ECONOMIC RESEARCH FORUM
منتدى البحوث الاقتصادية
2017

working paper series

SELF-RATED HEALTH AND PRIMARY CARE UTILIZATION: IS SELECTION INTO HEALTHCARE ENDOGENOUSLY DETERMINED?

Firat Bilgel and Burhan Can Karahasan

Working Paper No. 1079

SELF-RATED HEALTH AND PRIMARY CARE UTILIZATION: IS SELECTION INTO HEALTHCARE ENDOGENOUSLY DETERMINED?

Firat Bilgel and Burhan Can Karahasan

Working Paper 1079

April 2017

This work has benefited from a financial grant from the Economic Research Forum (ERF) under grant no. 2015-145 of the Global Development Network Research Competition on the Economics of Healthcare in the ERF Region. The content and the recommendations of this research do not necessarily reflect the views of the ERF.

Send correspondence to:
Firat Bilgel
Okan University
firat.bilgel@okan.edu.tr

First published in 2017 by
The Economic Research Forum (ERF)
21 Al-Sad Al-Aaly Street
Dokki, Giza
Egypt
www.erf.org.eg

Abstract

This study assesses the causal effects of primary care utilization on subjective health status in Turkey using individual-level data from the 2012 Health Research Survey. Employing recursive bivariate models that take into account the possibility that selection into healthcare might be correlated with the subjective health status of the respondent, we find that selection into primary care is endogenously determined and that the utilization of preventive care significantly improves one's self-rated health after controlling for sociodemographics, socioeconomic status, health behavior and risk factors and access to healthcare. The distribution of treatment effects suggests significant between- and within-inequalities in health gains from preventive care utilization in disfavor of chronic patients. Analysis also points out that barriers to healthcare access are associated with lower self-rated health and that significant location-based inequalities exist in the utilization of preventive care among chronic patients. GP care utilization however, only exerts a trivial causal effect on self-rated health exclusively among females, rural residents and chronic patients.

JEL Classification: I12; I14; C31

Keywords: Recursive bivariate models, ordinal outcomes, treatment effects, primary care, Turkey

ملخص

تقيم هذه الدراسة الآثار السببية لاستخدام الرعاية الأولية على الحالة الصحية الذاتية في تركيا باستخدام بيانات على مستوى الأفراد من مسح البحوث الصحية لعام 2012. باستخدام نماذج متكررة ثنائية المتغيرات والتي تأخذ في الاعتبار إمكانية أن الاختيار في الرعاية الصحية قد تكون مرتبطة مع الحالة الصحية الذاتية للمستجيب، نجد أن الاختيار في الرعاية الأولية يتم تحديدها داخليا، وأن استخدام الرعاية الوقائية يحسن بشكل كبير من الصحة الذاتية التصنيف بعد السيطرة على الديموغرافيا الاجتماعية، والوضع الاجتماعي والاقتصادي، والسلوك الصحي وعوامل الخطر والحصول على الرعاية الصحية. ويشير توزيع آثار العلاج إلى وجود تفاوتات كبيرة وسوء توزيع للمكاسب الصحية من جراء استخدام الرعاية الوقائية واستياء المرضى المزمنين. ويشير التحليل أيضا إلى أن العوائق التي تحول دون الوصول إلى الرعاية الصحية ترتبط بصحة أقل تصنيفا، وأن أوجه التفاوت الكبيرة القائمة على الموقع موجودة في استخدام الرعاية الوقائية بين المرضى المزمنين. ومع ذلك، فإن استخدام الرعاية الصحية العامة لا يؤدي إلا إلى تأثير سببي لا يذكر على الصحة ذات التصنيف الذاتي حصرا بين الإناث وسكان الريف والمرضى المزمنين.

1. Introduction

Self-rated health (henceforth SRH) has been extensively used in epidemiological studies not only as a complex, inclusive and a reliable measure of population health but also a powerful predictor of mortality, morbidity (Idler and Benyamini, 1997, Jylha, 2009) and healthcare utilization (Fylkesnes, 1993; Miilunpalo et al., 1997; Geitona et al., 2007). While the literature on the determinants of SRH in developed countries is extremely rich using sociodemographical factors, socioeconomic status, physical functioning, psychosocial wellbeing, chronic history and health behaviors & risk factors as predictors of SRH, a number of studies exclusively focuses on particular subpopulations by gender and gender differences (Unden and Elofsson, 2006), ethnicity (Komar et al., 2006; Min et al., 2014), adolescent population (Vingilis et al., 2002), elderly population (Hoeymans et al., 1997; Kivinen et al., 1998; Menec and Chippereld, 2001), immigrants (Gross et al., 2001; Newbold, 2005) and those with chronic health problems and disability (Cott et al., 1999). Other studies focus on the effects of particular health behaviors and risk factors such as obesity (Ferraro and Yu, 1995; Prosper et al., 2009; Okosun et al., 2001; Cullinan and Gillespie, 2015), cholesterol (Goldman et al., 2004), exercise (Lamb et al., 1990; Jylha et al., 2001), alcohol consumption (Poikolainen et al., 1996; Guallar-Castillon et al., 2001; Stranges et al., 2006; Frisher et al., 2015), medical insurance (Hullegie and Klein, 2010), socioeconomic status (Bobak et al., 1998; Nicholson et al., 2005), income inequality (Kennedy et al., 1998) and social capital (Schultz et al., 2008; Snelgrove et al., 2009).[1]

Although the empirical strand of literature that aims at identifying the determining factors of healthcare utilization and the impact of utilization on health outcomes in Turkey is thin, a number of important studies attempt to explain inequalities in healthcare utilization and SRH. Using the 1993 Demographic and Health Survey, Celik and Hotchkiss (2000) investigate the socioeconomic determinants of prenatal care utilization among ever married women in Turkey and find maternal education to be one of the most important determinants of prenatal care utilization along with health insurance and household wealth. The analysis also points out to significant geographical (East/West), location- (rural/urban) and ethnicity-based inequalities in prenatal care utilization.

Using the 2003 World Health Survey data, Sozmen et al. (2012) measure the major sources of socioeconomic inequalities in SRH in Turkey using the concentration index (CI). The CI for SRH suggests that suboptimal health is more prevalent among less wealthy individuals and that low educational attainment, not being married, geographical disadvantage and poor economic status pose increased risks of having poor SRH. Two follow-up studies in Turkey using the 2008 Health Survey data focus on inequalities in non-communicable diseases, SRH and healthcare utilization. In the former study, Sozmen and Unal (2014) find that the majority of diseases are prevalent among less wealthy individuals and they are associated with low wealth and low educational attainment. In the latter study, Sozmen and Unal (2016) assess the inequalities in healthcare utilization among Turkish adults through the measurement of CI and the Horizontal Inequity Index (HI) for General Practitioner (GP) care, specialist care, inpatient care, dental care and emergency care. The results indicate that the poor are more likely to utilize emergency, inpatient and GP care while there is significant inequality in specialist care and oral heath care services in favor of the rich. Ergin and Kunst (2015) assess regional inequalities in SRH using 2002 World Health Survey for Turkey and find that regional differences are mostly driven by socioeconomic status. While the location-based (rural/urban) differences dissappear after controlling for the education level and the household wealth, regional

[1] Recently, a number of studies use comparative methods to identify the differential effects of the determinants of SRH across countries (Zimmer et al., 2000; Kasmel et al., 2004; Prus, 2011; French et al., 2012; Park and Lee, 2013; Campos et al., 2015).

differences remain only among older women with a higher prevalence of poor SRH particularly in the Eastern regions.

Our study distinguishes from the related literature to the extent of specifying (1) healthcare utilization as a causal determinant of SRH and (2) an endogenous relationship between SRH and healthcare utilization. First, we posit that if there is an association that runs from SRH to utilization as the empirical literature suggests; that is individuals are likely to resort to receive healthcare as a result of poor or ill-health, then naturally, a reverse causation from utilization to SRH exists. Upon the screening of the vast literature on the determinants of SRH, only very few studies do actually consider utilization as an important predictor of SRH (Damian et al., 1999; Darviri et al., 2011). Even in the presence of a unidirectional relationship that runs from SRH to utilization and that rejects reverse causation, it is evident that a measure as complex and inclusive as SRH will be correlated with a number of unobservable dimensions of healthcare utilization (i.e. SRH will be endogenous to utilization). We are not aware of any empirical study on the determinants of healthcare utilization that accounts for the endogenous nature of SRH despite the fact that it casts doubts on the identifiability and the validity of inference. In a nutshell, this problem requires the use of an instrument that can plausibly be viewed as randomly moving around one's SRH, which, we think, is extremely difficult to summon. Second, under a reverse causation scenario, unobservable confounders may be important to the selection and outcome processes and thus selection into care is likely to be correlated with the respondent's subjective assessment of health status. If so, a univariate specification of the relationship between SRH and healthcare utilization will be biased and inconsistent and a bivariate model that takes into consideration the possibility that the unobservable determinants of SRH will be correlated with the unobservable determinants of healthcare utilization has to be specified.

Figure 1 conceptualizes our framework and visualizes our research design. It is highly similar to those of House (2002) and Prus (2011) in terms of the direction of the effects of social determinants on SRH except for mainly two structural differences. First, healthcare utilization is central to our framework in understanding the social determinants of SRH. Therefore, some of the social determinants of SRH have indirect effects through utilization. For example, sociodemographic factors such as age, gender, location and marital status have direct effects on both SRH and utilization as well as indirect effects through shaping health behaviors & risk factors and socioeconomic status, albeit whether all the direct effects we do and do not conjecture exist is ultimately an empirical question. In our framework, sociodemographics, health behaviors & risk factors and access measures act as common causes of both SRH and utilization.

The second structural difference is related to our research design. Non-random selection into care is represented by the presence of the unoberservables in figure 1 that simultaneously determine SRH and utilization. Consequently, this makes impossible to identify the causal effect of utilization on SRH by univariate methods and/or covariate adjustment and prompt us to resort to bivariate models where we ideally search for an exogenous source of variation (i.e. instrumental variable) to help identify selection. If our conjecture is empirically supported, that is if utilization is in fact endogenous and that selection into care is non-random, the processes that determine SRH and utilization cannot be thought independently.

In the light of our elaborations, we attempt to identify the causal impact of primary care utilization on SRH. Since selection into healthcare is not random for some unobservable factors may determine healthcare utilization and SRH simultaneously, we employ an endogenous, latent-factor treatment effects model to infer the causal relationship between healthcare utilization and SRH where primary healthcare utilization is instrumented with the individual's knowledge of their family physician. Through estimating treatment effects, we find that

preventive care utilization significantly improves one's SRH after controlling for age, gender, location, marital status, chronic disease history, body mass index, frequency of exercise, juice, tobacco and alcohol consumption, education, income, type of health insurance and unmet healthcare needs. An advantage of employing ordered models is that it allows us to examine the distribution of treatment effects, suggesting significant between- and within-inequalities in health gains from preventive care utilization in disfavor of chronic patients. We also find that unmet healthcare needs as barriers to healthcare access are associated with lower self-rated health and that significant location-based inequalities exist in the utilization of preventive care among chronic patients. GP care utilization however, only exerts a trivial causal effect on SRH exclusively among females, rural residents and chronic patients.

Section 2 outlines our empirical strategy where healthcare utilization is treated as an endogenous covariate of interest when SRH is measured on an ordinal scale. Alternatively, we explore the case in which SRH is defined as an optimal/suboptimal dichotomous variable. Section 3 reports and discusses the results of the analysis by employing bivariate models in the presence of endogenous healthcare utilization. Section 4 concludes.

2. Empirical Strategy

2.1 Data and sample

We use individual-level data obtained from the Health Research Survey administered by the Turkish Statistics Office (TurkStat) in 2012. The survey comprises of a nationally representative sample of 37,979 respondents in Turkey and keeps track of a detailed account of socioeconomic, demographic, environmental and health related indicators, lifestyle and habits, chronic diseases, other physical and mental health problems, daily activities, primary and secondary healthcare utilization including oral, dental and hospital care.

The survey employs a two-stage stratified cluster sampling method. The urban-rural distinction is used as an external stratification criterion. In the first stage, the sampling units of blocks were chosen from clusters in which the average number of households is hundred. In the second stage, households are systematically selected from each cluster. In order to obtain parameter estimates through sampling with a representative power of the population, the dataset was subject to weighting where the sampling weights represent the inverse probability of being selected into the sample. All the variables used in our analysis were obtained through face-to-face interview and self-report. The raw data was not designed to readily serve our research objectives. We therefore recoded the entire data, excluded cases with missing values and cases for which respondents refused to answer and discarded all respondents under the age of 15. This procedure results in 24,022 (63.3 percent of the original sample size) observations to be used in the analysis.

Among different forms, primary care remains to be the most important form of healthcare to maintain the overall population health for it is the first line of defence; it is inexpensive, easily delivered than specialty or inpatient care and effective in the containment of diseases (Perry and Gesler, 2000; Guagliardo, 2004). A well-functioning primary healthcare system can prevent or reduce unnecessarily expensive specialty care. We analyze the causal effect of two types of primary care utilization on SRH: Preventive and GP care. Both utilization measures are coded as binary variables, taking the value of one if the respondent ever used the type of healthcare under consideration and zero otherwise. The SRH variable is coded on an ordinal scale from 1 to 5, 1 representing "very poor" health and 5 representing "very good" health. The control variables used in the analysis consist of indicators of sociodemographics (age, gender, location, marital status), socioeconomic status (education, income), health behavior and risk factors (chronic disease history, obesity, frequency of physical exercise, consumption of fruit and vegetable juice, tobacco and alcohol consumption) and access to healthcare (health

insurance, unmet healthcare needs). The questions and the coding of the variables can be found in table 1.

2.2 Descriptive statistics

Table 2 reports the descriptive statistics for the full sample and by utilization and SRH status respectively. Overall, 1646 participants (6.85 percent) rated their health status as either "poor" or "very poor", 5581 (23.23 percent) participants rated their health status as "average" and 16795 participants (69.92 percent) rated their health status as "good" or "very good". In terms of healthcare utilization, 87 percent of the participants declared to have utilized GP care. Strikingly, only 20 percent of the respondents in the sample declared to have ever used preventive care. In the sample, 3.7 percent of the respondents are underweight (BMI $<$ 18.5), 42.3 percent are normal weight (18.5 $<$ BMI $\leq$ 25), 35.1 percent are overweight (25 $<$ BMI $<$ 30) and 18.9 percent are obese (BMI $>$ 30)[2]. In terms of risk factors, half of the sample suffered from at least one chronic health problem.

In table 2, the first column shows the means and standard deviations for the entire sample and the remaining columns show the means by preventive and GP care utilization status respectively. A naïve comparison of SRH by utilization status reveals that those who utilize primary care have statistically and significantly lower levels of perceived health (poorer health) compared to non-utilizers. However, comparison by utilization status is problematic because of selection bias; that is those who utilize healthcare do so because they are different from non-utilizers. Thus the difference in SRH we observe in table 2 consists of an average causal effect of healthcare utilization on SRH for those who actually utilize healthcare plus a selection bias that consists of the difference between the SRH that would be observed among utilizers had they not utilized healthcare (counterfactual) and the SRH of those who actually do not utilize healthcare. A comparison of SRH by utilization status also points out that having lower (poorer) health ratings among utilizers might be the reason they utilize healthcare in the first place.

A comparison of other respondent characteristics by utilization status indicates that utilizers of healthcare have consistently and significantly higher incidence of chronic disease history and obesity over non-utilizers and that males are more reluctant to utilize heathcare. With respect to healthcare access, preventive care utilization is more (less) prevalent among publicly insured (out-of-pocket payers) and for those with unmet healthcare needs[3].

The last five columns in table 2 report the descriptive statistics by SRH status. There is a clear gradient for preventive care utilization, individuals over the age of 55, chronicity and obesity, being more prevalent at poorer SRH levels. However, GP care utilization does not exhibit observable differences in means across SRH levels. On the other hand, an optimistic self-assessment of health is more prevalent for singles and individuals below the age of 34. With respect to socioeconomic status and access to healthcare, there is a clear gradient for illiterate and poorest individuals and those who have unmet healthcare needs being more prevalent at poorer SRH levels.

---

[2] The survey does not directly ask/measure the individual's body mass index (BMI). We constructed this index using the height and the weight of the respondent by the following formula: $\text{BMI} = weight/(height)^2$.

