Health Risk Realization Versus Warning: Impact on Lifestyle Behaviours*

Zoey Verdun†

June 2022

Abstract

Using individual-level panel data from *Understanding Society* I estimate the response to a health risk realization — heart attack or diabetes diagnosis — on a healthy lifestyle index. To overcome the endogeneity of a diagnosis, I match on initial health risks. I find individuals improve their overall lifestyle healthiness when faced with a large negative health event such as a diagnosis (heart attack or diabetes) whereas they do not respond to solely receiving information about certain disease risk factors, via a diagnosis of high blood pressure or chest pain. The drivers of the overall effect are a decrease in the number of cigarettes smoked and an increase in the probability to quit drinking alcohol; there is no significant effect found for either diet or exercise. I find some heterogeneity by sex, but only when looking at individual lifestyle behaviours. Overall, the findings suggest that the realization of a health risk leads individuals to improve their lifestyle behaviours, while only a signal about their health risks leads to no such change.

Keywords: *health risks; disease diagnosis; lifestyle behaviours; behavioural change*

JEL codes: *I12, D83*

---

*I am grateful to Michèle Belot, Jonathan James and Seetha Menon for their valuable comments and suggestions, as well as Brown Bag seminar participants in the department of economics at the Erasmus University Rotterdam for their helpful comments and discussion. A pre-analysis was submitted and published on the American Economic Association's RCT registry in October 2019 (reference: AEARCTR-0004943), for deviations see Appendix A.

†zoey.verdun@eui.eu, European University Institute (EUI)

1. Introduction

Obesity, cardiovascular disease and diabetes are currently a source of much concern, with rising costs globally, from both direct healthcare costs and indirect productivity loss (Bloom et al., 2011), and cardiovascular disease and diabetes being among the top ten global causes of death (World Health Organization, 2018). That being said, there is evidence that a healthier lifestyle can prevent such chronic diseases (Willett et al., 2006). Studies have shown that adopting a healthier lifestyle — such as reducing or quitting smoking, improving diet, exercising and reducing alcohol consumption — can improve quality of life by both extending an individual's lifespan and increasing the quality of the years to come (Chou et al., 2012; Rizza et al., 2014). Finally, and quite remarkably, various medical scholars have found evidence suggesting that the progress of these diseases can be stopped or in some cases even reversed through lifestyle changes.[1]

Governments and other institutions have placed a great deal of emphasis on encouraging the adoption of better lifestyles, often through information campaigns. Despite these efforts, the public health and economics literature suggests that information alone is often not sufficient to achieve long term changes in lifestyle (e.g. Kelly and Barker, 2016; Carrera et al., 2020). This appears to be particularly true of diet, whereas the evidence is more mixed in the case of exercise, smoking and alcohol consumption. Nevertheless, the medical literature has some well-documented cases of individuals who *do* make successful lifestyle changes, when provided with more than just information (Esselstyn et al., 1995; Ornish et al., 1998; Lanza et al., 2001).[2] Finally, there is a strand of literature, discussed further below, that looks at whether individuals re-optimize their lifestyle habits after important life changes (e.g. Wood and Neal, 2016; Hut and Oster, 2022); the results are mixed.

[1] Buttar et al. (2005); Sagner et al. (2014); McMacken and Shah (2017).

[2] Aspects associated with such changes include having: new information or knowledge, support, certain personality traits, or had a health shock/diagnosis. I study this last one.

This paper makes two contributions. First, there is little to no causal evidence on the impact of a realization of a health risk on lifestyle behaviours. Therefore, in this paper I estimate what impact a realization of a health risk — the diagnosis of a heart attack or diabetes — has on lifestyle behaviours, where the behaviours relate to diet, exercise, smoking and alcohol consumption. Given a realization of a health risk is often not exogenous, I use a matching approach to overcome this endogeneity problem. The matching approach is very well suited to my setting because the variable of interest is not a choice variable, though it is endogenous. I match on the initial health risk of individuals, and then subsequently some individuals experience the realization of the health risk while others do not. For the analysis, I create a healthy lifestyle index, which captures four lifestyle behaviours, consisting of eight behaviour variables: one for diet,[3] two each for exercise and smoking, and three for alcohol consumption. The exact descriptions of the eight variables are provided in Section 3.1.

The second contribution of this paper is the comparison of the response from a diagnosis of a realized health risk versus a diagnosis of risk factors. The former being the realization of an actual health condition (diagnosis of heart attack or diabetes), whereas the latter is only a signal on health status (diagnosis of high blood pressure or angina (chest pain)).Finally, this paper also explores the heterogeneity of response by sex. This heterogeneity is of particular interest given the literature on the differences between men and women both for the biological and social determinants and consequences of chronic disease and lifestyle behaviours (see for example Vlassoff, 2007).

The first finding of this paper is a positive association between a diagnosis of a realized health risk and a subsequent change in lifestyle. On average, such a diagnosis leads to a statistically significant increase in the healthy behaviour lifestyle index by 0.227 units. When looking at the behaviour variables that make up the index, two of the behaviours are driving most of the result. The

[3]Henceforth, in the analysis "diet" or "dietary change" only refers to a change in fruit and vegetable (F&V) consumption. Due to few diet-related variables, F&V is a proxy for diet.

number of daily cigarettes smoked decreases by 4.53 cigarettes, significant at the 1% level, and the probability to quit drinking alcohol increases by 6.32 percentage points, significant at the 5% level.

This paper's second finding is that only experiencing a realization of a health risk (heart attack or diabetes diagnosis) has a large and significant effect on improving overall lifestyle behaviour, whereas the overall effect of only receiving information about certain health risk factors is little to none. The positive effect of the health risk realization on the index (0.534 units) is more than six times larger than the insignificant effect of a risk factor diagnosis (0.086 units). Looking at individual behaviours, the improvement in the lifestyle index following a realization of a health risk is driven by three statistically significant variables: an increase in daily fruit and vegetable consumption (0.42 servings), an increase in the probability to quit smoking (9.46 percentage points) and a decrease in the number of cigarettes smoked per day (6.34 cigarettes). While the overall effect of the diagnosis of risk factors is minimal, one underlying driver does increase statistically significantly: the number of days abstaining from drinking alcohol increases (0.24 days). There is also weak evidence that the diagnosis of risk factors increases the probability to quit smoking and to quit drinking and decreases the daily number of cigarettes smoked.

Finally, I find heterogeneous effects by sex. The average effect on the index is similar for women and men and not statistically different. However, when I decompose the effect by the index's components, it appears women respond with larger effects and across more behaviours than men. This larger response by women is hidden in the average effect because some of their responses have opposite effects on the index. For two of these underlying drivers, women differ in a statistically significantly way from men in their response: women increase their probability to quit drinking (13.8 percentage points) and to quit smoking (7.7 percentage points) while men do not. Looking beyond statistically significant differences, women *decrease* the number of days per month walked at least 10 or 30 minutes by 2.3 and 2.5 days respectively, while men at most decrease by just over half a day. There is also tentative evidence that men more

than women decrease the daily number of cigarettes smoked and increase the number of daily servings of fruits and vegetables consumed.

In the economics literature there are only a few studies that investigate what lifestyle changes, if any, an individual undertakes following a health related diagnosis. Furthermore, there is no consensus yet as to which lifestyle behaviours do in fact change. The three studies closest to this paper are Oster (2018), Bhalotra et al. (2020) and Hut and Oster (2022), that all look at either a response to a diagnosis of a realized health risk or that of a risk factor (see next Section for a discussion on their results and how their papers compare to mine). Similarly, there are only a few studies in the medical literature investigating behavioural change in response to a health risk realization or disease diagnosis; the two most related being Chong et al. (2017) and Fassier et al. (2017).

My paper also relates to the economics literature on perceived risks and subjective probabilities of (realized) health risks (and also relates to the equivalent medical literature). Theoretically, a realized health risk can be conceptualized as a signal an individual receives about his or her own health status. Related papers, discussed in more detail in the next section, look at the response of individuals to generic or tailored health information and the updating of beliefs about their (perceived) risk of developing a disease or medical condition (e.g Bhalotra et al., 2020; Belot et al., 2020). My paper relates to this literature by studying *both* the impact of receiving information about a health risk factor (noisier signal) and the impact of the health risk being realized (clearer signal) on changes in lifestyle behaviour. Additionally, by studying the numerous lifestyle behaviours concurrently, it contributes to the literature on behavioural change and the likely heterogeneous costs of such change. Finally, this paper also contributes to the (economics) behavioural change literature and the findings that often behavioural change is difficult to achieve (e.g. Simon, 1955; Tversky and Kahneman, 1985; Frederick et al., 2002; Condon and McCarthy, 2006; Kelly and Barker, 2016; Oster, 2018; Bhalotra et al., 2020). My paper discusses how changes in different lifestyle behaviours may or may not occur depending on the costs that an individual faces when

considering making one or more changes. By better understanding whether, and if so, which lifestyle behaviours change after a realization of a health risk, certain health interventions can take this paper's findings into account — that individuals respond significantly to a realized health risk but not to a risk factor diagnosis, and that not all lifestyle behaviours are (equally) responsive — when trying to encourage changes in behaviour.

The rest of the paper is structured as follows: Section 2 summarizes the related literature, both in economics and in medicine. Section 3 describes the data. Section 4 describes the results from analysing the impact of a realized health risk on lifestyle and Section 5 describes and compares the impacts on lifestyle of solely a realized health risk versus only a diagnosis of a risk factor. Section 6 discusses the heterogeneity of the results by sex. Section 7 concludes.

2. Related Literature

The majority of the health economics literature on behavioural change builds upon the notion that the provision of (new) information, whether it be generic or tailored to the individual, and the subsequent updating of beliefs influences behaviour. However, a growing body of literature, not just in economics but also in health and psychology, finds that knowledge or information is usually necessary but not sufficient for change (Bartiaux, 2008; Mathis and Steffen, 2015; Kelly and Barker, 2016). For example, Kelly and Barker (2016) highlight several common errors made when attempting to foster behavioural change that relate to information, whether it be changes in diet, alcohol consumption or physical activity. Two such errors are assuming "knowledge and information drive behaviour" and "it is about getting the message across" (p.111). There is also a large body of literature in behavioural economics that looks at behavioural biases and self-control issues such as time-inconsistent and present-biased preferences (see for example Laibson, 1997; O'Donoghue and Rabin, 1999, 2015; Benhabib et al., 2010).

That being said, when looking across different individual health behaviours the evidence of the impact of information on behavioural change is mixed. For diet, it is the most clear, generic information is not sufficient (Brownell and Cohen, 1995; Worsley, 2002). There is even evidence that individuals are willing to forgo health benefits to prevent having to modify their diet (Atkin, 2016; Oster, 2018). For exercise, Young et al. (1996) find a health education campaign has little to no effect on physical activity levels. In contrast, Craig et al. (2006) find a public-awareness campaign to increase walking among adults, using an objective self-monitoring tool, to be effective. For smoking, two integrative reviews find mass media campaigns effective when it comes to fostering awareness and behavioural change (Flay, 1987; Durkin et al., 2012). However, Strecher et al. (1994), using computer-tailored targetted messaging, find positive effects only for light to moderate smokers. Finally, for alcohol consumption Fleming et al. (1999) find interventions by a physician can be effective in reducing different outcomes of alcohol use. In contrast, a review of the existing literature on prevention and treatment of college students finds little support for educational or awareness programs, but more support for other (non-informational) interventions in reducing alcohol use and its negative consequences (Larimer and Cronce, 2002).

Compared to the limited success of broader information campaigns there are more examples of successful change in both the behavioural change literature and the medical literature. However, these examples often involve more than just the general provision of information. Significant behavioural change, when it happens, often occurs in specific subgroups of individuals or individuals in certain situations. For example, in the economics literature, Hut and Oster (2022) find evidence that dietary change only occurs in a specific sub-group of individuals, those whose baseline diet consists of a small number of foods. In the medical literature, such specific groups of individuals or situations are usually patients of certain doctors, intensive programs or interventions (Esselstyn et al., 1995; Lanza et al., 2001). For example, in the Lifestyle Heart

Trial patients sustained intensive lifestyle changes — diet, exercise, smoking and stress management — over several years (Ornish et al., 1998).

There is also a growing literature in economics on the impact of the diagnosis of risk factors, such as hypertension or high blood pressure, and the provision of (tailored) health status information on changes in dietary behaviour (Zhao et al., 2013; Belot et al., 2020; Carrera et al., 2020) and other lifestyle behaviours such as smoking, exercise and alcohol use (Edwards, 2018; Bhalotra et al., 2020). In this literature, the paper closest to mine is Bhalotra et al. (2020). They find a high blood pressure diagnosis leads to a reduction in smoking but it has no impact on either exercise or alcohol consumption; furthermore, they find possible but imprecise evidence that diets marginally improve after diagnosis. My paper differs from Bhalotra et al. (2020) in that where they only look at the impact of a risk factor diagnosis, high blood pressure, I look at the diagnosis of risk factors *and* the corresponding diagnosis of the realized health risk itself.

Regarding the other related papers, Zhao et al. (2013) studies the impact of a high blood pressure diagnosis on diet and finds that individuals reduce fat intake, with richer individuals reducing more. Carrera et al. (2020) look at changes in workplace cafeteria spending in response to a high cholesterol diagnosis, they find that only those unaware and at high risk respond, but the spending changes are only modest and temporary. Belot et al. (2020) via an experiment, test whether providing either generic or tailored information on the risk of developing a heart attack or diabetes facilitates the adoption of a healthier dietary habits. They find evidence of change in the short run, but predominantly for the generic treatment. They also test but find no effect on healthiness of food choices when additional time is provided. Finally, Edwards (2018) looks at the impact of notification of results from submitted biomarkers on health behaviours. In particular, they find that notification of high blood glucose levels is associated with individual weight loss and increased exercise, and changes in health behaviour of spouses as well.

A smaller related economics literature is on the impact of disease diagnoses, such as cancer, heart disease or diabetes, on dietary change (Fassier et al., 2017; Oster, 2018; Hut and Oster, 2022). One of the two papers in this area closest to mine, Oster (2018), investigates the impact of a diabetes diagnosis on diet changes and finds a small but significant effect on calorie reduction in the month right after diagnosis, though the effect is no longer significant in the months following. The other related paper, Hut and Oster (2022), also find no effect of a diabetes diagnosis on diet, but rather that baseline diet and dietary concentration are the predictors of dietary change. My paper builds upon Oster (2018) and Hut and Oster (2022) by investigating the impact of diagnosis on several lifestyle behaviours, not only diet. By looking at several behaviours, and not just one, I allow for individuals to change certain behaviours but not others. This paper differs from Oster (2018) by including heart attack, alongside diabetes, as the diagnosis being studied; Hut and Oster (2022) are more similar to my paper since they look at the diagnosis of three different disease categories, which include diabetes and heart disease. Finally, Fassier et al. (2017) look at changes in diet and alcohol consumption before and after a cancer diagnosis. They find both healthy changes, decreased alcohol and sweetened drinks consumption, and unhealthy changes, decrease in vegetable consumption and reduction in the intake of many vitamins and minerals.

