97	82.21	77.73	45.46	47.71	44.73
CVD	0.0364	0.180***	-0.126*	0.0297	0.214**	-0.172*	0.0541	0.281**	-0.224
	(0.0478)	(0.0654)	(0.0697)	(0.0664)	(0.0920)	(0.0988)	(0.0990)	(0.141)	(0.151)
N	39,202	83,637	65,127	28,289	61,947	48,538	18,407	41,111	32,650
First stage *F-stat*	106.3	107.6	103.5	69.82	68.67	66.34	40.90	38.34	36.87
Congestive heart failure	-0.0400***	-0.0170	-0.0636***	-0.0575***	-0.0397	-0.0817**	-0.0923***	-0.0837**	-0.125**
	(0.0116)	(0.0210)	(0.0229)	(0.0176)	(0.0284)	(0.0332)	(0.0279)	(0.0427)	(0.0546)
N	39,202	83,637	65,127	28,289	61,947	48,538	18,407	41,111	32,650
Coronary heart disease	-0.0370*	-0.0365	-0.106***	-0.0462	-0.0400	-0.136**	-0.0847*	-0.109	-0.196**
	(0.0214)	(0.0328)	(0.0382)	(0.0315)	(0.0482)	(0.0561)	(0.0462)	(0.0728)	(0.0864)
N	39,202	83,637	65,127	28,289	61,947	48,538	18,407	41,111	32,650
Angina	0.00265	0.0113	-0.0154	0.0122	0.0154	-0.00120	0.0384	0.0403	-0.00596
	(0.0280)	(0.0352)	(0.0377)	(0.0406)	(0.0504)	(0.0540)	(0.0627)	(0.0789)	(0.0832)
N	39,202	83,637	65,127	28,289	61,947	48,538	18,407	41,111	32,650
Heart Attack	0.0135	0.0431	-0.00223	0.0127	0.0479	-0.00989	-0.0139	0.0477	-0.0267
	(0.0324)	(0.0403)	(0.0441)	(0.0448)	(0.0571)	(0.0617)	(0.0651)	(0.0872)	(0.0939)
N	39,202	83,637	65,127	28,289	61,947	48,538	18,407	41,111	32,650
Stroke	0.0211	0.117***	-0.0442	0.0128	0.147**	-0.0713	-0.000181	0.185**	-0.148
	(0.0305)	(0.0425)	(0.0419)	(0.0416)	(0.0601)	(0.0587)	(0.0601)	(0.0905)	(0.0923)
N	39,202	83,637	65,127	28,289	61,947	48,538	18,407	41,111	32,650
First stage *F-stat*	106.3	107.6	103.5	69.82	68.67	66.34	40.90	38.34	36.87

Note: for estimation details see note to table 5

References

Barrett-Connor, Elizabeth, John Z. Ayanian et al., *A Nationwide Framework for Surveillance of Cardiovascular and Chronic Lung Diseases*, Institute of Medicine of the National Academies, 2011.

Benjamin, Emelia J., Michael J. Blaha, Stephanie E. Chiuve, Mary Cushman, Sandeep R. Das, Rajat Deo, Sarah D. de Ferranti, James Floyd, Myriam Fornage, Cathleen Gillespie, Carmen R. Isasi, Monik C. Jiménez, Lori Chaffin Jordan, Suzanne E. Judd, Daniel Lackland, Judith H. Lichtman, Lynda Lisabeth, Simin Liu, Chris T. Longenecker, Rachel H. Mackey, Kunihiro Matsushita, Dariush Mozaffarian, Michael E. Mussolino, Khurram Nasir, Robert W. Neumar, Latha Palaniappan, Dilip K. Pandey, Ravi R. Thiagarajan, Mathew J. Reeves, Matthew Ritchey, Carlos J. Rodriguez, Gregory A. Roth, Wayne D. Rosamond, Comilla Sasson, Amytis Towfighi, Connie W. Tsao, Melanie B. Turner, Salim S. Virani, Jenifer H. Voeks, Joshua Z. Willey, John T. Wilkins, Jason HY. Wu, Heather M. Alger, Sally S. Wong, and Paul Muntner, "Heart Disease and Stroke Statistics 2017 Update: A Report From the American Heart Association," *Circulation*, 2017, *135* (10), e146–e603.

Benjamin, Emelia J, Salim S Virani, Clifton W Callaway, Alanna M Chamberlain, Alexander R Chang, Susan Cheng, Stephanie E Chiuve, Mary Cushman, Francesca N Delling, Rajat Deo et al., "Heart disease and stroke statistics-2018 update: a report from the American Heart Association.," *Circulation*, 2018, *137* (12), e67.

Blanchflower, David G. and Andrew J. Oswald, "Hypertension and happiness across nations," *Journal of Health Economics*, March 2008, *27* (2), 218–233.

Bollinger, Bryan, Phillip Leslie, and Alan Sorensen, "Calorie posting in chain restaurants," *American Economic Journal: Economic Policy*, 2011, *3* (1), 91–128.

Carrera, Mariana, Syeda A Hasan, and Silvia Prina, "The Effects of Health Risk Assessments on Cafeteria Purchases: Do New Information and Health Training Matter?," *Journal of Health Economics*, Forthcoming.

Courtenay, Will H, Donald R McCreary, and Joseph R Merighi, "Gender and ethnic differences in health beliefs and behaviors," *Journal of health psychology*, 2002, *7* (3), 219–231.

Crossley, Thomas F and Steven Kennedy, "The reliability of self-assessed health status," *Journal of health economics*, 2002, *21* (4), 643–658.

Delavande, Adeline and Hans-Peter Kohler, "The impact of HIV testing on subjective expectations and risky behavior in Malawi," *Demography*, 2012, *49* (3), 1011–1036.

Dustan, HP, "Gender differences in hypertension.," *Journal of human hypertension*, 1996, *10* (6), 337–340.

Edwards, Ryan, "If my blood pressure is high do I take it to heart? Behavioral impacts of biomarker collection in the health and retirement survey," *Unpublished Manuscript*, 2016.

Ettehad, Dena, Connor A Emdin, Amit Kiran, Simon G Anderson, Thomas Callender, Jonathan Emberson, John Chalmers, Anthony Rodgers, and Kazem Rahimi, "Blood pressure lowering for prevention of cardiovascular disease and death: a systematic review and meta-analysis," *The Lancet*, 2016, *387* (10022), 957–967.

Fat, Linda Ng, "Health Survey for England 2017 Adult health," Technical Report, NHS Digital 2018.

Griffith, Rachel, Martin O'Connell, and Kate Smith, "The importance of product reformulation versus consumer choice in improving diet quality," *Economica*, 2017, *84* (333), 34–53.

Grossman, Michael, "On the concept of health capital and the demand for health," *Journal of Political economy*, 1972, *80* (2), 223–255.

Hoorens, Vera, "Self-enhancement and superiority biases in social comparison," *European review of social psychology*, 1993, *4* (1), 113–139.

Kawano, Yuhei, "Diurnal blood pressure variation and related behavioral factors," *Hypertension research*, 2011, *34* (3), 281.

Kuneš, Jaroslav, Johanne Tremblay, François Bellavance, and Pavel Hamet, "Influence of environmental temperature on the blood pressure of hypertensive patients in Montreal," *American journal of hypertension*, 1991, *4* (5_Pt_1), 422–426.

Maryon-Davis, A, "Hypertension–'The Silent Killer'. Faculty of Public Health of the Royal Collegue of Physicians United Kingdom," *Briefing Statement*, 2005, pp. P1–6.

Motta, Matthew, Timothy Callaghan, and Steven Sylvester, "Knowing less but presuming more: Dunning-Kruger effects and the endorsement of anti-vaccine policy attitudes," *Social Science Medicine*, 2018, *211*, 274 – 281.

Organization, World Health, "Global health risks: mortality and burden of disease attributable to selected major risks. 2009," Technical Report 2013.

_ , "Global health observatory data repository," Technical Report 2017.

Oster, Emily, "Diabetes and Diet: Behavioral Response and the Value of Health," *American Economic Journal: Applied Economics*, 2018, *10* (4), 308–48.

_ , "Health Recommendations and Selection in Health Behaviors," *American Economic Review: Insights*, 2019.

_ **, Ira Shoulson, and E Dorsey**, "Limited life expectancy, human capital and health investments," *The American Economic Review*, 2013, *103* (5), 1977–2002.

_ **,** _ **, and** _ , "Optimal expectations and limited medical testing: evidence from Huntington disease," *The American Economic Review*, 2013, *103* (2), 804–830.

_ , _ , **Kimberly Quaid, and E Ray Dorsey**, "Genetic adverse selection: Evidence from long-term care insurance and Huntington disease," *Journal of Public Economics*, 2010, *94* (11), 1041–1050.