[3] In the survey, the health insurance variable was categorized into uninsured (out-of-pocket), green card holders (health benefits to those who cannot afford to pay for healthcare), Social Security Institution (SGK) insureds, Ba$\breve{g}$-Kur (covers the self-employed), SSK (covers blue and white collar), Government Employees Retirement Fund, private insurance and private pension. We grouped SGK, Ba$\breve{g}$-Kur, green card holders, SSK and Government Employees Retirement Fund into the category "Public" and private insurance and private fund into "private". See table 1.

2.3 Primary care utilization as an endogenous treatment

2.3.1 SRH as an ordinal outcome

Our outcome of interest is the individual's SRH, measured on an ordinal scale with J possible ordered outcomes, $j = 1,\ldots,J$. Let $T_i \in \{0,1\}$ be the binary treatment variable that takes the value of one if the individual utilizes primary care and zero if the individual does not. The selection equation is:

$$T_i = \begin{cases} 1 & \text{if} \quad T_i^* = Z_i\gamma + \upsilon_i > 0 \\ 0 & \text{if} \quad T_i^* = Z_i\gamma + \upsilon_i \leq 0 \end{cases} \tag{1}$$

where Z_i is the set of covariates in the selection equation and υ_i is the error term. The outcome equation can be written as:

$$Y_i = \begin{cases} 1 & \text{if} \quad -\infty < Y_i^* \leq \mu_1 \\ 2 & \text{if} \quad \mu_1 < Y_i^* \leq \mu_2 \\ \vdots & \qquad\vdots \\ J-1 & \text{if} \quad \mu_{J-1} < Y_i^* \leq \mu_J \\ J & \text{if} \quad \mu_J < Y_i^* \leq \infty \end{cases} \tag{2}$$

where μ_j's are the cutpoints. The latent outcome variable Y_i^* is defined as:

$$Y_i^* = X_i\beta + \delta T_i + \varepsilon_i \tag{3}$$

where ε_i is the idiosyncratic error term and X_i are the covariates of the outcome equation that include individual characteristics such as age, gender, chronic disease history, location, lifestyle, habits and educational attainment.

In order to obtain consistent estimates, the latent errors υ_i and ε_i should follow a bivariate joint normal distribution with correlation ρ. If $\rho = 0$, then equation (2) can be estimated by generalized ordered probit. If $\rho \neq 0$, then the unobservable determinants of health care selection are said to be correlated with the unobservable determinants of SRH, rendering health care utilization endogenous for the fact that individuals may observe a deteriorating health status and may choose to resort to receive healthcare or may receive healthcare and observe health status. The possibility of endogeneity requires a joint estimation of equations (1) and (3) via maximum-likelihood to obtain consistent and asymptotically efficient estimates.

To account for the possibility that healthcare utilization may be endogenous to SRH, a source of exogenous variation should be found such that it might plausibly be viewed as randomly moving around the choice to resort to primary care. In practice, this source of exogenous variation helps identify selection into care. It should be (strongly) correlated with utilization (i.e. relevant), should exhibit an impact on SRH through and only through the choice to resort to healthcare and should not be directly related to SRH (i.e. excluded) or the latent errors of the model (i.e. clean). A primal candidate that satisfies such properties is the individual's knowledge of their family physician. We posit that individuals who are aware of their family physician are more likely to utilize preventive and GP care and that the physician knowledge has no direct, evident relation to one's subjective health status.

First, such knowledge does not necessarily presuppose healthcare utilization in Turkey because individuals are informed online of the identity of their GP as long as the individual is registered in the district of residence. In fact, in the sample, 9.9 percent of the respondents do know the identity of their family physician but have never utilized healthcare, 22.4 percent does not know

this information and have not received care and the remaining 67.7 percent know their family physician and have actually received care.

Second, we argue that such designation facilitates to come into contact with the healthcare provider when necessary. Since the socialization law of 1964 until 2010, a designated family physician did not exist and patients had to receive treatment from whoever was attending when care was sought. Especially for healthcare services that require follow-up, patients were reluctant to receive care or unwilling to be monitored knowing that a different physician will be assigned at each visit. The implementation of a designated family physician by the Ministry of Health was exactly meant to rectify this situation. From the patient's point of view, this creates a sense of belonging and that they will be able to receive treatment from the same physician at every visit without having to repeatedly brief about their medical or follow-up history. We therefore expect the coefficient of family physician knowledge to exert a statistically and significantly positive impact on primary healthcare in the selection equation.

For any j, Let $Y_i(1)$ be the SRH for individual i, if the individual utilized healthcare and let $Y_i(0)$ be the SRH if the individual did not utilize healthcare. The unit-level causal effect of healthcare utilization on SRH for any j, is the difference between the observed SRH under healthcare utilization and the missing potential SRH that would be observed had the individual not utilized healthcare, that is $\tau_i = Y_i(1) - Y_i(0)$.

The average treatment effect (ATE) shows the expected effect of healthcare utilization on SRH for a randomly drawn individual from the population for a given level of j:

$$\begin{aligned} ATE_j &= E[Y_i(1) - Y_i(0) \mid j] = E[Y_i(1) \mid T = 1, j] - E[Y_i(1) \mid T = 0, j] \\ &= \frac{1}{N}\frac{1}{S}\sum_{i=1}^{N}\sum_{s=1}^{S}[\Phi\{\mu_k - (X_i\beta + \delta + \lambda\eta_{is})\} - \Phi\{\mu_{k-1} - (X_i\beta + \delta + \lambda\eta_{is})\}] \\ &\quad - [\Phi\{\mu_k - (X_i\beta + \lambda\eta_{is})\} - \Phi\{\mu_{k-1} - (X_i\beta + \lambda\eta_{is})\}] \end{aligned} \tag{4}$$

where $k = 1, \ldots, K$, $K = J + 1$, $\mu_K = \infty$, N is the number of observations, S is the number of simulation draws, Φ is the standard normal cumulative distribution and δ is the coefficient of our endogenous binary treatment, healthcare utilization.

The other measure of interest is the average treatment effect on the treated (ATT). The ATT shows the expected effect of healthcare utilization for a randomly drawn individual only from those individuals who utilized healthcare for a given level of j:

$$\begin{aligned} ATT_j &= E[Y_i(1) - Y_i(0) \mid T = 1, j] = E[Y_i(1) \mid T = 1, j] - E[Y_i(0) \mid T = 1, j] \\ &= \frac{1}{N}\frac{1}{S}\sum_{i=1}^{N}\frac{1}{E\{\Phi(Z_i\gamma)\}}\Big(\sum_{s=1}^{S}\Phi(Z_i\gamma + \eta_{is})[\Phi\{\mu_j - (X_i\beta + \delta + \lambda\eta_{is})\} - \Phi\{\mu_{j-1} - (X_i\beta + \delta + \lambda\eta_{is})\} \\ &\quad - \Phi\{\mu_j - (X_i\beta + \lambda\eta_{is})\} + \Phi\{\mu_{j-1} - (X_i\beta + \lambda\eta_{is})\}]\Big) \end{aligned} \tag{5}$$

A crucial implication of the bivariate model is that if the unobservables in the outcome and selection equations are not jointly distributed normally, the estimates of the model will be inconsistent and biased. We relax the assumption of joint normality of the unobservables in the treatment and the selection equations by following a latent-factor approach and reformulate the error-generating process in the following way (Gregory, 2015):

$$\begin{aligned} \upsilon_i &= \lambda_T\eta_i + \varsigma_i \\ \varepsilon_i &= \lambda_Y\eta_i + \iota_i \end{aligned} \tag{6}$$

where λ_T and λ_Y are the loading factors describing the dependence of the latent errors for the treatment and the outcome respectively and only the marginal distributions of ς and ι are assumed to be normal. Simulating the distribution of by taking random draws form its chosen distribution, the likelihood function can be written as:

$$L_i = \frac{1}{S}\prod_{i=1}^{N}\sum_{s=1}^{S}\Phi\{\tau\times((Z_i\gamma+\lambda_T\eta_i))\} \times\sum_{k=1}^{K}\{I\times(Y=k)\}\{\Phi(\mu_k - X_i\beta+\lambda_Y\eta_i)-\Phi((\mu_{k-1}-X_i\beta+\lambda_Y\eta_i))\} \quad (7)$$

2.3.2 SRH as a binary outcome

Following Sozmen et al. (2012) and Sozmen and Unal (2016), we alternatively define our SRH measure by classifying it as a binary variable where "very good", "good" and "average" are coded as "optimal health" taking the value of 0 and "poor" and "very poor" are coded as "suboptimal health" taking the value of 1. In this case, our outcome of interest becomes the likelihood to report suboptimal SRH. We assume that SRH, Y_i is determined by the latent index

$$Y_i = 1\left[X_i'\beta+\delta T_i > \varepsilon_i\right] \quad (8)$$

where $1[.]$ is the indicator function taking the value of 1 if the statement in the brackets is true and 0 otherwise.

The treatment equation is given by the following:

$$T_i = 1\left[Z_i'\gamma_1+\gamma_0 W_i > \upsilon_i\right] \quad (9)$$

where Z_i is the set of covariates, W_i is the instrumental variable, and υ_i is the error term.

The model is identified by assuming that the instrumental variable W_i is independent of ε_i, υ_i and X. Given normality, the model above can be estimated via maximum likelihood which yields consistent and asymptotically efficient estimates. The ATE and the ATT are given by:

$$\begin{aligned}\text{ATE} &= E\left[Y_i(1)-Y_i(0)\right] = E(Y_i(1)\mid T=1)-E(Y_i(1)\mid T=0)\\ &= E\left\{1\left[X_i'\beta+\delta>\varepsilon_i\right]-1\left[X_i'\beta>\varepsilon_i\right]\right\}\end{aligned} \quad (10)$$

$$\begin{aligned}\text{ATT} &= E\left[Y_i(1)-Y_i(0)\mid T=1\right] = E(Y_i(1)\mid T=1)-E(Y_i(0)\mid T=1)\\ &= E\left\{1\left[X_i'\beta+\delta>\varepsilon_i\right]-1\left[X_i'\beta>\varepsilon_i\right]\mid Z_i'\gamma_1+\gamma_0 W_i>\upsilon_i\right\}\end{aligned} \quad (11)$$

where $Y_i(1)-Y_i(0)$ denotes the difference in SRH due to healthcare utilization.

Another approach, advocated by Angrist and Pischke, is to employ an instrumental variable (IV) estimation on equation (8) disregarding the binary nature of the outcome and simply use W_i to instrument T_i (Angrist and Pischke, 2009). As pointed out by Imbens and Angrist however, linear IV methods capture local average treatment effects (LATE) independent of whether the outcome variable is binary, non-negative, or continuous, but do not guarantee an accurate measurement of ATE (Imbens and Angrist, 1994). In our context, LATE can be interpreted as the expected effect of healthcare utilization on SRH for a randomly drawn individual who could be prompted to utilize healthcare by making them aware of their family physician, W. For any two values, W_0 and W_1, of the instrument, the corresponding LATE is: $E\left[Y_i(1)-Y_i(0)\mid T_i(W_1)=1, T_i(W_0)=0\right]$, where $T_i(W_1)$ is the potential treatment status when the

instrument takes the value of W_1 and $T_i(W_0)$ is the potential treatment status when the instrument takes the value of W_0.

3. Results

In order to gain some preliminary knowledge on the endogeneity of primary care utilization, we first follow the Angrist-Pischke approach and estimate equation (8) through the limited information maximum likelihood (LIML) method and instrument T_i (preventive or GP care utilization) by W_i (the family physician knowledge). The results are displayed in table 3. Our aim is not to draw inferences regarding the impact of non-utilization variables on binary SRH or to estimate the LATE. Therefore, we confine ourselves to the diagnostic tests reported at the bottom of the table. In column (1) of table 3, we estimate equation (8) for preventive care. The underidentification test and the associated p-value suggest that the excluded instrument (family physician knowledge) is extremely relevant. Further, the results of the endogeneity test suggests that preventive care utilization should be treated as endogenous at conventional test levels. In column (2), we repeat the same exercise for GP care. In contrast to the preventive care model, the result of the endogeneity test suggests that GP care can actually be treated as exogenous, although the instrument is still strongly relevant.

3.1 Preventive care

The possiblity that selection into healthcare might be correlated with the respondent's SRH status is first captured by a bivariate ordered probit model whose results are reported in table 4 for preventive care. As these models rely on the assumption that the unobservables in the outcome and the selection equations are jointly normally distributed yet the unobservables rarely exhibit a bivariate normal distribution in practice, we follow a latent-factor approach[4]. We assume that the distribution of the unobservables is skewed and that the coefficient of the loading factors in the error-generating process in equation (6) follow a gamma distribution[5]. In table 4, we estimate the parameters of the model using maximum simulated likelihood.

We first confine ourselves to the paramater estimates of non-utilization variables. Column (1) of table 4 displays the full sample results. Respondents tend to report an increasingly pessimistic self-assessment of health as they grow older and become exposed to multiple health problems and illnesses associated with ageing (Franks et al., 2003; Szwarcwald et al., 2005). On the other hand, those aged below 65 are less likely to utilize preventive care and the relation between the age gradient and healthcare utilization follows a U-shaped pattern in the full sample, with an effect getting stronger in the middle-age groups (35 to 54). It has been reported that poor SRH is associated with females while holding other factors constant, regardless of the socioeconomic and cultural differences in the samples being studied (Cott et al., 1999; Molarius and Janson, 2002; Asfar et al., 2007; Moradi-Lakeh et al., 2015). While males are less likely to use preventive care, they tend to be more optimistic in their SRH relative to females. Those who reside in rural areas are more likely to report a pessimistic self-assessment of health but location appears not to affect preventive care utilization. Interestingly, our findings indicate that marital status does not have any impact on SRH in the full sample; however we find that singles are less likely to use preventive care. One explanation for this

[4] We use **treatoprobitsim** command in Stata (Gregory, 2015).

[5] Other distributions for the density of the latent factor are uniform, normal, lognormal, logistic and chi-square. While normal and lognormal distributions are direct transformations of each other, the chi-square is a special case of gamma distribution. On the other hand, the logistic and the uniform are notably distant to and have no direct or asymptotic relationships with the rest of these continuous distributions. In the preliminary analysis, we estimated the bivariate ordered probit model under each of these distributional assumptions. While the chi-square and the gamma distributions yield almost identical results both in terms of the direction as well as the statistical significance of all parameters, the ATE and the ATT under the normal and the uniform distributions have signs contradictory to our expectations. As for the logistic, lognormal and the uniform distributions, selection into care is found to be independent of the unobservables that determine SRH; a finding that contradicts with our initial conjecture.

lower likelihood is that, singles tend to be less conscious and caring of their health relative to couples who can monitor each other's health status and inculcate their partner to seek medical advice; an interaction undefined or absent among singles. Another explanation is that singles tend to be much younger in Turkey compared to non-singles, they are less likely to experience health problems and are therefore in less need to resort to healthcare.

Respondents suffering from chronic health problems are more likely to use preventive care but also more likely to report a pessimistic self-assessment of health. Regular smokers tend to have a pessimistic health assessment relative to non-smokers (Manderbacka et al., 1999) and expectedly they are more likely to use preventive care services. Regular alcohol consumption on the other hand, does not affect SRH whereas alcohol consumption and suboptimal health typically follows a J-shaped pattern in Nordic countries (Poikolainen et al., 1996) and a monotonically inverse relationship in the Mediterranean (Guallar-Castillon et al., 2001).

The relationship between obesity and SRH has proved to be unequivocal in developed countries (Ferraro and Yu, 1995; Manderbacka et al., 1999; Darviri et al., 2011), even in the absence of chronic conditions (Okosun et al., 2001). Consistently, we find that those with a body mass index (BMI) greater than 30 (i.e. obese) are more likely to report a pessimistic self-assessment of health. With respect to other health behaviors and risk factors, a very low frequency (once or twice a week) of ten-minute walks has no significant impact on one's SRH. However, respondents tend to be more optimistic about their SRH when exercised at least thrice a week; a finding consistent with those reported by reported by Lamb et al. (1990); Cott et al. (1999) and Piko (2000). Similarly, those who consume fruit or vegetable juices are more likely to be optimistic about their perceived health. Both of these effects become stronger with an increasing frequency of exercise and juice consumption.