This paper contributes to both the above literatures — diagnosis of risk factors and diagnosis of realized health risk — by looking at *both* the receiving of information, a signal, on health status via the diagnosis of risk factors (high blood pressure and angina) and the actual diagnosis of the realized health risk (heart attack and diabetes). To the best of my knowledge, this is the first paper to compare the differences in response to a diagnosis of risk factors versus the realized health risk.

Other related literature this paper ties into is on incorrect knowledge and uncertainty about the risks and risk factors for certain health conditions, such as heart attacks, and how they correspond to health status. Individuals may under or overestimate their perceived risk from engaging in certain lifestyle behaviours

and hence also their subjective probabilities of having or getting a health condition or disease. For example, Belot et al. (2020), a paper closely related to this one finds, using an experimental approach, that individuals are often too pessimistic about their health status and thus the provision of tailored health information does not lead to improvements in lifestyle behaviours (for literature examples on incorrectly perceived risks for different lifestyles see Appendix B). Similarly, Hurd and McGarry (2002) find that individuals respond to new information, such as the onset of a disease or other health condition (which can be considered a more precise signal), by updating their subjective probabilities of survival. The lack of clear signals on health status is one of the reasons why individuals may not be willing to make any changes to their behaviour (Sanderson et al., 2009; Logie-MacIver et al., 2012).

In the medical literature, for example, Condon and McCarthy (2006) find that some individuals believed heart attacks only occurred in older people and therefore thought they could postpone changing their lifestyle to a later time. They also find heart attack patients had already been aware of their poor lifestyles — whether it be smoking, stress or poor diet — and yet, for a variety of reasons and beliefs, many did not change their lifestyle prior to the event. Many individuals were waiting for an initial "warning sign" in order to motivate themselves to improve their lifestyle (Condon and McCarthy, 2006). Similar to the economics literature, this need for a warning sign can be interpreted as a need to receive a more precise signal on an individual's health status before knowing what is the optimal level of a lifestyle (change) to implement.

3. Data and Analysis

The following section describes the data, the main variables, the (endogeneity) controls and the criteria for exclusion; it also discusses the descriptive statistics.

Data. The data is from *Understanding Society (UKHLS)*, a longitudinal survey of approximately 40,000 households from the United Kingdom, which includes data on lifestyle behaviours and health topics, such as if and when

an individual was last diagnosed with a heart attack or diabetes. The analysis uses the first five waves to provide pre-, during and post-treatment waves (for additional details, see Appendix B). In the main analysis the diagnosis is that of a heart attack or diabetes and the outcome is an index measure of changes in lifestyle. After applying all criteria for exclusion nearly 16,000 observations remain, of which just over 230 receive a diagnosis.[4] In the secondary analysis I split individuals into receiving exclusively either a diagnosis of a realized health risk or a diagnosis of certain health risk factors (high blood pressure or chest pain). After applying all restrictions, there remain approximately 12,200 control observations, just under 100 individuals who receive only a realized health risk and just over 450 who receive only a diagnosis of risk factors.

3.1. Description of the Variables

Dependent Variable. The dependent variable is a healthy lifestyle behaviour index, henceforth lifestyle index, that captures four lifestyle behaviours: diet, exercise, smoking and alcohol consumption. The index is based on eight behaviour variables, which are listed in Table 1.

TABLE 1. Four Lifestyle Behaviours Consisting of Eight Lifestyle Variables

Diet	**Walking**	**Smoking**	**Drinking Alcohol**
z_1: Number of servings of fruit and vegetables consumed (per day)	z_2: Number of days walked at least 10 minutes (past four weeks)	z_4: Dummy if currently smokes (extensive margin)	z_6: Dummy if currently drinks (at least one alcoholic drink in past 12 months) (extensive margin)
	z_3: Number of days walked at least 30 minutes (past four weeks)	z_5: Number of cigarettes smoked (per day) (intensive margin)	z_7: Number of days did not drink alcohol (past 7 days) (intensive margin)
			z_8: Number of drinks consumed on heaviest drinking day (past 7 days) (intensive margin)

I calculate the index, following Kling et al. (2007), at the individual level (i) using an equally-weighted sum of the z-scores of the four lifestyle

[4]If not clearly specified, the diagnosis always refers to the diagnosis in the main analysis.

behaviours: diet (B_1), exercise (B_2), smoking (B_3), and alcohol consumption (B_4). Each behaviour is an average of the z-scores (z_j) of the one to three behaviour variables (j) that make up that lifestyle behaviour: $B_1 = z_1$, $B_2 = (1/2)(z_2 + z_3)$, $B_3 = (1/2)(z_4 + z_5)$, and $B_4 = (1/3)(z_6 + z_7 + z_8)$. The z-score for each behaviour variable for each individual (z_{ij}) is obtained by subtracting the mean of that variable (μ_j) from the individual's observed behaviour of that variable (x_{ij}) and then dividing it by that variable's standard deviation (σ_j): $z_{ij} = (x_{ij} - \mu_j)/\sigma_j$. The index equation is:

$$Index_i = \sum_{k=1}^{4} B_{ik}$$

Independent Variables. The diagnosis is the diagnosis of a heart attack or diabetes,[5] which in the main analysis does not exclude the possibility of getting a diagnosis of a health risk factor in the same time period. Since these diagnoses of realized events are relatively rare, for reasons of power, I pool the two diagnoses and treat them as one.[6] For a discussion on statistical power, small sample of 'treated' individuals, and minimum detectable effect size calculations, see Appendix D. The realized health risk diagnosis variable takes a value of 1 if the individual was "newly diagnosed" with at least one of either a heart attack or diabetes or 0 if not newly diagnosed with either condition. I consider a medical condition as newly diagnosed if the individual responded "yes" to being diagnosed in waves 3 or 4 and "no" in the previous waves (1 and 2); I consider a medical condition as "not diagnosed" if the individual responded "no" in all four waves (1 through 4).[7]

The secondary analysis splits the receiving of a diagnosis into having only the realized health event (diagnosis of heart attack or diabetes) versus only

[5] Only 1% of diagnosed individuals get both. While the effect of experiencing two diagnoses is plausibly stronger, this study treats receiving two as the same as receiving one.

[6] Main analysis split by diagnosis, as a robustness check, finds similar effects (Appendix C).

[7] As a robustness check, I run the main regression separately for individuals diagnosed in wave 3 versus wave 4 to confirm no significant difference in responses (see Appendix C).

receiving information on health risk factors (diagnosis of high blood pressure or angina). This differentiation is important because a large realized health event gives a clear signal of health status; the health status signal from the information received from a risk factor diagnosis is much less precise. There is a significant share of individuals who receive one or both of these risk factor diagnoses but do not go on to experience either a heart attack or diabetes diagnosis. I run two analyses, each using one of the two diagnosis outcome variables: the "realized diagnosis only" variable and the "risk factor diagnosis only" variable. The former takes a value of 1 if the individual received a diagnosis of a realized health risk and did not receive a risk factor diagnosis. It takes a value of 0 if neither diagnosis occurred. The latter variable takes a value of 1 if the individual received a risk factor diagnosis and did not experience a realized health risk diagnosis. Again, it takes a value of 0 if neither occurred. For both variables I exclude individuals who experience both kinds of diagnosis.

Controls. An individual's likelihood of receiving any health diagnosis is not exogenous, it is correlated with their pre-diagnosis behaviours. Therefore, it is important to account for an individual's initial diagnosis risk level (i.e. the probability of being diagnosed with either a heart attack or diabetes). This risk level is partly determined by an individual's previous diet, exercise, smoking behaviour and possibly alcohol consumption.[8] To address the endogeneity problem, I control for the risk factor variables that make up this initial risk, which in the UKHLS are: age, sex, high blood pressure, smoking (extensive and intensive margins), fruit and vegetable consumption, and physical activity.[9] These risk factors come from risk assessment tools such as "Your Disease Risk" (Siteman Cancer Center, 2019) and organizations such as the American Heart Association (2017). In the secondary analysis I exclude high blood pressure as a risk factor since it is part of the outcome variable.

[8] I exclude alcohol as a risk factor because public health associations, such as the American Heart Association (2014), are less clear on its impact on the risk of heart attack or diabetes.

[9] Body Mass Index (BMI) and cholesterol are not included, see Appendix A.3 for details.

Other controls include education, ethnicity, employment status and geography. Education is included using the derived variable "highest education ever reported". The six original categories are merged into four: bachelor's degree or above, high school completion (A-level), high school completion or equivalent, and no qualifications. Ethnicity is included as a binary variable: white and non-white; the non-white category consists of the following ethnic groups: mixed; Asian or Asian British; black/African/Caribbean/black British; and other. Employment status is included as a proxy for income: individuals are split into either full-time employed, part-time employed or inactive. An urban dummy is included, which indicates whether an individual lives in a rural or urban region. Finally, a categorical variable is included that indicates in which of the twelve UK Government Office Regions (GOR) an individual resides.

Exclusion Criteria. The realized health risk diagnosis (in wave 3 or 4) is the main dependent variable in the analysis, therefore the following exclusion criteria are necessary. Exclude individuals who have been diagnosed with a heart attack or diabetes previously (at any time prior to wave 3). Exclude individuals if information about their diagnosis, or lack thereof, is not available in prior waves, to prevent any confounding effects.[10] Exclude individuals from the control group if they are in a household who has a member who is part of the treated group (i.e. has received a diagnosis), for details see Appendix A.4. Finally, exclude individuals if the diagnosis variable, at least one of the index variables, or any of the main controls are missing (for additional discussion, including lack of attrition, see Appendix B). For the secondary analysis an analogous set of restrictions for inclusion are applied, however this time for any previous diagnosis of high blood pressure or angina.

[10]Including individuals for whom it is not known if they were diagnosed prior to the first wave of observation; in wave 1 individuals are asked if they were ever previously diagnosed.

3.2. Descriptive Statistics

Descriptive statistics for the sample used for the analysis are shown in Table 2. In the sample, 44% of individuals have a bachelor's degree or above, about half have some form of high school degree and 10% have no qualifications. The sample consists of 89% white individuals, is nearly 60% female and has an average age of 48. In terms of health behaviours, individuals, prior to treatment, on average consume 3.4 daily servings of fruits and vegetables, walk 15.4 days per month at least 10 minutes per day, walk 9.5 days per month at least 30 minutes per day, 81% do not smoke and only 12% do not drink. In the full sample, the average number of daily cigarettes smoked is 2.4, whereas among smokers the average is 11.7. For alcohol consumption, in the full sample, individuals consume 2.8 drinks on their heaviest drinking day in a week, whereas looking only at drinkers, it is 3.1. Finally, in the full sample, individuals abstain from drinking 5 out of 7 days per week, whereas drinkers abstain 4.8 days.

TABLE 2. Descriptive Statistics

	count	mean	sd
Demographics			
Education: GCSE or other school qualification	15,853	0.28	0.45
Education: A-level etc	15,853	0.18	0.38
Education: Bachelor's degree or above	15,853	0.44	0.50
Non-white	15,853	0.11	0.31
Female	15,853	0.59	0.49
Age	15,853	47.84	16.07
Health Behaviours (pre-treatment)			
Number of servings of fruit/veg consumed per day	15,853	3.38	1.58
Number of days walked at least 10 minutes, past 4 weeks	15,853	15.43	10.80
Number of days walked at least 30 minutes, past 4 weeks	15,853	9.47	10.12
Does not smoke	15,853	0.81	0.39
Number of cigarettes smoked per day (for all)	15,853	2.38	6.20
Number of cigarettes smoked per day (for smokers)	3,234	11.69	8.94
Does not drink (at least in past 12 months)	15,853	0.12	0.32
Total drinks on heaviest drinking day, past 7 days (for all)	15,853	2.82	3.66
Total drinks on heaviest drinking day, past 7 days (for drinkers)	14,311	3.12	3.73
Number days did not drink, past 7 days (for all)	15,853	5.04	2.08
Number days did not drink, past 7 days (for drinkers)	14,311	4.83	2.08

4. Main Analysis: Impact of Realized Health Risk Diagnosis

The methodology used for both the main and secondary empirical analysis is kernel matching based on an estimated propensity score, comparing the average treatment effect on the treated (ATT) for treated and control units, using the first-differencing method.[11] I estimate the propensity score using the following variables: ethnicity, education, employment status (as a proxy for income), urban/rural dummy, a categorical variable for regions and the previously described, in Section 3.1, initial health diagnosis risk factors. For additional discussion on the construction and interpretation of the propensity score estimation, see Appendix B. Finally, I bootstrap the standard errors.[12]

4.1. Summary Statistics: Matched Treated versus Control

Table 3 shows the difference in means for various demographics and pre-treatment behaviours including the healthy lifestyle behaviour index. It compares the means for these variables for individuals who have been matched, both treated and controls. For example, the average age of the treated individuals is 59.4 years whereas the average age of matched controls is 58.1 years; 46% of treated individuals are female, whereas 49% of matched controls are female; and the pre-treatment index value, a weighted sum of the pre-treatment health behaviours, is very similar across the treated and matched control groups. For all variables in the table there are no statistically significant differences between the two groups; of particular note the alcohol consumption variables are not included in the propensity score matching algorithm and yet these variables are also statistically not different from one

[11] The propensity score estimations satisfy the three necessary conditions: balancing property, unconfoundedness assumption and common support (see Appendix E). Appendix F discusses the matching estimator selection procedure and assessment of the balance. Appendix G explains different matching approaches, including kernel matching.

[12] Propensity scores are not known, but rather estimated, prior to matching; by default, standard errors from kernel matching do not take this into account.

another. This observation suggests that the matching procedure works not just for observables, but also for unobservables in the sense that some other non-matched variables are also "matched" without having to be included in the matching procedure.