Papadopoulos, Dimitris P., Iordanis Mourouzis, Costas Thomopoulos, Thomas Makris, and Vasilios Papademetriou, "Hypertension crisis," *Blood Pressure*, 2010, *19* (6), 328–336.

Pereira, Marta, Nuno Lunet, Ana Azevedo, and Henrique Barros, "Differences in prevalence, awareness, treatment and control of hypertension between developing and developed countries," *Journal of hypertension*, 2009, *27* (5), 963–975.

Public Accounts Committee, "Access to General Practice inquiry," 2017.

Sanderson, Saskia C, Jo Waller, Martin J Jarvis, Steve E Humphries, and Jane Wardle, "Awareness of lifestyle risk factors for cancer and heart disease among adults in the UK," *Patient education and counseling*, 2009, *74* (2), 221–227.

Slark, J, MS Khan, P Bentley, and P Sharma, "Knowledge of blood pressure in a UK general public population," *Journal of human hypertension*, 2014, *28* (8), 500.

SPRINT Research Group, "A randomized trial of intensive versus standard blood-pressure control," *New England Journal of Medicine*, 2015, *373* (22), 2103–2116.

Thornton, Rebecca L, "The demand for, and impact of, learning HIV status," *The American economic review*, 2008, *98* (5), 1829–1863.

Viscusi, W and Jahn K Hakes, "Risk beliefs and smoking behavior," *Economic Inquiry*, 2008, *46* (1), 45–59.

Whelton, Paul K., Robert M. Carey, Wilbert S. Aronow, Donald E. Casey, Karen J. Collins, Cheryl Dennison Himmelfarb, Sondra M. DePalma, Samuel Gidding, Kenneth A. Jamerson, Daniel W. Jones, Eric J. MacLaughlin, Paul Muntner, Bruce Ovbiagele, Sidney C. Smith, Crystal C. Spencer, Randall S. Stafford, Sandra J. Taler, Randal J. Thomas,

Kim A. Williams, Jeff D. Williamson, and Jackson T. Wright, "2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines," *Hypertension*, 2018, *71* (6), e13–e115.

Wisdom, Jessica, Julie S Downs, and George Loewenstein, "Promoting healthy choices: Information versus convenience," *American Economic Journal: Applied Economics*, 2010, *2* (2), 164–178.

Woodhouse, Peter R, Kay-Tee Khaw, and Martyn Plummer, "Seasonal variation of blood pressure and its relationship to ambient temperature in an elderly population.," *Journal of hypertension*, 1993, *11* (11), 1267–1274.

Zhao, Meng, Yoshifumi Konishi, and Paul Glewwe, "Does information on health status lead to a healthier lifestyle? Evidence from China on the effect of hypertension diagnosis on food consumption," *Journal of Health Economics*, 2013, *32* (2), 367–385.

Ziebarth, Nicolas R., "Lung cancer risk perception biases," *Preventive Medicine*, 2018, *110*, 16 – 23.

The Impact of a Personalised Blood Pressure Warning on Health Behaviours

Sonia Bhalotra[1,2], Adeline Delavande[1], Paul Fisher[1] and Jonathan James[3]

[1] Institute for Social and Economic Research, University of Essex

[2] Department of Economics, University of Essex

[3] Department of Economics, University of Bath

December 2019

Supplementary Material

Appendix A: Survey vs. clinical measures of blood pressure

BP would ideally be measured under controlled conditions to remove the influence of environmental factors that affect it and that make comparisons across different individuals in the population more complicated. Environmental factors that affect BP are: the time of day, room temperature, behaviours around the time of measurement, presence of a medical practitioner (white coat effects) and random fluctuations. If the extent of the factors varies between the survey interview and the clinic then, for the same sample of individuals, measured BP will differ across the settings.

While both are objective measures of BP, compared to a clinical measure, the one from the survey could on average either over–estimate or under–estimate BP. Whether the issue is of an important magnitude, say when estimating the presence of undiagnosed HT in the population, is an empirical question. In principle, estimated HT prevalence could be sensitive to small deviations in BP related to the environmental setting of the measurement. This occurs because HT is a discontinuous function of BP and therefore HT status may be sensitive to small perturbations in BP. While the environmental factors affecting BP have been documented in the medical literature, the impact of them on estimates of HT prevalence is typically not discussed in the social science literature.

More importantly for the present paper, the measurement issues could result in misclassification (eg. being classified as high in the survey when BP would be normal under the conditions in a clinic) even if on average measured BP was the same in both settings. This is very relevant for the interpretation of the results in the paper (and in particular the interpretation of our first stage IV estimates), as it relates to the share of respondents who should not show a response to a 'high' survey feedback in formal diagnoses of HT. In other words, a proportion measured high in the survey HT will turn out not to be HT when measured by their GP.

Contrasting to the systematic determinants of BP is random noise that occurs from reading–to–reading. The same individual measured twice under control conditions, is

likely to receive two differing measurements. For this reason, usual practice is to average over multiple measurements at a given visit, as is done in both the household survey and in the clinic. Still, averaging over a small number of readings will reduce, but not remove noise. As above, random noise becomes very important for marginal cases near to the HT cutt-off ie. a person measured as 139 and normal could 'become' high under a different reading if just a small pertubation raised their BP to 140. While random noise is not expected to lead to differences in average BP as measured in the survey interview and clinic, it could lead to misclassification (say where a respondent was measured as 140 and high in the survey but then found to be 139 and so normal by their GP).

This appendix describes the environmental factors that lead to systematic differences between BP measured in the survey interview and the clinic. It presents results from adjusting the survey measures in an attempt to replicate measurements that would be obtained in the clinical setting such as a GP's office. The estimates give us an indication of the extent to which survey respondents who were promopted to visit their GP would turn out to be HT according to their GP. We find that on average, the survey slightly over-estimates BP levels compared to ones taken in a GP's office, but it more seriously overestimates the share of people classified as being high. Applying our adjustments to the raw survey measures indicates that 16.7 percent of participants have high BP compared to 17.7 percent in the unadjusted data. Our adjustments imply misclassification where 8.4 percent of the survey high have normal BP in the adjusted data. We further simulate additional BP readings by introducing small amounts of random noise to the survey measures and find that this further increases misclassification.

Below we discuss each of the environmental factors that affects BP and discuss the effects of our adjustments for each of them:

i) Diurnal variation: BP varies over the day or diurnally. BP is usually highest in the early morning and evening, and falls in the early afternoon with the lowest readings being during the night (Kawano (2011)). Survey nurse visits occur across the day but there is a spike in visits mid-morning and early evening (figure A1). The latter spike is concurrent with the known evening peak in BP. Approximately 50 percent of nurse

visits occurred after 15:00 and 25 percent after 18:00 with 99 percent of visits starting before 8pm (figure A2). As the use of GP services is expected to be spread uniformally throughout the working day[14], measured BP is expected to systematic differ between the GP and survey setting.

To explore the presence of time of day effects in our data, we estimated a linear regression model for nurse measured blood pressure where we control for time of day up to the fourth power; room temperature and its square; whether eaten/smoked/drunk alcohol/done exercise in the 30 minutes prior to the readings (dummy variable) and demographic controls. Figure A3 plots the predicted BP (for a female with a-levels, aged 65, retired and room temperature of 18C). The figure fails to reproduce an early morning spike in BP, as we have very few early morning observations in our data, however, it does recreate the other expected patterns. Predicted BP rises steadily from the late afternoon but with a plateau through the late morning/early afternoon period. Predicted systolic BP ranges from a minimum of 125.7 at 08:04 and a maximum of 131.0 at 19.23.

Using the regression estimates we adjust each reading to correspond to one taken at 12:00 noon (say in the middle of GP office hours). While the adjustments lead to only small differences in the mean levels, they lead to larger differences in the shares high and in misclassification. The mean adjusted systolic (diastolic) is 122.5 (70.9) compared to 123.6 (71.4) in the unadjusted data. Row two of table A1 shows that 15.8 are high according to adjusted BP but only 17.7 percent are high in the raw measures. Misclassification is non-trivial with 10.7 percent of individuals originally classified as high being normal once we adjust for time of day. Conversely, only 0.03 percent of individuals originally classified as normal are high once we adjust for time of day.

ii) Temperature variation: Blood pressure falls with temperature (Kuneš et al. (1991)). Temperature differences between a respondents home and the medical setting are therefore a source of difference between clinical and survey measures of BP.[15] In princi-

[14]In the UK, the most commonly used contract is the General Medical Services contract, which stipulates that 'core hours' are 8 am to 6.30 pm and 18% closed at or before 3 pm on at least one weekday (Public Accounts Committee (2017))

[15]Related to the above is the seasonal pattern in BP (Woodhouse et al. (1993))

ple, the survey measures could over estimate (cold home setting) or underestimate (warm home setting) relative to a GP practice.

