Estimating the eect of retirement on mental health via panel discontinuity designs

Fe, Eduardo and Hollingsworth, Bruce

2 March 2012

Estimating the effect of retirement on mental health via panel discontinuity designs

Eduardo Fé*†
University of Manchester

Bruce Hollingsworth‡
Lancaster University

March 2, 2012

Abstract

This article explores the potential effects of retirement on mental health and health care utilisation isolating sources of potential heterogeneity in treatment effect. To estimate the effects of retirement, we devise a new identifying strategy that combines kink and regression discontinuity designs with panel data methods. Our method is then applied to the British Household Panel Survey, a rich representative longitudinal survey. It is found that retirement has a small impact on primary care use, but overall has little effect on mental health.

Key Words: *Discontinuity design, weak identification, retirement, health.*

JEL Classification: *C21, C30, C90, J26, I18.*

*The authors thank the many helpful comments from James Banks, Thomas Mason and seminar participants at the Centre for Health Economics at Monash University and the joint HESG-CES at Aix-en-Provence

†Health Economics. Room 4.305, Jean McFarlane Building. University of Manchester. Manchester, M13 9PL, UK. email:e.fe-rodriguez@manchester.ac.uk.

‡Division of Health Research, University of Lancaster, Bowland Tower East, Lancaster LA1 4YT. email: b.hollingsworth@lancaster.ac.uk

1 Introduction.

The causal effect that retirement might have on health and health care is a topic of considerable interest, not only for economists, but also for legislators, due to the potential consequences on public finance and the welfare system. Before the proliferation of debates and policies affecting the delay of the default retirement age, there has been a trend towards early retirement, especially in OECD countries. Several of the key explaining such a trend have been identified, among which special prominence has been given to the implicit disincentives faced by workers concerning retirement age (Krueger and Pischke, 1992, Blundell et al., 2002, Gruber and Wise, 2002, 2004), the disadvantage faced by older workers during industrial restructuring and periods of technological change (Banks and Smith, 2006), as well as the impact of an ageing population. The trend has been exacerbated by a rise in the age at which people leave school (Banks and Smith, 2006) and its ultimate consequence has been a rise in the ratio of retirees per person of working age. The implications are well known: in pay-as-you-go systems it can compromise the solvency of pensions and welfare funds, while in the context of fully funded systems, a too high proportion of retirees would affect the amount of savings available for investment and, therefore, economic growth.

To counter balance the negative effects of retirement dynamics, policy makers in developed countries are now debating or implementing increases in the retirement age and the age at which people are eligible for pensions. Most policies that have been put forward contemplate increases of one to two years in the default retirement age, which constitutes a 2.5 % - 5% increase in a working life of 40 years. However, it has been argued that retirement might affect health, in which case the valuation of policies that prolong retirement should account for such an effect. In the presence of health effects, a longer retirement horizon, by postponing changes in health outcomes, could affect the utilization of health services by older adults (conditional on life expectancy), and this would impact on any projected variation in health care expenditures.

When it comes to unveiling any causal relations between retirement and health, economic theory can often be inconclusive as pointed out by Dhaval et al. (2008). These authors outline that, in accordance with the Grossman model (Grossman, 1972), if the marginal value of time increases after retirement, individuals will increase their expenditure in preventive health care because of the consumption value of health but, simultaneously, this effect might be mitigated by the fact that the cost of visiting a medical practice or health provider increases. Therefore, there is a need for empirical studies to articulate the potential relationship between retirement and health.

In practice, selection bias complicates the identification of the effect of retirement on health and practitioners have addressed this problem using a variety of instrumental variable strategies. However existing results tend to be contradictory, even after taking into account the potential endogeneity of retirement. Seminal work by Charles (2004) concludes that the "... *direct effect of retirement on well-being is positive once the fact that retirement and well being are simultaneously determined is accounted for...*". Dhaval et al. (2008) finding that retirement leads to a 6-9% decrease in mental health, and a 5-6% increase in illness conditions. Coe and Lindeboom (2008) conclude that there are no negative health effects of retirement. Neuman (2008) finds that there is strong evidence dismissing the idea that retirement harms health, while Johnston and Lee (2009) conclude that retirement improves individuals' sense of well-being and mental health, but not necessarily physical health. In this article we present further evidence regarding the causal effect of retirement on health introducing a number of methodological novelties.

The focus of this article is mental health and utilisation of health services. The focus on mental health is justified because, as suggested by medical and sociological research, mental health is potentially the most vulnerable dimension of a person's health upon retirement. Retirement may imply a loss for the employee, but it may also signify a relief from strenuous conditions. Either effect is likely to manifest in a change mental fitness. Our study tries to take into account that retirement is likely to have heterogeneous effects. Previous research by Gruber and Wise (2002),Gruber and Wise (2004) and Banks and Smith (2006) suggests that there are different experiences of retirement, and in line with these results we try to isolate different sources of heterogeneity. In particular we pay special attention to education, occupation and job satisfaction.

Finally, our identification strategy is also new. It is well know that, in the United Kingdom, the conditional distribution of retirement on age exhibits a discontinuity at the, recently changed, default retirement age (DRA) of 65 (for men). For instance, Blundell et al. (2002), Smith (2006), Banks and Smith (2006) report 20%-25% jumps in the proportion of retirees at that age. Johnston and Lee (2009) exploit this feature[1] using regression discontinuity (Thistlethwaite and Campbell, 1960, van der Klaauw, 1996, Hahn et al., 1999), using an indicator of exceeding the default retirement age as an instrument. We build upon this approach, but we introduce two methodological innovations. Firstly, we note that the distribution of retirees exhibits a second feature, namely

[1] It is convenient to remark that Johnston and Lee (2009) use the Health Survey for England, a repeated cross-section, and the reported jump in their work greatly exceeds the 20/25% gap typically reported in most studies.

there is a kink at the default retirement age. The rate of retirees grows fast before the default retirement age, but this rate slows down considerably afterwards. Following recent work by Card et al. (2009) and Dong (2011) we exploit this discontinuity in the density function of retirement in order to produce an additional set of instruments, which allows us to add identification power to our estimator, as well as circumvent potential weak instrumental variable (IV) problems. Secondly, when panel data are available, it is a wide-spread practice to take the average cluster approach described in Lemieux and Milligan (2008), Battistin et al. (2009), or Dong (2011). However, as pointed out recently by Petterson-Lidbom (forthcoming), one can exploit the two stage least square interpretation of discontinuity designs (Angrist and Pischke, 2008) and combine the discontinuity approach with standard I.V. panel methods. This has the advantage that comparisons are drawn between the same individual at consecutive points in time (right before and right after the DRA), making it less contentious to use the main assumptions underpinning nonparametric discontinuity techniques, namely comparability of distributions around the threshold and continuity of potential outcomes. One of the important implications of this is that the choice of bandwidth, though important, becomes a less critical aspect of this work.

The structure of the article is as follows. Section 2 revisits the relationship between retirement and health, explaining the challenges faced when estimating causal effects. We outline some of the mechanisms through which retirement can affect health and comment on important sources of heterogeneity in outcomes. Section 3 introduces the econometric methodology, providing an overview of existing discontinuity designs. Section 4 describes the data and present the main results, while section 5 concludes.

2 Health and Retirement.

Labour supply outcomes and health status are simultaneously determined. Health is a form of human capital valued by employer and employee alike. People invest in health to alter the depreciation of health capital, but this requires time and income. Employees' preferences between work and leisure may change following variations in health status while, at the same time, health status can affect the time horizon over which labour decisions are taken. Furthermore, labour market activity may affect health through stress, working conditions, boredom or even complacency... This intricate set of relationships greatly complicates the estimation of the causal effect of health on labour supply (and vice versa).

Further complications are faced when studying the retirement decisions of older workers, not only because the correlation between health and employment status intensifies around retirement (with health becoming a major reason to retire, as has been pointed out by Disney et al., 2006), but also due to the nature of the data normally available (Stern, 1989, Bound, 1991). Survey data often only contains self-reported indicators of health, which are imperfect representations of actual health status[2]. These indicators are also subject to measurement error because a person's perceptions of his own health status is often not directly comparable with those of other individuals (see, Bound, 1991 or Lindeboom and Kerkhofs, 2002). Furthermore, respondents may rationalise their retirement decisions, given that health is one of the few reasons why a person of working age might be out of work and early retirement benefits might be available only for those deemed incapable of work. When the emphasis of the research is, as in this article, the effect of retirement on health, comparability of outcomes and the relevance of measures of health remain an issue in so far as one needs to interpret what the dependent variable is actually capturing. However, retirement status is normally well approximated by survey data, conditional on the researchers' definition of retirement[3].