With respect to socioeconomic status, we find that those with an educational attainment below higher education tend to have pessimistic SRH levels with an effect getting stronger as we move towards lower education brackets. However, we do not find conclusive and strong evidence as to the impact of education on healthcare utilization, yet observe that those with a primary education are less likely to use preventive care services relative to those in the remaining brackets. In all income quintiles, respondents tend to report pessimistic health levels with an effect getting stronger as we move towards the bottom income quintiles[6]. Several factors such as malnutrition due to lack of purchasing power, improper care and health risks associated with residing in areas that heavily use low-cost but eco-unfriendly means of heating might explain the relationship between poor economic status and poor health in Turkey.

With respect to healthcare access, those who are publicly insured under the auspices of the Social Security Institution (SGK) are more likely to use preventive care relative to private or no insurance but regardless of the type, insurance does not affect one's SRH. In a striking contrast, those with unmet healthcare needs report significantly pessimistic health assessment, suggesting that barriers in access to healthcare pose higher risks of reporting poor or ill-health. However, having unmet healthcare needs do not affect the probability of utilizing preventive care. A possible explanation is that the supply of preventive care services in Turkey are continuously monitored; patient follow-up is strictly enforced by the Ministry of Health and physicians that fail patient monitoring are sanctioned by an administrative fine.

Our results, in line with our expectations, suggest that individuals who are aware of their family physician are more likely to use preventive care services. From table 4, the individual's

[6] Income does not appear in the final selection equation since healthcare costs in Turkey are compensated by the public healthcare system to a significant extent. Our preliminary findings also confirm that household income has no bearing on one's propensity to use healthcare services.

knowledge of their family physician has a positive and statistically significant impact on the probability of preventive care utilization in the selection equation.

To assess the causal effects of healthcare utilization on SRH, we report the ATE and the ATT at the bottom of table 4. Since the outcome variable is measured on an ordinal scale, the ATE and the ATT are reported for every level of health among J possible ordered health outcomes. The ATE indicates that in the full sample reported in column (1), preventive care utilization decreases the probability for an individual to report very poor, poor and average health by 0.5, 2.7 and 6.1 percent respectively and increases the probability to report very good health by 7.8 percent. Among those who actually utilize healthcare, ATT shows that preventive care utilization decreases the probability for an individual to report very poor, poor and average health by 2, 6.7 and 7.5 percent respectively and increases the probability to report good and very good health by 11.3 and 4.9 percent respectively.

To understand the differential impact of preventive care utilization on the male versus female respondents, rural versus urban residents and chronic versus non-chronic patients, we divide the full sample by gender, location and chronic status and report the regression results in columns (2) through (7) respectively. In contrast to the full sample estimates, both single men and women are more likely to report an optimistic health assessment. We find that females with a BMI > 30 tend to be more pessimistic about their health whereas males are not. Regular alcohol consuming males are also more likely to report pessimistic health levels and more likely to resort to preventive care utilization. Males (females) are also more (less) likely to utilize preventive care at ages above (below) 64. Much of the effect of marital status and education on healthcare utilization we observe in the full sample is driven by females who are single and who attained primary education.

With respect to treatment effects, preventive care utilization increases the probability to report very good health among males by about 15.3 percent, whereas this effect is slightly lower (9.1 percent) for females at conventional test levels. Among males (females) who actually utilize healthcare, preventive care utilization increases the probability to report good and very good health by 25.3 (13.1) and 7.6 (5.1) percent respectively.

One notable difference for rural and urban resident models in columns (4) and (5) of table 4 is that the former group consistently benefits from regular physical exercise regardless of the frequency whereas the latter group does so when exercised on a daily-basis. With respect to health insurance, while privately insured urban residents tend to report an optimistic health assessment, the type of insurance neither affect SRH among rural residents nor the choice to receive preventive care regardless of location. On the other hand, we find that being divorced negatively affects the choice to utilize preventive care among urban residents but not among rural residents. For either locations, singles are consistently less likely to utilize preventive care.

The last two columns in table 4 reports the estimates by chronic status. Chronic patients tend to benefit from physical exercise whereas non-chronic patients do not. We also find that chronic patients living in rural areas are less likely to utilize preventive care; however location does not affect non-chronic respondents' propensity to utilize healthcare. With respect to marital status, consistent with the findings of Cott et al., (1999), singles and married individuals who do not suffer from a chronic health problem tend to report an optimistic health assessment and are less likely to utilize preventive care; a link unobserved among chronic patients. Expectedly, chronic patients over the age of 64 are morely likely to utilize healthcare. However, seniors (75+) and those below the age of 55 who did not experience any chronic health problem are less likely to use preventive care. While obesity is associated with a pessimistic self-assessment of health for chronic patients, it has no effect on SRH in the absence of chronic disease conditions.

Expectedly, obesity can lead to a number of chronic health problems and its effect on SRH appears to be mediated through chronic conditions.

In all seven models in table 4, the largest gains from preventive care utilization favor males who are also more optimistic about their SRH. We further observe a clear gradient of age, healthy consumption and socioeconomic status on SRH throughout all the models considered. Unmet healthcare needs as a barrier to healthcare access, have adverse effects on SRH regardless of gender, location or chronicity but they do not affect the likelihood to receive preventive care. With respect to other determinants of utilization, the clearest picture appears for females and chronic patients who are more likely to utilize preventive care. The relation between ageing and healthcare utilization follows a U-shaped pattern among urban and non-chronic patients and a J-shaped pattern among females and rural residents. For males and chronic patients, this relationship appears to be monotonically positive for the elderly (i.e. ageing increases the probability to utilize care). Further, consistent with the diagnostics of the Angrist-Pischke model, the likelihood-ratio (LR) test of λ reported at the bottom of table 4 indicates that the unobservables that determine preventive care utilization tend to move with the unobservables that affect perceived health, rendering preventive care endogenous.

Based on table 4, we plot the distribution of the probability of selection into preventive care for the full sample and by gender, location and chronicity in figure 2. For the full sample distribution, shown in figure 2a, the thick left tail of the distribution suggests that selection into preventive care is very unlikely (in the chunks of 1 percent) for an overwhelming proportion of the respondents and ranges up to 40 percent for a very small group of respondents with an average probability of selection of 6.7 percent.

With respect to the differential impact, the right-skewing distribution in figure 2c suggests that the probability of selection into preventive care does not differ significantly by location, with a probability of 6.3 and 8 percent for rural and urban residents respectively. However, we observe some differentiation in the selection probability by gender and that females are more likely (10 percent) to be selected into preventive care than males (3.4 percent). The most striking case concerns the distribution of the probability of selection by chronic status. Expectedly, patients without a chronic disease history are less likely (1.9 percent) on average to be selected into preventive care than those with at least one chronic health problem (11.7 percent). Overall, figure 2 suggests that within-inequality in the utilization of preventive care is more dramatic for males and rural residents and that there exists a between-inequality in the utilization of preventive care with respect to gender and notably with respect to chronic status.

Figures 3 to 6 respectively show the distribution of treatment effects (TE) and treatment effects on the treated (TT) for the full sample, by gender, by location and finally by chronic status. For the full sample, the distribution pattern suggests that the gains from preventive care utilization are distributed highly unequally. The distribution of TE and TT by gender and location in figures 4 and 5 respectively show no apparent between-gender- or between-location-inequality in health gains from preventive care utilization as much as it shows within-gender- and within-location inequalities. On the other hand, figure 6 suggests significant between- and within-inequalities in health gains from preventive care utilization, overwhelmingly in favor of patients without a chronic disease history.

In order to identify the causal effect of healthcare utilization on the probability of reporting suboptimal SRH, we further employ a bivariate probit model to account for the endogenous nature of healthcare utilization where SRH is now defined as a dichotomous variable taking

the value of 1 for suboptimal health and 0 otherwise[7]. Table 5 reports the results. The standard errors are computed via bootstrapping in all specifications.

With respect to the effect of sociodemographics, socioeconomic status, health behaviors & risk factors and access to healthcare on SRH and preventive care utilization, the direction of the impact of the covariates included in the model depicts a consistent picture to that reported in table 4. The only exceptions are the impact of marital status on SRH and the impact of location on the probability to utilize preventive care. While we find no impact of marital status on SRH for the full sample in table 4, the bivariate probit estimates in table 5 suggest that singles and divorced are more likely to report suboptimal SRH. Similarly, the selection equation in table 5 shows that rural residents are less likely to use preventive care whereas this effect is not statistically significantly different from zero at conventional test levels in the selection equation in table 4.

For the full sample shown in column (1), the ATE or the average causal effect of reporting suboptimal SRH indicates that preventive care utilization decreases the probability to report suboptimal SRH by 3.5 percent. On the other hand, the ATT shows that preventive care utilization decreases the probability to report suboptimal SRH for those who actually utilize preventive care by about 7.7 percent. However, this effect is not statistically distinguishable from zero at conventional test levels. In the sample, the probability or the relative frequency of having utilized preventive care, p_T is 0.1294. This implies that the average treatment effect on the control (i.e. non-utilizers), ATC is -0.029[8].

At the bottom of column (1) of table 5, the size of the error correlation and the Wald test indicate that the error terms, ε_i and υ_i are highly and statistically significantly positively correlated, confirming that preventive care utilization is endogenous. Violation of the assumption of joint normal distribution of error terms invalidates the estimation of a bivariate probit model and the associated statistical inference. Therefore, Murphy's score test results are reported at the bottom of table 5 (Murphy, 2007; Chiburis et al., 2012). The results show that the error terms are not jointly bivariate standard normal at conventional test levels for all models except that for chronic patients, invalidating statistical inference. However, for chronic patients the Murphy's score test indicates that the assumption of joint bivariate normal error distribution cannot be rejected. Therefore valid inference can only be made regarding column (6) of table 5. Accordingly, the ATE and the ATT in column (6) respectively suggest that preventive care utilization among those with a chronic disease history reduces the probability to report suboptimal health by about 6.3 percent and by 9.9 percent among those who actually utilize preventive care. However, neither treatment effects are statistically distinguishable from zero at conventional test levels.

3.2 GP care

For the GP care models where the outcome variable is the ordinal SRH, the direction of the impact of need and non-need variables in bivariate ordered models are in line with those of the preventive care model. However, neither the ATE nor the ATT are statistically distinguishable from zero at conventional test levels. The likelihood-ratio (LR) test of λ also indicates that the outcome and the selection equations are not dependent for the GP care model, confirming our previous findings that GP care utilization can actually be treated as exogenous. Further, the ordered probit models under exogeneity of GP care utilization fail to meet the parallel trends assumption even under a partial proportional odds model[9]. We therefore report here only the

[7] We use the biprobittreat command in Stata (Chiburis et al., 2012), available at: https://webspace.utexas.edu/rcc485/www/code.html (last visited Dec. 20, 2013)

[8] $ATE = p_T(ATT) + (1 - p_T)ATC = -0.035 = (0.1294)(-0.077) + (1 - 0.1294)(ATC)$

[9] The results of these models are available from the authors upon request.

results for the probit models where the outcome variable is the binary SRH taking the value of 1 for "suboptimal health" and 0 for "optimal health" as defined in section 2.3.2.

The results are reported in table 6. As in the previous section, column (1) reports the full sample results; columns (2) to (7) assess the differential impact of GP care utilization by gender, location and by chronic disease history respectively. In contrast to preventive care models, gender and location do not affect the probability to report suboptimal SRH in the GP care model. For the impact of the rest of the covariates, a picture similar to that reported in tables 4 and 5 emerges. In table 6, the adverse effect of being a single on SRH is driven by males. While pro-health behaviors such as exercise and juice consumption are associated with a lower likelihood to report suboptimal SRH, risk factors still play a negative role on SRH throughout the subgroups by gender, location and chronicity. We observe that individuals in the bottom quintiles as well as those in lower education brackets are also more likely to report suboptimal SRH levels relative to those in the top with an effect getting stronger as we move to bottom brackets.

For the full sample shown in column (1), the AME or the average marginal effect reported at the bottom of the table shows that GP care utilization decreases the probability to report suboptimal SRH by about 1 percent. With respect to the differential impact of GP by gender, location and chronicity, while GP care utilization decreases the probability to report suboptimal SRH by 2.3, 1.2 and 2.1 percent for females, rural residents and patients with a chronic disease history, we find that GP care utilization does not exert a causal effect on males, urban residents and non-chronic patients. In table 6, we further report the AME by gender, location and chronicity to assess whether the causal effect of GP care utilization differs within the subgroups. We find that respondents in higher age groups are more likely to benefit from GP care. This patterns also repeats within females, rural residents and chronic patients.

The gender-specific AME of GP care utilization in columns (1), (4) and (6) and the location-specific AME in columns (1), (3) and (6) in table 6 suggest that the probability to report suboptimal SRH does not differ across gender and location. Consistent with the implications of the models for preventive care in table 4, the gain in self-perceived health as a result of GP care utilization is much larger for chronic patients than for those without any chronic health problem.

4. Conclusion

This study investigated the causal impact of primary care utilization on SRH and their correlates in Turkey using the 2012 Health Research Survey dataset. Our empirical strategy is grounded on the premise that selection into healthcare is not random and that both healthcare utilization and the individual's subjective health status may be simultaneously determined by unobservable factors. The possibility that selection into healthcare might be correlated with the respondent's SRH and therefore endogenous, did not prove to be unsubstantiated for preventive care services. To address the endogeneity concerns, we employed an endogenous latent-factor treatment effects model and used the individual's knowledge of their family physician as an instrument for primary healthcare utilization even though the nonlinear models such as the ones we consider in this study are not bound to be predicated on the use of instruments. We find that for ordered SRH levels, preventive care utilization significantly increases one's subjective health after controlling for a number of factors that are thought to affect SRH and selection into primary care.

On the other hand, the preliminary analysis as well as the diagnostics of the recursive bivariate models did not show any compelling evidence on the endogeneity of GP care utilization and we employed univariate probit to model our binary SRH measure under GP care. Our findings suggest that although the utilization of GP care affects one's SRH, this causal effect is very small and only marginally statistically significant.

For both type of healthcare services, the associations we found between SRH or healthcare utilization and their correlates are consistent with the empirical findings of the broader literature we surveyed. We find that residents in rural areas, singles, undereducated and the poor are more likely to report pessimistic or suboptimal SRH but they are less likely to utilize healthcare. In all models we considered, pro-health behaviors are found to be associated with an optimistic self-assessment and such individuals are less likely to report suboptimal SRH. On the other hand, risk factors are consistently associated with poor or ill-health and they prompt a higher likelihood to come into contact with the healthcare system.

With respect to healthcare access, one dimension (unmet healthcare needs) significantly lowers one's SRH but does not impede selection into preventive care; the other dimension (public health insurance) does not affect SRH yet yields a higher likelihood of selection into preventive care. Location-based inequalities in the utilization of preventive care seem to exist among patients with a chronic disease history and these inequalities disfavor rural areas.

Albeit somewhat complicated in their intepretation, ordered models provide a great deal of advantage in assessing the distribution of both selection into care and the subjective gains of utilizing healthcare. The distribution patterns point out to between-inequalities in selection into preventive care with respect to gender and chronic status and between- and within-inequalities in health gains from utilization in disfavor of patients with a chronic disease history.

A number of caveats of our analysis are in order. First, we are unable to observe the exact location of the respondents. The lack of georeferenced data prevents us from exploring the spatial dimensions of the relationship between healthcare utilization and health status as well as the impact of spatial accessibility on healthcare utilization. With such data, one could further assess regional inequalities in both SRH and healthcare utilization and identify regions of priority. Second, the Health Research Survey we use is not originally designed to study healthcare utilization and the measurement of SRH and various measures on health behaviors & risk factors are based on self-reporting that may introduce information and recall bias. Finally, the comparison of SRH across gender, location or chronic status may be subject to different cut points. For example, what may be perceived as "very good health" for chronic patient could be regarded as "good health" for those without any chronic disease history.

References

Angrist, J. and J. Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Asfar, T. B. Ahmad, S. Rastam, T.P. Mulloli, K.D. Ward, and W. Maziak. 2007. Self-rated health and its determinants among adults in syria: a model from the middle east. *BMC Public Health*, 7(1):1.

Bobak, M., H. Pikhart, C. Hertzman, R. Rose, and M. Marmot. 1998. Socioeconomic factors, perceived control and self-reported health in russia. a cross-sectional survey. *Social Science & Medicine*, 47(2):269–279.

Campos, A.C.V.C. Albala, L. Lera, H. Sánchez, A.M.D. Vargas, and E.F. Ferreira. 2015. Gender differences in predictors of self-rated health among older adults in brazil and chile. *BMC public health*, 15(1):1.