TABLE 3. Summary Statistics — Comparing Matched Treated and Controls

Variable	Mean Treated	Mean Control
Demographics		
Age	59.41	58.08
Female	0.46	0.49
Non-white	0.09	0.12
Education: No qualification	0.22	0.22
Education: GCSE or other school qualification	0.34	0.33
Education: A-level etc	0.15	0.16
Education: Bachelor's degree or above	0.29	0.29
Urban	0.24	0.25
Employment status: Inactive	0.66	0.65
Employment status: PT Employed	0.11	0.11
Employment status: FT Employed	0.23	0.24
Health Behaviours (pre-treatment)		
Number of servings of fruit/veg consumed per day	3.20	3.24
Number of days walked at least 10 minutes, past 4 weeks	12.87	13.65
Number of days walked at least 30 minutes, past 4 weeks	7.83	8.03
Does not smoke	0.76	0.76
Number of cigarettes smoked per day	3.52	3.28
Does not drink (at least in past 12 months)	0.13	0.14
Total drinks on heaviest drinking day, past 7 days	2.39	2.58
Number days did not drink, past 7 days	5.25	4.93
Pre-treatment Healthy Lifestyle Behaviours Index Score		
Healthy Behaviour Index	−0.45	−0.41

Note: None of the differences between treated and control are statistically significantly different from one another at the 5% level (or even at the 10% level). For the smoking and alcohol variables the full sample is used, regardless of whether individuals smoke or drink.

4.2. Empirical Strategy

The empirical analysis uses first-differences for two reasons.[13] First, the outcome variable is a difference between two periods. Second, the use of fixed

[13]I use the term "first-differences" even though the difference is between waves 2 and 5.

effects controls for unobservable time-invariant individual heterogeneity.[14] To evaluate the effects of the realized health risk diagnosis, I estimate the following model:

$$\Delta Index_{it} = \beta \Delta RealizedDiagnosis_{it} + \Delta u_{it}$$

where $\Delta Index_{it} = Index_{i,t=5} - Index_{i,t=2}$ denotes the difference between the lifestyle behaviour index in wave 2 and in wave 5. Similarly, the independent variable, $\Delta RealizedDiagnosis_{it} = RealizedDiagnosis_{i,t=5} - RealizedDiagnosis_{i,t=2}$, is the difference in diagnosis occurrence between wave 2 and wave 5, since in either wave $RealizedDiagnosis_i$ is a dummy for whether the diagnosis has occurred in waves 3 or 4. In wave 2, the dummy takes a value of 0 (no diagnosis) for all individuals; in wave 5 it takes either 0 or 1. Finally, $\Delta u_{it} = \Delta u_{i,t=5} - \Delta u_{i,t=2}$ is the differenced error term.

4.3. Findings

Index. The effect of a realized health risk diagnosis on changes to the lifestyle index is shown in column 1 of Table 4; the realized health risk diagnosis leads to an increase in the index by 0.227 units. I interpret an increase in the index as one or more lifestyle behaviours having become healthier. Therefore, the finding suggests that an individual who experiences a realized health risk diagnosis improves their lifestyle.[15] A more intuitive understanding of the size of the effect, using concrete numerical examples, is provided below.

Recall, there are four lifestyle behaviours that together consist of eight behaviour variables. Each lifestyle behaviour is an average of the z-scores (standardized with mean 0 and variance 1) of its corresponding behaviour variable(s). Each of the four lifestyle behaviours is then combined with equal weight to make up the index. If an individual were to increase the healthiness of one lifestyle behaviour by one unit, while keeping the other behaviours unchanged, then the overall value of the index would increase by one unit.

[14] Using first-differences is equivalent to using fixed effects since I only compare two periods.

[15] Though not part of the pre-analysis plan, I find a significant difference, at the 10% level, between older and younger treated individuals. For more details see Appendix C.

The found effect, a change in the index of 0.227 units, can be the result of, for example, an increase of 0.36 servings of fruits and vegetables consumed per day. Of course combinations of different amounts of change are also possible; for example a 0.23 change in the index can result from an increase in 0.18 servings of fruits and vegetables consumed daily and an increase of 2.45 days walked at least 10 minutes per day in a month.

For a better understanding of which and how many lifestyle behaviours may be driving the findings, I also discuss the individual effects of the behaviour variables that make up the index, shown in the remaining columns 2-9 of Table 4.

TABLE 4. Realized Health Risk Diag.: ATT of Change in Lifestyle Index and Components

	Index	Fruit/Veg	Walk 10	Walk 30	Smoke	Nr Cigs	Drink	Heavy	Days
Realized Diag.	0.227**	0.222*	-1.489*	-1.326*	0.0361*	4.525***+	0.0632**+	0.174	0.0836
	(0.111)	(0.122)	(0.853)	(0.736)	(0.0193)	(1.582)	(0.0254)	(0.223)	(0.129)
Observations	15,853	15,853	15,853	15,853	15,853	3,234	15,853	14,311	14,311

Realized health risk diagnosis definition: the diagnosis of heart attack or diabetes in wave 3 or 4. Independent variable: difference between the realized health risk diagnosis variable in wave 2 (pre-treatment) and wave 5 (post-treatment). Each dependent variable: difference between its value in wave 2 (pre-treatment) and wave 5 (post-treatment). Bootstrap standard errors in parentheses, 1000 reps. Kernel matching using 0.0075 bandwidth for index and 0.0009375 bandwidth for individual index components. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, $^{+}$ $p < 0.05$ using a Hochberg correction with FDR of 0.1 (correction not applicable for index coefficient).

Decomposition of index variables. To analyse the effect of the lifestyle index's eight variables, I apply the necessary correction to control the Type I error rate given the testing of multiple hypotheses. To this end I use the Hochberg correction; it generates a critical value for each outcome to test if that outcome is statistically significant at the 5% significance level. The critical values are calculated using a 0.1 false discovery rate (FDR).

The interpretation of the remaining eight columns of Table 4 is as follows, a positive coefficient is: an increase in the quantity of a healthy behaviour, a decrease in the quantity of an unhealthy behaviour, or an increase in the probability of quitting an unhealthy behaviour between wave 2 and wave 5.

The difference in the average number of servings of fruit and vegetables consumed per day (Fruit/Veg) increases by 0.22 servings. The number of days per month walked at least 10 minutes per day (Walk 10) or at least 30 minutes

per day (Walk 30) decreases by 1.49 days and 1.33 days, respectively. The negative point estimates for both the days walked variables hint at possible physical limitations from experiencing a negative health event. The probability that an individual quits smoking (Smoke) increases by 3.61 percentage points. For all four of these variables, the effects are statistically significant at the 10% level, however none remain statistically significant after applying the Hochberg correction.

Looking at the next two variables, the number of cigarettes smoked in a day (Nr Cigs) decreases significantly at the 1% level by 4.53 cigarettes; the probability to quit drinking alcohol (Drink) decreases statistically significantly at the 5% level by 6.32 percentage points.[16] Both variables remain significant after applying the Hochberg correction. For the final two variables, the reduction in drinks consumed on the heaviest drinking day of the week (Heavy) and the number of days in a week an individual abstained from drinking alcohol (Days), though neither statistically significant, suggest if anything a possible decrease in the frequency and intensity of alcohol consumed.

In summary, the key findings from this decomposition analysis are that the behavioural changes driving the main finding of the primary analysis are a decrease in the daily number of cigarettes smoked and an increase in the probability to quit drinking alcohol. Possible, but less precisely estimated, drivers also include an increase in fruit and vegetable consumption, an increase in the probability to quit smoking and *decreases* in the number of days in a month individuals walked at least 10 and at least 30 minutes. A decomposition by sex is discussed in Section 6.

[16] The number of observations for the alcohol consumption variables are lower than for the other variables because questions regarding alcohol consumption habits were asked in a self-completion part of the survey, which quite some individuals did not fill out.

5. Secondary Analysis: Impact of Diagnosis: Realized vs Risk Factor

The secondary analysis follows the same set-up as the main analysis: kernel matching using propensity score estimation with first-differences. It studies the impact of the precision of the health status information received by analysing two separate cases: "realized diagnosis only" and "risk factor diagnosis only", as defined in Section 3.1, and then compares them. I estimate the propensity score separately for each case.[17]

5.1. Empirical Strategy

The analysis of the health status signal precision is the investigation of the impact of a receiving either a signal from a health risk realization (preciser signal) or from a risk factor diagnosis (noisier signal) relative to no signal received on the lifestyle index. To evaluate the effects of only experiencing a health risk realization (heart attack or diabetes diagnosis), I estimate the following model:

$$\Delta Index_{it} = \beta \Delta RealizedOnly_{it} + \Delta u_{it}$$

where $\Delta Index_{it}$, as defined in the main analysis, denotes the difference in the lifestyle index. Analogous to the main analysis, the independent variable, $\Delta RealizedOnly_{it} = RealizedOnly_{i,t=5} - RealizedOnly_{i,t=2}$, is the difference in the realized health risk diagnosis variable between wave 2 and wave 5. Finally, Δu_{it} is also defined as in the main analysis. To evaluate the effects of only receiving a risk factor diagnosis, I estimate the following model:

$$\Delta Index_{it} = \beta \Delta RiskFactor_{it} + \Delta u_{it}$$

[17] The matching strategy selection procedure and assessment of the balance are discussed in Appendices F.2 and F.3, respectively. For the realized diagnosis only case the optimal kernel bandwidth is 0.00375, for the risk factor diagnosis only case it is 0.0075. I select these optimal bandwidths based on what leads to the best balance using the same assessment criteria as the matching strategy selection procedure.

where the risk factor diagnosis equation is the same as the realized health risk diagnosis equation except that the independent variable, $\Delta RiskFactor_{it} = RiskFactor_{i,t=5} - RiskFactor_{i,t=2}$, is the difference in the risk factor diagnosis variable between wave 2 and wave 5.

5.2. Findings – Realized Health Risk Diagnosis Only

Index. Column 1 of Table 5 shows the positive impact of a realized health risk diagnosis on the lifestyle index: an increase of 0.534 units, statistically significant at the 1% level. This finding suggests that individuals respond to receiving only the realized health risk diagnosis by significantly improving the overall healthiness of their lifestyle.[18]

TABLE 5. Realized Health Risk Diagnosis Only: ATT of Change in Lifestyle Index and Components

	Index	Fruit/Veg	Walk 10	Walk 30	Smoke	Nr Cigs	Drink	Heavy	Days
Realized Diag. Only	0.534***	0.418**+	-0.504	-1.090	0.0946**+	6.340**+	0.0538	0.143	0.139
	(0.194)	(0.190)	(1.246)	(1.034)	(0.0383)	(2.525)	(0.0341)	(0.261)	(0.183)
Observations	12,339	12,339	12,339	12,339	12,339	2,612	12,339	11,176	11,176

Realized health risk diagnosis only definition: the diagnosis of heart attack or diabetes in wave 3 or 4, without a concurrent or existing diagnosis of high blood pressure or angina. Independent variable: difference between the realized health risk diagnosis variable in wave 2 (pre-treatment) and wave 5 (post-treatment). Each dependent variable: difference between its value in wave 2 (pre-treatment) and wave 5 (post-treatment). Bootstrap standard errors in parentheses, 1000 reps. Kernel matching using 0.00375 bandwidth for index and 0.0009375 bandwidth for individual index components. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, $^{+}$ $p < 0.05$ using a Hochberg correction with FDR of 0.1. (correction not applicable for index coefficient).

Decomposition of index variables. Columns 2-9 of Table 5 show the impact of receiving only the diagnosis (heart attack or diabetes) on the eight index variables with and without multiple hypothesis testing correction. The following three variables have effects that are statistically significant at the 5% level and remain significant even after the correction: increase in fruit and vegetable consumption (0.42 servings), decrease in the probability to quit smoking (9.46 percentage points) and decrease in the number of cigarettes smoked (6.34 cigarettes).

[18]While not in the pre-analysis plan, I find a statistically significant difference between individuals with or without some higher education. For more details, see Appendix C.

To put these findings in perspective, I compare them to the existing literature. The increase in fruit and vegetable consumption is in line with Oster (2018) who finds a diabetes diagnosis leads to a small but significant improvement in diet, including increased fruit and vegetable consumption. The significant impact of a diagnosis on the probability to quit smoking is in line with Chong et al. (2017) who find a significant impact of a diabetes diagnosis on the probability of quitting, just under 20 percentage points. However, where I also find a significant effect in the reduction of the number of cigarettes smoked, Chong et al. (2017) do not find a significant effect, though the sign of their coefficient still suggests a reduction (of 1.82 cigarettes).

Looking beyond statistical significance, except for the walking variables, all the point estimates have the expected non-negative sign. Given that there are fewer treated individuals in this realized health risk diagnosis only case, it is possible that decomposed there is not enough power to statistically detect all the separate effects of the individual behaviour variables that make up the index.

Due to concerns of insufficient power, and with the aim to still provide some insights, I provide a more descriptive discussion of the differences in magnitudes between the realized health risk diagnosis only and risk factor diagnosis only cases below, after the risk factor diagnosis only case (see Appendix D for a table of minimum detectable effect sizes).

5.3. Findings – Risk Factor Diagnosis Only

Index. Column 1 of Table 6 shows the impact of receiving only a risk factor diagnosis on the lifestyle index. In this case, the information does not lead to a statistically significant increase in the index (p-value: 0.322). This finding suggests that individuals either do *not* increase their overall lifestyle healthiness after receiving such risk factor information (a null result) or individuals do react to it but the effect is not detectable due to insufficient statistical power (a discussion on statistical power, minimum detectable effect size and precision of estimates is in Appendix D). However, note that statistically significant

or not, the response from the risk factor diagnosis only case is small in absolute magnitude and significantly smaller than the response from receiving the realized health risk diagnosis only, discussed further below. Based on the 95% confidence interval, I can rule out an effect size larger than 0.255, which is still twice as small an effect compared to the realized health risk diagnosis.

TABLE 6. Risk Factor Diagnosis Only: ATT of Change in Lifestyle Index and Components

	Index	Fruit/Veg	Walk 10	Walk 30	Smoke	Nr Cigs	Drink	Heavy	Days
Risk Factor Diag. Only	0.0856	0.0230	-0.419	-0.864	0.0272**	1.553*	0.0326**	0.115	0.240**+
	(0.0865)	(0.0833)	(0.571)	(0.529)	(0.0138)	(0.855)	(0.0160)	(0.163)	(0.0948)
Observations	12,701	12,701	12,701	12,701	12,701	2,697	12,701	11,496	11,496

Risk factor diagnosis only definition: only the diagnosis of high blood pressure or angina. Independent variable: difference between risk factor diagnosis variable in wave 2 (pre-treatment) and wave 5 (post-treatment). Each dependent variable: difference between its value in wave 2 (pre-treatment) and wave 5 (post-treatment). Bootstrap standard errors in parentheses, 1000 reps. Kernel matching using 0.0009375 bandwidth. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, $^{+}$ $p < 0.05$ using a Hochberg correction with FDR of 0.1 (correction not applicable for index coefficient).

Decomposition of index variables. Columns 2-9 of Table 6 show the impact of a risk factor diagnosis (high blood pressure or angina) on the eight index variables, both with and without the Hochberg multiple hypothesis testing correction. All the point estimates have the expected positive sign except for the walking variables. Other than an increase of 0.24 days in the number of days in a week individuals abstain from drinking alcohol, none of the variables have a statistically significant effect after applying the Hochberg correction. This latter finding is in line with one of the main findings in this paper that overall only receiving risk factor information (via diagnosis) has little to no effect on lifestyle change.