Retirement can affect health through a number of mechanisms. A more well known argument suggests that, upon retirement people's consumption patterns might be affected if retirement implies an unexpected negative shock in a person's wealth (thus violating the conditions of life-cycle models that support the idea of smooth consumption pathways over a life time). In that case, individuals might lower their intake of food or, perhaps more likely, they might change the quality of their diets so that the intake of, for example, saturated fats is increased. For example, ready meals, diary and red meat are more readily available, and generally cheaper, than whole grain food, vegetables or fish. However the former are worse nutritional packages than the latter, and can have dramatic short term effects on health markers such as cholesterol levels -see Hu et al. (1997) or Willett (1998). The effect of retirement on consumption has been the object of a wealth of research, but evidence is mixed (see, for instance, Banks et al., 1998, Smith, 2004, Smith, 2006, Hurst, 2008, Haider and Stephens, 2007, Battistin et al., 2009). In terms of food preparation and quality of food intake Aguiar and Hurst (2005) find that retirement is innocuous, although some evidence exists that contradicts this as well (see Brzozowski and Lu, 2010).

[2]Even when objective measures are available, one must ensure that these are actually related to labour supply

[3]When the emphasis of the research is the causal effect of health on retirement, the above problems are particularly relevant. Researchers have devised different strategies to tackle some of the problems, including instrumental variable approaches (as in Stern, 1989, Lindeboom and Kerkhofs, 2002, Smith, 2006, Disney et al., 2006) and simultaneous equation models (Sickles and Taubman, 1986, Sickles and Yazbeck, 1998, Cai, 2010, Bound et al., 2010).

A second argument, which is the first focus of this article, emphasises the impact of retirement on mental health. Medical and sociological research has argued that mental health is the most vulnerable dimension of a person's health upon retirement. The so called *role theory* would play a major role to explain this line of causality. As noted in Kim and Moen, 2002, role theory conjectures that people who have retired from their careers are vulnerable to feeling a role loss, which could lead to psychological distress. Clearly, the contrary could be true: retirement from demanding roles might lead to an improvement of mental well-being. Either way, mental health scales would become the relevant variables when studying the potential health-impact of retirement. Any other variable that might indicate a deterioration (or improvement) of mental health would be useful to evaluate this theory. For instance, recent medical research suggests that high blood pressure and migraine are highly correlated with poor mental health (Ahn and Ashizawa, 2010, Stam et al., 2010). Some of these measures have the advantage of being relatively more objective than self-reported mental health -in particular, medical examination is normally required prior to knowing if one has high blood pressure.

Whatever mechanism might explain any causality of retirement on mental health, economists and policy makers are ultimately interested in the effects that changes in health status can have on the demand for health care, and therefore the question needs to be asked concerning whether individuals' utilisation of health care changes upon retirement. Clearly, if retirement affects any dimension of health, survey data would reflect a variation in patterns of utilisation. However, utilisation and need for health care need not go hand-in-hand, especially when considering primary care (to which individuals normally self-refer). Individuals might have incentives or contractual advantages during their working lives that promote absenteeism, part of which can be rationalised through consumption of medical care. Statutory or occupational sick leave and sick pay can create these type of incentives, especially when the monitoring of absenteeism in the firm is not strict. Another reason can be found in the Shapiro and Stiglitz model of efficiency wages (Shapiro and Stiglitz, 1984). These authors propose a model by which efficiency wages[4] exceed the pure competition market rate. The excess payment increases with the employees' cost of effort, the ease of finding a job if fired and the rate at which worker's discount future utility, and decreases as the probability of detecting shirking in the firm increases. Even in firms tightly control the health and health care of their employees, once the decision of retirement has been taken the consequences of getting caught shirking might become mild or disappear altogether, in which case shirking might increase just before retirement, to drop again after leaving the work force. Therefore,

[4]Firms pay efficiency wages when the unit wage elasticity of employees' effort equals one.

the second focus of this article will be whether utilisation of health care changes upon retirement.

2.1 Heterogeneous effects.

As pointed out in Blundell et al. (2002), Gruber and Wise (2002, 2004) and Banks and Smith (2006), there are very different experiences of retirement, and different groups of the population are likely to behave differently when leaving paid employment permanently. The defining features of each of these groups are elusive. This article explores the effect of three potential sources of heterogeneity.

In accordance with human capital theories, a person's schooling decisions maximise the present value of lifetime earnings. The higher the rate of discount, the less likely a worker will invest in education, as they discount heavily the receipt of future income generated by further education. Fuchs (1982) suggests that differences in time preferences are established early in life and remain relatively stable, which suggests that these discount rates will still be high upon retirement. In accordance with Grossman (1972), health promoting activities have a consumption and an investment dimension, the later due in part to the ability of health to generate future income by enabling a person to work. Individuals about to retire might postpone health care utilization if that might incur foregone income or opportunity costs. However, this effect would be more evident among the population of people that discount income at a higher rate. In practice, discount rates are unobservable but in view of the relationship between discounting and educational attainment, people with similar level of schooling will be relatively homogeneous in terms of discount rates. Therefore, level of schooling can be used to identify how individuals trade health for income around retirement.

Education also determines, to a certain degree, people's skill mix and later occupation in life. A number of recent articles (see Autor et al., 2003, Acemoglu and Autor, 2011 and references therein) have noted that occupational mix can explain evolution of wage differentials via, for example, the introduction of new technologies and off-shoring, both of which can depress returns to certain tasks while boosting returns to others. For instance, Firpo et al. (2011) take the example of debugging and updating software. Consider a country with a group of firms producing software. The firms can subcontract debugging and updating its software overseas, where it can be completed promptly and cheaply. This frees time for the firms' own programmers to concentrate on developing new applications, thus boosting their productivity. However, off-shoring will depress returns to debugging, and increase returns for developers in the country were the firm is located. This channel can further induce different retirement patterns among occupations, an argument that is implicit in (Banks and Smith, 2006), when they mention the disadvantage of older workers during industrial restructuring and technological change as a potential determinant of retirement. People are endowed with heterogeneous mixes of skills, but the argument could be made that people self-select into occupations as to maximise their returns given their knowledge about their own mix of skills. This would result in specific occupations having workers with similar social, emotional or even physical abilities, so that common patterns of reaction to retirement might arise in such occupations[5]. Therefore, this article explores the causal relationship of retirement on health by occupational groups.

The last source of heterogeneity that we consider in this article is job satisfaction. Bad contractual arrangements or a destructive work environment are likely to affect, not only productivity, but also staff moral and sense of well-being at work so that retirement can relieve workers from an activity that could be seen as a burden. Alternatively, people might enjoy a healthy, enriching working environment; their job satisfaction might be high and they might develop an attachment for their posts, in which case retirement might lead to a sense of loss of welfare with consequences for individual's moral. If any of these pathways is dominant, one would observe that retirement alters mental health, and so it becomes important to consider job satisfaction prior to retirement when studying the relationship between retirement and health outcomes.

3 Estimation Methods: Discontinuity Designs

This analysis uses the British Household Panel Survey (BHPS hereafter), a representative longitudinal survey that contains employment and retirement histories for each surveyed person as well as other information at individual and household level. We consider the first 16 waves of the data set, corresponding to the period 1991 to 2006, and focus on the sub-population of male individuals. In the analysis, a person was taken as a retiree if (*i*) that was his self-reported job market status, (*ii*) he reported not to have undertaken any paid work during the two weeks prior to the interview, and (*iii*) he did not re-enter the job market after retiring, regardless of whether their subsequent job status is employed or unemployed.

The identifying strategy followed here exploits that, during the period under consideration, there was a default state pension age (for men) in the U.K. of 65, which

[5]Of course this argument presupposes that individuals have a good understanding of their own abilities, which is a contentious assumption -though perhaps empirically testable.

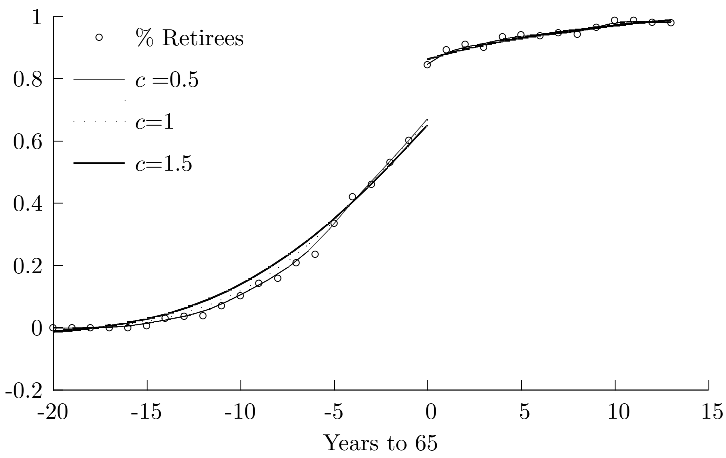

Figure 1: Discontinuities in the proportion of male retirees. The abscissa is $Age - 65$. Local linear estimator with bandwidth $h = c\sigma_x n^{-0.2}$, for $c = 0.5, 1$ and 1.5.

induces a number of peculiarities in the distribution of retirees. The estimate of this distribution is plotted in figure 1. The average ratio of retirees at age 64 for our sample is approximately 60% confirming the findings of previous studies[6] that most people have already retired by the time they reach the DRA. However, the first interesting feature of the distribution is the discontinuity at age 65, where the proportion of retirees jumps to about 85%. This discontinuity in the distribution of retirees can be exploited using regression discontinuity (Thistlethwaite and Campbell, 1960, van der Klaauw, 1996, Hahn et al., 2001) to obtain a first set of instrumental variables for retirement in the health-retirement equation[7]. The procedure can be formalised as follows.