Celik, Y. and D.R. Hotchkiss. 2000. The socio-economic determinants of maternal health care utilization in turkey. *Social Science & Medicine*, 50(12):1797–1806.

Chiburis, J. R. Das, and M. Lokshin. 2012. A practical comparison of the bivariate probit and linear iv estimators. *Economic Letters*, 117:762–66.

Cott, C.A., M.A. Gignac, and E.M. Badley. 1999. Determinants of self rated health for canadians with chronic disease and disability. *Journal of Epidemiology and Community Health*, 53(11):731–736.

Cullinan, J. and P. Gillespie. 2015. Does overweight and obesity impact on self-rated health? evidence using instrumental variables ordered probit models. *Health Economics*, DOI: 10.1002/hec.3215.

Damian, J., A. Ruigomez, V. Pastor, and J.M. Martin-Moreno. 1999. Determinants of self assessed health among spanish older people living at home. *Journal of Epidemiology and Community Health*, 53(7):412–416.

Darviri, C., A. K Artemiadis, X. Tigani, and E.C. Alexopoulos. 2011. Lifestyle and self-rated health: a cross-sectional study of 3,601 citizens of athens, greece. *BMC Public Health*, 11(1):1.

Ergin, I. and A.E. Kunst. 2015. Regional inequalities in self-rated health and disability in younger and older generations in turkey: the contribution of wealth and education. *BMC public health*, 15(1):1.

Ferraro, K.F. and Y. Yu. 1995. Body weight and self-ratings of health. *Journal of Health and Social Behavior*, pages 274–284.

Franks, P., M.R. Gold, and K. Fiscella. 2003. Sociodemographics, self-rated health, and mortality in the us. *Social Science & Medicine*, 56(12):2505–2514.

French, D.J., C. Browning, H. Kendig, M.A. Luszcz, Y. Saito, K. Sargent-Cox, and K.J. Anstey. 2012. A simple measure with complex determinants: investigation of the correlates of self-rated health in older men and women from three continents. *BMC Public Health*, 12(1):1.

Frisher, M., M. Mendonça, N. Shelton, H. Pikhart, C. de Oliveira, and C. Holdsworth. 2015. Is alcohol consumption in older adults associated with poor self-rated health? cross-sectional and longitudinal analyses from the english longitudinal study of ageing. *BMC public health*, 15(1):1, 2015.

Fylkesnes, K. 1993. Determinants of health care utilization - visits and referrals. *Scandinavian Journal of Public Health*, 21(1):40–50.

Geitona , M., D. Zavras, and J. Kyriopoulos. 2007. Determinants of healthcare utilization in greece: implications for decision-making. *The European Journal of General Practice*, 13(3):144–150.

Goldman, N., D.A. Glei, and M. Chang. 2004. The role of clinical risk factors in understanding self-rated health. *Annals of epidemiology*, 14(1):49–57.

Gregory, C.A. 2015. Estimating treatment effects for ordered outcomes using maximum simulated likelihood. *The Stata Journal*, 15(3):756–774, 2015.

Gross, R., S. Brammli-Greenberg, and L. Remennick. 2001. Self-rated health status and health care utilization among immigrant and non-immigrant israeli jewish women. *Women & Health*, 34(3):53–69.

Guagliardo, M.F. 2004. Spatial accessibility of primary care: concepts, methods and challenges. *International Journal of Health Geographics*, 3(1):1–13.

Guallar-Castillón, P., F. Rodrguez-Artalejo, L.D. Gañán, P.L Banegas, J.R.and Urdinguio, and R.H. Cabrera. 2001. Consumption of alcoholic beverages and subjective health in spain. *Journal of Epidemiology and Community Health*, 55(9):648–652.

Hoeymans, N., E.J.M. Feskens, D. Kromhout, and G.A.M Van Den Bos. 1997. Ageing and the relationship between functional status and self-rated health in elderly men. *Social science & medicine*, 45(10):1527–1536.

House, J.S. 2002. Understanding social factors and inequalities in health: 20th century progress and 21st century prospects. *Journal of health and social behavior*, 43:125–142.

Hullegie, P. and T.J. Klein. 2010. The effect of private health insurance on medical care utilization and self-assessed health in germany. *Health Economics*, 19(9):1048–1062.

Idler, E.L and Y. Benyamini. 1997. Self-rated health and mortality: a review of twenty-seven community studies. *Journal of Health and Social Behavior*, pages 21–37.

Imbens, G. and J. Angrist. 1994. Identification and estimation of local average treatment effects. *Econometrica*, 62:467–75.

Jylhä, M. , J.M. Guralnik, J. Balfour, and L.P. Fried. 2001. Walking difficulty, walking speed, and age as predictors of self-rated health the women's health and aging study. *The Journals of gerontology Series A: Biological Sciences and Medical Sciences*, 56(10):M609–M617.

Jylhä, M. 2009. What is self-rated health and why does it predict mortality? towards a unified conceptual model. *Social Science & Medicine*, 69(3):307–316.

Kasmel, A., V. Helasoja, A. Lipand, R. Prättälä, J. Klumbiene, and I. Pudule. 2004. Association between health behaviour and self-reported health in estonia, finland, latvia and lithuania. *The European Journal of Public Health*, 14(1):32–36.

Kennedy, B.P. , I. Kawachi, R. Glass, and D. Prothrow-Stith. 1998. Income distribution, socioeconomic status, and self rated health in the united states: multilevel analysis. *British Medical Journal*, 317(7163):917–921.

Kivinen, P., P. Halonen, M. Eronen, and A. Nissinen. 1998. Self-rated health, physician-rated health and associated factors among elderly men: the finnish cohorts of the seven countries study. *Age and Ageing*, 27(1):41–47.

Kómár, M., L. Nagymajtényi, T. Nyári, and E. Paulik. 2006. The determinants of self-rated health among ethnic minorities in hungary. *Ethnicity and Health*, 11(02):121–132.

Lamb, K.L., K. Roberts, and D.A. Brodie. 1990. Self-perceived health among sports participants and non-sports participants. *Social Science & Medicine*, 31(9):963–969.

Manderbacka, K., O. Lundberg, and P. Martikainen. 1999. Do risk factors and health behaviours contribute to self-ratings of health? *Social Science & Medicine*, 48(12):1713–1720.

Menec, V.H. and J.G. Chipperfield. 2001. A prospective analysis of the relation between self-rated health and health care use among elderly canadians. *Canadian Journal on Aging*, 20(03):293–306.

Miilunpalo, S., I. Vuori, P. Oja, M. Pasanen, and H. Urponen. 1997. Self-rated health status as a health measure: the predictive value of self-reported health status on the use of physician services and on mortality in the working-age population. *Journal of Clinical Epidemiology*, 50(5):517–528.

Min, J.W., S. Rhee, S.E. Lee, J. Rhee, and T. Tran. 2014. Comparative analysis on determinants of self-rated health among non-hispanic white, hispanic, and asian american older adults. *Journal of Immigrant and Minority Health*, 16(3):365–372.

Molarius, A. and S. Janson. 2002. Self-rated health, chronic diseases, and symptoms among middle-aged and elderly men and women. *Journal of Clinical Epidemiology*, 55(4):364–370.

Moradi-Lakeh, M., C. El Bcheraoui, M. Tuffaha, F. Daoud, M. Al Saeedi, M. Basulaiman, Z.A. Memish, M.A. Al Mazroa, A. Al Rabeeah, and A.H. Mokdad. 2015. Self-rated health among saudi adults: findings from a national survey, 2013. *Journal of community health*, 40(5):920–926.

Murphy, A. 2007. Score tests of normality in bivariate probit models. *Economics Letters*, 95(3):374–379.

Newbold, K.B. 2005. Self-rated health within the canadian immigrant population: risk and the healthy immigrant effect. *Social Science & Medicine*, 60(6):1359–1370.

Nicholson, A., M. Bobak, M. Murphy, R. Rose, and M. Marmot. 2005. Socio-economic influences on self-rated health in russian men and women - a life course approach. *Social Science & Medicine*, 61(11):2345–2354.

Okosun, I.S., S. Choi, T. Matamoros, and G.E. Dever. 2001. Obesity is associated with reduced self-rated general health status: evidence from a representative sample of white, black, and hispanic americans. *Preventive Medicine*, 32(5):429–436.

Park, J.H. and K.S. Lee. 2013. Self-rated health and its determinants in japan and south korea. *Public Health*, 127(9):834–843.

Perry, B. and W. Gesler. 2000. Physical access to primary health care in andean bolivia. *Social Science & Medicine*, 50(9):1177–1188.

Piko, B. 2000. Health-related predictors of self-perceived health in a student population: the importance of physical activity. *Journal of Community Health*, 25(2):125–137.

Poikolainen, K., E. Vartiainen, and H.J. Korhonen. 1996. Alcohol intake and subjective health. *American Journal of Epidemiology*, 144(4):346–350.

Prosper, M., V.L. Moczulski, and A. Qureshi. 2009. Obesity as a predictor of self-rated health. *American Journal of Health Behavior*, 33(3):319–329.

Prus, S.G. 2011. Comparing social determinants of self-rated health across the united states and canada. *Social Science & Medicine*, 73(1):50–59.

Schultz, J., A.M. O'Brien, and B. Tadesse. 2008. Social capital and self-rated health: Results from the us 2006 social capital survey of one community. *Social Science & Medicine*, 67(4):606–617.

Snelgrove, J.W., H. Pikhart, and M. Stafford. 2009. A multilevel analysis of social capital and self-rated health: evidence from the british household panel survey. *Social science & medicine*, 68(11):1993–2001.

Sözmen, K., H. Baydur, H. Simsek, and B. Ünal. 2012. Decomposing socioeconomic inequalities in self assessed health in turkey. *International Journal for Equity in Health*, 11:73.

Sözmen, K. and B. Ünal. 2014. Socioeconomic inequalities in non-communicable diseases and self assessed health in Turkey. *Iranian Journal of Public Health*, 43(6):736–748.

Sözmen, K. and B. Ünal. 2016. Explaining inequalities in health care utilization among Turkish adults: Findings from health survey 2008. *Health Policy*, 120:100–110,.

Stock, J. H. and M. Yogo. 2005. *Testing for Weak Instruments in Linear IV Regression. In Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. Cambridge University Press.

Stranges, S., J. Notaro, J.L. Freudenheim, R.M. Calogero, P. Muti, E. Farinaro, M. Russell, T.H. Nochajski, and M. Trevisan. 2006. Alcohol drinking pattern and subjective health in a population-based study. *Addiction*, 101(9):1265–1276.

Szwarcwald, C.L., Paulo Roberto Borges de Souza-Júnior, M.A.P Esteves, G.N. Damacena, and F. Viacava. 2005. Socio-demographic determinants of self-rated health in brazil. *Cadernos de Saúde Pública*, 21:S54–S64.

Undén, A. and S. Elofsson. 2006. Do different factors explain self-rated health in men and women? *Gender Medicine*, 3(4):295–308.

Vingilis, E.R., T.J. Wade, and J.S. Seeley. 2002. Predictors of adolescent self-rated health: analysis of the national population health survey. *Canadian Journal of Public Health*, 93:193–197.

Zimmer, Z., J. Natividad, H. Lin, and N. Chayovan. 2000. A cross-national examination of the determinants of self-assessed health. *Journal of health and Social Behavior*, 41:465–481.

Appendix

Figure 1: Conceptual Framework on the Social Determinants of Healthcare Utilization and SRH

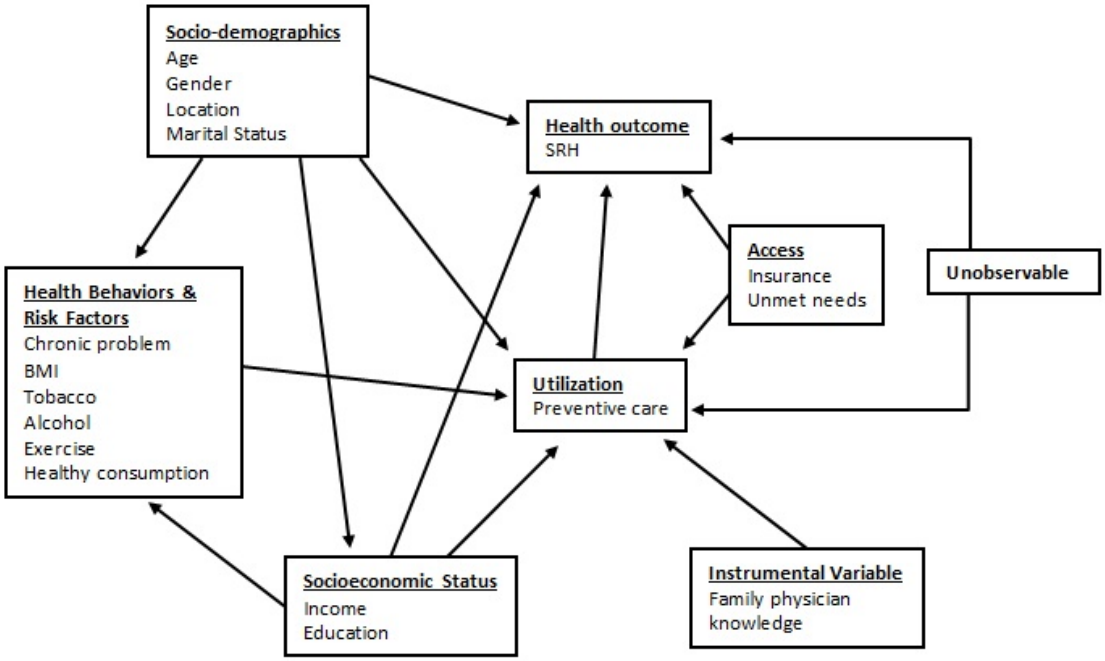
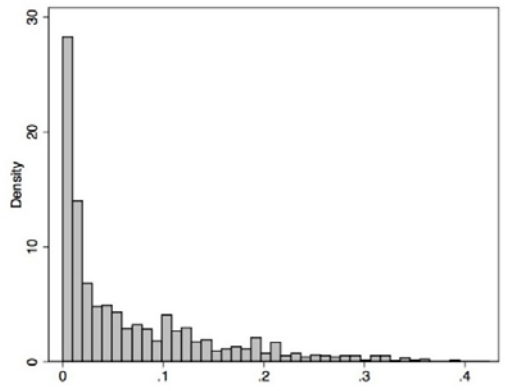
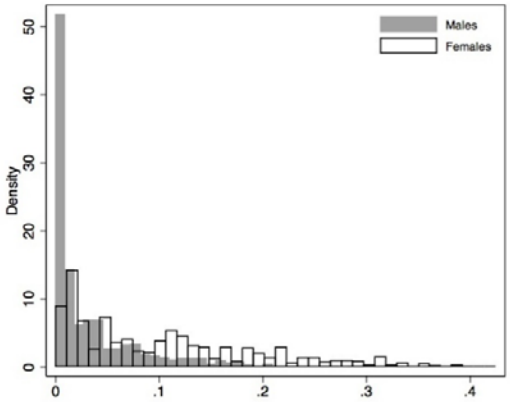