Looking at the individual behaviours, fruits and vegetables consumed and the number of days in a month walked at least 10 or at least 30 minutes are not statistically significant. The following three effects were statistically significant prior to the correction: the probability to quit smoking increases by 2.72 percentage points, the number of cigarettes smoked per day decreases by 1.55 cigarettes and the probability to quit drinking alcohol decreases by 3.26 percentage points. The decrease in the number of drinks consumed on the heaviest drinking day of the week is not significant. Finally, the number of days in a week from which alcohol is abstained increases statistically significantly

by 0.24 days and is the only variable to remain so after applying the Hochberg correction.

I compare my findings to recent evidence by Bhalotra et al. (2020). Whereas I find that a receiving information on certain health risk factors has a small but significant impact on increasing the number of days individuals abstain from drinking alcohol and a possible impact on the probability to quit drinking alcohol, Bhalotra et al. (2020) find that a high blood pressure diagnosis has no impact on alcohol consumption. Bhalotra et al. (2020) also find there is a significant effect on the probability to quit smoking, an increase of 9 percentage points, similar to my findings that there is a possible increase in the probability of quitting. Finally, my finding that risk factor information has no impact on exercise is in line with Bhalotra et al. (2020).

5.4. Findings: Comparing Realized Only & Risk Factor Only Cases

This section compares the two cases briefly discussed above: only receiving a realized health risk diagnosis versus only receiving the risk factor diagnosis. First it looks at the index, second the decomposition of the index by individual behaviours.

Index. First, looking at index analysis, the realized health risk diagnosis only case has only 94 treated individuals whereas the risk factor diagnosis only case has 456 treated individuals; both cases have over 12,000 control individuals. I compare the two cases by comparing the first column of Tables 5 and 6. In the first case, the realized health risk diagnosis leads to a 0.534 unit increase in the lifestyle index, whereas in the second case, the risk factor diagnosis has a point estimate of 0.0856, which is not statistically significant. The point estimate of the realized health risk diagnosis only case is more than six times larger than that of the risk factor diagnosis only case with the difference being statistically significant at the 5% level. Although it seems fairly intuitive that the realization of a health risk has a larger effect than the smaller (non-significant) effect found for receiving of risk factor information, this paper, to the best of my knowledge,

is the first to confirm such intuition empirically. One explanation for this finding is that receiving a clear(er) signal about an individual's health status may lead to (more precise) updating of their beliefs regarding their current health status, and thus make an individual (more) willing or motivated to improve one or more of their lifestyle behaviours.

The very large difference in effect size found for the index when comparing the different signal precisions is reinforced in the decomposition of the index into its behaviour variables, which is discussed next. The overall finding is that for all the behaviour variables the magnitude of the realized health risk diagnosis point estimate is either significantly greater (i.e. healthier) or not statistically different from the risk factor diagnosis one.

Decomposition of individual behaviours. This section compares the realized health risk diagnosis only and risk factor diagnosis only cases by comparing columns 2-9 of, respectively, Tables 5 and 6. As mentioned, the concern is given that there are only a select number of individuals who receive any type of diagnosis, especially a realized health risk diagnosis, some effects are not detected due to being underpowered. For the realized health risk diagnosis only case there are at most 141 treated individuals, whereas 569 for the risk factor diagnosis only case.[19]

There are three variables — fruit and vegetable consumption, probability to quit smoking and number of cigarettes — for which the difference is statistically significant between the effect size of the realized health risk diagnosis and the risk factor diagnosis; for each of these variables there is a statistically significant effect for the realized health risk diagnosis only case, at the 5% level, and no such effect for the risk factor diagnosis only case. The rest of the variables, except for days drinking alcohol, are not statistically significant in either case (realized health risk diagnosis or risk factor diagnosis) and are

[19] I say "at most" because for each behaviour variable the number of individuals diagnosed (realized health risk or risk factor) varies. The numbers reported here are for smoking status, the variable with the highest number of diagnosed individuals.

also not statistically different between the two cases. For the days abstained from drinking alcohol, there is a statistically significant effect for the risk factor case whereas there is no detectable effect for the realized health risk diagnosis case; however, the difference is still not statistically different from zero. Taking a more descriptive approach, the magnitudes of the effects, except for the number of days abstained from drinking alcohol, are all larger in the realized health risk diagnosis only case. This observation is in line with my overall finding that experiencing a realized health risk diagnosis has a significantly larger impact on lifestyle changes than only receiving information on certain health risk factors.

6. Heterogeneous Effects by Sex

Index. The effect of a realized health risk diagnosis on the lifestyle index by females and males is reported, respectively, in Column 1 of Table 7 and Column 1 of Table 8. The effect of the diagnosis for women leads to a 0.22 standard deviation increase in the index, slightly lower than the statistically significant 0.25 standard deviation increase by men; however, the difference is not statistically significant.

TABLE 7. ATT of Change in Lifestyle Index and Components, by sex — Female

	Index Female	Fruit/Veg Female	Walk 10 Female	Walk 30 Female	Smoke Female	Nr Cigs Female	Drink Female	Heavy Female	Days Female
Realized diag.	0.220	0.220	-2.493^{**+}	-2.288^{**+}	0.0771^{**+}	3.857	0.138^{***+}	0.130	0.00279
	(0.180)	(0.163)	(1.211)	(1.109)	(0.0313)	(2.470)	(0.0434)	(0.266)	(0.171)
Observations	9,311	9,311	9,311	9,311	9,311	1,857	9,311	8,266	8,266

Realized health risk diagnosis definition: the diagnosis of heart attack or diabetes in wave 3 or 4. Independent variable: difference between the realized health risk diagnosis variable in wave 2 (pre-treatment) and wave 5 (post-treatment). Each dependent variable: difference between its value in wave 2 (pre-treatment) and wave 5 (post-treatment). Bootstrap standard errors in parentheses, 1000 reps. Kernel matching using 0.00046875 bandwidth for index and 0.0009375 bandwidth for individual index components. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, $^{+}$ $p < 0.05$ using a Hochberg correction with FDR of 0.1 (correction not applicable for index coefficient).

A further decomposition of the effect of the diagnosis on the index's eight lifestyle behaviour variables provides insights into the differences in behavioural change between women and men across these variables, respectively Columns 2-9 of Table 7 and Table 8). First, I provide a brief overview of those findings, then a more detailed discussion follows.

TABLE 8. ATT of Change in Lifestyle Index and Components, by sex — Male

	Index Male	Fruit/Veg Male	Walk 10 Male	Walk 30 Male	Smoke Male	Nr Cigs Male	Drink Male	Heavy Male	Days Male
Realized diag.	0.248*	0.287*	-0.605	-0.361	0.00347	5.190**	0.00301	0.242	0.155
	(0.138)	(0.170)	(1.172)	(1.008)	(0.0216)	(2.158)	(0.0281)	(0.344)	(0.181)
Observations	6,542	6,542	6,542	6,542	6,542	1,377	6,542	6,045	6,045

Realized health risk diagnosis definition: the diagnosis of heart attack or diabetes in wave 3 or 4. Independent variable: difference between the realized health risk diagnosis variable in wave 2 (pre-treatment) and wave 5 (post-treatment). Each dependent variable: difference between its value in wave 2 (pre-treatment) and wave 5 (post-treatment). Bootstrap standard errors in parentheses, 1000 reps. Kernel matching using 0.005625 for index and 0.0009375 bandwidth for individual index components. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, $^{+}$ $p < 0.05$ using a Hochberg correction with FDR of 0.1 (correction not applicable for index coefficient).

Decomposition – Overview. For women four of the eight lifestyle behaviour variables have changes that statistically differs from zero: decrease in the number of days walked at least 10 minutes in a month, decrease in the number of days walked at least 30 minutes in a month, increase in the probability to quit smoking, and increase in the probability to quit drinking alcohol. For men, none of the variables remain statistically significant, however there is some weaker evidence suggesting men increase their fruit and vegetable consumption and decrease the number of cigarettes they smoke.[20] The only differences that are statistically significant between women and men is the probability to quit drinking alcohol and the probability to quit smoking.

Finally, while at first glance it appears that women and men react approximately equally to the realized health risk diagnosis, the decomposition tells a different story. Women react more strongly than men, but with regard to the index the decreases in walking counteracts for the most part the increase in other behaviours. One possible reasoning for this finding is that the medical literature finds that heart attacks for women are harder to detect and diagnose (see for example Brush Jr et al., 2020). Therefore, I suggest that the heart attacks for women in the sample may be more severe than for men and thus elicit a larger response. This possible reasoning and more details on the

[20] Lack of a significant effect could be due to insufficient statistical power rather than a null result. Recall, Appendix D discusses statistical power and min. detectable effect sizes.

statistically significant effects and their magnitudes are discussed further in the next section.

Decomposition – In detail. The two statistically significant differences between men and women are the probability to quit smoking, at the 10% significance level, and the probability to quit drinking alcohol, at the 1% significance level. What follows is a comparison of the statistically significant effects for either men or women, but which are *not* statistically significant *between* men and women.

Only for men are the following behaviours statistically significant different from zero, though it should be noted that neither of the following effects remain significant after applying the Hochberg multiple hypothesis testing correction: at the 10% level, men increase their fruit and vegetable consumption by 0.29 servings and they decrease the number of cigarettes they smoke, by 5.19 cigarettes at the 5% significance level.[21] Only for women are the following behaviours statistically significant different from zero: decrease their number of days walked at least 10 and at least 30 minutes in a month, by 2.49 days and 2.29 days respectively, both at the 10% level; increase their probability to quit smoking by 7.71 percentage points at the 10% level; and increase their probability to quit drinking alcohol by 13.8 percentage points at the 5% level. Neither sex has a detectable statistically significant decrease in the intensity or number of days drinking alcohol.

As briefly mentioned above, although the difference in the overall effect of a realized health risk diagnosis for women and men is similar, it appears that women react much more than men. However, this difference is not visible in the main analysis because women both improve certain health behaviours (increase in the index) and suffer a reduction in physical activity/walking (decrease in the index), which is why these effects partially cancel out. According to the medical literature (see for example Brush Jr et al., 2020), heart attacks for women are

[21] The larger magnitude decrease in cigarettes by men may be partly mechanical; men smoke more cigarettes than women prior to the diagnosis, thus have greater scope for reduction.

harder to detect and thus less likely diagnosed than for men. Therefore, one possible explanation for my findings is that the heart attack diagnoses for women in my sample may be on average stronger than those for men, as those are the heart attacks more likely to be detected and thus reported in my data. The larger reaction by women is visible both through the larger response to improving healthier behaviours (such as quitting smoking and drinking alcohol) but also via the significant decrease in walking. The latter may come about in part because those (women) who experience a more severe heart attack may be more likely to be physically disabled and/or reduce their physical activity afterwards.[22]

7. Conclusion

This paper finds that individuals improve their overall lifestyle, measured using a lifestyle index, after experiencing a diagnosis by improving some of their lifestyle behaviours. It also empirically confirms the intuition that some individuals make larger lifestyle changes when the diagnosis received is more severe, interpreted as receiving a more precise signal about their health status. Examples of such diagnoses include a heart attack or diabetes diagnosis. This paper finds the main driving the lifestyle change are increasing the probability to quit smoking and drinking. There is little evidence of lifestyle changes when individuals do not experience a realization of a health risk but rather only receive (less precise) information about their health status, such as being diagnosed with a health risk factor such as high blood pressure or angina.

These findings are in line with the literature (e.g. Oster, 2018; Hut and Oster, 2022) that finds that diet is generally little or not responsive to realized health risks such as diabetes diagnosis. The other results also fit within the mixed results of the literature that finds that health risk signals are typically not effective in changing certain behaviours. For example Bhalotra et al. (2020)

[22]For note on heterogeneity analysis by realized health risk and sex, see Appendix B.

and Carrera et al. (2020) find temporary or no effects of high blood pressure diagnosis on diet changes, however Zhao et al. (2013) does find a response for dietary fat. Similarly, Bhalotra et al. (2020) does find an effect of high blood pressure diagnosis on smoking, something I find weak evidence for as well.

Furthermore, although the overall effect on lifestyle changes between men and women is of similar size, the heterogeneity lies in the changes made to the behaviours that make up the overall lifestyle index. In particular, women generally make more and larger changes, however some of these changes have opposite effects on the lifestyle index. One reason for this sex differential may be attributable to the increased difficulty of diagnosing heart attacks in women.

More generally, a possible reason for the heterogeneous findings across different factors (such as health status signal precision or demographics) is that individuals may face varying costs when considering behavioural change, both across individuals but also across behaviours within an individual. These findings suggest that policy-makers interested in fostering certain lifestyle changes must take into account both what lifestyle behaviour they are trying to change and which individuals they are targetting, as heterogeneous costs may play a role in the effectiveness of policies or interventions being considered. Future research should explore what are the different costs faced by individuals when considering lifestyle changes, whether they be financial, social or personal, with the latter including personality characteristics. Finally, the changes to the lifestyle behaviours found in this paper are for the time horizon of one to two years after a diagnosis; it would be very interesting and societally relevant to investigate in future research if such changes are also sustained in the longer run.

Appendix A: Deviations from Pre-Analysis Plan

The plan is located: https://www.socialscienceregistry.org/trials/4943/history/56027. The following sections of Appendix A discuss the deviations.

A.1. Additional Outcome Variable

My pre-analysis plan mentions seven health behaviour outcome variables, whereas this paper uses eight. The additional outcome variable captures the extensive margin of alcohol consumption: a dummy for if an individual is a drinker or not. Furthermore, the original pre-analysis plan did not include that I would summarize these behaviours into a lifestyle index. I introduced an index to have only a single outcome variable.

The reason for including this additional alcohol consumption variable is the lack of an extensive margin variable for alcohol consumption. In the case of the smoking variables there is one intensive and one extensive margin variable. In the case of alcohol consumption, there were only two intensive margin variables. However, I realized it is important to have both an extensive and an intensive margin variable for this behaviour as it is possible that individuals behave heterogeneously after experiencing a diagnosis (whether it be for a realized health risk or a risk factor), changing one of these margins but maybe not both.

Finally, I keep both the intensive margin variables for alcohol consumption: the number of days an individual did not drink alcohol in the past seven days and the total number of drinks an individual consumed on the heaviest drinking day in the past seven days. The reason is that the literature on harmful alcohol consumption and behaviours suggests that when measuring the most harmful aspects of alcohol consumption on health (i.e. binge drinking) it is *both* the intensity with which an individual drinks in a given period of time (about two hours) and how often per week an individual drinks that matter.