Let Y be an outcome of interest with conditional expected value $G(x) = E(Y|X = x)$. In our case Y is health status measured, for instance, as an indicator of depression or the average score in a general health questionnaire. X is a random variable, typically correlated with Y, that determines allocation to treatment status -the latter represented by T, a binary indicator taking value 1 if a person in the sample is treated, and whose conditional expectation is $P(x) = E(T|X = x)$. In a typical regression discontinuity application, allocation to treatment depends on the running variable X exceeding a threshold value x_o. In our case, X denotes an individual's age, T indicates retirement

[6]See, Banks and Smith (2006), Gruber and Wise (2002, 2004).

[7]A number of authors have exploited this type of discontinuity in the distribution of retirees, including Battistin et al. (2009), Stancanelli and van Soest (2011), Johnston and Lee (2009), Dong (2011), to mention but a representative few.

status and $x_o = 65$ (the default retirement age). Perfect compliance with treatment is not a requirement for successful identification -and in our case it is not satisfied[8]. Using Rubin's potential outcomes framework let $Y(1)$ and $Y(0)$ refer to an individual's outcomes with and without the treatment, only one of which is revealed in empirical research,

$$Y = TY(1) + (1 - T)Y(0) = \alpha + \tau T.$$

where $\tau = Y(1) - Y(0)$ is the treatment effect parameter. A number of sources[9] show that, if

$$\lim_{x \to x_o^+} P(x) \neq \lim_{x \to x_o^-} P(x)$$

then,

$$E(\tau | X = x_o) = \frac{\lim_{x \to x_o^+} G(x) - \lim_{x \to x_o^-} G(x)}{\lim_{x \to x_o^+} P(x) - \lim_{x \to x_o^-} P(x)} \tag{3.1}$$

The necessary condition for identification is that $E(Y(0)|X = x)$ and $E(Y(1)|X = x)$ are continuous at x_o -a condition we will refer to as continuity of potential outcomes (CPO)[10]. Under additional weak restrictions, Lee (2008) shows that CPO naturally leads to quasi-randomization in small neighbourhoods to the left and right the cut-off point, and this provides the power (and justifies the use) of regression discontinuity. Therefore, evaluation of empirical distributions of pre-determined characteristics just above and below x_o becomes a valid check of the otherwise non-testable CPO assumption.

Estimation in RD designs in this setting admits use of Instrumental Variable (IV) interpretation[11]. From a parametric point of view, for small $e(n) > 0$, one has the following outcome model in $(x_o - e, x_o + e)$,

$$Y = \beta_0 + h(X) + \tau T + \varepsilon \tag{3.2}$$

where the unknown $h(X)$ can be approximated at x_o by a polynomial series expansion

[8] Traditionally one distinguishes between designs with perfect compliance, "sharp"RD, and designs with imperfect compliance, "fuzzy". In this paper we are interested in the second type of design.

[9] See Hahn et al. (1999), Lee (2008), Angrist and Pischke (2008) or Lee and Lemieux (2010)

[10] Under homogeneous treatment effect the ratio in equation (3.1) identifies τ itself, and in this case only continuity of $E(Y(0)|X)$ is required (Hahn et al., 1999)

[11] Without loss of generality, in what follows we adopt a parametric approach. The equivalent nonparametric take on RD designs can be found in numerous references, including Hahn et al. (1999), Angrist and Pischke (2008) or Dong (2011).

(Angrist and Pischke, 2008), so that[12]

$$Y = \beta_0 + \beta_1(X - x_o) + \ldots + \beta_p(X - x_o)^p + \tau T + \varepsilon \tag{3.3}$$

where ε is uncorrelated with X by assumption. In our case, this implies that age looses its detrimental effect on average health if we restrict attention to a group of people whose age differs only slightly. If the probability of treatment is discontinuous at x_o, then $D_i = \mathbb{I}(X_i \geq x_o)$ is correlated with T_i but, by hypothesis, uncorrelated with ε. Therefore, D_i stands as a valid instrument for T_i in $(x_o - e, x_o + e)$ and τ can be consistently estimated via 2SLS using D_i and a polynomial in X_i as instruments.

If the discontinuity in the distribution of T given X is small, researchers face a problem of weak identification (Feir et al., 2011). It is well know that weak instruments exacerbate the small sample bias of 2SLS estimators and will render misleading inferential procedures (Bound et al., 1995, Staiger and Stock, 1997, Feir et al., 2011, Kleibergen, 2002). However, a recent paper by Dong (2011) (see also Card et al., 2009), has pointed out that, should a discontinuity in the distribution of treatment be too small, identification of treatment effects is possible if there is a discontinuity in the first derivative of the distribution of treatment -that is, a kink in the distribution of T given X- so that:

$$\lim_{e \to 0} \frac{\partial P(x)}{\partial x}\Big|_{x = x_o + e} \neq \lim_{e \to 0} \frac{\partial P(x)}{\partial x}\Big|_{x = x_o - e}. \tag{3.4}$$

The essential condition for identification is continuity in the first order partial derivatives of potential outcomes. Dong (2011) shows that, under this assumption, it is possible to identify treatment effects when there is a kink in the conditional distribution of T, or a jump, or both -in which case the estimator is a weighted average of the estimator used in RD designs and Kink designs, namely

$$\tau = \frac{\lim_{x \to x_o^+} G(x) - \lim_{x \to x_o^-} G(x) + w_n(\lim_{x \to x_o^+} G'(x) - \lim_{x \to x_o^-} G'(x))}{\lim_{x \to x_o^+} P(x) - \lim_{x \to x_o^-} P(x) + w_n(\lim_{x \to x_o^+} P'(x) - \lim_{x \to x_o^-} P'(x))} \tag{3.5}$$

for a given sequence $\{w_n\}$ such that $\lim_{n \to \infty} w_n = 0$. Continuous differentiability of potential outcomes is not a directly testable assumption. Dong suggests that the adequacy of the kink design can be tested by checking the existence of kinks and jumps in the conditional means of pre-determined variables, although she does not provide a formal discussion regarding the nature of these conditions. A formal analysis can be undertaken

[12] For convenience we abuse the notation slightly and use the same symbol for the error term in equations (3.2) (3.3)

drawing from Lee (2008) and Card et al. (2009), and is undertaken in the Appendix.

Looking at Figure 1 it is a priori unclear whether the jump in the discontinuity is sufficiently large to identify the causal effect of retirement on health. However, the figure exhibits a change in the slope after crossing default retirement age. The existence of the kink can be confirmed by a simple OLS regression of the retirement indicator on $D = I(X_i \geq 65)$, $D_i * X_i$ and a fifth polynomial in X_i. The coefficient for the interaction term in such regression was -.0568 ($p = 0.000$). It was mentioned above that the retirement rate at age 64 is about 60% for a typical year and jumps to about 85% at age 65. The increase in the proportion of retirees up to age 64 is fast, reflected in a large positive, increasing slope of the half of the distribution to the left of x_o. However, the right half of the distribution exhibits only a moderate slope, suggesting that retirement beyond the DRA converges slowly to 1. This feature allows us to build a second set of instrumental variables. Dong (2011) shows that her kink design approach also admits a local IV interpretation. Under the premises underpinning the local outcome model in equation (3.2) and its approximation in (3.3), a kink at x_o can be captured by the interaction of D and X, in which case D and $D*X$ are both valid instruments to identify τ via 2SLS using the feasible first stage equation

$$T = \alpha_o + \alpha_1(X - x_o) + \ldots \alpha_p(X - x_o)^p + \pi_1 D + \pi_2 D * (X - x_o) + u.$$

In this case, Dong (2011) shows that the estimator of τ is an empirical version of (2.6) with weights $w_1 = cov(T, D)$, $w_2 = cov(T, X * D)$.

The above discontinuity designs have been developed with cross-sectional data in mind. When repeated cross-section and panel data are available, it is a wide-spread practice to take the average cluster approach described in Lemieux and Milligan (2008), Battistin et al. (2009), or Dong (2011), whereby the relevant endogenous variables are averaged within each wave and across the panel conditional on clusters of the running variable. As noted by Lemieux and Milligan (2008), in a regression discontinuity framework, the resulting estimator of the treatment effect is identical to weighted estimates of individual wave estimators, when the weights used are the number of observations per cell.