Figure 2: Distribution of the Probability of Selection into Preventive Care

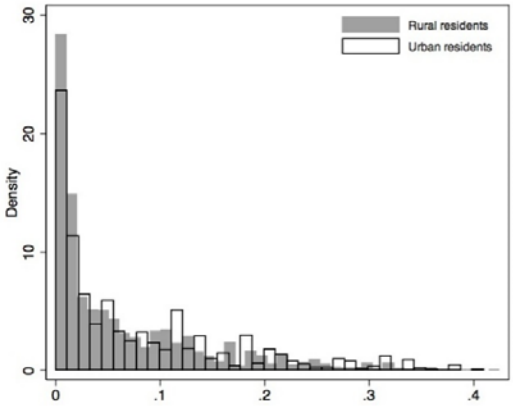
(a) Full Sample



(b) By Gender



(c) By Location



(d) By Chronicity

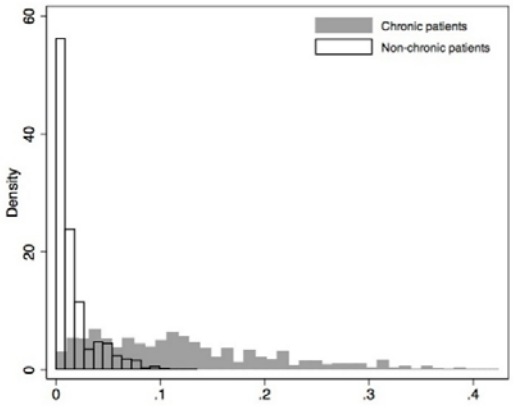


Figure 3: Distribution of Treatment Effects, Preventive Care, Full Sample

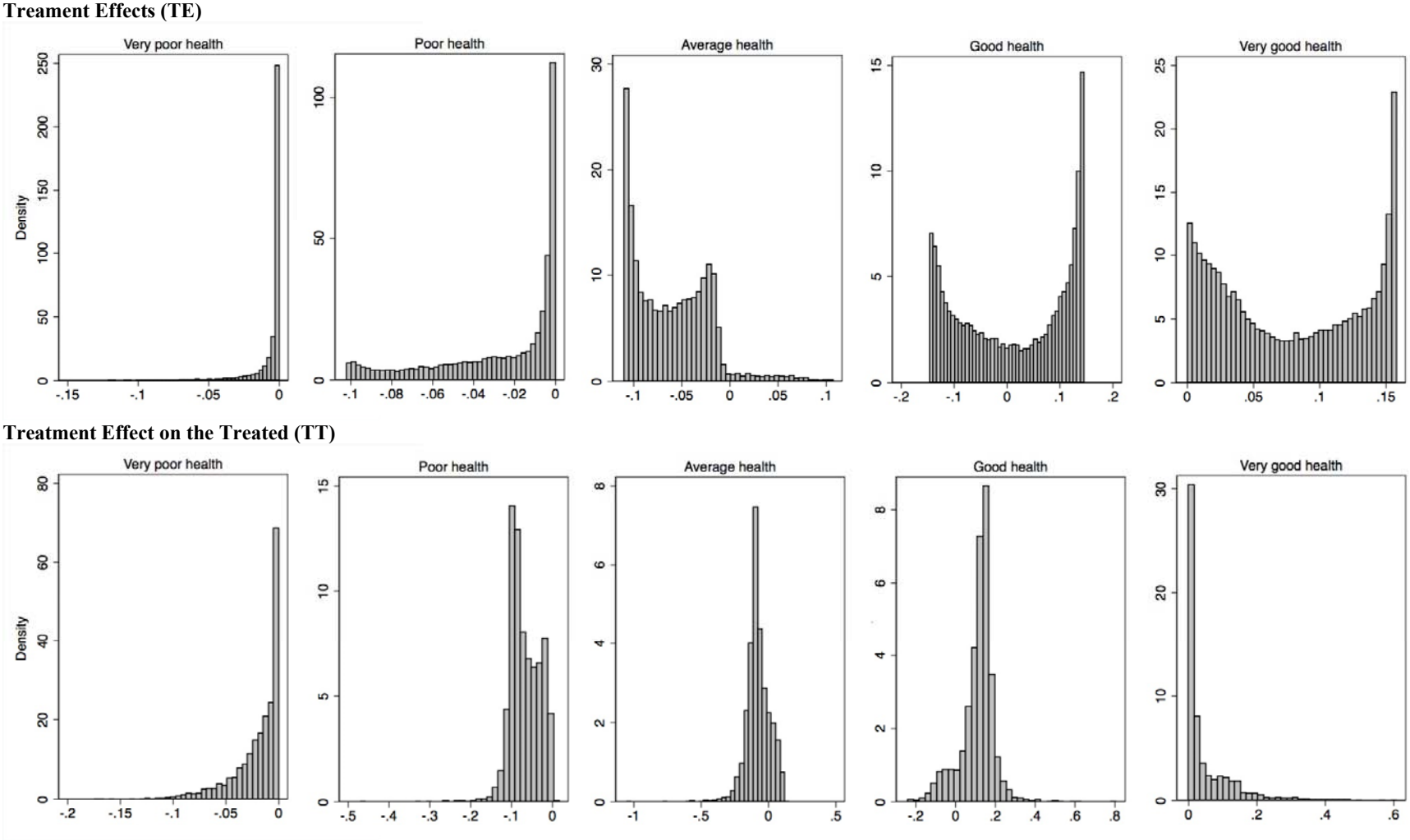
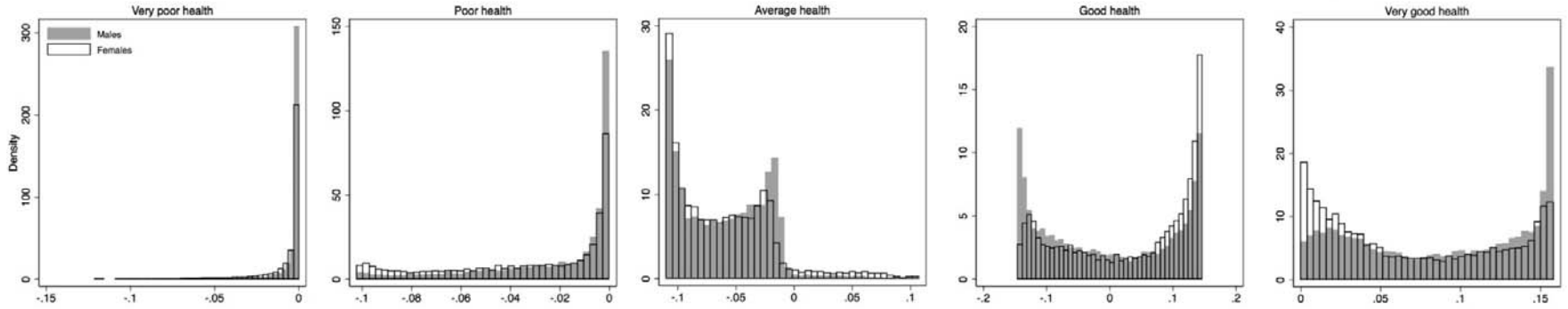


Figure 4: Distribution of Treatment Effects by Gender, Preventive Care

Treament Effects (TE)



Treatment Effects on the Treated (TT)

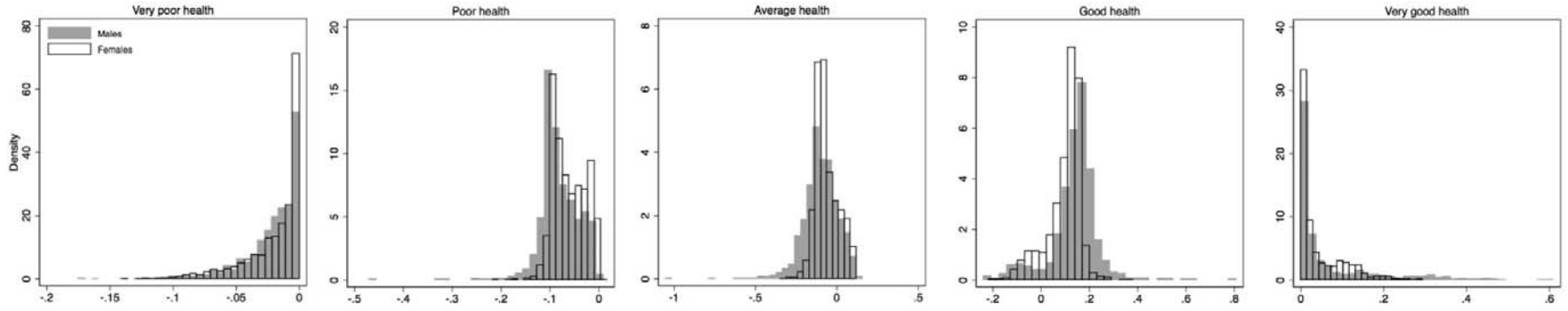
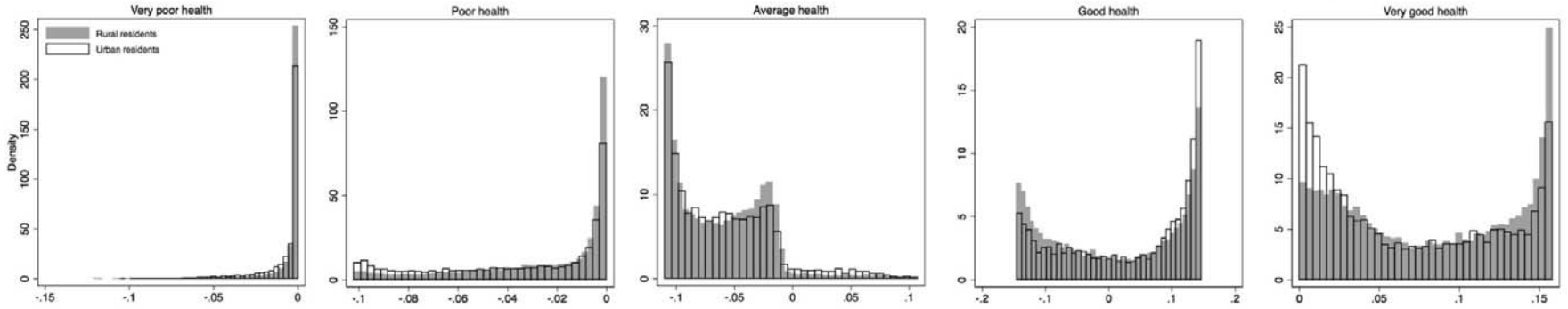


Figure 5: Distribution of Treatment Effects by Location, Preventive Care

Treament Effects (TE)



Treatment Effects on the Treated (TT)

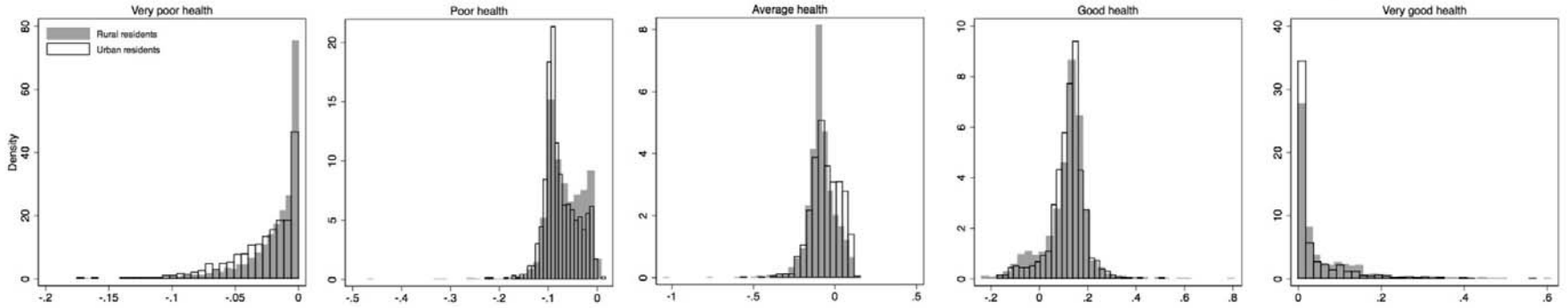
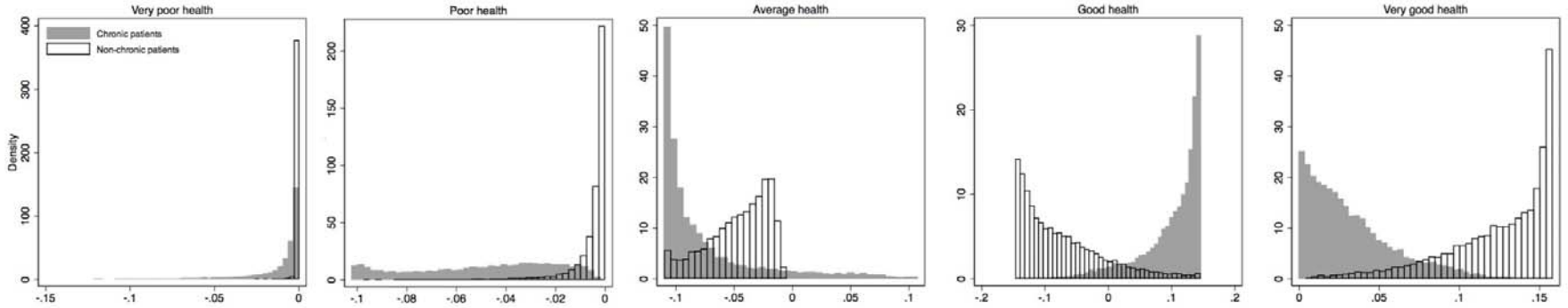


Figure 6: Distribution of Treatment Effects by Chronic Disease History, Preventive Care

Treament Effects (TE)



Treatment Effects on the Treated (TT)

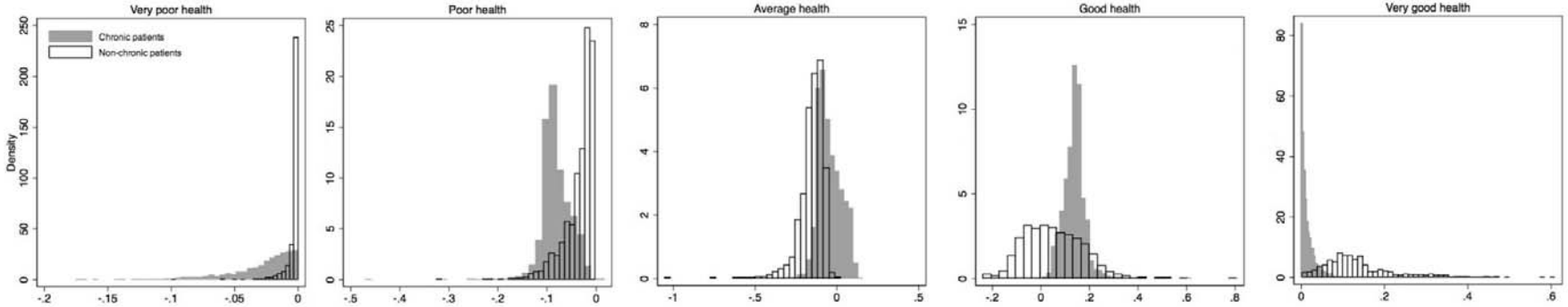


Table 1: Coding of the Variables Used in the Analysis, TurkStat Health Research Survey 2012 (*N=24,022*)

Survey Question	Our Coding	Variable name
What is your overall health status?	=1 if very poor; ... =5 if very good	Self-rated Health
Have you ever experienced or been experiencing any of the following chronic health problems? Asthma, Thyroid, Chronic obstructive lung disease, Hepatitis, Heart Attack, Coronary heart disease, Cerebro-vascular disease, Calcification, Arthritis, Back/Neck pain, Diabetes, Allergies, Gastric/Duodenal ulcer, Liver failure, Cancer, Migraine, Wetting, Anxiety, Depression, Anemia, Mental (other), Disability, Sinusitis	=1 if experienced any of the chronic problems; =0 otherwise	Chronic health problem
Have you received any of the following preventive healthcare service under primary care in the last 12 months from a community health center or a family physician? Vaccination (pregnant/women aged 15-46/other), family planning, monitoring (pregnant women aged 15-46/other), chronic disease monitoring, communicable disease monitoring health education, nutrition counselling, youth counselling, psychosocial support	=1 if received any of the preventive care services =0 otherwise	Preventive care utilization
What was the last time you received care from a family physician or a GP? <12 months ago, ≥ 12 months, Never	=1 if ever received =0 otherwise	GP care utilization
Respondent's location Respondent's gender	=1 if rural; =0 otherwise =1 if male; =0 if female	Location Gender
Do you know who your family physician is?	=1 if yes =0 otherwise	Family physician knowledge
Obesity	=1 if BMI $>$ 30; =0 otherwise	Obesity
How often do you drink fruit or vegetable juice in a week? 0, <1, ≥ 1, ≥ 4, 7, >7	A dummy variable is created for each of the six categories	Juice consumption
How many days within a week have you walked at least 10 min.? None to 7 days a week	A dummy variable is created for each of the eight categories	Exercise
Do you regularly use tobacco (alcohol)?	=1 if yes =0 otherwise	Tobacco (Alcohol) consumption
To which age group do you belong? 15-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75+	A dummy variable is created for each of the seven categories	Age group
What is your marital status? Single, Married, Widowed, Divorced	A dummy variable is created for each of the four categories	Marital Status
What is your highest education? Illiterate, literate (no scholling), primary school, primary education, secondary education, high school or equivalent, vocational school, master's or Ph.D	Four categories are constructed (none, primary, secondary, higher) and a dummy variable is created for each	Education
Which institution covers the cost of your medical care? Social Security Institution (retired, officer, worker, craftsmen), Green Card, Private Insurance/Pension, Out-of-pocket, Other	Four categories are constructed (public, private, out- of-pocket, other) and a dummy variable is created for each	Insurance
Which of the following brackets does your income fall into? $<$ 350, 351-500, 501-620, 621-750, 751-900, 901-1100, 1101-1300, 1301-1700. 1701-2300, $>$ 2301 TL	All ten categories are grouped into two (quintiles) and a dummy is created for each quintile	Income
Have you ever felt the need to receive outpatient care in the last 12 months but have not received it?	=1 if yes =0 otherwise	Unmet healthcare needs