A.2. No Splitting of Pooled Realized Health Risk or Risk Factor Variables

The separating of the pooled realized health risk diagnosis (or risk factor diagnosis) variable, as a robustness check, for either the main analysis or secondary analysis is not performed, though I mentioned it in the pre-analysis

plan. The reason is insufficient power, the main reason to pool to begin with. However, I do run the first regression of the main analysis split by diagnosis as a robustness check to confirm the effect magnitudes are similar, see Table C.8.

A.3. No BMI or Blood Cholesterol in Propensity Score

In the pre-analysis plan I wrote that if more than half the sample had the BMI and blood cholesterol variables non-missing, I would perform a robustness check to see if there is any impact of their inclusion on the final findings. However, for the blood cholesterol variable only 32% of the sample is non-missing and for BMI only 48% is non-missing. Therefore, these variables are not included in any analysis or robustness check.

A.4. Changes to Empirical Strategy

The pre-analysis plan specified I would match the treated with the controls in my sample and then do regression analysis on that matched sample, either matching using stratification or nearest neighbour 1:1. However, given that kernel matching provides the best balance, I use it instead. Therefore, instead of matching treated with control units and then running standard regression analyses using the matched sample, I use the matching commands directly. This means that the matching program will match the treated and controls using the selected strategy and then provide the ATT (or ATE, if desired). The consequence of changing the technical approach to regression analysis it is no longer possible to include household clustered standard errors (something I said I would do in my pre-analysis plan).[23] I can include the clustered standard errors in a regular regression analysis, but cannot when using matching

[23]In theory, the inclusion of household clustered standard errors is considered important because there are individuals in the sample that live in the same household together. It is important to take into account that such individuals may not be identically and independently distributed. One example of this possible dependence is if one household member receives a diagnosis another household member may adjust their behaviour, even without having received a diagnosis themselves.

analysis commands directly. However, the use of a different matching strategy and/or different standard errors have no effect on the findings, as discussed in Appendix G.1.

Furthermore, to address the likely lack of independence between household members, I omit all household members who would be classified as a control individual but who are part of a household that contains a treated individual. This is because other individuals inside the same household as a treated individual, especially in the case of heart attack, often improve their lifestyle behaviours significantly and persistently (Fadlon and Nielsen, 2019). Such omitted "treated-control" members make up less than 1% of the sample.

A.5. Multiple Hypothesis Correction

My pre-analysis plan stated the use of the Holm-Bonferroni procedure. The reasoning was that the more powerful Hochberg procedure only holds under non-negative dependence. However, that reasoning was incorrect since the correlations between the outcomes are all non-negative. Most of the correlations between the outcomes are positive and the select few that have a negative sign have such a small correlation coefficient that I consider those to be essentially zero. Therefore, ex-post I decided the use of the Benjamin-Hochberg correction procedure, henceforth Hochberg correction, instead was better; it is uniformly more powerful than the Holm-Bonferroni procedure.

Appendix B: Additional Details

Section 2: Related Literature – Literature on examples of incorrectly perceived risks for the different lifestyles. For examples of incorrectly perceived risks for the different lifestyles, see: diet (Condon and McCarthy, 2006), smoking (Heikkinen et al., 2010), and exercise (Fitgerald et al., 1994; Crombie et al., 2004). For alcohol consumption, incorrectly perceived risks may stem from conflicting recommendations between some public health associations, such as the American Heart Association (2014) and the current medical literature

(Stockwell et al., 2016), a likely source of confusion to the public and doctors alike.

Section 3: Data and Analysis – Choice and use of certain waves of data. There are currently nine waves available. The lifestyle behaviour variables are only available in waves 2, 5 and 7; however wave 7 is not used because some questions have been changed compared to those in waves 2 and 5. Wave 1 captures any diagnoses that may have occurred prior to the start of the UKHLS.

Section 3: Data and Analysis – Timing between measurement moments. Individuals are usually surveyed every 12 months. Thus, the time from the measurement of the pre-diagnosis behaviour to the diagnosis is typically 0-24 months; between the diagnosis and the post-diagnosis behaviour measurement is usually 12-36 months.

Section 3.1: Description of the Variables: Criteria for Exclusion – Exclusion and attrition. As mentioned in the main text, because I exclude individuals if the diagnosis variable, at least one of the index variables, or any of the main controls are missing, I therefore also exclude individuals missing (any variables) from wave 5. Attrition is not a concern: only six individuals experienced a diagnosis and then died before the wave 5 data collection; one of whom already misses a variable that leads to omission from the analysis.

Section 4: Main Analysis: Impact of Realized Health Risk Diagnosis – Construction and interpretation of propensity score estimation. The propensity score estimation, aside from being a linear function of these variables, includes for some of these variables higher order terms and interactions. The higher order terms chosen for inclusion are solely selected such that the estimated propensity score satisfies the balancing property. As such, no behavioural interpretation needs to be given.

Section 4.3: Findings – Additional examples of change in index value interpretation. In the case of walking, where the lifestyle behaviour consists of

two variables, a change in the index of 0.227 units is equivalent to: an increase of 4.90 days walked at least 10 minutes per day in a month, or an increase of 2.37 days walked at least 30 minutes per day in a month.[24]

Alternatively, it is also possible that some lifestyle behaviours worsen, for example that individuals walked fewer times per month. In such a case a combination of an increase in fruit and vegetable consumed by 0.51 servings and a *decrease* in the number of days walked per month at least 10 minutes by 1 day would still lead to a change in the index of 0.227 units.[25]

Section 6: Heterogeneous Effects by Sex – Details on heterogeneity also by realized health risk. Although not written in my pre-analysis plan, I run a heterogeneity analysis for the response on the index from a realized health risk diagnosis, splitting both by whether the diagnosis is a heart attack or a diabetes diagnosis and by sex. However, I do not have sufficient power to be able to detect whether the difference in the response between women and men for heart attacks is statistically different from that of women and men for diabetes.

Appendix C: Exploring Additional Heterogeneous Effects

I did not preregister any of the following heterogeneous effects in my pre-analysis plan. However, I investigate these different heterogeneous effects at the suggestion of a few readers of an earlier version of this paper.

[24]The difference in the necessary number of days walked comes from that reality that an increase the number of days walked at least 10 minutes per day has no impact on the number of days walked at least 30 minutes per day. However, the reverse is not true; an increase in the number of days walked at least 30 minutes per day also changes the number of days walked at least 10 minutes per day. The opposite is true for a decrease.

[25]Here again, a decrease in the variable walked 10 minutes a day also decreases the variable of walked 30 minutes a day, I account for this in the calculation for this example.

Split by Median Age of Treated. Table C.1 shows the heterogeneous effects of age on the change in health index in response to a realized health risk diagnosis. Tables C.2 and C.3 also show the heterogeneous effects of age but now in response to either exclusively the realized health risk diagnosis or the risk factor diagnosis, respectively. What is interesting to note is that for the main analysis realized health risk diagnosis (Table C.1) only the younger treated individuals respond significantly to the diagnosis; there is no significant response from older individuals; and the difference is statistically significant at the 10% level. By contrast, in the case of exclusively the realized health risk diagnosis (Table C.2) it appears that both younger and older treated individuals respond equally strong. Although not an exact comparison, this finding is in line with Oster (2018) who finds no effect of age on the impact of a diabetes diagnosis on diet. Finally, in the case of receiving only the risk factor diagnosis (Table C.3) there is no significant reaction to the diagnosis nor a significant difference in the response between younger and older treated individuals; this is not surprising since there is no (large) response to the risk factor diagnosis in the main analysis either.

TABLE C.1. Realized Health Risk Diagnosis on Index — Split by Median Age

	Younger (< 62 years)	Older (≥ 62 years)
Realized Diagnosis	0.452^{***}	0.0391
	(0.170)	(0.145)
Observations	12,231	3,622

Realized health risk diagnosis definition: the diagnosis of heart attack or diabetes in wave 3 or 4. Bootstrap standard errors in parentheses, 1000 reps. Kernel matching (0.0075 bandwidth). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Split by (no) Higher Education. Table C.4 shows the heterogeneous effects of (not) having some higher education on the change in health index in response to a realized health risk diagnosis. Tables C.5 and C.6 also show the heterogeneous effects of education but now in response to only the realized health risk diagnosis and only the risk factor diagnosis, respectively. For the

TABLE C.2. Realized Health Risk Diagnosis Only on Index — Split by Median Age

	Younger ($\leq$ 58 years)	Older ($>$ 58 years)
Realized Diagnosis Only	0.540**	0.551*
	(0.255)	(0.313)
Observations	9,820	2,519

Realized health risk diagnosis only definition: the diagnosis of heart attack or diabetes in wave 3 or 4, without a concurrent or existing diagnosis of high blood pressure or angina. Bootstrap standard errors in parentheses, 1000 reps. Kernel matching (0.00375 bandwidth). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

TABLE C.3. Risk Factor Diagnosis Only on Index — Split by Median Age

	Younger ($\leq$ 55 years)	Older ($>$ 55 years)
Risk Factor Diagnosis Only	0.173	0.0186
	(0.131)	(0.108)
Observations	9,392	3,309

Risk factor diagnosis only definition: the diagnosis of high blood pressure or angina. Bootstrap standard errors in parentheses, 1000 reps. Kernel matching (0.0075 bandwidth). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

case of education, the findings are clear that it is the individuals with no higher education (lower than a bachelor's degree) that react to the realized health risk diagnosis, both in the main analysis realized health risk diagnosis case (Table C.4) and the realized health risk diagnosis only case (Table C.5). There are no (heterogeneous) effects for the risk factor diagnosis only case (Table C.6).

Comparing my findings to that of the literature, my realized health risk diagnosis only finding is not in line with Oster (2018), who finds no effect of demographics (including education) on the impact of diet on a diabetes diagnosis. However, in my analysis the effect of the realized health risk diagnosis only is being driven not just by diet, but also by smoking behaviour (both the number of cigarettes and the probability of smoking), therefore it is not an ideal comparison. In the case of the risk factor diagnosis only, my finding is in line with Bhalotra et al. (2020), who study the impact of high blood pressure on different lifestyle behaviour, they also do not find heterogeneous effects by education.

TABLE C.4. Realized Health Risk Diagnosis on Index — Split by (no) Higher Education

	No Higher Education	Some Higher Education
Realized Diagnosis	0.294**	0.0603
	(0.139)	(0.177)
Observations	8,904	6,949

Realized health risk diagnosis definition: diagnosis of heart attack or diabetes in wave 3 or 4. Bootstrap standard errors in parentheses, 1000 reps. Kernel matching (0.0075 bandwidth). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

TABLE C.5. Realized Health Risk Diag. Only on Index — Split by (no) Higher Education

	No Higher Education	Some Higher Education
Realized Diagnosis Only	0.717***	0.114
	(0.262)	(0.254)
Observations	6,540	5,799

Realized health risk diagnosis only definition: the diagnosis of heart attack or diabetes in wave 3 or 4, without a concurrent or existing diagnosis of high blood pressure or angina. Bootstrap standard errors in parentheses, 1000 reps. Kernel matching (0.00375 bandwidth). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

TABLE C.6. Risk Factor Diagnosis Only on Index — Split by (no) Higher Education

	No Higher Education	Some Higher Education
Risk Factor Diagnosis Only	0.103	0.0530
	(0.109)	(0.150)
Observations	6,767	5,934

Risk factor diagnosis only definition: the diagnosis of high blood pressure or angina. Bootstrap standard errors in parentheses, 1000 reps. Kernel matching (0.0075 bandwidth). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Split by Timing of Diagnosis. Recall that the timeline of the data is pre-diagnosis lifestyles are measured in wave 2, realized health risk diagnoses (or risk factor diagnoses) occur in wave 3 and/or 4, and post-diagnosis lifestyles are measured in wave 5. The longer/shorter time since the diagnosis relative to the post-diagnosis moment of measurement of lifestyle change has two mechanisms that can affect the strength of the response in opposing manners. On the one hand an earlier occurring diagnosis (wave 3) may give individuals more time to adopt lifestyle changes, in which case I expect a larger effect for individuals

who experienced a diagnosis in wave 3 compared to wave 4. However, on the other hand, it could also be that the initial changes made after the diagnosis are not sustained in the longer run and therefore the (longer term) effect of a diagnosis in wave 3 on lifestyle change is weaker than a diagnosis that occurred in wave 4. I omit individuals who get a diagnosis in both waves, which is less than 10% of the treated sample. Table C.7 shows the heterogeneous effects of experiencing a diagnosis in wave 3 versus wave 4. These findings suggest the first mechanism plays a larger role — individuals need sufficient time to change their lifestyle after a diagnosis rather than that they lose (some of) their initial changes made over time.

TABLE C.7. Realized Health Risk Diagnosis on Index — Split by Timing of Diagnosis

	Wave 3	Wave 4
Disease Diagnosis	0.309*	0.187
	(0.187)	(0.155)
Total Observations	15,713	15,738
Treated Observations	92	117

Realized health risk diagnosis definition: the diagnosis of heart attack or diabetes in wave 3 or 4. Bootstrap standard errors in parentheses, 1000 reps. Kernel matching (0.0075 bandwidth). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Split by Realized Health Risk Diagnosis. Recall that the main realized health risk diagnosis variable consists of individuals who have had either a heart attack and/or a diabetes diagnosis, and that the main reason for pooling these diagnoses is to increase statistical power. Although both diagnoses require a similar change in lifestyle to alleviate or even reverse the condition, individuals may respond differently across the diagnoses. Table C.8 investigates any heterogeneous effects across the two diagnoses. I omit individuals who experience both diagnoses, about 1% of the treated sample.

TABLE C.8. Realized Health Risk Diagnosis on Index — Diagnosis Split by Medical Condition

	Heart Attack	Diabetes
Realized Diagnosis	0.270	0.180
	(0.193)	(0.136)
Total Observations	15,695	15,776
Treated Observations	74	155

Realized health risk diagnosis definition: the diagnosis of a realized health risk in wave 3 or 4. Each column corresponds to a diagnosis of that realized health risk only. Bootstrap standard errors in parentheses, 1000 reps. Kernel matching (0.0075 bandwidth). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Appendix D: Power Calculations

Given the relatively small number of treated individuals in my sample, for many of my regression results there is a concern of insufficient power. This is particularly a problem when splitting the results further to get at heterogeneous effects such as sex or to assess the impact of the individual behaviours that make up the index.