As pointed out by Petterson-Lidbom (forthcoming), RD relies on the cut-off value inducing local quasi-experimental randomization, which is the strength, but also Achille's heel, of the method. In practice nonparametric identification requires large amounts of data at the boundary of the threshold value, which is not always available. However if one exploits the panel structure of the data, then estimates via fixed effects (FE), and first-differences (FD) draw comparisons within the same individual, making the treatment and control groups comparable by definition, the assumption of continuous (differentiable) potential outcomes is less contentious regardless of the sample size, thus reducing the relevance of choosing a bandwidth. Following this rationale, we suggest the following local outcome model,

$$\begin{aligned} Y_{it} &= \theta_i + \lambda_t + \tau T_{it} + \beta_1(X_{it} - x_o) + \ldots + \beta_p(X_{it} - x_o)^p + \varepsilon_{it} && (3.6) \\ T_{it} &= \theta_i + \lambda_t + +\beta_1(X_{it} - x_o) + \ldots + \beta_p(X_{it} - x_o)^p \\ &+ \pi_1 D_{it} + \pi_2 D_{it} * (X_{it} - x_o) + \nu_{it}. && (3.7) \end{aligned}$$

where θ_i and λ_t are individual and period specific coefficients. This model can be estimated via first difference instrumental variables (FD-IV) on the structural equation, with instruments described in the first stage equation. As before the unknown control functions $q(.)$ and $r(.)$ are replaced by polynomial approximations in practice. To test the validity of the instruments, we follow Staiger and Stock (1997), and calculate the first stage test of joint significance (F test), expecting this to exceed a nominal value of ten.

4 Empirical Analysis.

We use the discontinuity design just described to evaluate the effect of retirement on a collection of health-related indicators which can be classified in three groups: mental health indicators, measures of health care utilisation and design check variables[13]. We describe these variables next.

In the BHPS, mental health is measured with Goldberg's General Health Questionnaire[14], GHQ-12, (Goldberg, 1978). The amount of strain and sense of a role felt by interviewees constitute two of the domains explored by GHQ-12. Firstly, individuals are asked if they have recently felt that they "*were playing a useful part in things*". We use this as our role-variable. We created a binary indicator taking value 1 if an individual answered "*less so*"or "*much less*"to this question and 0 otherwise. Secondly, GHQ-12 contains the question "*Have you recently felt constantly under strain?*". We adopt this

[13]Our analysis was broader, and we also studied the average score in Goldberg's General Health Questionnaire (GHQ-12), self-reported life satisfaction, and unhappiness, but each of these outcomes appeared statistically insignificant, so we do not include them in the final set of results.

[14]The answers to the questionnaire are summarised in a twelve point scale that measures the inability to carry out normal functions and the appearance of new and distressing psychological phenomena. Large average scores denote deteriorating mental health.

as our strain-variable. A binary indicator was created that took the value 1 if an individual answered "*rather more*"or " *much more*". Apart from the GHQ-12 and the role and strain dummies, BHPS includes a binary indicator taking value 1 if a person reported to "*have any of the health problems or disabilities listed ... anxiety or depression*". This is a more direct indicator of (self-reported) mental health, and was also included in the analysis.

In addition to the above mental health scores, we also considered indicators of migraine and high blood pressure. These can be seen a more objective indicators of general health, but they can also proxy mental health. Medical evidence suggests that there is a genetic association between migraine and depression (Ahn and Ashizawa, 2010, Stam et al., 2010). The medical literature has also recently identified a correlation between depression and high-blood pressure (Licht et al., 2008, Dawood et al., 2009). These indicators might be directly affected by retirement or, given the result in the medical literature, if retirement affects mental health, it is likely that these indicators will also be affected.

Our measures of health care utilisation are a dummy variable indicating if an individual had at least one inpatient stay in hospital during the previous year and a dummy variable indicating if an individual visited a general practitioner during the previous year on six or more occasions. In the typical year, the proportion of individuals in the panel that visited a general practitioner (family doctor) three or more times over the previous 12 month period was around 40% while the proportion of individuals who visited a general practitioner six or more times ranges between 15-20%, so the latter is a relatively infrequent event. Note that individuals normally make appointments with their general practitioner on their own initiative whenever they consider that medical care is required, therefore this measure of utilization might not reflect actual medical need. Self-referral is based on the subjective evaluation of a patient but, as we argued earlier, the number of visits to a family doctor may simply reflect incentives faced by individuals at their work place. On the contrary, inpatient stays require the qualified opinion of a physician, and therefore apart from measuring utilization, this variable is a more objective proxy of medical need and overall health status. If retirement seriously affects health in the short run, we should observe some variation in the levels of these variables.

In order to address potential heterogeneity in the causal effect of retirement, we undertook the analysis by education level (higher education, secondary education and no qualifications), occupation group (white collar or manual worker) and self-reported job satisfaction the year before retirement (whether individuals were *completely satisfied* with their jobs or not).

Following previous practice in the literature (see for example Battistin et al. (2009), Dong (2011), Petterson-Lidbom, forthcoming) model (3.6) was estimated using four different bandwidths, namely $\pm\infty, \pm 10, \pm 8$ and ± 6. However it is convenient to remark that, since we are exploiting the panel structure of the data, comparisons are drawn at the level of the individual, so there are not major issues regarding comparability of treatment and controls; since our design further includes a set of time dummies, any trends are being taken care of, and therefore we expect the results to be relatively insensitive to the choice of bandwidth.

4.1 Results.

The analysis began by assessing whether the distribution of the running variable (actual age minus the default retirement age) exhibited any kink or jump for each year in the panel. Following the theory given in Appendix A, any such discontinuities would imply that our results could be confounded by variation caused by reasons other than retirement status. Figure 2 exhibits the distribution of the running variable for three years (1995, 2000 and 2005). It is clear from this figure that there are no obvious discontinuities around the cut-off implied by the default retirement age, so that the first necessary identifying condition seems to be satisfied.

To further test the validity of the design, we sought kinks or jumps in the conditional distributions of a number of pre-determined variables, namely the proportion of individuals without qualifications and the proportion of manual workers in the sample. Once again, the theory of discontinuity designs suggests that these must not exist, otherwise inference might be biased due to variation in factors other than treatment itself. Figures 3 and 4 represent scatter plots of the average (across the panel) number of manual workers and people without qualifications at each age cell. The figures include local linear estimators[15] of the conditional means above and below the cut off set by the default retirement age. Neither figure suggests the existence of discontinuities around the default retirement age. As a last evaluation, we also included as outcome variables in our study the proportion of individuals with diabetes. In accordance with the medical literature, diabetes is related to certain diseases and genetic endowment. There is evidence, however, that Type II diabetes is related to sedentary life styles (Risérus et al., 2009), and therefore, in the long-run, retirement might lead to diabetes if it conveys a change to a less active life-style. Even then, it is unlikely that the prevalence of diabetes increases in the short run immediately after retirement, and therefore our discontinuity

[15]The bandwidths were chosen as $h = c\sigma_x n^{-1/5}$, where $c = 0.5, 1, 1.5, 2$ and σ_x is the sample standard deviation of the running variable.

design should not reveal any variation in the prevalence of this disease in the sample. Table 2 contains the results produced by the FD-IV estimator for a variety of groups[16]. In all instances the causal effect of retirement on diabetes was not significant, as we had expected. Therefore, we find sufficient evidence to support the robustness of the discontinuity design.

Further evidence regarding the strength of the design is provided by the first stage estimates collected at the bottom of Tables 3 to 7. Table 3 contains the reduced form estimates for the whole sample. The dummy variable capturing the jump in the proportion of retirees appears with a significant coefficient ranging between 15.9% and 18.0%, which is just below the expected 20/25% jump in the proportion of retirees reported by our descriptive analysis and previous published results (e.g. Smith (2006), Banks and Smith (2006)). The discontinuity in the first derivative of the conditional mean, captured by the interaction term in equation (3.7) ranges between -0.009 and -0.037, although is only statistically significant for the larger bandwidths considered in the exercise. The F-test for the joint significance of the set of instruments is always significant and never below 19.1 and therefore, in accordance withStaiger and Stock (1997), weak instruments does not appear to be a concern. The pattern observed for the whole sample stays approximately invariable when different groups of individuals are considered, although the magnitude of the discontinuity in the distribution of retirees varies up to a maximum of between around 40% to 45% (for the group of manual workers). However, we find that for the groups of people with secondary and higher education, the jump in the distribution of retirees is small (10 to 16% in the secondary education group and 5 to 10% in the higher education group), and so is the F test (which remains significant in all instances, but can be as low a 3). In these examples, the significance of the kink in the first stage equation is also lost, and therefore this raises some doubts about the validity of level of schooling as a classification system to study retirement via discontinuity designs in the BHPS.

The estimated causal effect of retirement in the outcomes of interest are collected in Tables 3 to 7. We first focus our attention on the self reported indicators of mental health. Table 3 shows that the causal effect of retirement on high blood pressure, migraine and depression is statistically insignificant. This is also found in all the remaining tables, and the finding is invariant to the choice of bandwidth. Therefore, we fail to detect a short term effect of retirement on our indicators of mental health. The next question is whether

[16] We report the results for all sample, individuals with a higher education degree, those who were completely satisfied with their jobs the year before retirement and while collar workers. The results for the remaining groups were similar and are available from the authors upon request.

the role and strain theories can be corroborated by our study. As with the indicators of mental health, we do not find any statistically significant short term effect of retirement on the GHQ's indicators of sense of role loss and strain. Once again, the results are invariant irrespective of the choice of bandwidth, educational and occupational groups or the level of job satisfaction prior to retirement.