Figure A4 confirms BP falls with room temperature in our data, by plotting predicted BP against temperature from our regression models above. The model predicts that reducing room temperature from 18 degrees to 17 degrees would increase systolic (diastolic) BP by 0.46 (0.39) points.

We use our regression coefficients to adjust the survey measures to a common temperature (say the temperature in a GP's practise). Our estimates imply a temperature threshold of 19.7 degrees below which the unadjusted high prevalence is below the adjusted high prevalence and above which the pattern reverses. We first estimate a lower bound for the maximum underestimation of the survey by assuming a lower bound for the room temperature expected in the clinical setting. We set the lower bound to 18 degrees– UK medical guidelines for minimum home room temperature.[16] Row three of table A1 shows the results from applying the adjustment. 19 percent of respondents are high once the adjustment is applied.[17] Our upper-bound for the average underestimation of the survey is therefore 1.3 percentage points. Moreover, 1.5 percent of those normal in the unadjusted data turn out to be high when we adjust to 18 degrees, whereas 1.9 percent of those normal in the unadjusted measures are high once the adjustment is applied.

A more plausible room temperature for the clinical setting is 20 degrees – the minimum recommended by the WHO for the sick, disabled, very old or very young. Row four of the table shows that when we adjust to 20 degrees 17.5 percent are high, very similar to the unadjusted value. However, misclassification remains fairly common with 4.88 percent of the unadjusted high being normal.

iii) Behaviours before the measurement: Behaviours in the period directly preceding a measurement can also influence the reading. These include: eating; smoking; drinking alcohol or vigorous exercise. All of these raise blood pressure for a temporary

[16]NHS choices website (`www.nhs.uk/Livewell/winterhealth/Pages/KeepWarmKeepWell.aspx`). Accessed 24/5/17.

[17]13.1 percent of the sample have a room temperature that is measured by the nurse as below 18 degrees. It is expected that their BP would be overestimated relative to the GP setting. Respondents with room temperature measured above 18 degrees would be overestimated, relative to the GP setting.

period. 16.7 percent of our sample mentioned that they had participated in one of the behaviours in the 30 minutes prior to the survey measurement. We find that as a matter of practise, these effects turnout to be rather unimportant. Our regression estimates show the effect of doing one of these is to increase systolic (diastolic) BP by 0.26 (0.68).

It is hard to quantify how common the behaviours are before a GP visit but it seems plausible that they are more common in the survey setting where respondents were not instructed to avoid these activities and therefore the survey would tend to overestimate BP relative to a GP. To estimate an upper bound for the difference between the survey measures and clinical measures, we assume none who did the activity in the survey would do it in the GP setting and adjust the survey measures using the coefficients from our regression model. The share calculated as high is unchanged when the adjustments are implemented and there is no misclassification (row five of table A1).

iv) White coat hypertension: It is usually thought that BP is higher when measured in a medical compared to home setting due to nervousness in the presence of medical experts that may raise BP (white coat effect). (CITE) estimate the size of the effect. In the case of our survey measure of BP, measurement occurs in the home but in the presence of a health care specialist (survey nurse). Therefore, it is not clear that BP should be any lower in the survey setting. Quantifying white coat effects is an interesting challenge for future research but it is outside the scope of our research design.

v) Combined effect of all adjustments: Table A1 shows the results from applying all of the adjustments.[18] We estimate 16.7 percent are high which is one percentage point below the unadjusted figure. 91.6 percent of those measured high by the survey nurse are high according to the adjusted measure. Conversely, 0.6 percent of the normal according to the survey nurse are high on the adjusted measure.

vi) Random fluctuations in BP: Aside from the systematic factors affecting BP, there is random variation meaning it varies from reading-to-reading within an individual. For this reason, usual practise is to average blood pressure over multiple readings at a

[18]We adjust to the WHO minimum room temperature for the elderly.

given visit, as is done in the survey.[19] Still, it is possible that a survey respondent gets a different BP reading from their GP than in the survey interview, purely for the reason that there is noise in any single or group of BP measures.

We simulate for each individual, two additional BP readings and apply the feedback rules of the survey nurse. The simulated feedback can be thought of as what a respondent might get at a GP's office hours, ceteris paribus. The simulations proceed as follows: For each individual, we estimate 'true' BP (separately for systolic and diastolic) by taking the mean of the three BP scores observed by the survey nurse. We then draw error terms and add these to the 'true' value to give us simulated BP readings measured with error. The simulated error terms are drawn from a normal distribution with mean and variance set equal to that in the observed data (that is the distribution of true' BP minus the observed BP). We allow the simulated error terms to be drawn from distributions that are reading specific to allow for the fact that BP systematically falls across readings.

Table A2 shows the results of the simulation. The overall share of those measured high is hardly changed (17.9 percent) but misclassification is prevalent. Of those classified as high by the survey nurse, only 87.1 percent are high according to the simulated new measures. Conversely, 3.0 percent of those normal according to the survey nurse are high according to the simulated BP readings.

We also examined the combined effect of our simulations for the non-systematic variations in BP (noise) and the adjustments for the systematic variations (ie. time, temperature, other behaviours). Overall, we find that the survey overestimates the share high compared to the clinical setting and that misclassification is large. When applying both the simulations and adjustments we estimate 17.2 percent are high, 84.6 percent of the survey high are high and 2.8 percent of the survey normal are high.

[19]For each individual in our data, we observe 3 BP readings that were taken at one minute intervals. For example, 24.1 percent of those with mildly raised BP according to reading 2, are normotensive according to reading 3; and 1.4 percent of those raised on reading number 2 are in fact not hypertensive according to the third reading. Conversely, 4.6 of those normotensive on the second reading where actually hypertensive according to the third.