The following section provides the minimum detectable effect (MDE) sizes that I expect given the data in order to assist in understanding whether insignificant results are due to insufficient power. If the effect size is smaller than the MDE then I cannot exclude that the effect is significant but only that the estimate is not precisely measured to due insufficient power. In contrast, if the estimate is larger than the MDE, then most likely the insignificant effect is actually a null result. For each regression presented in the paper the MDE size is shown in Table D.1. I calculate the minimum detectable effect size using the following equation:

$$MDE = (t_{\alpha/2} + t_{1-\kappa})\sqrt{\frac{\sigma^2}{np(1-p)}} = (t_{\alpha/2} + t_{1-\kappa})\frac{\sigma}{\sqrt{np(1-p)}}$$

Where using conventional values $t_{\alpha/2} = 1.96$ if $\alpha = 0.05$ and $t_{1-\kappa} = 0.84$ if $\kappa = 0.80$, n is the sample size, p is the proportion of the sample that is treated, and σ is the standard deviation (SD) of the dependent variable.

TABLE D.1. Power Calculations

Table	Table Description	Dependent Variable	MDE
Table 4	Main realized health risk diagnosis	Index	0.304
Table 4	Main realized health risk diagnosis	Fruit/Veg	0.300
Table 4	Main realized health risk diagnosis	Walk 10	2.119
Table 4	Main realized health risk diagnosis	Walk 30	2.014
Table 4	Main realized health risk diagnosis	Smoke	0.048
Table 4	Main realized health risk diagnosis	Nr Cigs	3.036
Table 4	Main realized health risk diagnosis	Drink	0.056
Table 4	Main realized health risk diagnosis	Heavy	0.775
Table 4	Main realized health risk diagnosis	Days	0.336
Table 5	Realized health risk diagnosis only	Index	0.479
Table 5	Realized health risk diagnosis only	Fruit/Veg	0.467
Table 5	Realized health risk diagnosis only	Walk 10	3.330
Table 5	Realized health risk diagnosis only	Walk 30	3.196
Table 5	Realized health risk diagnosis only	Smoke	0.077
Table 5	Realized health risk diagnosis only	Nr Cigs	4.097
Table 5	Realized health risk diagnosis only	Drink	0.084
Table 5	Realized health risk diagnosis only	Heavy	1.253
Table 5	Realized health risk diagnosis only	Days	0.516
Table 6	Risk factor diagnosis only	Index	0.221
Table 6	Risk factor diagnosis only	Fruit/Veg	0.215
Table 6	Risk factor diagnosis only	Walk 10	1.534
Table 6	Risk factor diagnosis only	Walk 30	1.473
Table 6	Risk factor diagnosis only	Smoke	0.035
Table 6	Risk factor diagnosis only	Nr Cigs	2.040
Table 6	Risk factor diagnosis only	Drink	0.084
Table 6	Risk factor diagnosis only	Heavy	0.574
Table 6	Risk factor diagnosis only	Days	0.239
Table 7	Main realized h. risk diagnosis (only females)	Index	0.445
Table 7	Main realized h. risk diagnosis (only females)	Fruit/Veg	0.441
Table 7	Main realized h. risk diagnosis (only females)	Walk 10	3.092
Table 7	Main realized h. risk diagnosis (only females)	Walk 30	2.914
Table 7	Main realized h. risk diagnosis (only females)	Smoke	0.070
Table 7	Main realized h. risk diagnosis (only females)	Nr Cigs	3.962
Table 7	Main realized h. risk diagnosis (only females)	Drink	0.088
Table 7	Main realized h. risk diagnosis (only females)	Heavy	1.034
Table 7	Main realized h. risk diagnosis (only females)	Days	0.476
Table 8	Main realized h. risk diagnosis (only males)	Index	0.419
Table 8	Main realized h. risk diagnosis (only males)	Fruit/Veg	0.410

Continued on next page

Table D.1 – *Continued from previous page*

Table	Table Description	Dependent Variable	MDE
Table 8	Main realized h. risk diagnosis (only males)	Walk 10	2.924
Table 8	Main realized h. risk diagnosis (only males)	Walk 30	2.812
Table 8	Main realized h. risk diagnosis (only males)	Smoke	0.066
Table 8	Main realized h. risk diagnosis (only males)	Nr Cigs	4.703
Table 8	Main realized h. risk diagnosis (only males)	Drink	0.067
Table 8	Main realized h. risk diagnosis (only males)	Heavy	1.179
Table 8	Main realized h. risk diagnosis (only males)	Days	0.481

Appendix E: Propensity Score Estimation Conditions

The propensity score is the probability an individual is treated given a set of selected observables. For proper application two lemmas must hold. First, the balancing property: observations with the same propensity score have the same distribution of observable covariates independently of treatment status. Second, the unconfoundedness assumption: the assignment to treatment is independent given the propensity score (Rosenbaum and Rubin, 1985). Finally, there must be common support between treated and control units.

To ensure that the balancing property is satisfied, after estimating the propensity score, the propensity score estimation program takes the full sample of treated and controls, sorts the individuals by their estimated propensity score, and divides them into bins such that within each bin the mean propensity score is not statistically different between the treated and controls groups. Similarly, the balancing property also requires that the mean of each covariate used in the estimation of the propensity score is balanced within each bin between treated and control groups. I choose the exact specification for the propensity score estimation such as to meet these requirements. Finally, I confirm common support by looking at the overlap in estimated propensity scores between treated and control units.

Main Analysis. The balancing property is satisfied using seven bins. It is clear from Figure E.1 that there is common support between treated and control units, as none of the treated units are marked as "off-support". Common support is further verified by Table E.1, which shows the descriptive statistics

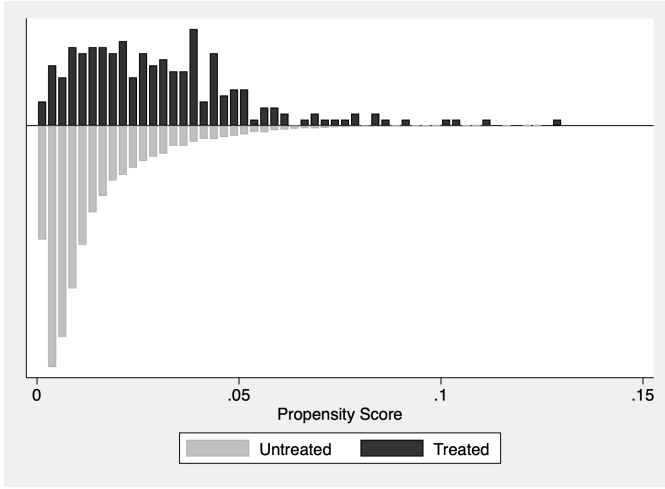

FIGURE E.1. Estimated Propensity Score Histogram (Full Sample)

of the estimated propensity score for the full sample, only the treated, and only the controls. It is important to note that both the minimum and the maximum for the treated group fall within the minimum and maximum of the control group. Therefore, common support for all treated units is further verified.

TABLE E.1. Descriptive Statistics of Estimated Propensity Score (Full Sample)

Sample	Count	Mean	SD	Min	Max
Full Sample	22,158	0.0160	0.0157	0.00055	0.1446
Only Treated	354	0.0316	0.0226	0.00114	0.1286
Only Controls	21,804	0.0157	0.0154	0.00055	0.1446

Realized Health Risk Diagnosis Only. The estimated propensity score for the "realized diagnosis only" case satisfies the balancing property and has common support, shown in Figure E.2, since once again there are no treated units marked as off-support. Table E.2, analogous to Table E.1, shows the descriptive statistics of the estimated propensity score for the three (sub)samples. Here as

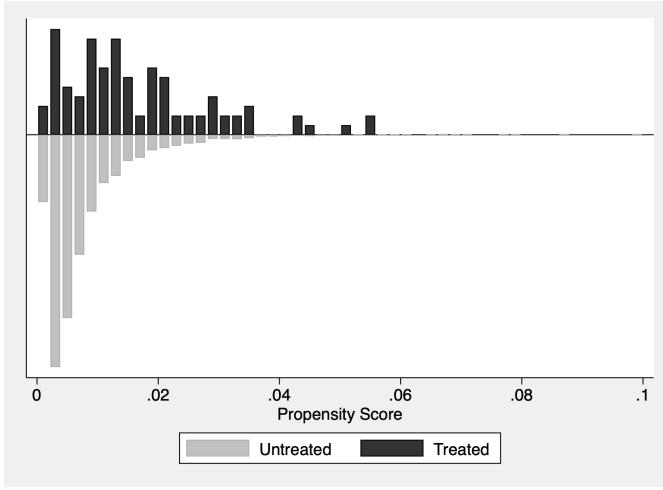

FIGURE E.2. Estimated Propensity Score Histogram (Realized Health Risk Diag. Only)

well, both the minimum and the maximum for the treated group fall within the minimum and maximum of the control group, thereby reinforcing that all treated units fall within the common support.

TABLE E.2. Descriptive Statistics of Estimated Propensity Score (Realized Health Risk Diagnosis Only)

Sample	Count	Mean	SD	Min	Max
Full Sample	17,155	0.0090	0.0089	0.00069	0.1084
Only Treated	154	0.0178	0.0143	0.00111	0.0795
Only Controls	17,001	0.0089	0.0088	0.00069	0.1084

Risk Factor Diagnosis Only. The estimated propensity score for the "risk factor diagnosis only" variable satisfies the balancing property; and it has common support for all observations, as shown in Figure E.3. Table E.3 shows the descriptive statistics of the estimated propensity score for the full, treated-only, and control-only samples. Again, note that both the minimum and the maximum for the treated group fall within those of the control group, which confirms that the treated units fall within the common support.

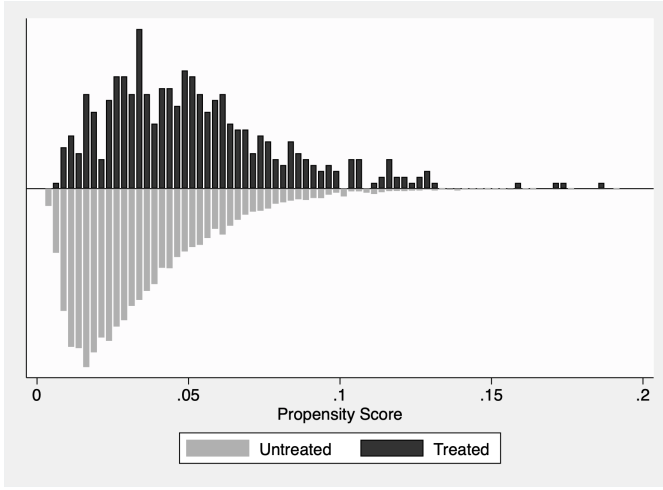

FIGURE E.3. Estimated Propensity Score Histogram (Risk Factor Diagnosis Only)

TABLE E.3. Descriptive Statistics of Estimated Propensity Score (Risk Factor D. Only)

Sample	Count	Mean	SD	Min	Max
Full Sample	17,625	0.0354	0.0249	0.00263	0.1967
Only Treated	624	0.0519	0.0297	0.00697	0.1869
Only Controls	17,001	0.0348	0.0244	0.00263	0.1967

Appendix F: Matching Strategy: Selection and Assessment

Due to the nature of matching, the propensity score is estimated and the matching performed *prior* to the analysis of the outcome variables. As a result, it is possible and encouraged to try several matching possibilities — such as nearest neighbour with or without replacement and with one or more controls, and radius or kernel with different bandwidths — and to run a range of balance diagnostics to assess which matching strategy leads to the best balance; this strategy is then used for the analysis of the outcomes (Garrido et al., 2014). The goal behind matching is to ensure that treated and control units are as similar as possible across observable covariates. Below are the selection and assessment criteria for both the main and secondary analyses.

F.1. Main Analysis

Selecting Matching Strategy with Best Balance. Table F.1 reports several possible matching strategies. In all cases, there are 232 treated individuals with a differing number of individuals used as controls depending on the matching strategy.[26] As shown in Table F.1 the kernel matching strategy with its selected bandwidth is the most balanced of the matches tried; the kernel matching with a bandwidth of 0.0075 is the best balanced match among the different kernel bandwidths tried. The kernel matching has the lowest mean and median percent standardized difference in covariates and the lowest Pseudo R^2. Furthermore, both the Rubin's B and Rubin's R fall within their desired cut-offs or ranges: below 25% and between [0.5,2], respectively. Therefore, I use this kernel matching strategy in the analysis. Below is a discussion of each summary measure of match quality and the characteristics of a good match. A comparison of the sensitivity of the findings to different matching strategies is provided in Appendix G.1.

TABLE F.1. Summary Measures of Match Quality for Original, Matched and Weighted Samples

Sample Type	Total Obs.	Total Treated	Total Controls	Pseudo R^2	Mean Standardized Difference (%)	Median Standardized Difference (%)	Rubin's B	Rubin's R
Original Sample	15,853	232	15,621	0.078	14.8	8.1	92.9*	0.92
NN 1:1 no replace.	464	232	232	0.033	5.7	4.7	42.9*	1.05
NN 1:1 with replace.	461	232	229	0.030	5.5	4.3	41.3*	1.04
Radius	15,853	232	15,621	0.013	4.3	3.2	26.5*	0.81
Kernel	**15,853**	**232**	**15,621**	**0.011**	**3.6**	**2.6**	**24.3**	**0.77**

Note: NN: Nearest Neighbour matching. Radius method has a caliper of 0.01. NN 1:1 with caliper omitted since match quality results are the same as the no caliper case. Kernel matching uses a bandwidth of 0.0075. If B>25% or R outside [0.5,2], marked with *.

Summary Measures of Match Quality. Table F.2 provides several summary measures of the overall balance of the variables used to estimate the propensity score and create the match, for both the unmatched and matched samples. The

[26] Pairs are created for those individuals who are not missing the index outcome variable. There are only 232 treated observations and not the 354 reported in Table E.1. This is because the "missing" individuals are missing one or more of the eight lifestyle index behaviour variables in either wave 2 or 5.

first column, the Pseudo-R^2, is the estimate from a probit of the propensity score equation. The closer the value is to zero the more the variables used to estimate the propensity score no longer have predictive power for the realized health risk (or risk factor) diagnosis, which implies better balance. Similarly, the second column shows the p-value for the likelihood ratio test that all covariates used for the estimation are jointly insignificant. Both these columns suggest that the match between treated and controls is quite balanced. The third and fourth columns show that the mean and median percent standardized difference between the treatment and controls groups have been reduced for a large extent by matching, respectively; the mean percent decreases from 14.8 to 3.6, the median from 8.1 to 2.6. A mean percent standardized difference of less than 10% is considered a good quality match, which is what I find. Finally, the last two columns are summary measures of matching quality suggested by Rubin (2001). Rubin's B is the absolute standardized difference of the means of the propensity score of the treated and control group. Rubin's R is the variance ratio of the treated and control groups' propensity score. Rubin (2001) specifies that groups are sufficiently balanced when the Rubin's B is less than 25%; similarly, the Rubin's R should be between 0.5 and 2. Both Rubin measures are within the desired range, indicating a good match. Summarizing, overall Table F.2 suggests the match is well balanced. A graphical interpretation of balance before and after matching at the individual covariate level is shown and discussed below.