Finally we evaluate the effect of retirement on utilization of medical care. As with the other measures, there are no significant effects of retirement on inpatient stays. However, we do find a significant effect of retirement on the proportion of patients that visited a GP on more than six occasions. This effect is only representative for the group of people who reported to be completely satisfied with their jobs the year before they took retirement. The estimated average treatment effect for this group is about -0.3, which is approximately equal to the effect estimated in all the other groups (although for the remaining groups the effect appears to be insignificant).

The question might arise if the lack of significance in the results presented here is due to poor identification, in which case the design would fail to capture variation in virtually any variable under consideration. We have already presented evidence supporting the strength of the discontinuity design, except in the case when individuals are classified by educational group. Nonetheless, we undertook a final check. Recent literature has identified a substantial rise in the amount of time spent in housework upon retirement -see, in particular, Aguiar and Hurst (2005). These authors find that individuals would substitute time in the workplace for time cooking and other tasks, which, in the view of Aguiar and Hurst (2005) compensates for the drop in personal expenditure, thus explaining part of the retirement-consumption puzzle (Banks et al. (1998)). Similar variation in housework has been detected in French data by Stancanelli and van Soest (2011). We evaluated the effect of retirement on time spent on housework in the BHPS. Figure 5 shows local linear regressions for the amount of time spent in housework in three different years (1995, 2000 and 2005). The plots consistently reflect a discontinuity on the amount of time spent in housework upon retirement. The figure suggests a jump of about 3-4 hours per week, on average. Table 9 collects the results of applying our discontinuity designs to the variable housework. It confirms a jump of between 3 and 5 hours per week in the amount to time devoted to housework, and the results are significant for all bandwidths and all groups considered, except when clusters are organised around education levels. However, given our findings regarding the weakness of the design when individuals are classified by educational group, this is not a surprising result.

Our results seem to suggest three hypotheses. Firstly, retirement does not affect mental health, high blood pressure or the incidence of migraine. This is in line with previous findings reported by Coe and Lindeboom (2008) and Neuman (2008). Our conclusion does not depend on education attainment, the occupational classifications or the level of job satisfaction a worker has immediately before retirement. Similarly, we find that inpatient stays remain unaffected by retirement. Inpatient stays normally denote an at least more-than-moderate health shock, and since referral to secondary care needs the qualified opinion of a physician, our results seems to suggest that other dimensions of health are not serious affected by retirement either. Despite these findings, our estimates suggested that utilization of primary care might be decreased by retirement. In particular, the proportion of people who visited a general practitioner on more than six occasions in a give year dropped by about 30% for the group of people who reported to be completely satisfied with their jobs prior to retirement. For the other groups considered in the analysis we also found a drop of around the same magnitude in this indicator, however the estimated causal effect for these groups was statistically insignificant. Assuming that this result can be extrapolated then, if as suggested here retirement is innocuous for health in the short run, the question arises of what could explain any variation in utilization of primary care. The hypothesis we have presented is that perhaps individuals have incentives to self-refer to a general practice during their working lives. For instance, contracts might contemplate a variety of arrangements of sick-leaves or sick-pay that might help to rationalise shirking. This is a question for further investigation.

5 Conclusion

In this article we have explored the causal effect of retirement on health using a representative sample of the population (the British Household Panel Survey). A number of researchers have pointed out that if retirement affects health the valuation of policies that prolong retirement should account for such effects, since a longer retirement horizon, by postponing changes in health outcomes, would affect the utilization of health services by older adults and this would impact upon medical expenditure. This notwithstanding, numerous published evidence suggests that there are different experiences of retirement, and therefore, any potential effect of retirement on health is likely to be heterogeneous. We have used a new identification strategy in order to capture the local average causal effect of retirement on a number of outcomes (namely mental health and health care utilisation). Our strategy combines discontinuity designs with panel data. Exploiting discontinuities in the distribution and the density functions of the proportion of retirees conditional on age, we have constructed instrumental variables that were fed into a first-difference instrumental variable panel data model. Proceeding this way, we could take into account the potential endogeneity in the treatment indicator. By exploiting the panel structure of the data, we had the advantage that comparisons are drawn within the same individual over time, so that the assumption of continuous differentiability of potential outcomes typically needed for the validity of discontinuity designs becomes much less contentious and, furthermore, the issue of selecting a bandwidth guaranteeing homogeneity of the treatment and control groups becomes also less fundamental. Our results suggest that the local average treatment effect of retirement on mental health is largely insignificant thus supporting previous evidence in Neuman (2008) and Coe and Lindeboom (2008) who suggests that this relationship is merely nominal. The effect of retirement on inpatient stays also appears to be insignificant.

We found some evidence suggesting that utilization of primary care might be decreased by retirement. In particular, the proportion of people who visited a general practitioner on more than six occasions in a give year dropped by around 30% for the group of people who reported to be completely satisfied with their jobs prior to retirement. For the other groups considered in the analysis we also found a drop in this indicator, however the estimated causal effect for these groups was statistically insignificant. We are not aware of similar results in the literature, so this specific finding needs further exploration to confirm its external validity. If retirement reduces the consumption of primary care, as our results suggest, then postponing the age of retirement might preclude the realisation of this benefit. Yet the gains derived from later retirement patterns are likely to more than compensate for that loss. However, the question of why the consumption of primary care might decrease upon retirement would maintain its interest. Given our results that suggest that retirement does not affect mental health or inpatient stays (and thus the prevalence of very severe health conditions), the drop in utilization of primary care must be due to external factors. One potential reason for this could be incentives faced by workers prior to retirement that might promote absenteeism, part of which can be rationalised through consumption of medical care. Statutory or occupational sick leave and sick pay can create this type of incentive when the monitoring of absenteeism is not too strict. Even in firms tightly monitor the health and health care use of their employees, or efficiency wages are paid once the decision of retirement has been taken the consequences of getting caught shirking might become mild or disappear altogether. Unfortunately, our data set does not provide sufficient information to explore this question further.

In common with most empirical work, one potential issue with our analysis is that it only considers average treatment effects. However, it is the overall distribution of health that is of interest for the policy maker because the average health status might remain constant over time, even when radical changes might be occurring in the distribution of health. To explore this, however, requires estimating distributional effects, and we leave this for future research.

A Appendix

Continuous differentiability of potential outcomes is a necessary condition for the validity of the kink design. However, this is not a directly testable assumption. Dong suggests that the adequacy of the kink design can be tested by checking the existence of kinks and jumps in the conditional means of pre-determined variables, although she does not provide a formal explanation. We provide next a formal analysis of this condition, borrowing from Lee (2008) and Card et al. (2009). Let us introduce the following general scenario.

Assumption 1. (i) (X, W) are random variables (only X is observable). The c.d.f. of W is denoted by $G(w)$. (ii) The conditional distribution of X given W, $F(x|w)$, is twice continuously differentiable at x_o for all w. (iii) The conditional density $f(x_o|w) > 0$ in $\mathcal{A}$, where $\int_{\mathcal{A}} dG(w) > 0$.

Assumption 2. $Y(0) = y_o(X, W)$, $Y(1) = y_1(X, W)$, where y_o, y_1 are real-valued and continuously differentiable at the cut-off, x_o. Finally Z (any pre-determined and observable variable) is such that $Z = z(W)$.

Assumption 3. $\beta_i = Y_i(0) - Y_i(1)$ is such that $\tau = E(\beta_i|X_i = x)$ is continuous differentiable at x_o, and X is independent if β_i conditional on X near x_o, for all w.

As in Lee (2008), W is the unobservable type of an individual -which we assume to be, without loss of generality, a single random variable. Identical individuals will share the same value of W. Assumption 1.ii implies that for each individual, the probability of obtaining an X just below and just above 0 are the same and that, in a neighbourhood about x_o, changes in this probability occur smoothly. The latter is not a requirement for identification in RD designs. By requiring $f(x_o|w) > 0$ we are ruling out situations when individuals can precisely manipulate their X score, what would invalidate the design. Assumption 2 allows outcome to vary with X directly, and not only indirectly through treatment status. Finally, Assumption 3 is required because we are assuming heterogeneous treatment effects. The condition is identical to that in Hahn et al. (1999) for RD designs. In the case of constant treatment effect is can be done without.

For an arbitrary function $H(x)$ denote $\lim_{x \to x_o^+} H'(x) = \lim_{e \to 0} \partial H(x)/\partial x|_{x = x_o + e}$. Given all three conditions, we can introduce the following result.