Table 2: Descriptive Statistics by utilization and SRH status, means and standard deviations (*N=24,022*)

Variable	Overall	Preventive Care		GP Care		Self-Rated Health Status				
	Mean (s.d)	Yes	No	Yes	No	Very poor	Poor	Average	Good	Very good
SRH	3.75 (0.78)	3.46***	3.83	3.73***	3.91	-	-	-	-	-
Preventive care	0.20 (0.40)	-	-	-	-	0.43	0.37	0.29	0.16	0.12
GP care	0.87 (0.34)	-	-	-	-	0.92	0.90	0.91	0.86	0.82
Age group										
15-24	0.18 (0.39)	0.11	0.20	0.18***	0.22	0.05	0.03	0.05	0.21	0.41
25-34	0.21 (0.41)	0.21	0.21	0.20***	0.23	0.06	0.06	0.12	0.25	0.27
35-44	0.20 (0.40)	0.17***	0.21	0.20***	0.23	0.07	0.12	0.19	0.23	0.17
45-54	0.18 (0.38)	0.17	0.18	0.18	0.17	0.12	0.19	0.23	0.17	0.10
55-64	0.12 (0.33)	0.16***	0.11	0.13***	0.08	0.27	0.22	0.20	0.09	0.04
65-74	0.07 (0.25)	0.11***	0.06	0.07***	0.04	0.22	0.21	0.13	0.04	0.009
75+	0.04 (0.19)	0.06***	0.03	0.04***	0.02	0.21	0.17	0.08	0.01	0.002
Gender	0.48 (0.50)	0.34***	0.51	0.46***	0.61	0.43	0.38	0.39	0.51	0.56
Location	0.75 (0.43)	0.73***	0.76	0.75***	0.76	0.70	0.66	0.71	0.78	0.76
Marital Status										
Single	0.23 (0.42)	0.11***	0.26	0.22***	0.31	0.11	0.06	0.07	0.26	0.49
Married	0.69 (0.46)	0.78***	0.67	0.70***	0.63	0.64	0.72	0.79	0.69	0.48
Divorced	0.02 (0.15)	0.02	0.02	0.02	0.03	0.02	0.03	0.03	0.02	0.02
Widowed	0.05 (0.23)	0.09***	0.05	0.06***	0.03	0.23	0.18	0.11	0.03	0.006
Chronic health problem	0.50 (0.50)	0.71***	0.45	0.52***	0.38	0.90	0.92	0.82	0.39	0.17
Obesity	0.19 (0.39)	0.25***	0.17	0.20***	0.14	0.30	0.34	0.29	0.15	0.09
Alcohol consumption	0.03 (0.17)	0.03	0.03	0.03***	0.04	0.05	0.04	0.03	0.03	0.02
Tobacco Consumption	0.36 (0.03)	0.34***	0.36	0.35***	0.43	0.32	0.35	0.37	0.37	0.30
Exercise										
None	0.39 (0.49)	0.38	0.39	0.39*	0.38	0.69	0.60	0.39	0.36	0.39
Once	0.02 (0.14)	0.03***	0.02	0.02	0.02	0.04	0.02	0.03	0.02	0.02
Twice	0.04 (0.21)	0.06***	0.04	0.04	0.05	0.02	0.04	0.05	0.05	0.03
Three days	0.06 (0.24)	0.07***	0.06	0.06**	0.05	0.07	0.05	0.07	0.06	0.05
Four days	0.04 (0.20)	0.05***	0.04	0.04*	0.03	0.02	0.03	0.04	0.04	0.03
Five days	0.08 (0.27)	0.07**	0.08	0.08**	0.07	0.02	0.04	0.07	0.08	0.09
Six days	0.04 (0.19)	0.03***	0.04	0.03***	0.05	0.006	0.01	0.02	0.04	0.04
Seven days	0.33 (0.47)	0.31***	0.34	0.33***	0.35	0.14	0.20	0.32	0.35	0.34
Juice consumption										
Never	0.18 (0.38)	0.18	0.17	0.17	0.18	0.31	0.28	0.23	0.15	0.13
Less than once	0.29 (0.45)	0.29	0.29	0.30***	0.25	0.26	0.31	0.32	0.28	0.25
At least once	0.21 (0.40)	0.20	0.21	0.20***	0.22	0.18	0.15	0.19	0.22	0.20
At least four times	0.16 (0.36)	0.15	0.16	0.16	0.15	0.12	0.13	0.13	0.17	0.18
Seven times	0.15 (0.36)	0.14**	0.15	0.15**	0.16	0.09	0.12	0.11	0.15	0.20
＞seven times	0.02 (0.15)	0.04***	0.02	0.02	0.03	0.05	0.02	0.01	0.02	0.04
Education										
None	0.13 (0.34)	0.18***	0.12	0.14***	0.11	0.48	0.38	0.22	0.09	0.05
Primary	0.55 (0.50)	0.53***	0.55	0.56***	0.50	0.43	0.51	0.58	0.55	0.52
Secondary	0.19 (0.39)	0.17***	0.19	0.18***	0.22	0.06	0.07	0.12	0.21	0.25
Higher	0.13 (0.34)	0.12***	0.13	0.12***	0.17	0.02	0.04	0.07	0.15	0.17
Income										
1st quintile	0.07 (0.27)	0.09**	0.07	0.08	0.08	0.17	0.17	0.10	0.06	0.05
2nd quintile	0.12 (0.32)	0.13**	0.11	0.12**	0.11	0.14	0.17	0.14	0.11	0.09
3rd quintile	0.25 (0.43)	0.26*	0.24	0.25***	0.21	0.32	0.28	0.27	0.24	0.22
4th quintile	0.23 (0.42)	0.21***	0.23	0.23	0.22	0.20	0.19	0.23	0.23	0.24
5th quintile (top 20%)	0.33 (0.47)	0.32	0.33	0.32***	0.38	0.17	0.18	0.27	0.36	0.40
Health Insurance										
Public	0.93 (0.24)	0.96***	0.93	0.94***	0.90	0.90	0.93	0.95	0.93	0.92
Private	0.01 (0.12)	0.01	0.01	0.01***	0.03	0.006	0.009	0.01	0.01	0.02
Out-of-pocket	0.04 (0.20)	0.03***	0.05	0.04***	0.07	0.05	0.04	0.03	0.04	0.05
Other	0.02 (0.14)	0.01	0.02	0.02	0.03	0.06	0.02	0.02	0.02	0.03
Unmet healthcare needs	0.12 (0.32)	0.14***	0.11	0.12***	0.13	0.32	0.22	0.18	0.10	0.06

Notes: All variables are binary except SRH. *, ** and *** denote difference in means between utilizers and non-utilizers at the 10, 5. and 1 percent significance level respectively.

Table 3: Limited Information Maximum Likelihood (LIML) Estimates

	Preventive Care	GP Care
Healthcare utilization	-0.108 (0.082)	-0.020 (0.015)
Age groups		
25-34	0.007 (0.005)	0.009** (0.005)
35-44	0.006 (0.009)	0.014** (0.006)
45-54	0.026***(0.009)	0.033***(0.006)
55-64	0.070***(0.010)	0.074***(0.009)
65-74	0.133***(0.015)	0.132***(0.014)
75+	0.225***(0.025)	0.220***(0.024)
Gender	-0.012 (0.009)	-0.003 (0.004)
Location	0.003 (0.005)	0.004 (0.005)
Marital Status		
Single	-0.003 (0.019)	0.006 (0.017)
Married	-0.007 (0.017)	-0.010 (0.017)
Divorced	-0.017 (0.020)	-0.016 (0.019)
Chronic health problem	0.079***(0.011)	0.066***(0.004)
Obesity	0.024***(0.007)	0.022***(0.007)
Exercise		
Once	-0.024 (0.016)	-0.029** (0.015)
Twice	-0.032***(0.011)	-0.036***(0.011)
Three days	-0.041***(0.009)	-0.044***(0.008)
Four days	-0.048***(0.008)	-0.051***(0.008)
Five days	-0.037***(0.007)	-0.037***(0.007)
Six days	-0.037***(0.009)	-0.040***(0.009)
Seven days	-0.045***(0.005)	-0.048***(0.005)
Consumption of juice		
Less than once	-0.014* (0.007)	-0.016** (0.007)
At least once	-0.025***(0.008)	-0.028***(0.007)
At least four times	-0.008 (0.008)	-0.012* (0.008)
Seven times	-0.018** (0.008)	-0.021***(0.007)
＞ seven times	0.003 (0.018)	-0.010 (0.015)
Alcohol consumption	0.029* (0.016)	0.024 (0.015)
Tobacco consumption	0.003 (0.005)	0.001 (0.005)
Education		
None	0.078***(0.010)	0.078***(0.010)
Primary	0.018***(0.005)	0.020***(0.005)
Secondary	0.010* (0.005)	0.010** (0.005)
Income		
1st	0.037***(0.011)	0.035***(0.010)
2nd	0.021** (0.008)	0.021** (0.008)
3rd	0.007 (0.006)	0.007 (0.006)
4th	0.001 (0.005)	0.002 (0.005)
Health Insurance		
Public	-0.008 (0.012)	-0.010 (0.011)
Private	-0.019 (0.015)	-0.022 (0.014)
Out-of-pocket	-0.007 (0.014)	-0.006 (0.014)
Unmet healthcare needs	0.059***(0.008)	0.058***(0.008)
Constant	0.027 (0.027)	0.028 (0.026)
Underidentification test (p-value)	70.52 (0.0000)	1022.26 (0.0000)
Weak identification test	71.833	1262.83
Endogeneity test (p-value)	3.247 (0.0716)	1.032 (0.3097)

Notes: The outcome variable is SRH (binary). The underidentification test reports the Kleibergen-Paap rk LM statistic and the p-value for the null hypothesis that the equation is underidentified (i.e. the excluded instruments are irrelevant). The weak identification test reports the Kleibergen-Paap rk Wald F-statistic for the null hypothesis that the equation is weakly identified. Weak identifcation test critical values for 10% and 15% maximal LIML size are 16.38 and 8.96 respectively (Stock and Yogo, 2005). The endogeneity test reports the chi-square and the p-value for the null hypothesis that healthcare utilization can be treated as exogenous. . All specifications use sampling weights provided by TurkStat. *, ** and *** denote statistical significance at 10, 5 and 1 percent respectively.

Table 4: Preventive Care, Bivariate Ordered Probit Estimates

	Full Sample (1)	Male (2)	Female (3)	Rural (4)	Urban (5)	Chronic (6)	Non-chronic (7)
Preventive care utilization	**0.422* (0.231)**	**0.734*** (0.223)**	**0.572*** (0.179)**	**0.422* (0.239)**	**0.465** (0.224)**	**0.263* (0.148)**	**0.502* (0.279)**
Age groups							
25-34	-0.313*** (0.042)	-0.315*** (0.065)	-0.327*** (0.060)	-0.295*** (0.047)	-0.356*** (0.088)	-0.336*** (0.073)	-0.280*** (0.050)
35-44	-0.490*** (0.047)	-0.486*** (0.076)	-0.510*** (0.067)	-0.467*** (0.054)	-0.542*** (0.093)	-0.635*** (0.077)	-0.357*** (0.060)
45-54	-0.706*** (0.051)	-0.676*** (0.085)	-0.767*** (0.071)	-0.680*** (0.058)	-0.773*** (0.103)	-0.837*** (0.081)	-0.579*** (0.066)
55-64	-0.961*** (0.068)	-1.022*** (0.102)	-1.002*** (0.087)	-0.946*** (0.077)	-1.015*** (0.116)	-1.062*** (0.091)	-0.833*** (0.085)
65-74	-1.255*** (0.089)	-1.336*** (0.119)	-1.315*** (0.111)	-1.224*** (0.102)	-1.338*** (0.133)	-1.350*** (0.103)	-1.045*** (0.117)
75+	-1.580*** (0.110)	-1.669*** (0.156)	-1.668*** (0.127)	-1.579*** (0.130)	-1.617*** (0.151)	-1.634*** (0.119)	-1.491*** (0.142)
Gender	0.190*** (0.034)	-	-	0.191*** (0.037)	0.207*** (0.057)	0.168*** (0.039)	0.189*** (0.039)
Location	-0.053** (0.025)	-0.055 (0.038)	-0.045 (0.037)	-	-	-0.045 (0.034)	-0.059 (0.036)
Marital Status							
Single	0.106 (0.069)	0.253* (0.134)	0.164* (0.091)	0.147* (0.079)	-0.017 (0.123)	-0.062 (0.084)	0.347*** (0.109)
Married	0.042 (0.053)	0.207* (0.121)	0.007 (0.068)	0.076 (0.064)	-0.068 (0.091)	-0.003 (0.059)	0.173* (0.096)
Divorced	0.119 (0.090)	0.215 (0.179)	0.157 (0.112)	0.180* (0.102)	-0.282 (0.179)	0.076 (0.106)	0.242 (0.153)
Risk factors							
Chronic health problem	-0.970*** (0.074)	-1.053*** (0.081)	-1.032*** (0.076)	-0.984*** (0.076)	-0.961*** (0.091)	-	-
Alcohol consumption	-0.088 (0.066)	-0.155** (0.074)	0.228 (0.250)	-0.121 (0.081)	-0.034 (0.109)	-0.155* (0.089)	0.009 (0.093)
Tobacco consumption	-0.081*** (0.025)	-0.045 (0.033)	-0.168*** (0.044)	-0.064** (0.029)	-0.144*** (0.053)	-0.064* (0.034)	-0.104*** (0.035)
Obesity	-0.136*** (0.029)	-0.061 (0.046)	-0.200*** (0.040)	-0.127*** (0.034)	-0.171*** (0.052)	-0.156*** (0.034)	-0.043 (0.049)
Exercise							
Once	0.024 (0.072)	-0.017 (0.126)	0.035 (0.094)	-0.001 (0.084)	0.053 (0.146)	0.009 (0.089)	0.092 (0.104)
Twice	0.036 (0.052)	-0.023 (0.087)	0.066 (0.070)	0.029 (0.062)	0.101 (0.098)	0.184*** (0.065)	-0.168** (0.082)
Three days	0.138*** (0.047)	0.037 (0.076)	0.199*** (0.063)	0.148*** (0.054)	0.082 (0.085)	0.219*** (0.061)	0.021 (0.068)
Four days	0.175*** (0.057)	0.072 (0.087)	0.250*** (0.078)	0.160** (0.065)	0.242** (0.114)	0.300*** (0.076)	-0.008 (0.080)
Five days	0.117*** (0.043)	0.095 (0.064)	0.138** (0.062)	0.117** (0.049)	0.102 (0.086)	0.219*** (0.061)	-0.010 (0.058)
Six days	0.148*** (0.057)	0.201*** (0.070)	0.092 (0.102)	0.154** (0.063)	0.112 (0.133)	0.270*** (0.094)	0.008 (0.066)
Seven days	0.121*** (0.025)	0.155*** (0.037)	0.093** (0.038)	0.125*** (0.030)	0.126*** (0.046)	0.222*** (0.035)	-0.006 (0.034)
Juice consumption							
Less than once	0.084*** (0.033)	0.107** (0.048)	0.077 (0.047)	0.065 (0.040)	0.137** (0.055)	0.099** (0.041)	0.066 (0.052)
At least once	0.148*** (0.035)	0.143*** (0.052)	0.164*** (0.052)	0.142*** (0.043)	0.155** (0.062)	0.206*** (0.046)	0.074 (0.053)
At least four times	0.107*** (0.038)	0.099* (0.055)	0.126** (0.056)	0.085* (0.045)	0.163** (0.066)	0.100** (0.050)	0.105* (0.056)
Seven times	0.227*** (0.039)	0.242*** (0.057)	0.232*** (0.058)	0.192*** (0.046)	0.362*** (0.075)	0.206*** (0.052)	0.235*** (0.059)
＞ seven times	0.377*** (0.088)	0.268** (0.117)	0.511*** (0.137)	0.396*** (0.099)	0.251 (0.199)	0.385*** (0.120)	0.339*** (0.117)
Education							
None	-0.644*** (0.056)	-0.698*** (0.097)	-0.640*** (0.077)	-0.650*** (0.067)	-0.552*** (0.132)	-0.719*** (0.070)	-0.507*** (0.080)
Primary	-0.319*** (0.038)	-0.292*** (0.053)	-0.325*** (0.062)	-0.307*** (0.041)	-0.251** (0.119)	-0.401*** (0.055)	-0.208*** (0.049)
Secondary	-0.171*** (0.040)	-0.132** (0.055)	-0.185*** (0.067)	-0.192*** (0.044)	0.034 (0.127)	-0.276*** (0.062)	-0.066 (0.052)
Income							
1st	-0.193*** (0.042)	-0.258*** (0.066)	-0.225*** (0.065)	-0.231*** (0.064)	-0.238*** (0.069)	-0.252*** (0.057)	-0.172*** (0.067)
2nd	-0.084** (0.037)	-0.190*** (0.061)	-0.070 (0.057)	-0.097* (0.050)	-0.166** (0.068)	-0.105* (0.055)	-0.120** (0.054)
3rd	-0.070*** (0.026)	-0.110** (0.044)	-0.099** (0.044)	-0.066* (0.034)	-0.190*** (0.060)	-0.107*** (0.041)	-0.070* (0.041)
4th	-	-0.078* (0.043)	-0.041 (0.043)	-0.057* (0.033)	-0.051 (0.064)	-0.087** (0.041)	-0.004 (0.040)
Health Insurance							
Public	0.033 (0.066)	-0.030 (0.098)	0.064 (0.103)	0.028 (0.078)	0.061 (0.124)	0.054 (0.100)	0.013 (0.085)
Private	0.140 (0.100)	0.067 (0.139)	0.166 (0.158)	0.070 (0.110)	0.648** (0.260)	0.249* (0.141)	-0.043 (0.130)
Out-of-pocket	-0.005 (0.081)	-0.133 (0.121)	0.162 (0.123)	0.050 (0.098)	-0.089 (0.143)	-0.068 (0.121)	0.047 (0.106)
Unmet healthcare needs	-0.411*** (0.039)	-0.394*** (0.056)	-0.457*** (0.053)	-0.410*** (0.044)	-0.438*** (0.067)	-0.397*** (0.041)	-0.421*** (0.062)