Figure A1: Time of nurse visit

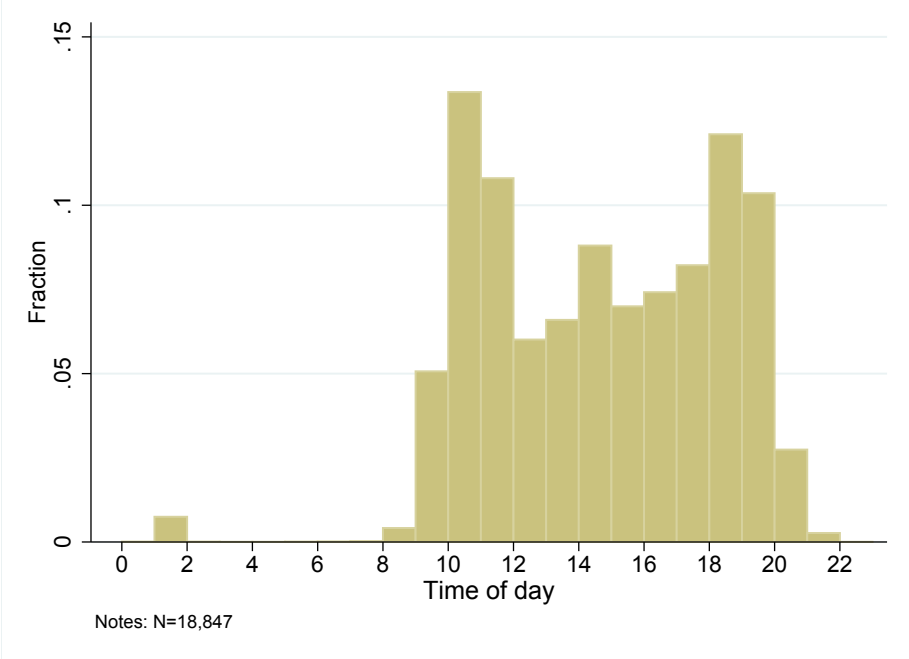


Figure A2: CDF of time of nurse visit

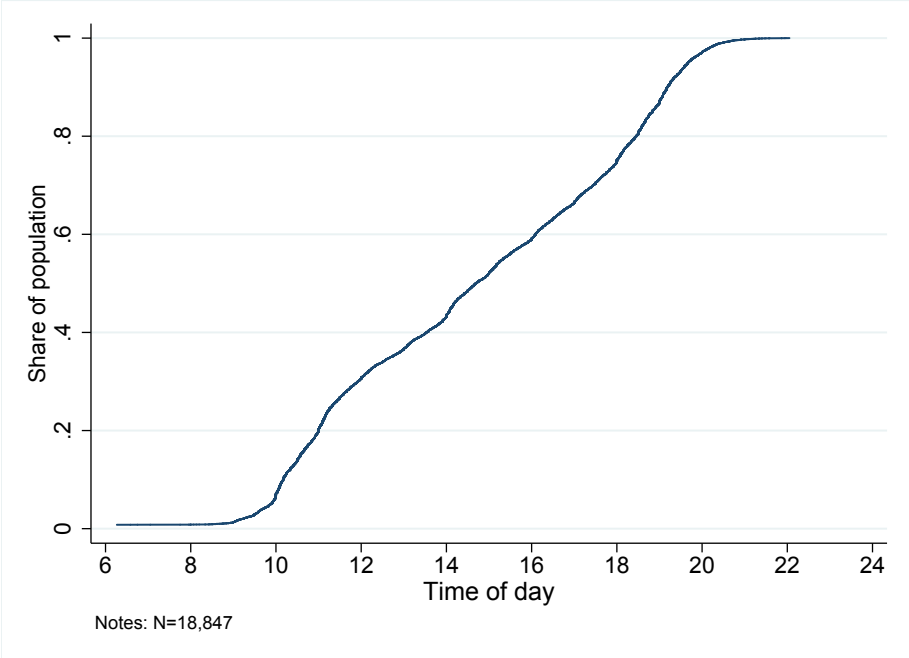
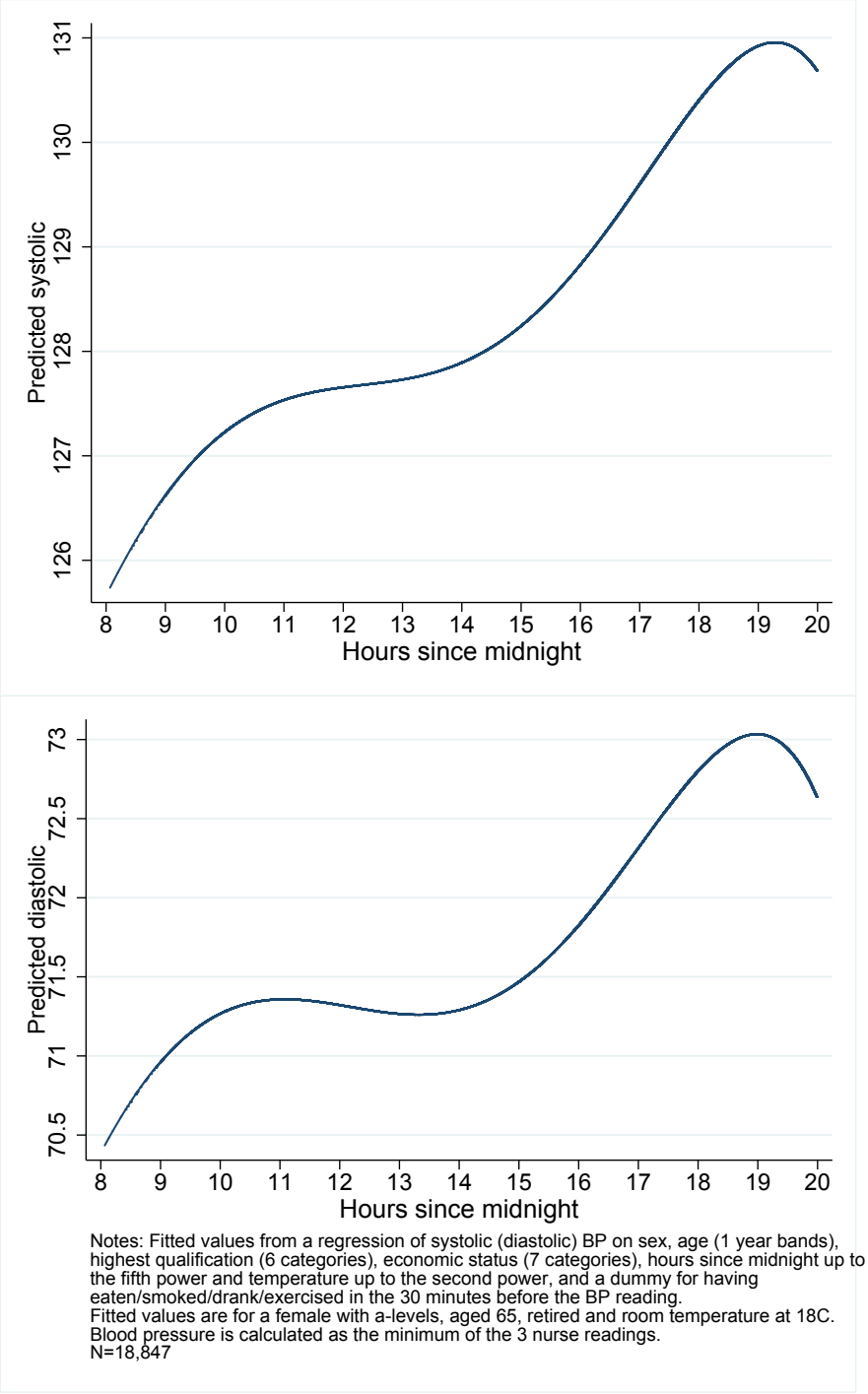
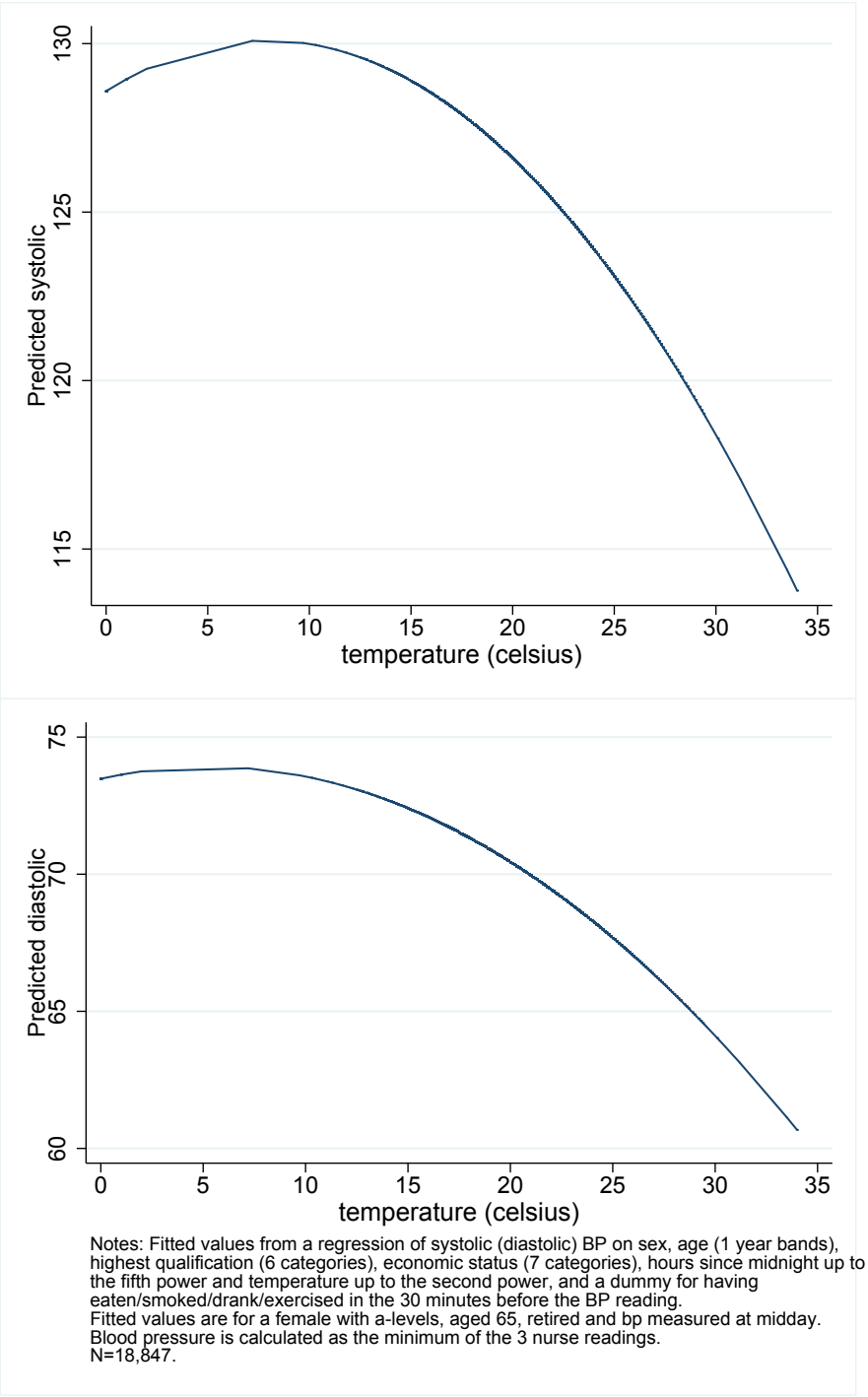


Figure A3: Predicted BP by time of day



Notes: Fitted values from a regression of systolic (diastolic) BP on sex, age (1 year bands), highest qualification (6 categories), economic status (7 categories), hours since midnight up to the fifth power and temperature up to the second power, and a dummy for having eaten/smoked/drank/exercised in the 30 minutes before the BP reading.
Fitted values are for a female with a-levels, aged 65, retired and room temperature at 18C.
Blood pressure is calculated as the minimum of the 3 nurse readings.
N=18,847

Figure A4: Predicted BP by room temperature



Notes: Fitted values from a regression of systolic (diastolic) BP on sex, age (1 year bands), highest qualification (6 categories), economic status (7 categories), hours since midnight up to the fifth power and temperature up to the second power, and a dummy for having eaten/smoked/drank/exercised in the 30 minutes before the BP reading.
Fitted values are for a female with a-levels, aged 65, retired and bp measured at midday.
Blood pressure is calculated as the minimum of the 3 nurse readings.
N=18,847.