Proposition 1 *Under Assumptions 1 to 4, (i) $F(W \leq w|X = x)$ is continuously differentiable at x_o, (ii) $F(Z \leq z|X = x)$ is continuously differentiable at x_o for any* *predetermined variable* Z, *and* (iii)

$$\tau = \frac{\lim_{x \to x_o^+} G'(x) - \lim_{x \to x_o^-} G'(x)}{\lim_{x \to x_o^+} P'(x) - \lim_{x \to x_o^-} P'(x)} \tag{A-1}$$

This result is proved in the appendix. The key innovation of the proposition is result (ii). Dong (2011) establishes that identification in kink designs requires continuous differentiability of potential outcomes. But this is an untestable assumption. Result (ii) in the proposition establishes a way to verify the validity of the kink design by evaluating that the conditional c.d.f of observable pre-determined variables do not exhibit jumps or kinks about the threshold value x_o. If these exist, the it is unlikely that outcomes will be continuously differentiable. The proof of (i) and (ii) can be found in Card et al. (2009). To prove (iii) note that f $Y_i = \alpha_i + \beta_i T_i$, where $\alpha_i = Y(0)$, $\beta_i = Y_i(1) - Y_i(0)$. Then,

$$\begin{aligned} E(Y_i|X = x_o + e) &= E(\alpha_i|X = x_o + e) + E(\beta_i T_i|X_i = x_o + e) \\ &= E(\alpha_i|X = x_o + e) \\ &+ E(\beta_i|X_i = x_o + e)E(T_i|X_i = x_o + e), \end{aligned} \tag{A-2}$$

under conditional independence. Differentiating with respect to x and taking limits, we have

$$\begin{aligned} \frac{\partial}{\partial x}E(Y_i|X = x_o + e) &= \frac{\partial}{\partial x}E(\alpha_i|X = x_o + e) \\ &+ \frac{\partial}{\partial x}E(\beta_i|X_i = x_o + e)E(T_i|X_i = x_o + e) \\ &= \int \frac{\partial y_o(w,v)}{\partial x}\frac{f(x|w)}{f(x)}dF(w) \\ &+ \int y_o(w,v)\frac{\partial}{\partial x}\left(\frac{f(x|w)}{f(x)}\right)dF(w) \\ &+ \frac{\partial E(\beta_i|X = x_o + e)}{\partial x}E(T_i|X = x_o + e) \\ &+ E(\beta_i|X = x_o + e)\frac{\partial E(T_i|X = x_o + e)}{\partial x} \end{aligned} \tag{A-3}$$

A similar result follows for $\frac{\partial}{\partial x} E(Y_i|X = x_o - e)$. Taking limits as $e \to 0$, it follows that,

$$\begin{aligned} & \lim_{e \to 0} \frac{\partial}{\partial x} E(Y_i|X = x_o + e) - \lim_{e \to 0} \frac{\partial}{\partial x} E(Y_i|X = x_o - e) \\ = \quad & E(\beta_i|X = x_o + e) \lim_{e \to 0} \left(\frac{\partial E(T_i|X = x_o + e)}{\partial x} - \frac{\partial E(T_i|X = x_o - e)}{\partial x} \right) \end{aligned} \tag{A-4}$$

because of the continuous differentiability of potential outcomes and $f(v|w)$ and the continuity of $E(\beta_i|X)$ at x_o.

References

Acemoglu, D. and D. Autor (2011). Skills, tasks and technologies: impliations for employment and earnings. *in O. Ashenfleter and D. Card, eds, Handbook of Economics. IV.B.*, 1043–1172.

Aguiar, M. and E. Hurst (2005). Consumption versus expenditure. *Journal of Political Economy 113*, 919–948.

Ahn, A. and T. Ashizawa (2010). Headaches and the blues: a shared genetic association between migraine and depression. *Neurology. 74*, 278–279.

Angrist, J. and J. Pischke (2008). *Mostly Harmless Econometrics.* Princeton University Press.

Autor, D., F. Levy, and R. Murnane (2003). The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics 118(4)*, 1279–1333.

Banks, J., R. Blundell, and S. Tanner (1998). Is there a retirement-saving puzzle? *American Economic Review. 88*, 769–788.

Banks, J. and S. Smith (2006). Retirement in the U.K. *Oxford Review of Economic Policy 22*, 40–56.

Battistin, E., A. Brugiavini, E. Rettore, and G. Weber (2009). The retirement consumption puzzle: Evidence from a Regression Discontinuity approach. *American Economic Review 99*, 2209–2226.

Blundell, R., C. Meghir, and S. Smith (2002). Pension incentives and the pattern of early retirement. *Economic Journal 112*, C153–C170.

Bound, J. (1991). Self-reported versus objective measures of health in retirement models. *Journal of Human Resources 26*, 106–138.

Bound, J., D. Jaeger, and R. Baker (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association 90*, 443–450.

Bound, J., T. Stinebrickner, and T. Waidmann (2010). Health, economic resources and the work decision of older men. *Journal of Econometrics 156*, 106–129.

Brzozowski, M. and Y. Lu (2010). Home cooking, food consumption and food production among retired Canadian households. *Canadian Public Policy 36*, 107–128.

Cai, L. (2010). The relationship between health and labour force participation: Evidence from a panel data simultaneous equation model. *Labour Economics 17*, 77–90.

Card, D., D. Lee, and Z. Pei (2009). Quasi-experimental identification and estimation in the Regression Kink Design. *Working Paper 553, Princeton University*.

Charles, K. (2004). Is retirement depressing?: labor force inactivity and psychological well-being in later life. *Research in Labor Economics 23*, 269–299.

Coe, N. and M. Lindeboom (2008). Does retirement kill you? evidence from early retirement windows. Technical report, Tilburg University, Center for Economic Research, Discussion Paper 2008-93.

Dawood, T., M. Schlaich, A. Brown, and G. Lambert (2009). Depression and blood pressure control: all antidepressants are not the same. *Hypertension: Journal of the American Heart Association doi: 10.1161/HYPERTENSIONAHA.109.133272*.

Dhaval, D., R. I. Rashad, and J. Spasojevic (2008). The effects of retirement on physical mental health outcomes. *Southern Economic Journal 75*, 497–523.

Disney, R., C. Emmerson, and M. Wakefield (2006). Ill health and retirement in Britain: a panel data-based analysis. *Journal of Health Economics. 25*, 621–649.

Dong, Y. (2011). Jumpy or kinky? regression Discontinuity without the discontinuity. Technical report, MPRA Paper 25427.

Feir, D., T. Lemieux, and V. Marmer (2011). Weak identification in Fuzzy Regression Discontinuity desings. Technical report, Micro Theory Working Paper, Microeconomics.ca.

Firpo, S., N. Fortin, and T. Lemieux (2011). Occupational tasks and changes in the wage structure. *IZA Discussion Papers 55442*.

Fuchs, V. (1982). *Time preference and health: an exploratory analysis*. The University of Chicago Press.

Goldberg, D. (1978). *Manual of the General Health Questionnnaire*. NFER Publishing.

Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political Economy 82*, 233–255.

Gruber, J. and D. Wise (2002). *Social Security and retirement around the World.* University of Chicago Press.

Gruber, J. and D. Wise (2004). *Social Security programs and retirement around the World: Micro-estimation.* University of Chicago Press.

Hahn, J., P. Todd, and W. van der Klaauw (1999). Evaluating the effect of an antidiscrimination law using a regression-discontinuity design. *NBER Working Paper 7131.*

Hahn, J., P. Todd, and W. van der Klaauw (2001). Identification and estimation of treatment effects with a regression discontinuity design. *Econometrica 69*, 201–209.

Haider, S. and M. Stephens (2007). Is there a retirement-consumption puzzle? evidence using subjective retirement expectations. *Review of Economics and Statistics. 89*, 247–264.

Hu, F., M. Stampfer, J. Manson, E. Rimm, G. Colditz, B. Rosner, C. Hennekens, and C. Willett (1997). Dietary fat intake and the risk of coronary heart disease in women. *New England Journal of Medicine 337*, 1491–1499.

Hurst, E. (2008). The retirement of a consumption puzzle. *NBER Working Paper 13789.*

Johnston, D. and W. Lee (2009). Retiring to the good life? the short term effects of retirement on health. *Economic Letters 103*, 8–11.

Kim, J. and P. Moen (2002). Retirement transitions, gender, and psychological wellbeing: a life-course, ecological models. *Journal of Gerontology: Psychological Sciences 57B*, 212–222.

Kleibergen, F. (2002). Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica 70*, 1781–1803.

Krueger, A. and J. S. Pischke (1992). The effect of social security on labour supply: a cohort analysis of the notch generation. *Journal of Labor Economics 10*, 412–437.

Lee, D. (2008). Randomized experiments from non-random selection in U.S. House elections. *Journal of Econometrics 142*, 675–674.

Lee, D. and T. Lemieux (2010). Regression discontinuity desings in economics. *Journal of Economic Literature 48*, 281–355.

Lemieux, T. and K. Milligan (2008). Incentive effects of social assistance: a regression discontinuity approach. *Journal of Econometrics 142*, 807–828.

Licht, C., E. de Geus, F. Zitman, W. Hoogendijk, R. van Dyck, and B. Pennix (2008). Association between major depressive disorder and heart rate variability in the Netherlands Study of Depression and Anxiety (NESDA). *Archives of General Psychiatry 65*, 1358–1367.

Lindeboom, M. and M. Kerkhofs (2002). Health and work of the elderly. subjective health measures, reporting erros and the endogenous relationship between health and work. *IZA Discussion Paper Series 457*.

Neuman, K. (2008). Quit your job and get healthier? the effect of retirement on health. *Journal of Labor Research 29*, 177–201.