* Outcome equation: Self-rated health

Table 4: Continued

		Full Sample (1)	Male (2)	Female (3)	Rural (4)	Urban (5)	Chronic (6)	Non-chronic (7)
	Preventive care utilization	**0.422* (0.231)**	**0.734*** (0.223)**	**0.572*** (0.179)**	**0.422* (0.239)**	**0.465** (0.224)**	**0.263* (0.148)**	**0.502* (0.279)**
* Selection equation: Healthcare Utilization	Family physician knowledge	0.445***(0.055)	0.484*** (0.072)	0.378*** (0.081)	0.486*** (0.068)	0.358*** (0.093)	0.468*** (0.080)	0.436*** (0.070)
	Age groups							
	25-34	-0.217** (0.096)	-0.047 (0.116)	-0.203 (0.130)	-0.145 (0.109)	-0.460** (0.202)	-0.015 (0.170)	-0.317*** (0.121)
	35-44	-0.615*** (0.105)	-0.018 (0.136)	-0.772*** (0.139)	-0.534*** (0.120)	-0.882*** (0.215)	-0.226 (0.177)	-0.818*** (0.135)
	45-54	-0.566*** (0.113)	0.122 (0.148)	-0.782*** (0.149)	-0.508*** (0.130)	-0.769*** (0.219)	-0.094 (0.185)	-0.916*** (0.146)
	55-64	-0.287** (0.123)	0.271* (0.156)	-0.432** (0.171)	-0.179 (0.144)	-0.634*** (0.226)	0.199 (0.194)	-0.768*** (0.176)
	65-74	0.032 (0.148)	0.620*** (0.197)	-0.195 (0.210)	0.221 (0.185)	-0.488** (0.240)	0.449** (0.211)	-0.259 (0.268)
	75+	0.217 (0.183)	0.948*** (0.248)	-0.180 (0.252)	0.434* (0.239)	-0.372 (0.278)	0.684*** (0.245)	-0.809*** (0.272)
	Gender	-0.623*** (0.052)	-	-	-0.640*** (0.060)	-0.617*** (0.093)	-0.520*** (0.076)	-0.751*** (0.067)
	Location	-0.068 (0.052)	-0.045 (0.069)	-0.075 (0.075)	-	-	-0.175** (0.075)	0.052 (0.071)
	Marital Status							
	Single	-0.665*** (0.159)	-0.367 (0.308)	-0.962*** (0.196)	-0.573*** (0.190)	-0.979*** (0.274)	-0.352 (0.215)	-0.633** (0.278)
	Married	0.125 (0.128)	-0.316 (0.287)	0.149 (0.145)	0.170 (0.157)	-0.016 (0.207)	0.055 (0.147)	0.454* (0.256)
	Divorced	-0.067 (0.200)	-0.196 (0.366)	-0.132 (0.228)	-0.161 (0.225)	0.889** (0.448)	-0.026 (0.255)	0.104 (0.346)
	Risk factors							
	Chronic health problem	0.826*** (0.048)	0.915*** (0.067)	0.756*** (0.068)	0.760*** (0.057)	1.018*** (0.085)	-	-
	Alcohol consumption	0.282** (0.138)	0.315** (0.139)	-0.421 (0.480)	0.358** (0.172)	0.060 (0.219)	0.327* (0.194)	0.197 (0.199)
	Tobacco consumption	0.100* (0.056)	0.088 (0.068)	0.128 (0.090)	0.081 (0.065)	0.215** (0.102)	0.110 (0.078)	0.089 (0.077)
	Education							
	None	-0.069 (0.106)	-0.070 (0.167)	-0.130 (0.146)	-0.052 (0.127)	0.018 (0.232)	0.017 (0.149)	-0.114 (0.151)
	Primary	-0.162** (0.074)	-0.094 (0.096)	-0.261** (0.115)	-0.194** (0.081)	0.009 (0.207)	-0.119 (0.114)	-0.139 (0.095)
	Secondary	-0.064 (0.083)	-0.032 (0.108)	-0.116 (0.127)	-0.033 (0.089)	-0.169 (0.232)	0.072 (0.134)	-0.141 (0.101)
	Health Insurance							
	Public	0.286* (0.156)	0.269 (0.215)	0.363 (0.228)	0.324 (0.199)	0.237 (0.256)	0.372 (0.234)	0.176 (0.199)
	Private	0.278 (0.226)	0.217 (0.265)	0.366 (0.344)	0.333 (0.249)	-0.145 (0.472)	0.404 (0.336)	-0.009 (0.286)
	Out-of-pocket	-0.012 (0.187)	-0.166 (0.244)	0.047 (0.279)	-0.026 (0.239)	-0.002 (0.304)	0.152 (0.288)	-0.175 (0.230)
	Unmet healthcare needs	0.062 (0.073)	0.019 (0.100)	0.100 (0.100)	0.015 (0.088)	0.192 (0.127)	-0.011 (0.090)	0.250** (0.118)
	Constant	-2.120*** (0.247)	-3.015*** (0.401)	-1.869*** (0.335)	-2.326*** (0.295)	-1.868*** (0.434)	-1.771*** (0.357)	-2.123*** (0.371)
	λ	-0.272*** (0.101)	-0.422*** (0.093)	-0.328*** (0.080)	-0.262** (0.103)	-0.320*** (0.096)	-0.225*** (0.067)	-0.242** (0.106)
* Causal Effects	ATE							
	Very poor	-0.005***(0.002)	-0.005***(0.001)	-0.007***(0.001)	-0.004***(0.001)	-0.007***(0.002)	-0.006***(0.002)	-0.001**(0.001)
	Poor	-0.027***(0.009)	-0.032***(0.007)	-0.039***(0.008)	-0.024***(0.008)	-0.039***(0.014)	-0.033***(0.012)	-0.005***(0.002)
	Average	-0.061**(0.027)	-0.089***(0.021)	-0.082***(0.024)	-0.061**(0.028)	-0.065**(0.025)	-0.047**(0.022)	-0.045***(0.015)
	Good	0.014(0.046)	-0.027(0.048)	0.037(0.039)	0.009(0.043)	0.033(0.04)	0.06*(0.031)	-0.097(0.067)
	Very good	0.078**(0.037)	0.153***(0.039)	0.091***(0.022)	0.081**(0.041)	0.077**(0.031)	0.025**(0.011)	0.148**(0.075)
	ATT							
	Very poor	-0.02***(0.006)	-0.048***(0.01)	-0.026***(0.005)	-0.017***(0.005)	-0.03***(0.009)	-0.014**(0.005)	-0.012***(0.004)
	Poor	-0.067***(0.023)	-0.144***(0.03)	-0.079***(0.019)	-0.063**(0.027)	-0.085**(0.034)	-0.051***(0.02)	-0.035***(0.012)
	Average	-0.075*(0.041)	-0.136**(0.055)	-0.077**(0.032)	-0.084**(0.04)	-0.052(0.043)	-0.029(0.026)	-0.159**(0.067)
	Good	0.113*(0.063)	0.253***(0.069)	0.131***(0.041)	0.112*(0.059)	0.125**(0.063)	0.08**(0.033)	0.034(0.069)
	Very good	0.049**(0.025)	0.076***(0.023)	0.051***(0.014)	0.052*(0.028)	0.041**(0.018)	0.014**(0.006)	0.171*(0.097)
Prob. of selection into care		0.067	0.034	0.101	0.063	0.080	0.117	0.019
LR test of λ (p-value)		7.25 (0.0070)	20.76 (0.0000)	16.57 (0.0000)	6.51 (0.0110)	11.09 (0.0010)	11.15 (0.0010)	5.24 (0.0220)
Number of observations		24022	11477	12545	18129	5893	11993	12029

Notes: The likelihood-ratio test of λ reports the test statistic and the p-value for the null hypothesis of independent equations. All specifications use sampling weights provided by TurkStat; employ gamma latent-factor structure and are based on 100 Halton sequence-based quasi-random draws per observation. Robust standard errors in parentheses. ATE: Average treatment effect, ATT: Average treatment effect on the treated. *, ** and *** denote statistical significance at 10, 5 and 1 percent level respectively.

Table 5: Preventive Care, Bivariate Probit Estimates

	Full Sample (1)	Male (2)	Female (3)	Rural (4)	Urban (5)	Chronic (6)	Non-chronic (7)
Healthcare utilization	-0.367* (0.188)	-0.369 (0.234)	-0.431 (0.435)	-0.356* (0.213)	-0.511 (0.436)	-0.369 (0.240)	-0.267 (0.385)
Age groups							
25-34	0.218***(0.079)	0.308** (0.134)	0.127 (0.104)	0.191* (0.104)	0.304* (0.177)	0.198* (0.116)	0.413***(0.143)
35-44	0.379***(0.091)	0.563***(0.138)	0.259* (0.155)	0.347***(0.125)	0.477** (0.231)	0.406***(0.122)	0.501**(0.199)
45-54	0.618***(0.098)	0.747***(0.144)	0.541***(0.178)	0.551***(0.133)	0.791***(0.268)	0.648***(0.107)	0.740***(0.237)
55-64	0.884***(0.101)	1.042***(0.143)	0.794***(0.173)	0.836***(0.124)	1.025***(0.282)	0.901***(0.107)	1.128***(0.249)
65-74	1.041***(0.101)	1.218***(0.140)	0.932***(0.172)	0.991***(0.119)	1.186***(0.276)	1.064***(0.110)	1.175***(0.244)
75+	1.312***(0.108)	1.552***(0.148)	1.134***(0.201)	1.286***(0.133)	1.415***(0.295)	1.341***(0.118)	1.374***(0.299)
Gender	-0.098** (0.041)	-	-	-0.090* (0.053)	-0.134 (0.089)	-0.071 (0.049)	-0.144 (0.120)
Location	0.083***(0.030)	0.108** (0.050)	0.045 (0.044)	-	-	0.077** (0.034)	0.099 (0.090)
Marital Status							
Single	0.238** (0.098)	0.448***(0.139)	0.006 (0.182)	0.210** (0.106)	0.291 (0.222)	0.214** (0.104)	0.420 (0.281)
Married	0.045 (0.054)	0.014 (0.108)	0.022 (0.049)	0.040 (0.058)	0.054 (0.086)	0.041 (0.057)	0.039 (0.186)
Divorced	0.178* (0.094)	0.352* (0.213)	0.036 (0.121)	0.159 (0.112)	0.310 (0.213)	0.206** (0.103)	0.025 (1.002)
Risk factors							
Chronic health problem	0.900***(0.044)	0.984***(0.064)	0.835***(0.058)	0.922***(0.054)	0.862***(0.077)	-	-
Alcohol consumption	0.186** (0.076)	0.212** (0.087)	0.189 (2.493)	0.225** (0.098)	0.106 (0.120)	0.192** (0.075)	0.146 (1.172)
Tobacco consumption	0.061* (0.036)	-0.037 (0.047)	0.197***(0.053)	0.068 (0.045)	0.060 (0.068)	0.067* (0.039)	0.029 (0.092)
Obesity	0.120***(0.038)	0.034 (0.054)	0.165***(0.039)	0.093** (0.040)	0.180***(0.067)	0.134***(0.039)	0.033 (0.100)
Exercise							
Once	-0.249** (0.105)	-0.240 (0.186)	-0.255** (0.106)	-0.232* (0.128)	-0.280 (0.244)	-0.220** (0.100)	-4.948***(0.196)
Twice	-0.308***(0.076)	-0.305** (0.131)	-0.297***(0.089)	-0.269***(0.094)	-0.399***(0.154)	-0.335***(0.082)	-0.145 (0.232)
Three days	-0.354***(0.063)	-0.263** (0.110)	-0.395***(0.078)	-0.319***(0.070)	-0.446***(0.140)	-0.320***(0.066)	-4.951***(0.200)
Four days	-0.387***(0.079)	-0.382***(0.120)	-0.388***(0.122)	-0.405***(0.098)	-0.315** (0.155)	-0.356***(0.088)	-0.864 (2.061)
Five days	-0.406***(0.067)	-0.542***(0.105)	-0.326***(0.081)	-0.438***(0.080)	-0.304***(0.118)	-0.417***(0.072)	-0.360 (0.168)
Six days	-0.469***(0.105)	-0.660***(0.170)	-0.276** (0.109)	-0.522***(0.123)	-0.258 (0.197)	-0.402***(0.122)	-4.874***(0.211)
Seven days	-0.454***(0.048)	-0.489***(0.057)	-0.415***(0.065)	-0.449***(0.053)	-0.459***(0.093)	-0.483***(0.045)	-0.318 (0.088)
Juice consumption							
Less than once	-0.157***(0.044)	-0.103 (0.064)	-0.189***(0.057)	-0.129***(0.045)	-0.201***(0.065)	-0.158***(0.036)	-0.132 (0.107)
At least once	-0.255***(0.047)	-0.218***(0.073)	-0.277***(0.072)	-0.223***(0.055)	-0.304***(0.086)	-0.275***(0.050)	-0.155 (0.125)
At least four times	-0.182***(0.053)	-0.068 (0.077)	-0.259***(0.075)	-0.151** (0.061)	-0.235***(0.083)	-0.164***(0.052)	-0.288 (0.128)
Seven times	-0.128** (0.051)	-0.070 (0.074)	-0.170***(0.063)	-0.055 (0.051)	-0.344***(0.102)	-0.120** (0.051)	-0.149 (0.120)
＞ seven times	-0.103 (0.096)	0.076 (0.159)	-0.248* (0.137)	-0.031 (0.097)	-0.373 (0.922)	-0.114 (0.106)	-0.040 (1.259)
Education							
None	0.529***(0.071)	0.542***(0.093)	0.539***(0.134)	0.560***(0.078)	0.417 (0.280)	0.454***(0.082)	1.169 (1.412)
Primary	0.267***(0.056)	0.232***(0.074)	0.300***(0.115)	0.273***(0.067)	0.204 (0.259)	0.212***(0.068)	0.762 (1.384)
Secondary	0.116* (0.065)	0.109 (0.094)	0.120 (0.114)	0.125* (0.069)	0.031 (0.275)	0.113 (0.090)	0.344 (1.390)
Income							
1st	0.393***(0.054)	0.471***(0.090)	0.353***(0.071)	0.438***(0.079)	0.354***(0.113)	0.425***(0.068)	0.263**(0.108)
2nd	0.234***(0.052)	0.296***(0.095)	0.200***(0.067)	0.235***(0.064)	0.240***(0.090)	0.273***(0.056)	0.058 (0.108)
3rd	0.184***(0.044)	0.200***(0.073)	0.182***(0.052)	0.173***(0.045)	0.220** (0.087)	0.200***(0.051)	0.133 (0.092)
4th	0.122***(0.035)	0.151* (0.078)	0.100* (0.055)	0.117***(0.040)	0.146* (0.085)	0.137***(0.043)	0.082 (0.112)
Health Insurance							
Public	-0.102 (0.098)	0.073 (0.144)	-0.191 (0.145)	-0.033 (0.118)	-0.176 (0.167)	-0.050 (0.104)	-0.296 (0.247)
Private	-0.139 (0.143)	-0.051 (0.203)	-0.228 (0.230)	-0.064 (0.168)	-0.393 (2.235)	-0.092 (0.158)	-0.305 (2.296)
Out-of-pocket	-0.013 (0.113)	0.194 (0.179)	-0.181 (0.171)	0.031 (0.139)	-0.042 (0.170)	0.076 (0.138)	-0.336 (0.262)
Unmet healthcare needs	0.374***(0.036)	0.422***(0.056)	0.337***(0.059)	0.333***(0.049)	0.463***(0.081)	0.360***(0.041)	0.429***(0.093)
Constant	-2.791***(0.198)	-3.255***(0.274)	-2.487***(0.432)	-2.787***(0.238)	-2.640***(0.572)	-1.928***(0.215)	-3.183**(1.464)