Table A1: Share HT under different BP adjustments

Adjustment type	Full population	Hypertensive^U	Normotensive^U
None	17.7	100	0
Time of day (12 noon)	15.8	89.3	0
Temperature (18 degrees)	19	98.5	1.9
Temperature (20 degrees)	17.5	95.1	0.8
Eat/drank/smoked/exercised	17.7	100	0
All adjustments*	16.7	91.6	0.6

Notes: U superscript indicates the unadjusted data. *Temperature is adjusted to 30 degrees.

Table A2: Share HT when introducing random noise to BP

Adjustment type	Full population	Hypertensive^U	Normotensive^U
Unadjusted	17.7	100	0
Simulated + unadjusted	17.9	87.1	0.03
Simulated + adjusted	17.2	84.6	2.8

Notes: U superscript indicates the unadjusted data.
*adjustments are for the items in table X.

Appendix B: Additional figures and tables

Figure B1: Months between second interview and nurse visit

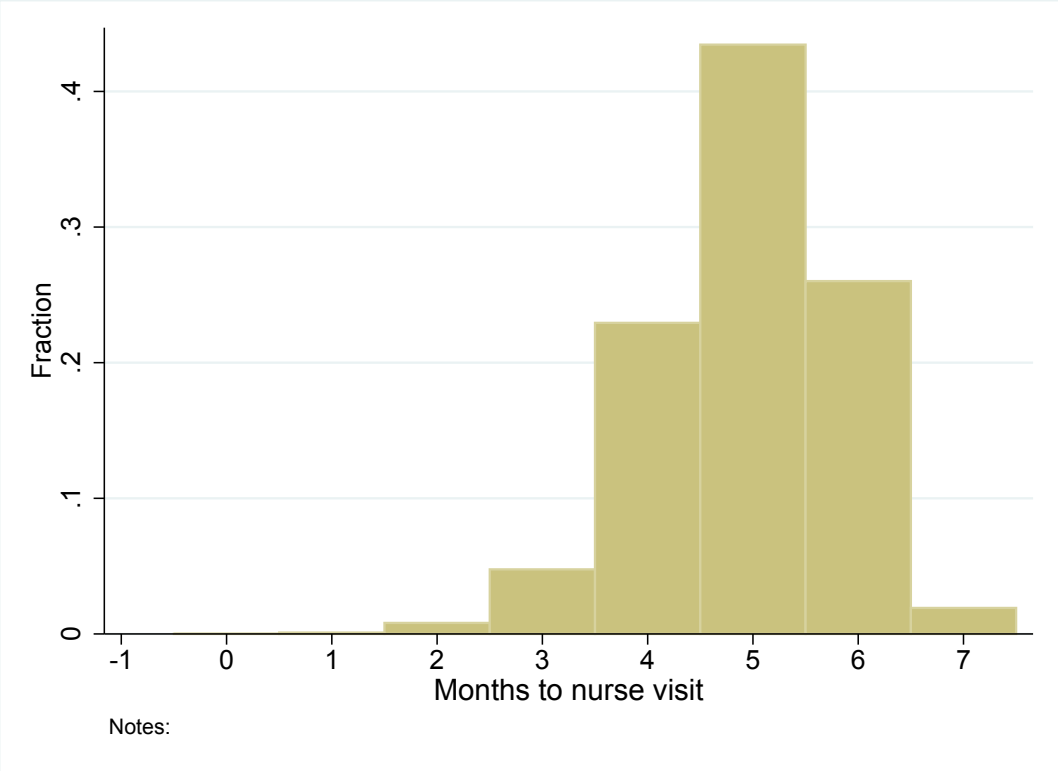
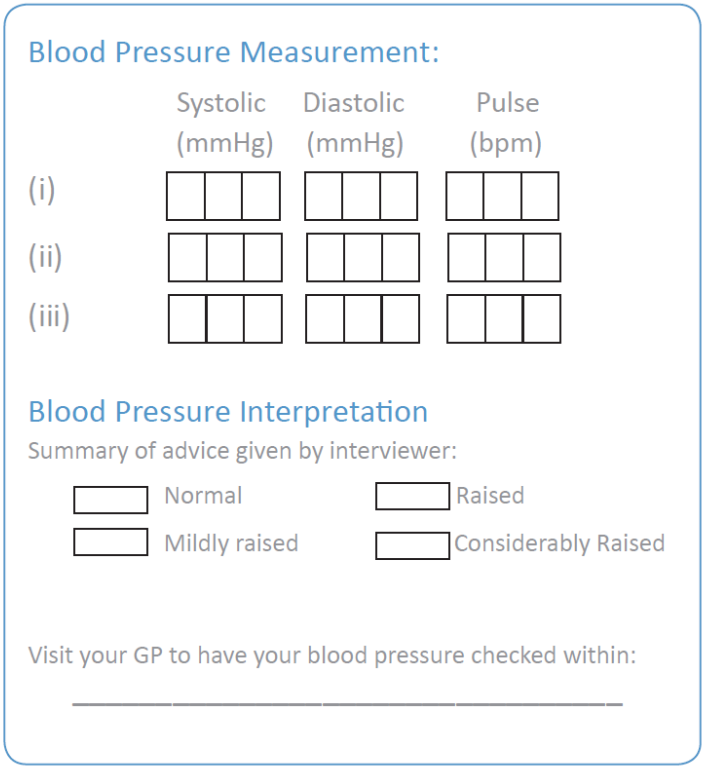


Figure B2: Sample measurement record card and oral feedback



Normal: 'Your blood pressure is normal.'

Mildly raised: 'Your blood pressure is a bit high today. Blood pressure can vary from day to day and throughout the day so that one high reading does not necessarily mean that you suffer from high blood pressure. 'You are advised to visit your GP within 2 months to have a further blood pressure reading to see whether this is a one-off finding or not.'

Raised: 'Your blood pressure is a bit high today. Blood pressure can vary from day to day and throughout the day so that one high reading does not necessarily mean that you suffer from high blood pressure. You are advised to visit your GP within 2 weeks to have a further blood pressure reading to see whether this is a one-off finding or not.'

Considerably raised: 'Your blood pressure is high today. Blood pressure can vary from day to day and throughout the day so that one high reading does not necessarily mean that you suffer from high blood pressure. You are strongly advised to visit your GP within 5 days to have a further blood pressure reading to see whether this is a one-off finding or not.'