Petterson-Lidbom, P. (Forthcoming). Does the size of the legislature affect the size of government? evidence from two natural experiments. *Journal of Public Economics*.

Risérus, U., W. Willett, and F. Hu (2009). Dietary fats and prevention of type 2 diabetes. *Progress in Lipid Reseach 48*, 44–51.

Shapiro, C. and J. Stiglitz (1984). Equilibrium unemployment as a worker discipline device. *American Economic Review. 74*, 433–444.

Sickles, R. and P. Taubman (1986). An analysis of the health and retirement status of the elderly. *Econometrica 54*, 1339–1356.

Sickles, R. and A. Yazbeck (1998). On the dynamics of demand for leisure and production of health: Evidence from the Retirement History Survey. *Journal of Business and Economic Statistics 16*, 187–197.

Smith, S. (2004). Can the retirement consumption puzzle be resorlved? Evidence from U.K. panel data. *The Institute for Fiscal Studies Working Paper WP04/07*.

Smith, S. (2006). The retirement-consumption puzzle and involuntary early retirement: evidence from the British Household Panel Survey. *Economic Journal 116*.

Staiger, D. and J. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65*, 557–586.

Stam, A., B. de Vries, A. Janssens, K. Vanmolkot, Y. Aulchenko, P. Henneman, B. Oostra, R. Frants, A. van den Maagdenberg, M. Ferrari, C. van Duijn, and G. Terwindt (2010). Shared genetic factors in migraine and depression: evidence from a genetic isolate. *Neurology 74*, 288–294.

Stancanelli, E. and A. van Soest (2011). Retirement and time use in couples: A regression discontinuity approach. *Working Paper*.

Stern, S. (1989). Measuring the effect of disability on labour force participation. *Journal of Human Resources 24*, 361–395.

Thistlethwaite, D. and D. Campbell (1960). Regression-Discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology 51*, 309–317.

van der Klaauw, W. (1996). A Regression-Discontinuity evaluation of the effect of financial aid offers on college enrollment. Technical report, New York University, Economics Department.

Willett, W. (1998). *Nutritional Epidemiology.* Oxford University Press.

Table 1: Proportion of Retirees by occupation, education and job satisfaction (488 retirements in total).

Education.		
	N	%
No qualifications:	142	29.1 %
University degree:	185	37.91%
Occupation.		
	N	%
Manual workers:	156	31.97%
White collar:	149	30.5%
Job satisfaction		
	N	%
Completely satisfied:	216	44.0%

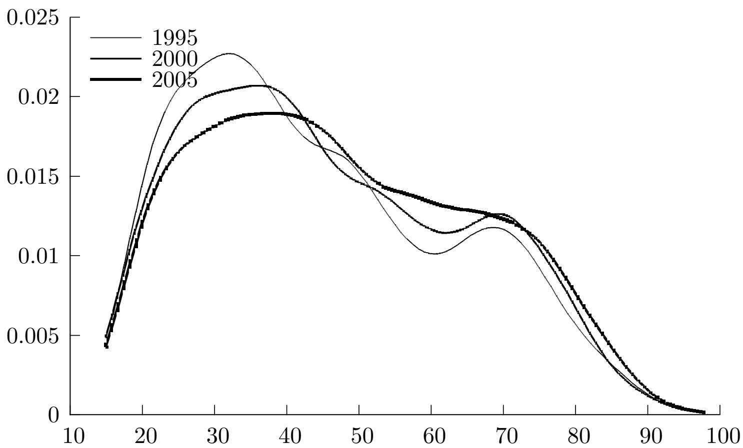

Figure 2: Density function of the running (age) variable for years 1995, 2000 and 2005.

Table 2: Local A.T.E. Diabetes

	(1)	h_{10}	h_8	h_6
All Sample	0.005	0.027	0.040	0.020
	(0.820)	(0.468)	(0.313)	(0.602)
Tot. Satisfied	−0.016	−0.019	−0.020	−0.024
	(0.390)	(0.397)	(0.418)	(0.361)
White collar	−0.043	0.004	0.011	0.011
	(0.339)	(0.886)	(0.759)	(0.756)
Higher education	−0.017	−0.055	0.089	0.025
	(0.748)	(0.626)	(0.487)	(0.815)

p-values in parentheses, * $p < 0.05$

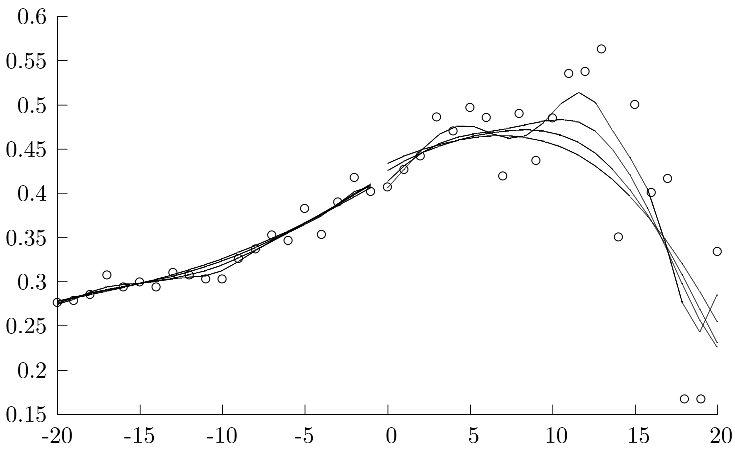

Figure 3: Discontinuities in the proportion of manual workers. The abscissa is $Age-65$. Local linear estimator with bandwidth $h = c\sigma_x n^{-0.2}$, for $c = 0.5, 1, 1.5$ and 2.

Table 3: Local A.T.E., all sample

	(1)	h_{10}	h_8	h_6
High blood pressure	−0.034	−0.024	−0.047	−0.068
	(0.596)	(0.816)	(0.667)	(0.532)
Depression/Anxiety	0.000	0.020	0.033	0.037
	(1.000)	(0.735)	(0.591)	(0.547)
Migraine	0.001	0.005	0.046	0.038
	(0.982)	(0.918)	(0.407)	(0.480)
GHQ-12	0.044	0.357	0.467	0.291
	(0.947)	(0.585)	(0.497)	(0.661)
Role	−0.090	−0.050	−0.014	−0.017
	(0.296)	(0.614)	(0.894)	(0.871)
Strain	−0.014	−0.020	0.031	0.000
	(0.904)	(0.856)	(0.786)	(0.999)
Inpatient	0.051	0.091	0.105	0.154
	(0.499)	(0.392)	(0.347)	(0.166)
G.P. Often	−0.154	−0.194	−0.275*	−0.252*
	(0.053)	(0.096)	(0.026)	(0.039)
	First Stage			
Dummy	0.165*	0.159*	0.180*	0.165*
	(0.000)	(0.000)	(0.000)	(0.000)
Interaction	−0.037*	−0.009*	−0.027	−0.013
	(0.000)	(0.000)	(0.065)	(0.543)
F test	136.330*	28.250*	24.880*	19.910*
	(0.000)	(0.000)	(0.000)	(0.000)

p-values in parentheses, * $p < 0.05$

Table 4: Local A.T.E., manual workers

	(1)	h_{10}	h_8	h_6
High blood pressure	-0.113^{*}	-0.105	-0.126	-0.122
	(0.024)	(0.222)	(0.168)	(0.188)
Depression	-0.059	-0.033	-0.042	-0.052
	(0.152)	(0.527)	(0.427)	(0.311)
Diabetes	0.037^{*}	0.004	0.011	0.011
	(0.006)	(0.886)	(0.759)	(0.756)
Inpatient	0.010	0.108	0.123	0.166
	(0.882)	(0.232)	(0.215)	(0.110)
Migraine	0.037	0.005	0.027	0.011
	(0.383)	(0.915)	(0.528)	(0.785)
Role	-0.039	0.052	0.080	0.048
	(0.606)	(0.539)	(0.359)	(0.600)
GHQ-12	0.472	0.641	0.714	0.533
	(0.422)	(0.290)	(0.244)	(0.388)
Strain	0.037	0.091	0.095	0.055
	(0.716)	(0.344)	(0.310)	(0.546)
GP Often	-0.037	-0.071	-0.089	-0.049
	(0.625)	(0.507)	(0.422)	(0.671)
First Stage				
Dummy	0.425^{*}	0.438^{*}	0.445^{*}	0.408^{*}
	(0.000)	(0.000)	(0.000)	(0.000)
Interaction	-0.214^{*}	-0.041	-0.030	-0.066
	(0.005)	(0.283)	(0.536)	(0.337)
F test	113.950^{*}	16.250^{*}	13.990^{*}	11.120^{*}
	(0.000)	(0.000)	(0.000)	(0.000)