TABLE F.2. Summary Measures of Match Quality

Sample	Pseudo R^2	Likelihood ratio test p-value	Mean Standardized Difference (%)	Median Standardized Difference (%)	Rubin's B	Rubin's R
Unmatched	0.078	0.000	14.8	8.1	92.9*	0.92
Matched	0.011	1.000	3.6	2.6	24.3	0.77

Note: * if B>25%, R outside [0.5; 2]

Match Quality of Individual Covariates. Figure F.1 shows graphically the percent standardized difference (bias) for the covariates used in the propensity

score estimation, both before and after matching. Overall, the figure shows that, except for ethnicity, the bias for all variables shown decreases with matching, often quite substantially (non-binary categorical variables are not shown for reasons of readability). In the case of ethnicity, the treated and untreated groups are not statistically different from each other in both the unmatched and matched cases, which means the increase is not of concern.

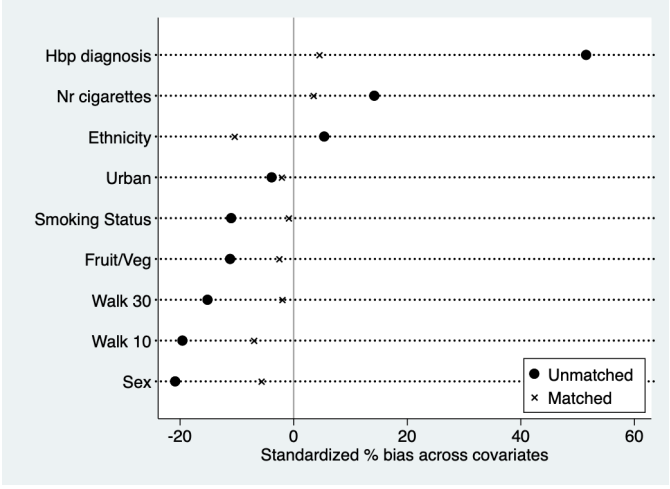

FIGURE F.1. Percent Standardized Difference (Bias) Across Propensity Score Estimation Covariates Before and After Matching

F.2. Secondary Analysis: Realized Health Risk Diagnosis Only

Selection Matching Strategy with Best Balance. Among the reported matching strategies, in Table F.3, the kernel matching strategy has the lowest mean and median percent standardized difference in covariates and the lowest Pseudo R^2. For the Rubin's B, although it does not fall within the desired 25% cut-off, the kernel strategy has the lowest Rubin's B value of all the matching strategies. Finally, the Rubin's R falls within the desired range of [0.5,2]. Although the kernel match is still not very well-balanced — as suggested by both the Rubin's B and the higher values of the other measures compared to both the main analysis match (shown previously) and the risk factor diagnosis only analysis match (shown subsequently) — the kernel strategy is still the best choice of the

available strategies and therefore I use it for this realized health risk diagnosis only analysis.

TABLE F.3. Summary Measures of Match Quality for Original, Matched and Weighted Samples (Realized Health Risk Diagnosis Only)

Sample Type	Total Obs.	Total Treated	Total Controls	Pseudo R^2	Mean Standardized Difference (%)	Median Standardized Difference (%)	Rubin's B	Rubin's R
Original Sample	12,339	94	12,245	0.063	13.7	8.5	84.9*	1.27
NN 1:1 no replace.	188	94	94	0.058	8.0	6.0	57.4*	1.14
NN 1:1 with replace.	187	94	93	0.059	8.0	6.1	57.9*	1.11
Radius	12,329	94	12,235	0.035	8.4	7.6	44.6*	0.76
Kernel	**12,323**	**94**	**12,229**	**0.025**	**5.7**	**4.8**	**37.1***	**0.65**

Note: NN: Nearest Neighbour matching. Radius method has a caliper of 0.01. NN 1:1 with caliper omitted since match quality results are the same as the no caliper case. Kernel matching uses a bandwidth of 0.00375. If B>25% or R outside [0.5,2], marked with *.

Summary Matching Quality Assessment. Table F.4 provides, for the realized health risk diagnosis only case, several summary measures of the overall balance of the variables used to estimate the propensity score and create the match, for both the unmatched and matched samples. The measures are the same as those described previously for the main analysis case. In this realized health risk diagnosis only case, the measures indicate that matching improves the balance between treated and controls. For example, the average percent standardized differences of the mean and median for the covariates from the propensity score estimation decreases: the mean from 13.7 to 5.7 and the median from 8.5 to 4.8. However, the Rubin's B measure suggests that although the matched sample is an improvement over the unmatched one, it is still not very well-balanced as the value (37.1%) falls above the 25% threshold. This less good match is likely in part attributable to the relatively small sample size of the treated for this realized health risk diagnosis only case compared to the other two cases. Nevertheless, the measures still suggest that the matched sample has more balance than the unmatched sample and therefore I use it for this "realized diagnosis only" analysis. A graphical interpretation of balance before and after matching at the individual covariate level is below.

Match Quality of Individual Covariates. Figure F.2 shows graphically the percent standardized difference (bias) for the covariates used in the propensity

TABLE F.4. Summary Measures of Match Quality (Realized Health Risk Diagnosis Only)

Sample	Pseudo R^2	Likelihood ratio test p-value	Mean Standardized Difference (%)	Median Standardized Difference (%)	Rubin's B	Rubin's R
Unmatched	0.063	0.000	13.7	8.5	84.9*	1.27
Matched	0.025	1.000	5.7	4.8	37.1*	0.65

Note: * if B>25%, R outside [0.5;2]

score estimation for the realized health risk diagnosis only, both before and after matching. In this case, the figure shows the matching is less successful in reducing the difference between the unmatched treated and unmatched control groups (i.e. the bias). For age and the interaction between sex and employment status matching reduces the bias quite drastically. However, for most other variables shown matching does not affect the bias much. Once again, non-binary categorical variables are not shown for readability reasons. However, none of the variables are statistically different between the treated and untreated groups after matching.

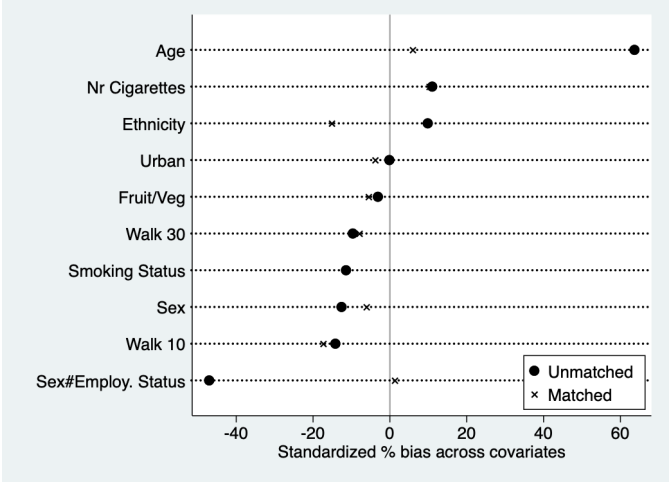

FIGURE F.2. Percent Standardized Difference (Bias) Across Propensity Score Estimation Covariates Before and After Matching (Realized Health Risk Diagnosis Only)

F.3. Secondary Analysis: Risk Factor Diagnosis Only

Selection Matching Strategy with Best Balance. Among the reported matching strategies in Table F.5 kernel matching has the lowest mean and median percent standardized difference in covariates and the lowest Pseudo R^2. Furthermore, both the Rubin's B and Rubin's R fall within their desired cut-offs or ranges, below 25% and between [0.5,2], respectively. Therefore, I use this kernel matching strategy in the analysis.

TABLE F.5. Summary Measures of Match Quality for Original, Matched and Weighted Samples (Risk Factor Diag.)

Sample Type	Total Obs.	Total Treated	Total Controls	Pseudo R^2	Mean Standardized Difference (%)	Median Standardized Difference (%)	Rubin's B	Rubin's R
Original Sample	12,701	456	12,245	0.052	12.0	6.3	71.2*	0.77
NN 1:1 no replace.	912	456	456	0.023	5.5	4.1	35.8*	1.50
NN 1:1 with replace.	895	456	439	0.022	5.3	4.5	34.9*	1.48
Radius	12,693	456	12,237	0.007	2.8	2.4	19.4	1.03
Kernel	**12,685**	**456**	**12,229**	**0.005**	**2.7**	**2.2**	**17.0**	**1.05**

Note: NN: Nearest Neighbour matching. Radius method has a caliper of 0.01. NN 1:1 with caliper omitted since match quality results are the same as the no caliper case. Kernel matching uses a bandwidth of 0.0075. If B>25% or R outside [0.5,2], marked with *.

Summary Match Quality Assessment. Table F.6 provides for the risk factor diagnosis case several summary measures of the overall balance of the variables used to estimate the propensity score and create the match, for both the unmatched and matched samples. The measures are the same as those described previously and all suggest a good match. Summarizing, overall Table F.6 suggests a well-balanced match. A graphical interpretation of balance before and after matching at the individual covariate level is below.

TABLE F.6. Summary Measures of Match Quality (Risk Factor Diagnosis Only)

Sample	Pseudo R^2	Likelihood ratio test p-value	Mean Standardized Difference (%)	Median Standardized Difference (%)	Rubin's B	Rubin's R
Unmatched	0.052	0.000	12.0	6.3	71.2*	0.77
Matched	0.005	1.000	2.7	2.2	17.0	1.05

Note: * if B>25%, R outside [0.5;2]

Match Quality of Individual Covariates. Figure F.3 shows graphically the percent standardized difference (bias) for the covariates used in the risk factor diagnosis only propensity score estimation, both before and after matching. Overall, the figure shows that except for Fruit/Veg the bias for all variables shown decreases with matching, often quite substantially. Once again, non-binary categorical variables are not shown for readability reasons. In the case of Fruit/Veg, the treated and untreated groups are not statistically different from each other in either the unmatched and matched cases; this slight increase is therefore of no concern.

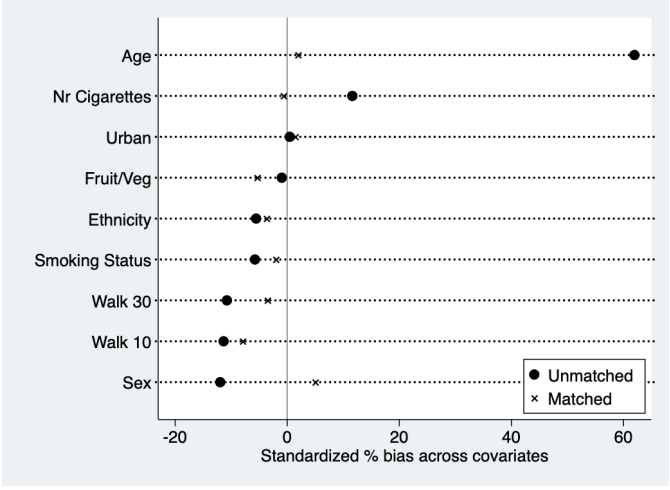

FIGURE F.3. Percent Standardized Difference (Bias) Across Propensity Score Estimation Covariates Before and After Matching (Risk Factor Diagnosis Only)

Appendix G: Matching and Weighting Strategies

Stratification. Stratification takes the full sample of treated and controls, sorts the individuals by estimated propensity score and then splits them into bins such that within each bin the mean propensity score is the same for the treated and controls groups. The analysis (the effect calculation) is performed for each bin with the assumption that within each bin individuals are similar.

Nearest Neighbour (NN) 1:1 Matching without Replacement. Nearest neighbour 1:1 matching can be implemented with or without replacement (a caliper to restrict the distance from which the nearest control is selected has no effect). In the case of no replacement, for each treated individual the control with the closest propensity score is selected, except where that control has already been assigned to another treated individual, in that case the next nearest control unit is selected.

NN 1:1 Matching with Replacement. Matching 1:1 with replacement is similar to matching 1:1 without replacement except that the control is returned to the pool each time, such that it can be used as a control for another treated individual. This means that for each treated individual the control with the closest propensity score is always selected.

NN 1:N Matching, N>1. Matching more than one control to the treated unit (1:N) is only possible with replacement. It is similar to matching 1:1 with replacement except that instead of matching a treated individual with its single nearest control, the treated individual is matched to its N nearest controls.

Kernel Matching. Kernel matching uses all observations within the common support, taking a weighted average of the inverse of the (propensity score) distance between each treated and control unit. In other words, control units nearer to the treated unit receive more weight.

Radius Matching. In radius matching for each treated unit within a specified radius all controls are used and assigned equal weight regardless of their (propensity score) distance to the treated unit.

G.1. Comparing Matching Strategies and Standard Errors

Table G.1 compares several matching strategies as well as the different options for standard errors. The main purpose of this table is to show that neither the matching strategy nor the choice for standard errors have a significant impact on the final findings. None of the estimates are statistically significantly different from one another at the 5% level. In each column the estimate reported

is the impact of the main analysis realized health risk diagnosis on the index, first differenced.

Column 1 presents the regression analysis matching using stratification and using regular robust standard errors. Column 2 presents the same as column 1 except that it clusters standard errors at the household level. From these two columns it is clear that such household-level clustering of standard errors has no effect on changing the reported standard errors.

When considering the matching strategy directly, rather than including matching as part of a regression analysis, there are several aspects to decide upon. First, the exact command to execute the matching (*psmatch*, *nnmatch* or *psmatch2*). Second, whether to provide the matching program/command with the propensity score directly (pscore) or whether to provide it with the covariates used to calculate the propensity score (cov) and allow the program to calculate its own propensity score that it will subsequently use to match on.[27] Third, decide on what kind of standard errors to use, both in terms of what the default standard errors are for each matching program/command but also how the standard errors can be adjusted to account for the number of matches used (if using nearest neighbour, N>1) and/or to use bootstrap standard errors, where applicable. For *psmatch* and *nnmatch* the default is robust Abadie-Imbens (A-I) standard errors, which takes into account that the propensity score is estimated rather than known.

In columns 3-5 I match 1:1 using A-I standard errors. Column 3 uses the *psmatch* command and provides the matching program with the previously calculated propensity score directly. Column 4 uses the *psmatch* command but now I provide the program with the covariates to calculate its own propensity score prior to matching. Column 5 uses the same approach as column 4, but now

[27] One of the main advantages of propensity score estimation using the *pscore* matching program, over other programs, is as part of the estimation this program checks and requires that the estimated propensity score, as well as the covariates used in such an estimation, are balanced among propensity score bins of treated and control groups; something that I would otherwise have to check manually.

using the *nnmatch* command. The *nnmatch* command only allows the provision of the covariates to calculate its own propensity score prior to matching; it does not allow the input of a previously calculated propensity score.