Table B1: Summary Statistics by Consent Status

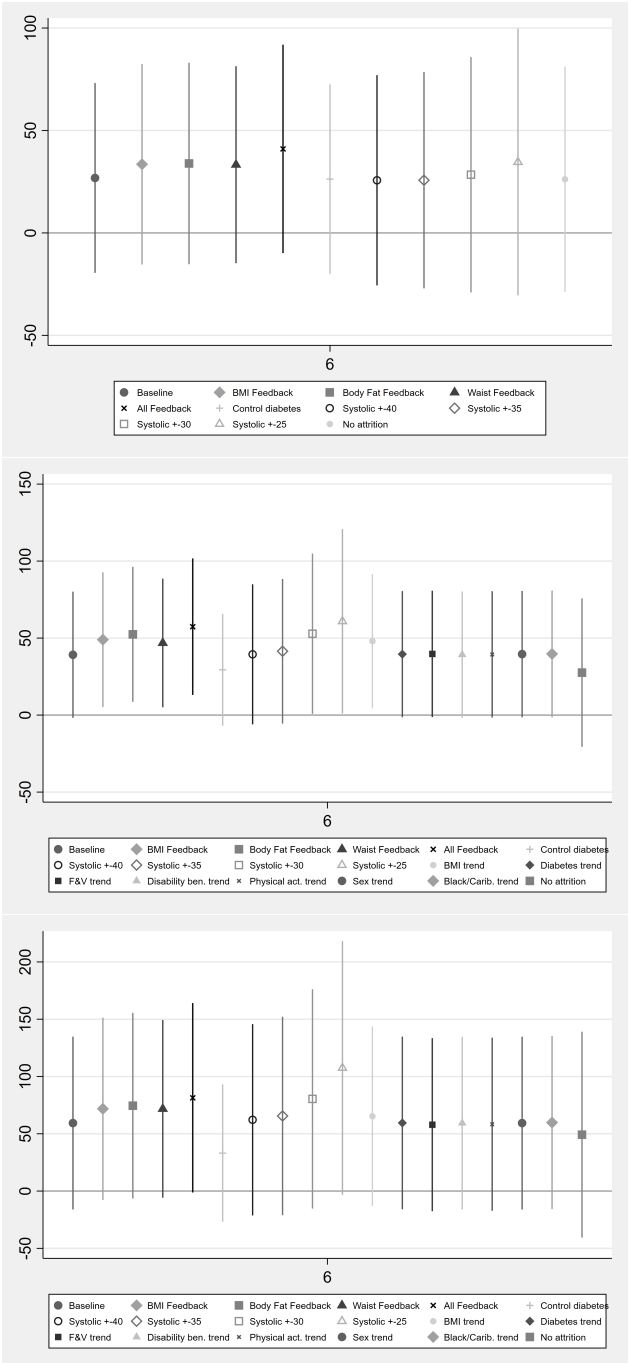
	Non-consenters	Consenters	Mean Diff	SE
Male	0.4455	0.4410	-0.0045	0.0053
Age	46.0141	50.2850	4.2709***	0.1959
Highest qualification:				
Degree	0.1933	0.2164	0.0231***	0.0044
Other higher degree	0.1048	0.1239	0.0191***	0.0034
A-level	0.2175	0.1948	-0.0227***	0.0043
GCSE	0.2245	0.2094	-0.0150***	0.0044
Other	0.1014	0.1107	0.0093**	0.0033
No qualification	0.1585	0.1448	-0.0137***	0.0039
General health:				
Excellent	0.1653	0.1535	-0.0118**	0.0039
Very good	0.3379	0.3435	0.0056	0.0051
Good	0.2843	0.2820	-0.0023	0.0048
Fair	0.1476	0.1563	0.0087*	0.0039
Poor	0.0649	0.0647	-0.0002	0.0026
Hypertension (self-reported)	0.2415	0.2608	0.0193***	0.0049
Economic status:				
Self-employed	0.0724	0.0717	-0.0007	0.0028
Employed	0.4812	0.4634	-0.0177***	0.0054
Unemployed	0.0571	0.0422	-0.0148***	0.0023
Retired	0.2136	0.2748	0.0612***	0.0046
Student	0.0717	0.0473	-0.0244***	0.0025
Long-term sick or disabled	0.0343	0.0362	0.0019	0.0020
Other	0.0697	0.0643	-0.0054*	0.0027
Gross household income (monthly)	3571.57	3540.06	-31.51	50.019
N	15,267	20,234		

Note: Non-consenters consist of those refusing a nurse visit (N=14,817) or consenting but not providing 3 valid blood pressure readings (N=450).

Table B2: Feedback thresholds based on the minimum of the second two nurse readings

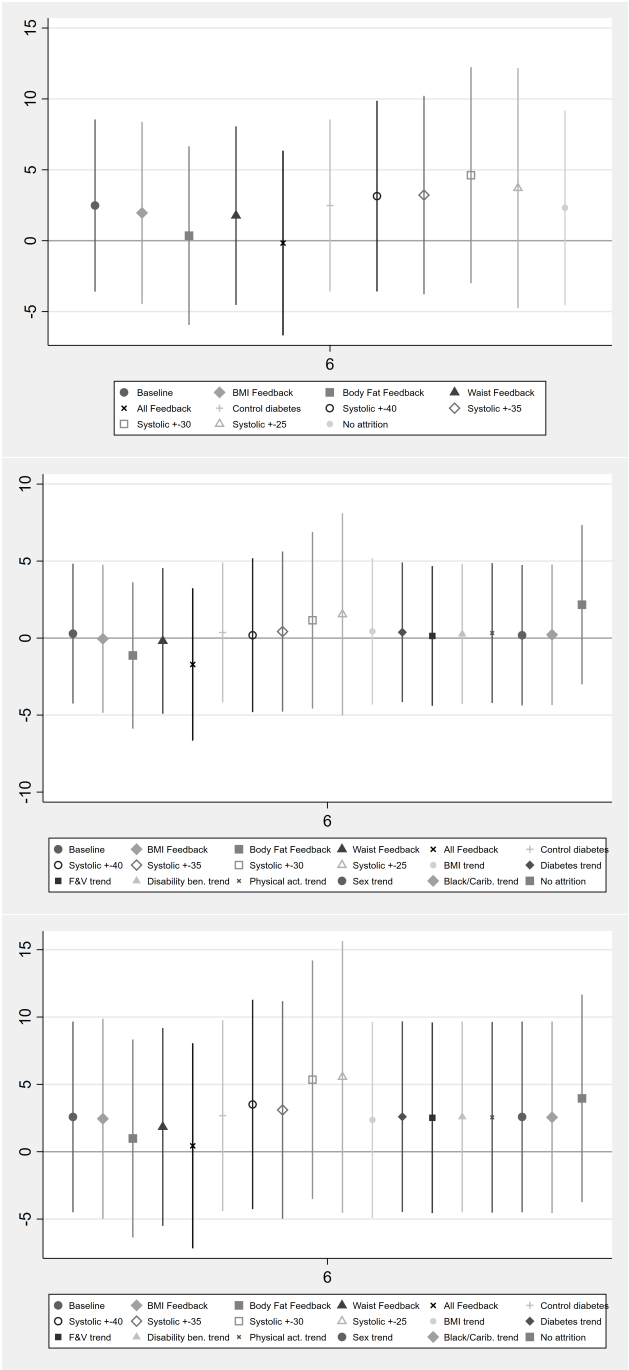
Classification	Systolic		Diastolic
Normal	≤ 140	and	≤ 90
Mildly raised	140-159	or	90-99
Raised	160-179	or	110-114
Considerably raised	$180\geq$	or	$115\geq$

Figure B3: The impact of hypertension diagnosis on alcohol spending - alternative specifications



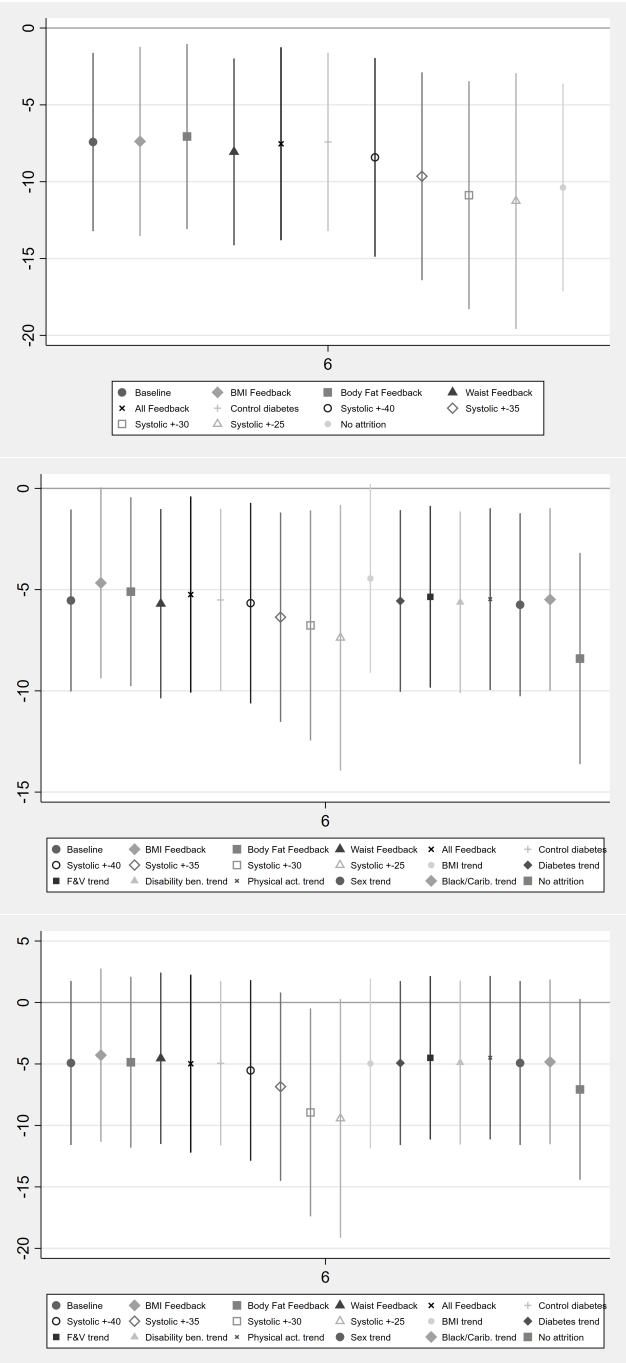
Note: Each dot is from a separate regression (FE-IV) set out in equation (2). The whiskers represent the 95% confidence interval calculated from standard errors clustered at the individual level. The top panel shows the immediate impact, the middle panel shows the long run impact not including trends, and the bottom panel is also for the long run and does include separate trends for those who have i. mildly raised hypertension and ii. raised and considerably raised hypertension.