p-values in parentheses, * $p < 0.05$

Table 5: Local A.T.E., no qualifications

	(1)	h_{10}	h_8	h_6
High blood pressure	−0.129	−0.146	−0.182	−0.140
	(0.174)	(0.214)	(0.142)	(0.262)
Depression	0.031	0.037	0.044	0.030
	(0.593)	(0.581)	(0.526)	(0.670)
Diabetes	0.037	0.046	0.039	0.024
	(0.313)	(0.317)	(0.446)	(0.644)
Inpatient	0.131	0.148	0.161	0.150
	(0.184)	(0.203)	(0.195)	(0.239)
Migraine	0.027	0.014	0.046	0.042
	(0.576)	(0.819)	(0.465)	(0.517)
Role	0.009	0.033	0.083	0.067
	(0.933)	(0.771)	(0.488)	(0.582)
GHQ-12	0.015	0.272	0.421	−0.095
	(0.982)	(0.707)	(0.584)	(0.904)
Strain	−0.001	0.037	0.045	0.011
	(0.994)	(0.743)	(0.701)	(0.924)
GP Often	−0.065	−0.119	−0.171	−0.159
	(0.552)	(0.370)	(0.225)	(0.266)
		First Stage		
Dummy	0.265*	0.226*	0.233*	0.195*
	(0.000)	(0.000)	(0.000)	(0.000)
Interaction	−0.044*	−0.070*	−0.054*	−0.107*
	(0.000)	(0.000)	(0.027)	(0.004)
F test	58.990*	27.670*	23.640*	18.360*
	(0.000)	(0.000)	(0.000)	(0.000)

p-values in parentheses, * $p < 0.05$

Table 6: Local A.T.E., Higher Education

	(1)	h_{10}	h_8	h_6
High blood pressure	0.101	0.073	−0.005	−0.105
	(0.581)	(0.853)	(0.989)	(0.788)
Depression	0.053	0.179	0.116	0.080
	(0.722)	(0.421)	(0.604)	(0.716)
Inpatient	−0.008	0.205	0.452	0.342
	(0.971)	(0.605)	(0.288)	(0.388)
Migraine	−0.122	−0.117	−0.099	−0.193
	(0.412)	(0.564)	(0.636)	(0.298)
Role	−0.108	0.049	−0.295	−0.300
	(0.705)	(0.896)	(0.452)	(0.420)
GHQ-12	0.883	2.416	0.472	0.917
	(0.694)	(0.340)	(0.842)	(0.672)
Strain	0.077	0.081	−0.043	−0.257
	(0.851)	(0.858)	(0.925)	(0.544)
GP Often	−0.355	−0.451	−0.423	−0.247
	(0.140)	(0.302)	(0.349)	(0.554)
	First Stage			
Dummy	0.056*	0.067*	0.091*	0.102*
	(0.000)	(0.000)	(0.000)	(0.000)
Interaction	−0.039*	0.016	0.053*	0.075*
	(0.000)	(0.415)	(0.033)	(0.004)
F test	29.200*	3.090*	3.000*	2.800*
	(0.000)	(0.000)	(0.000)	(0.000)

p-values in parentheses, * $p < 0.05$

Table 7: Local A.T.E., white collar

	(1)	h_{10}	h_8	h_6
High blood pressure	−0.153	−0.105	−0.126	−0.122
	(0.290)	(0.222)	(0.168)	(0.188)
Depression	−0.071	−0.033	−0.042	−0.052
	(0.562)	(0.527)	(0.427)	(0.311)
Inpatient	0.070	0.108	0.123	0.166
	(0.712)	(0.232)	(0.215)	(0.110)
Migraine	0.016	0.005	0.027	0.011
	(0.897)	(0.915)	(0.528)	(0.785)
Role	−0.082	0.052	0.080	0.048
	(0.740)	(0.539)	(0.359)	(0.600)
GHQ-12	−0.410	0.641	0.714	0.533
	(0.832)	(0.290)	(0.244)	(0.388)
Strain	−0.210	0.091	0.095	0.055
	(0.568)	(0.344)	(0.310)	(0.546)
GP Often	−0.057	−0.071	−0.089	−0.049
	(0.772)	(0.507)	(0.422)	(0.671)
	FIRST STAGE			
Dummy	0.117*	0.438*	0.445*	0.408*
	(0.000)	(0.000)	(0.000)	(0.000)
Interaction	−0.024*	−0.041	−0.030	−0.066
	(0.001)	(0.283)	(0.536)	(0.337)
F test	40.180*	16.250*	13.990*	11.120*
	(0.000)	(0.000)	(0.000)	(0.000)

p-values in parentheses, * $p < 0.05$

Table 8: Local A.T.E., completely Satisfied with job

	(1)	h_{10}	h_8	h_6
High Blood Pressure	0.025	0.032	0.048	0.097
	(0.805)	(0.778)	(0.675)	(0.374)
Depression	0.059	0.059	0.059	0.041
	(0.269)	(0.198)	(0.224)	(0.361)
Happy	−0.191	−0.223	−0.270	−0.290
	(0.271)	(0.176)	(0.087)	(0.060)
Inpatient	0.031	0.014	0.053	0.068
	(0.788)	(0.905)	(0.662)	(0.564)
Migraine	0.049	0.045	0.044	0.032
	(0.390)	(0.414)	(0.380)	(0.433)
Role	0.136	0.131	0.150	0.129
	(0.209)	(0.232)	(0.181)	(0.241)
GHQ-12	1.244	1.228	1.307	1.207
	(0.081)	(0.074)	(0.067)	(0.079)
Strain	0.123	0.105	0.140	0.143
	(0.347)	(0.390)	(0.238)	(0.201)
GP Often	−0.344*	−0.357*	−0.333*	−0.273*
	(0.006)	(0.008)	(0.015)	(0.043)
		First Stage		
Dummy	0.362*	0.361*	0.395*	0.330*
	(0.000)	(0.000)	(0.000)	(0.000)
Interaction	−0.067	−0.054	0.000	−0.085
	(0.060)	(0.308)	(0.994)	(0.335)
F test	12.020*	9.010*	8.630*	7.260*
	(0.000)	(0.000)	(0.000)	(0.000)

p-values in parentheses, * $p < 0.05$

Table 9: Local A.T.E. Housework

	(1)	h_{10}	h_8	h_6
All Sample	3.182*	3.327*	3.898*	5.079*
	(0.013)	(0.049)	(0.030)	(0.004)
Tot. Satisfied	4.810*	4.925*	4.819*	5.628*
	(0.003)	(0.003)	(0.005)	(0.001)
White collar	8.014*	4.657*	3.991*	3.714*
	(0.007)	(0.001)	(0.008)	(0.017)
University degre	5.511	5.301	6.138	6.191
	(0.144)	(0.341)	(0.296)	(0.281)

p-values in parentheses, * $p < 0.05$

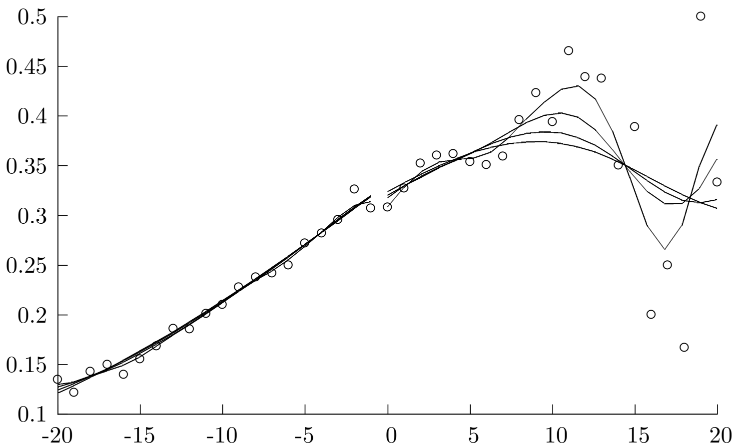

Figure 4: Discontinuities in the proportion of individuals with no qualifications. The abscissa is $Age - 65$. Local linear estimator with bandwidth $h = c\sigma_x n^{-0.2}$, for $c = 0.5, 1, 1.5$ and 2.

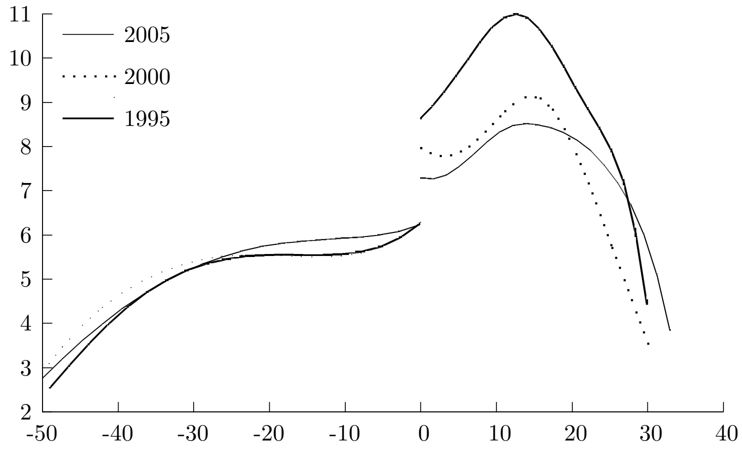

Figure 5: Discontinuities in the amount of time spent in housework (local linear regression), years 1995, 2000 and 2005. The abscissa is $Age - 65$.