Finally, columns 6 and 7 use the *psmatch2* command, which allows the use of the kernel matching strategy. Column 6 shows the effects using the default standard errors while column 7 shows the same effects but using bootstrap standard errors.

Summarizing, from this table neither the exact choice of matching strategy nor the standard error adjustment has any significant effect on the estimates or standard errors displayed. In general, the effect has a point estimate of approximately 0.23. The chosen specification is *psmatch2* with previously calculated propensity score, bandwidth of 0.0075 and bootstrap standard errors, corresponding to column 7. I already discussed the choice of optimal matching strategy in Appendix F.1.

TABLE G.1. Realized Health Risk Diagnosis on Index — Several Matching Strategies and Standard Errors

	(1) OLS stratification - no hh cl se	(2) OLS stratification - hh cl se	(3) psmatch (pscore) NN 1:1 - A-I se	(4) psmatch (cov) NN 1:1 - A-I se	(5) nnmatch (cov) NN 1:1 - A-I se	(6) psmatch2 (pscore) kernel 0.0075 bandwidth se	(7) psmatch2 (pscore) kernel 0.0075 bandwidth bootstrap se (1000reps)
Realized Diag.	0.230**	0.230**					
	(0.111)	(0.111)					
ATT			0.231**	0.221	0.172	0.227***	0.227**
			(0.111)	(0.142)	(0.141)	(0.111)	(0.111)
Constant	0.028	0.028					
	(0.027)	(0.028)					
Observations	15,853	15,853	15,853	15,853	15,853	15,853	15,853

Standard errors in parentheses. See column headers for type of standard error (se). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

References

American Heart Association (2014). "Alcohol and Heart Health." URL https://www.heart.org/en/healthy-living/healthy-eating/eat-smart/nutrition-basics/alcohol-and-heart-health. Article date: 14 Aug, 2014.

American Heart Association (2017). "What is Cardiovascular Disease?" URL https://www.heart.org/en/health-topics/consumer-healthcare/what-is-cardiovascular-disease. Article date: 31 May, 2017.

Atkin, David (2016). "The caloric costs of culture: Evidence from Indian migrants." *American Economic Review*, 106, 1144–1181.

Bartiaux, Françoise (2008). "Does environmental information overcome practice compartmentalisation and change consumers' behaviours?" *Journal of Cleaner Production*, 16, 1170–1180.

Belot, Michèle, Jonathan James, and Jonathan Spiteri (2020). "Facilitating healthy dietary habits: An experiment with a low income population." *European Economic Review*, 129, 103550.

Benhabib, Jess, Alberto Bisin, and Andrew Schotter (2010). "Present-bias, quasi-hyperbolic discounting, and fixed costs." *Games and Economic Behavior*, 69(2), 205–223.

Bhalotra, Sonia, Adeline Delavande, Paul Fisher, and Jonathan James (2020). "The Impact of a Personalised Blood Pressure Warning on Health Outcomes and Behaviours." *Institute of Social and Economic Research (ISER) Working Paper Series; No. 2020-2*.

Bloom, D.E., E.T. Cafiero, S. Abrahams-Gessel, L.R. Bloom, S. Fathima, A.B. Feigl, T. Gaziano, M. Mowafi, A. Pandya, K. Prettner, L. Rosenberg, B. Seligman, A.Z. Stein, and C. Weinstein (2011). "The Global Economic Burden of Non-communicable Diseases: A report by the World Economic Forum and the Harvard School of Public Health." Geneva, Switzerland: World Economic Forum.

Brownell, Kelly D. and Lisa R. Cohen (1995). "Adherence to dietary regimens 2: Components of effective interventions." *Behavioral Medicine*, 20, 155–164.

Brush Jr, John E., Harlan M. Krumholz, Erich J. Greene, and Rachel P. Dreyer (2020). "Sex differences in symptom phenotypes among patients with acute myocardial infarction." *Circulation: Cardiovascular Quality and Outcomes*, 13(2), e005948.

Buttar, Harpal S., Timao Li, and Nivedita Ravi (2005). "Prevention of cardiovascular diseases: Role of exercise, dietary interventions, obesity and smoking cessation." *Experimental & Clinical Cardiology*, 10, 229–249.

Carrera, Mariana, Syeda A. Hasan, and Silvia Prina (2020). "Do health risk assessments change eating habits at the workplace?" *Journal of Economic Behaviour and Organization*, 172, 236–246.

Chong, Shanley, Ding Ding, Roy Byun, Elizabeth Comino, Adrian Bauman, and Bin Jalaludin (2017). "Lifestyle changes after a diagnosis of type 2 diabetes." *Diabetes Spectrum*, 30, 43–50.

Chou, Chih-Hsuan, Chueh-Lung Hwang, and Ying-Tai Wu (2012). "Effect of exercise on physical function, daily living activities, and quality of life in the frail older adults: A meta-analysis." *Archives of Physical Medicine and Rehabilitation*, 93, 237–244.

Condon, Carol and Geraldine McCarthy (2006). "Lifestyle changes following acute myocardial infarction: Patients perspectives." *European Journal of Cardiovascular Nursing*, 5, 37–44.

Craig, Cora L., Catrine Tudor-Locke, and Adrian Bauman (2006). "Twelve-month effects of Canada on the Move: A population-wide campaign to promote pedometer use and walking." *Health Education Research*, 22, 406–413.

Crombie, Iain K., Linda Irvine, Brian Williams, Alison R McGinnis, Peter W. Slane, Elizabeth M. Alder, and Marion E. T. McMurdo (2004). "Why older people do not

participate in leisure time physical activity: A survey of activity levels, beliefs and deterrents." *Age and Ageing*, 33, 287–292.

Durkin, Sarah, Emily Brennan, and Melanie Wakefield (2012). "Mass media campaigns to promote smoking cessation among adults: An integrative review." *Tobacco Control*, 21, 127–138.

Edwards, Ryan D. (2018). "If My Blood Pressure Is High, Do I Take It to Heart? Behavioral Effects of Biomarker Collection in the Health and Retirement Study." *Demography*, 55, 403–434.

Esselstyn, Caldwell B., Jr., Stephen G. Ellis, Sharon V. Medendorp, and Timothy D. Crowe (1995). "A strategy to arrest and reverse coronary artery disease: A 5-year longitudinal study of a single physician's practice." *Journal of Family Practice*, 41, 560–568.

Fadlon, Itzik and Torben Heien Nielsen (2019). "Family health behaviors." *American Economic Review*, 109, 3162–3191.

Fassier, Philippine, Laurent Zelek, Lucie Lécuyer, Patrick Bachmann, Marina Touillaud, Nathalie Druesne-Pecollo, Pilar Galan, Patrice Cohen, Hélène Hoarau, Paule Latino-Martel, Emmanuelle Kesse-Guyot, Julia Baudry, Serge Hercberg, Mélanie Deschasaux, and Mathilde Touvier (2017). "Modifications in dietary and alcohol intakes between before and after cancer diagnosis: Results from the prospective population-based NutriNet-Santé cohort." *International Journal of Cancer*, 141, 457–470.

Fitgerald, James T., Steven P. Singleton, Anne Victoria Neale, Ananda S. Prasad, and Joseph W. Hess (1994). "Activity levels, fitness status, exercise knowledge, and exercise beliefs among healthy, older African American and white women." *Journal of Aging and Health*, 6, 296–313.

Flay, Brian R. (1987). "Mass media and smoking cessation: A critical review." *American Journal of Public Health*, 77, 153–160.

Fleming, Michael F., Linda Baier Manwell, Kristen Lawton Barry, Wendy Adams, and Ellyn A. Stauffacher (1999). "Brief physician advice for alcohol problems in older adults: A randomized community-based trial." *Journal of Family Practice*, 48, 378–386.

Frederick, Shane, George Loewenstein, and Ted O'donoghue (2002). "Time discounting and time preference: A critical review." *Journal of Economic Literature*, 40(2), 351–401.

Garrido, Melissa M., Amy S. Kelley, Julia Paris, Katherine Roza, Diane E. Meier, R. Sean Morrison, and Melissa D. Aldridge (2014). "Methods for constructing and assessing propensity scores." *Health Services Research*, 49, 1701–1720.

Heikkinen, Hanne, Kristiina Patja, and Piia Jallinoja (2010). "Smokers' accounts on the health risks of smoking: Why is smoking not dangerous for me?" *Social Science & Medicine*, 71, 877–883.

Hurd, Michael D and Kathleen McGarry (2002). "The predictive validity of subjective probabilities of survival." *The Economic Journal*, 112(482), 966–985.

Hut, Stefan and Emily Oster (2022). "Changes in household diet: Determinants and predictability." *Journal of Public Economics*, 208.

Kelly, Michael P and Mary Barker (2016). "Why is changing health-related behaviour so difficult?" *Public Health*, 136, 109–116.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz (2007). "Experimental analysis of neighborhood effects." *Econometrica*, 75, 83–119.

Laibson, David (1997). "Golden eggs and hyperbolic discounting." *The Quarterly Journal of Economics*, 112(2), 443–478.

Lanza, Elaine, Arthur Schatzkin, Cassandra Daston, Don Corle, Laurence Freedman, Rachel Ballard-Barbash, Bette Caan, Peter Lance, James Marshall, Frank Iber, Moshe Shike, Joel Weissfeld, Martha Slattery, Electra Paskett, Donna Mateski, Paul Albert, and the PPT Study Group (2001). "Implementation of a 4-y, high-fiber, high-fruit-and-vegetable, low-fat dietary intervention: Results of dietary changes in the Polyp Prevention Trial." *American Journal of Clinical Nutrition*, 74, 387–401.

Larimer, Mary E. and Jessica M. Cronce (2002). "Identification, prevention and treatment: A review of individual-focused strategies to reduce problematic alcohol consumption by college students." *Journal of Studies on Alcohol. Supplement*, s14, 148–163.

Logie-MacIver, Liz, Maria Piacentini, and Douglas Eadie (2012). "Using qualitative methodologies to understand behaviour change." *Qualitative Market Research: An International Journal*, 15, 70–86.

Mathis, Klaus and Ariel David Steffen (2015). "From rational choice to behavioural economics." In *European Perspectives on Behavioural Law and Economics*, edited by Klaus Mathis, pp. 31–48. Springer.

McMacken, Michelle and Sapana Shah (2017). "A plant-based diet for the prevention and treatment of type 2 diabetes." *Journal of Geriatric Cardiology*, 14, 342–354.

O'Donoghue, Ted and Matthew Rabin (1999). "Doing it now or later." *American Economic Review*, 89(1), 103–124.

O'Donoghue, Ted and Matthew Rabin (2015). "Present bias: Lessons learned and to be learned." *American Economic Review*, 105(5), 273–79.

Ornish, Dean, Larry W. Scherwitz, James H. Billings, K. Lance Gould, Terri A. Merritt, Stephen Sparler, William T. Armstrong, Thomas A. Ports, Richard L. Kirkeeide, Charissa Hogeboom, and Richard J. Brand (1998). "Intensive Lifestyle Changes for Reversal of Coronary Heart Disease." *Journal of the American Medical Association*, 280, 2001–2007.

Oster, Emily (2018). "Diabetes and diet: Purchasing behavior change in response to health information." *American Economic Journal: Applied Economics*, 10, 308–348.

Rizza, Wanda, Nicola Veronese, and Luigi Fontana (2014). "What are the roles of calorie restriction and diet quality in promoting healthy longevity?" *Ageing Research Reviews*, 13, 38–45.

Rosenbaum, Paul R. and Donald B. Rubin (1985). "Constructing a control group using multivariate matched sampling methods that incorporate the propensity score." *The American Statistician*, 39, 33–38.

Rubin, Donald B. (2001). "Using propensity scores to help design observational studies: Application to the tobacco litigation." *Health Services and Outcomes Research*

Methodology, 2(3-4), 169–188.

Sagner, Michael, D Katz, Gary Egger, Liana Lianov, KH Schulz, Marc Braman, Behrooz Behbod, Edward Phillips, Wayne Dysinger, and Dean Ornish (2014). "Lifestyle medicine potential for reversing a world of chronic disease epidemics: from cell to community." *International Journal of Clinical Practice*, 68, 1289–1292.

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

Simon, Herbert A (1955). "A behavioral model of rational choice." *The Quarterly Journal of Economics*, 69(1), 99–118.

Siteman Cancer Center (2019). "Your Disease Risk." URL https://siteman.wustl.edu/prevention/ydr/. Accessed on: 25 March, 2019.

Stockwell, Tim, Jinhui Zhao, Sapna Panwar, Audra Roemer, Timothy Naimi, and Tanya Chikritzhs (2016). "Do "moderate" drinkers have reduced mortality risk? A systematic review and meta-analysis of alcohol consumption and all-cause mortality." *Journal of Studies on Alcohol and Drugs*, 77, 185–198.

Strecher, Victor J., Matthew Kreuter, Dirk-Jan Den Boer, Sarah Kobrin, Harm J. Hospers, and Celette S. Skinner (1994). "The effects of computer-tailored smoking cessation messages in family practice settings." *Journal of Family Practice*, 39, 262–270.

Tversky, Amos and Daniel Kahneman (1985). "The framing of decisions and the psychology of choice." In *Behavioral decision making*, pp. 25–41. Springer.

Vlassoff, Carol (2007). "Gender differences in determinants and consequences of health and illness." *Journal of health, population, and nutrition*, 25(1), 47.

Willett, Walter C., Jeffrey P. Koplan, Rachel Nugent, Courtenay Dusenbury, Pekka Puska, and Thomas A. Gaziano (2006). "Prevention of chronic disease by means of diet and lifestyle changes." In *Disease Control Priorities in Developing Countries. 2nd edition.* The International Bank for Reconstruction and Development/The World Bank.

Wood, Wendy and David T Neal (2016). "Healthy through habit: Interventions for initiating & maintaining health behavior change." *Behavioral Science & Policy*, 2(1), 71–83.

World Health Organization (2018). "The top 10 causes of death." URL https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death. Article date: 24 May, 2018.

Worsley, Anthony (2002). "Nutrition knowledge and food consumption: Can nutrition knowledge change food behaviour." *Asia Pacific Journal of Clinical Nutrition*, 11(S6), S579–S585.

Young, Deborah Rohm, William L Haskell, C Barr Taylor, and Stephen P Fortmann (1996). "Effect of community health education on physical activity knowledge, attitudes, and behavior: The Stanford Five-City Project." *American Journal of Epidemiology*, 144, 264–274.

Zhao, Meng, Yoshifumi Konishi, and Paul Glewwe (2013). "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*, 32, 367–385.