Figure B4: The impact of hypertension diagnosis on SF-12 Physical - alternative specifications



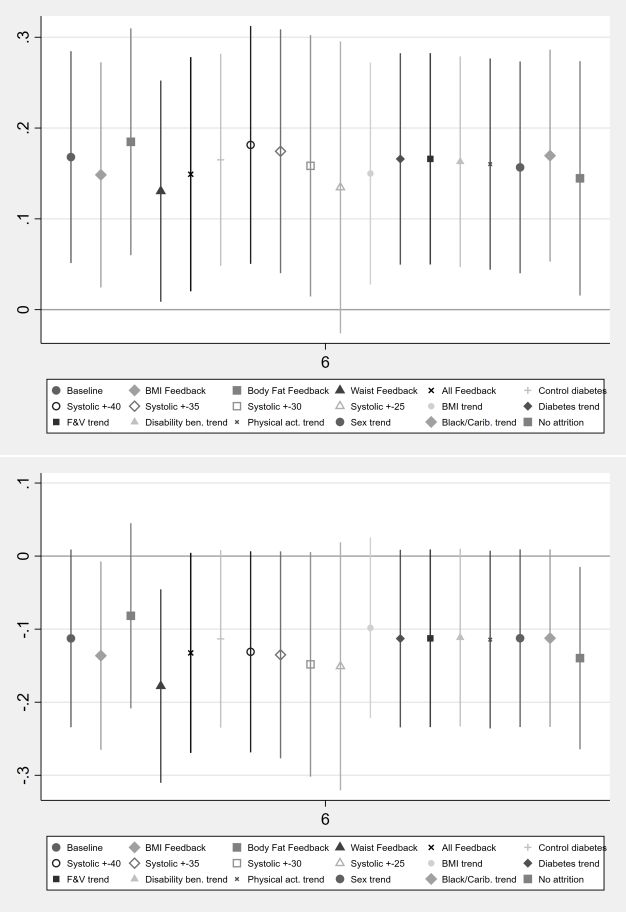
Note: Each dot is from a separate regression (FE-IV) set out in equation (2). The whiskers represent the 95% confidence interval calculated from standard errors clustered at the individual level. The top panel shows the immediate impact, the middle panel shows the long run impact not including trends, and the bottom panel is also for the long run and does include separate trends for those who have i. mildly raised hypertension and ii. raised and considerably raised hypertension.

Figure B5: The impact of hypertension diagnosis on SF-12 Physical - alternative specifications



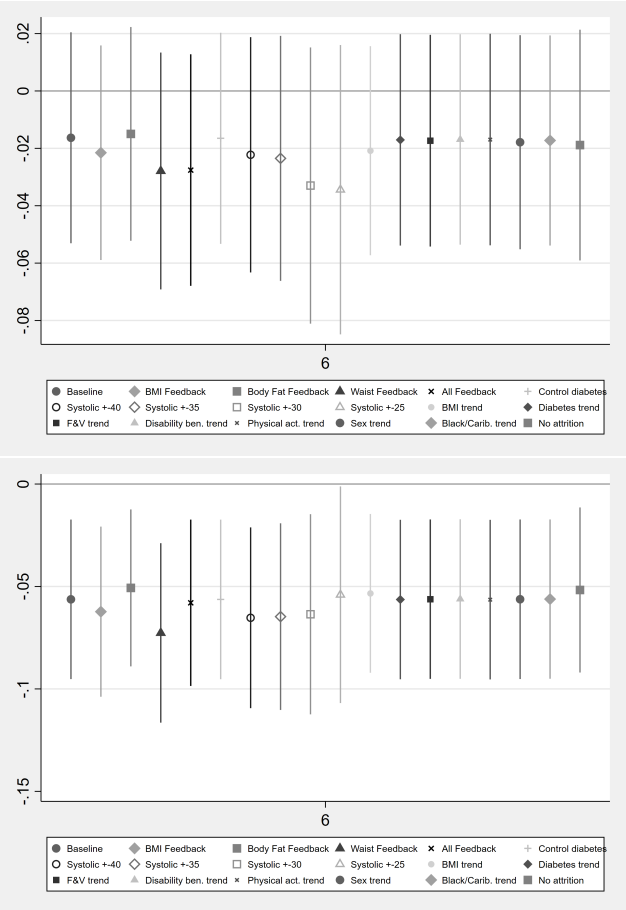
Note: Each dot is from a separate regression (FE-IV) set out in equation (2). The whiskers represent the 95% confidence interval calculated from standard errors clustered at the individual level. The top panel shows the immediate impact, the middle panel shows the long run impact not including trends, and the bottom panel is also for the long run and does include separate trends for those who have i. mildly raised hypertension and ii. raised and considerably raised hypertension.

Figure B6: The impact of hypertension diagnosis on CVD - alternative specifications



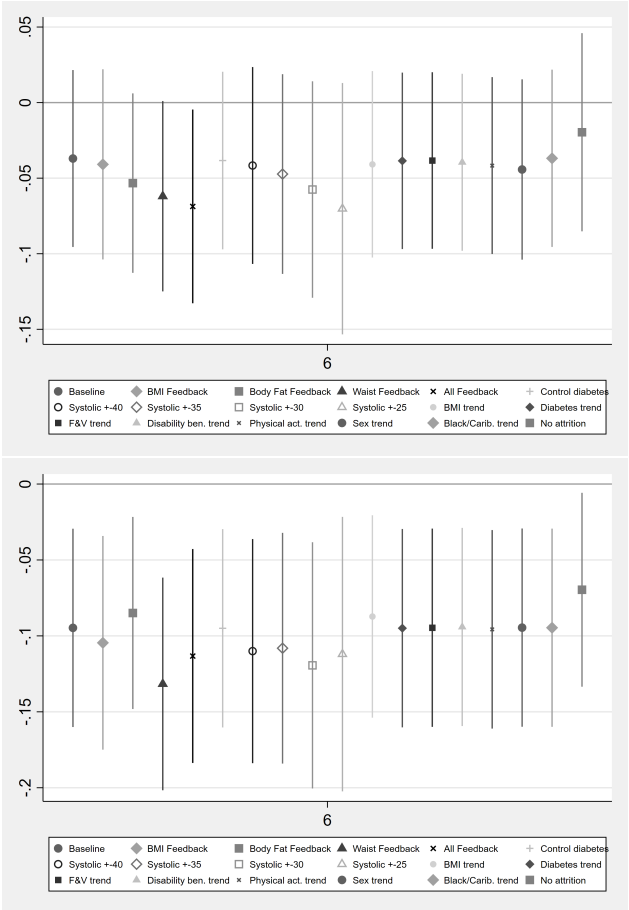
Note: Each dot is from a separate regression (FE-IV) set out in equation (2). The whiskers represent the 95% confidence interval calculated from standard errors clustered at the individual level. The top panel shows the shows the long run impact not including trends, and the bottom panel is for the long run and does include separate trends for those who have i. mildly raised hypertension and ii. raised and considerably raised hypertension.

Figure B7: The impact of hypertension diagnosis on congestive heart failure - alternative specifications



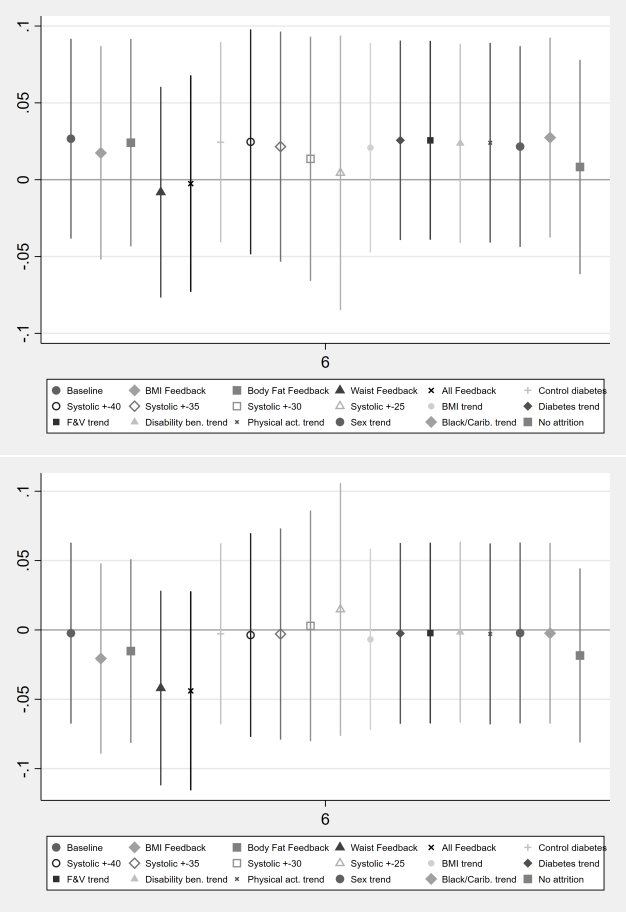
Note: Each dot is from a separate regression (FE-IV) set out in equation (2). The whiskers represent the 95% confidence interval calculated from standard errors clustered at the individual level. The top panel shows the shows the long run impact not including trends, and the bottom panel is for the long run and does include separate trends for those who have i. mildly raised hypertension and ii. raised and considerably raised hypertension.

Figure B8: The impact of hypertension diagnosis on coronary heart disease - alternative specifications



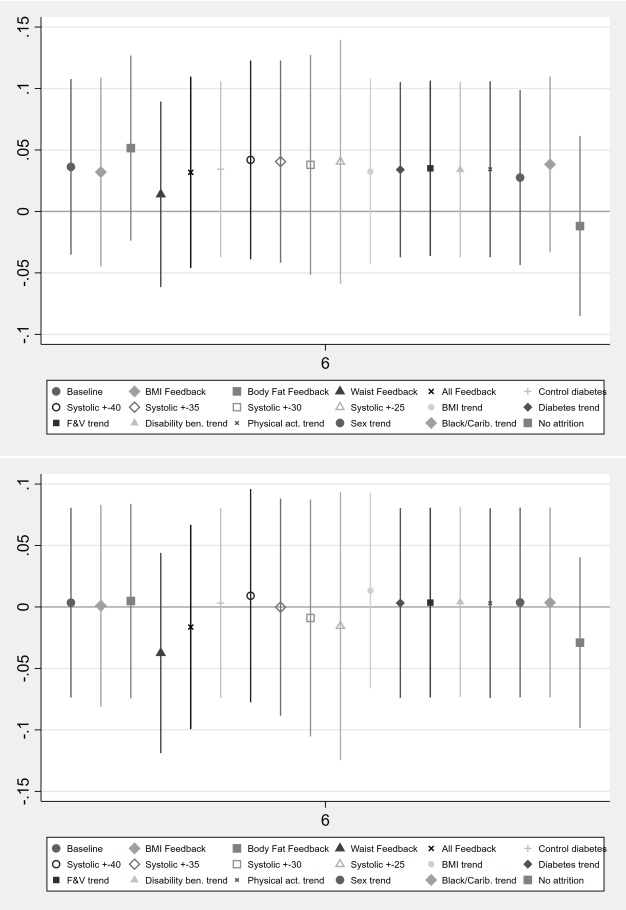
Note: Each dot is from a separate regression (FE-IV) set out in equation (2). The whiskers represent the 95% confidence interval calculated from standard errors clustered at the individual level. The top panel shows the shows the long run impact not including trends, and the bottom panel is for the long run and does include separate trends for those who have i. mildly raised hypertension and ii. raised and considerably raised hypertension.

Figure B9: The impact of hypertension diagnosis on angina - alternative specifications



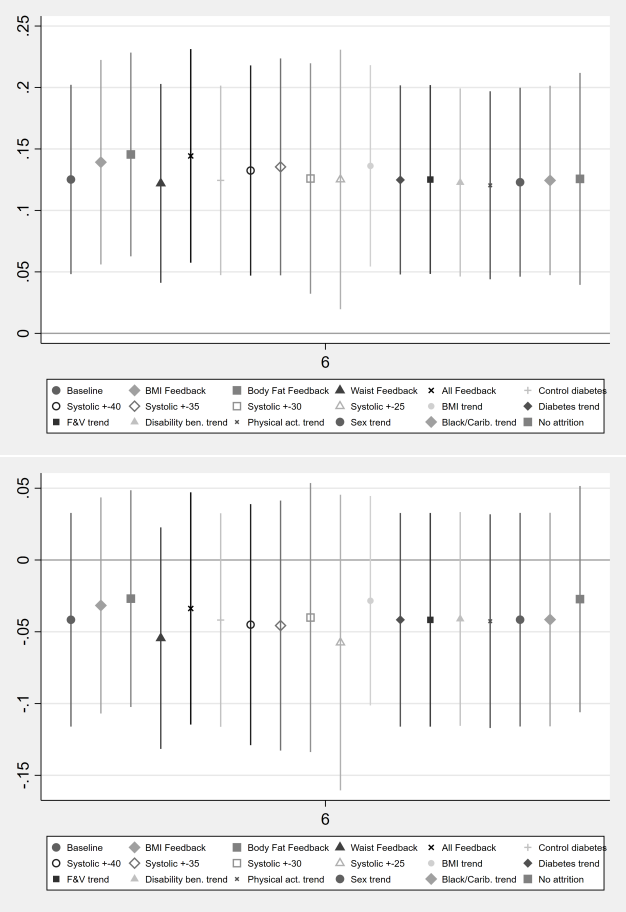
Note: Each dot is from a separate regression (FE-IV) set out in equation (2). The whiskers represent the 95% confidence interval calculated from standard errors clustered at the individual level. The top panel shows the shows the long run impact not including trends, and the bottom panel is for the long run and does include separate trends for those who have i. mildly raised hypertension and ii. raised and considerably raised hypertension.

Figure B10: The impact of hypertension diagnosis on heart attacks - alternative specifications



Note: Each dot is from a separate regression (FE-IV) set out in equation (2). The whiskers represent the 95% confidence interval calculated from standard errors clustered at the individual level. The top panel shows the shows the long run impact not including trends, and the bottom panel is for the long run and does include separate trends for those who have i. mildly raised hypertension and ii. raised and considerably raised hypertension.

Figure B11: The impact of hypertension diagnosis on stokes - alternative specifications



Note: Each dot is from a separate regression (FE-IV) set out in equation (2). The whiskers represent the 95% confidence interval calculated from standard errors clustered at the individual level. The top panel shows the shows the long run impact not including trends, and the bottom panel is for the long run and does include separate trends for those who have i. mildly raised hypertension and ii. raised and considerably raised hypertension.

Appendix C: Data Appendix

Self-reported GP diagnosis:

On entering the panel: Has a doctor or other health professional ever told you that you have any of the conditions listed on this card? Please just tell me the numbers that apply (see below).

At a second and subsequent interviews: Since [date of last interview] has a doctor or other health professional newly diagnosed you as having any of the conditions listed on this card? If so, which ones?

1. Asthma
2. Arthritis
3. Congestive heart failure
4. Coronary heart disease
5. Angina
6. Heart attack or myocardial infarction
7. Stroke
8. Emphysema
9. Hyperthyroidism or an over-active thyroid
10. Hypothyroidism or an under-active thyroid
11. Chronic bronchitis
12. Any kind of liver condition
13. Cancer or malignancy
14. Diabetes
15. Epilepsy
16. High blood pressure
17. Clinical depression

96. None of these

Timing of UKHLS data collection:

Table C1 shows the timing of data collection. Questions on household composition, happi-

Table C1: UKHLS content and timing of collection

Wave	**Module**	t_{UKHLS}	t_{BHPS}
All	household composition, general health, physical and mental health, health conditions including hypertension, hospitalisations, life satisfaction, employment, income and spending	All	All
2+	retirement planning	t=0	t=-1
2 and 5	Lifestyle: diet, exercise, alcohol consumption and smoking	t=0 and t=3	t=-1 and t=2

ness, general health, physical and mental health, health conditions including hypertension, hospitalisations, life satisfaction, employment, income and spending are asked directly at every wave. Information on retirement planning is asked at every wave from wave 2 onwards and savings data is collected in every other wave starting at wave 2. Lifestyle questions form part of rotating modules (waves 2 and wave 5) and include details of diet, exercise, alcohol consumption and smoking behaviour. Attitudes to risk were also collected (wave 1).

The lifestyle information usually has a short reference period referring to the month or week before the interview date, with some minor exceptions. For nutrition, information is collected on the usual number of days per week that fruit and vegetables are each eaten, and how many portions of each are eaten on a usual day. Alongside this, the main type of milk that is consumed (whole, semi-skimmed, skimmed, soya, other) and main type of bread (white, wholemeal, granary etc.) is recorded. For exercise, respondents are asked whether they have walked continuously for at least 10 minutes and 30 minutes in the past 4 weeks, the number of days for which this occurred and their self-reported average pace (slow, steady, brisk, fast). The type of any sports activities done are reported for the last 12 months. For alcohol consumption, respondents report whether they drank in the last week, for how many days, and the quantity on the day they drank the most. Separate variables also record the frequency of drinking in the past year and household spending on alcohol in the last 4 weeks. For smoking, we observe whether respondents have ever smoked, whether they currently smoke, the ages at which they started and gave-up, and

the usual quantity of cigarettes smoked per day.