* Outcome equation: Suboptimal SRH

Table 5: Continued

		Full Sample (1)	Male (2)	Female (3)	Rural (4)	Urban (5)	Chronic (6)	Non-chronic (7)
* Selection equation: Healthcare Utilization	Family physician knowledge	0.262***(0.025)	0.319***(0.039)	0.195***(0.041)	0.275***(0.030)	0.227***(0.049)	0.238***(0.034)	0.289***(0.039)
	Age groups							
	25-34	-0.120***(0.043)	-0.053 (0.068)	-0.122** (0.059)	-0.109** (0.053)	-0.146 (0.100)	0.001 (0.082)	-0.176***(0.057)
	35-44	-0.365***(0.047)	-0.038 (0.086)	-0.446***(0.054)	-0.351***(0.051)	-0.406***(0.111)	-0.133* (0.081)	-0.494***(0.066)
	45-54	-0.332***(0.046)	0.065 (0.085)	-0.455***(0.060)	-0.328***(0.059)	-0.356***(0.100)	-0.060 (0.079)	-0.552***(0.059)
	55-64	-0.174***(0.050)	0.208** (0.085)	-0.303***(0.067)	-0.138** (0.055)	-0.272** (0.107)	0.103 (0.083)	-0.499***(0.094)
	65-74	-0.030 (0.057)	0.350***(0.094)	-0.193** (0.080)	-0.008 (0.065)	-0.099 (0.109)	0.207** (0.090)	-0.261***(0.109)
	75+	0.024 (0.067)	0.495***(0.110)	-0.245** (0.101)	0.038 (0.085)	-0.045 (0.137)	0.274***(0.096)	-0.513***(0.173)
	Gender	-0.365***(0.024)	-	-	-0.377***(0.025)	-0.346***(0.051)	-0.264***(0.031)	-0.497***(0.036)
	Location	-0.055** (0.024)	-0.060* (0.036)	-0.041 (0.030)	-	-	-0.089***(0.029)	0.009 (0.043)
	Marital Status							
	Single	-0.475***(0.057)	-0.011 (0.133)	-0.767***(0.078)	-0.470***(0.064)	-0.478***(0.124)	-0.318***(0.080)	-0.542***(0.129)
	Married	0.048 (0.044)	0.032 (0.102)	-0.022 (0.047)	0.060 (0.052)	0.004 (0.080)	0.011 (0.046)	0.118 (0.124)
	Divorced	-0.040 (0.072)	0.148 (0.154)	-0.171* (0.089)	-0.073 (0.079)	0.138 (0.168)	0.017 (0.084)	-0.133 (0.154)
	Risk factors							
	Chronic health problem	0.508***(0.025)	0.650***(0.032)	0.412***(0.029)	0.474***(0.025)	0.615***(0.042)	-	-
	Alcohol consumption	0.162***(0.059)	0.151** (0.059)	-0.082 (0.245)	0.171***(0.061)	0.130 (0.109)	0.163** (0.070)	0.076 (0.123)
	Tobacco consumption	0.061** (0.025)	0.052 (0.034)	0.072** (0.031)	0.043 (0.027)	0.130** (0.053)	0.061** (0.029)	0.057 (0.039)
	Education							
	None	-0.054 (0.044)	0.024 (0.071)	-0.088 (0.053)	-0.055 (0.046)	0.034 (0.113)	-0.019 (0.046)	-0.083 (0.060)
	Primary	-0.077** (0.031)	-0.049 (0.049)	-0.121***(0.045)	-0.096***(0.035)	0.045 (0.106)	-0.048 (0.036)	-0.094**(0.040)
	Secondary	-0.016 (0.037)	0.046 (0.058)	-0.074 (0.055)	-0.001 (0.037)	-0.096 (0.132)	0.021 (0.048)	-0.041 (0.046)
	Health Insurance							
	Public	0.182** (0.077)	0.214* (0.127)	0.196** (0.096)	0.151 (0.098)	0.242* (0.128)	0.236***(0.089)	0.091 (0.115)
	Private	0.061 (0.090)	0.022 (0.153)	0.091 (0.137)	0.066 (0.094)	-0.023 (0.266)	0.133 (0.119)	-0.084 (0.178)
	Out-of-pocket	0.045 (0.088)	0.022 (0.137)	0.064 (0.119)	-0.041 (0.116)	0.220 (0.150)	0.099 (0.110)	-0.022 (0.141)
	Unmet healthcare needs	0.030 (0.029)	0.039 (0.048)	0.028 (0.029)	0.004 (0.035)	0.100** (0.045)	-0.018 (0.039)	0.163***(0.057)
	Constant	-1.063***(0.103)	-1.962***(0.189)	-0.754***(0.147)	-1.077***(0.119)	-1.223***(0.186)	-0.828***(0.132)	-0.902***(0.167)
	ATE	-0.035* (0.019)	-0.029 (0.019)	-0.048 (0.057)	-0.031 (0.019)	-0.063 (0.064)	-0.063 (0.042)	-0.006 (0.011)
	ATT	-0.077 (0.057)	-0.084 (0.070)	-0.089 (0.154)	-0.068 (0.059)	-0.137 (0.212)	-0.099 (0.079)	-0.017 (0.063)
Log likelihood		-15440.34	-5984.04	-9336.04	-11259.59	-4150.28	-10812.59	-4531.91
Bootstrap replications		100	100	100	100	100	100	50
Error correlation (ρ)		0.396	0.451	0.400	0.379	0.507	0.384	0.459
Wald test of ρ (p-value)		11.20 (0.0008)	8.71 (0.0032)	1.50 (0.2200)	9.02 (0.0027)	1.44 (0.2303)	5.51 (0.0189)	6.91 (0.0086)
Murphy's score test (p-value)		52.38 (0.0000)	23.96 (0.0044)	16.57 (0.0558)	44.29 (0.0000)	15.31 (0.0827)	13.95 (0.1242)	-
Number of observations		24022	11477	12545	18129	5893	11993	12029

Notes: The Wald test reports the chi-square statistic and p-values for the null hypothesis that the bivariate probit error correlation is zero (i.e. health- care utilization is exogenous). Murphy's score test reports chi-square statistic and the p-values for the null hypothesis that the error terms are bivariate standard joint normal. ATE: Average treatment effect. ATT: Average treatment effect on the treated. Bootstrapped standard errors in parentheses. *, ** and *** denote statistical significance at 10, 5 and 1 percent level respectively.

Table 6: GP Care, Probit Estimates

	Full Sample (1)	Male (2)	Female (3)	Rural (4)	Urban (5)	Chronic (6)	Non-chronic (7)
GP care utilization	-0.099 (0.060)	0.003 (0.080)	-0.212** (0.090)	-0.128*(0.073)	-0.020 (0.104)	-0.122*(0.071)	-0.019 (0.114)
Age groups							
25-34	0.168 (0.109)	0.352**(0.153)	0.027 (0.145)	0.104 (0.121)	0.299 (0.241)	0.062 (0.132)	0.507***(0.169)
35-44	0.332***(0.106)	0.614***(0.155)	0.184 (0.140)	0.253**(0.119)	0.495**(0.230)	0.304**(0.133)	0.427**(0.172)
45-54	0.552***(0.105)	0.739***(0.160)	0.475***(0.134)	0.459***(0.117)	0.759***(0.231)	0.520***(0.130)	0.642***(0.180)
55-64	0.798***(0.109)	1.020***(0.162)	0.703***(0.142)	0.703***(0.123)	1.020***(0.242)	0.736***(0.133)	1.118***(0.197)
65-74	0.986***(0.114)	1.236***(0.170)	0.883***(0.149)	0.885***(0.132)	1.210***(0.246)	0.956***(0.138)	1.010***(0.222)
75+	1.228***(0.127)	1.479***(0.190)	1.137***(0.168)	1.142***(0.153)	1.421***(0.253)	1.185***(0.149)	1.441***(0.266)
Gender	0.021 (0.050)	-	-	0.010 (0.060)	0.049 (0.087)	0.001 (0.055)	0.098 (0.107)
Location	0.035 (0.043)	0.026 (0.064)	0.024 (0.057)	-	-	0.043 (0.047)	0.005 (0.095)
Marital Status							
Single	0.214**(0.105)	0.581***(0.193)	-0.039 (0.135)	0.174 (0.124)	0.316 (0.205)	0.162 (0.120)	0.282 (0.225)
Married	0.046 (0.071)	0.112 (0.158)	0.054 (0.081)	0.050 (0.089)	0.052 (0.109)	0.064 (0.074)	-0.088 (0.193)
Divorced	0.001 (0.131)	0.431*(0.227)	-0.156 (0.162)	0.004 (0.151)	-0.012 (0.252)	0.007 (0.142)	-0.119 (0.298)
Risk factors							
Chronic health problem	0.921***(0.052)	0.982***(0.069)	0.893***(0.078)	0.972***(0.065)	0.812***(0.086)	-	-
Alcohol consumption	0.207*(0.106)	0.235**(0.108)	0.278 (0.433)	0.282**(0.130)	0.026 (0.153)	0.232**(0.115)	0.087 (0.300)
Tobacco consumption	0.022 (0.048)	-0.123**(0.060)	0.216***(0.068)	0.030 (0.056)	-0.012 (0.084)	0.056 (0.053)	-0.195*(0.106)
Obesity	0.132***(0.045)	0.070 (0.078)	0.162***(0.055)	0.098*(0.055)	0.219***(0.073)	0.142***(0.048)	0.062 (0.129)
Exercise							
Once	-0.202*(0.121)	-0.314 (0.206)	-0.210 (0.140)	-0.208 (0.146)	-0.200 (0.192)	-0.176 (0.126)	-
Twice	-0.238**(0.103)	-0.288*(0.168)	-0.207 (0.128)	-0.214*(0.119)	-0.312 (0.207)	-0.311***(0.103)	0.192 (0.249)
Three days	-0.399***(0.083)	-0.229 (0.140)	-0.494***(0.103)	-0.386***(0.095)	-0.463***(0.159)	-0.379***(0.088)	-
Four days	-0.461***(0.097)	-0.430***(0.156)	-0.484***(0.125)	-0.533***(0.116)	-0.298*(0.175)	-0.431***(0.105)	-0.932***(0.347)
Five days	-0.325***(0.091)	-0.460***(0.148)	-0.250** (0.118)	-0.372***(0.112)	-0.200 (0.154)	-0.336***(0.102)	-0.271 (0.187)
Six days	-0.440***(0.136)	-0.687***(0.195)	-0.222 (0.188)	-0.482***(0.160)	-0.268 (0.244)	-0.376** (0.150)	-
Seven days	-0.433***(0.049)	-0.470***(0.068)	-0.404***(0.068)	-0.452***(0.061)	-0.407***(0.078)	-0.460***(0.054)	-0.297***(0.102)
Juice consumption							
Less than once	-0.126**(0.054)	-0.089 (0.084)	-0.154** (0.070)	-0.070 (0.069)	-0.246***(0.079)	-0.132**(0.057)	-0.042 (0.138)
At least once	-0.256***(0.061)	-0.209**(0.096)	-0.291***(0.080)	-0.202***(0.077)	-0.374***(0.095)	-0.262***(0.067)	-0.167 (0.145)
At least four times	-0.102 (0.068)	0.050 (0.099)	-0.225** (0.093)	-0.046 (0.085)	-0.238**(0.105)	-0.084 (0.075)	-0.161 (0.158)
Seven times	-0.200***(0.067)	-0.114 (0.100)	-0.277***(0.092)	-0.126 (0.082)	-0.436***(0.113)	-0.177**(0.074)	-0.263*(0.153)
seven times	-0.073 (0.142)	0.161 (0.185)	-0.286 (0.205)	-0.039 (0.152)	-0.189 (0.354)	-0.120 (0.152)	0.161 (0.291)
Education							
None	0.717***(0.098)	0.663***(0.139)	0.812***(0.146)	0.725***(0.109)	0.650**(0.263)	0.656***(0.106)	1.287***(0.294)
Primary	0.416***(0.090)	0.321***(0.115)	0.543***(0.142)	0.424***(0.096)	0.335 (0.255)	0.367***(0.097)	0.902***(0.282)
Secondary	0.215** (0.100)	0.178 (0.128)	0.254 (0.159)	0.232** (0.106)	0.011 (0.281)	0.203*(0.108)	0.472 (0.325)
Income							
1st	0.259***(0.076)	0.361***(0.115)	0.218** (0.099)	0.301***(0.106)	0.231*(0.128)	0.268***(0.083)	0.212 (0.163)
2nd	0.183***(0.070)	0.298***(0.105)	0.111 (0.094)	0.167*(0.089)	0.214*(0.126)	0.231***(0.078)	-0.044 (0.146)
3rd	0.076 (0.059)	0.138 (0.090)	0.042 (0.078)	0.058 (0.068)	0.131 (0.121)	0.078 (0.065)	0.060 (0.122)
4th	0.011 (0.060)	0.122 (0.093)	-0.070 (0.078)	0.022 (0.067)	-0.025 (0.131)	0.043 (0.066)	-0.188 (0.134)
Health Insurance							
Public	-0.069 (0.109)	0.119 (0.163)	-0.220 (0.144)	-0.023 (0.157)	-0.161 (0.145)	-0.074 (0.122)	-0.105 (0.232)
Private	-0.271 (0.187)	0.030 (0.213)	-0.640** (0.262)	-0.235 (0.207)	-0.469 (0.335)	-0.250 (0.204)	-0.348 (0.399)
Out-of-pocket	-0.034 (0.142)	0.338*(0.201)	-0.498***(0.186)	-0.046 (0.195)	-0.007 (0.199)	-0.035 (0.158)	-0.102 (0.301)
Unmet healthcare needs	0.406***(0.050)	0.459***(0.075)	0.379***(0.066)	0.394***(0.063)	0.465***(0.076)	0.412***(0.054)	0.376***(0.119)
Constant	-2.944***(0.203)	-3.526***(0.324)	-2.598***(0.272)	-2.919***(0.244)	-2.919***(0.395)	-1.944***(0.227)	-3.447***(0.436)
AME	-0.010*(0.006)	-0.003 (0.007)	-0.023**(0.010)	-0.012*(0.007)	-0.002 (0.012)	-0.021*(0.012)	-0.0004 (0.002)
AME by age group							
25-34	-0.011*(0.007)	-0.0004 (0.007)	-0.024**(0.011)	-0.013*(0.008)	-0.003 (0.015)	-0.022*(0.013)	-0.0009 (0.006)
35-44	-0.013*(0.008)	-0.0005 (0.011)	-0.027**(0.012)	-0.015*(0.009)	-0.003 (0.017)	-0.026*(0.016)	-0.0009 (0.005)

45-54	-0.015*(0.009)	-0.0005 (0.012)	-0.032**(0.014)	-0.017*(0.010)	-0.004 (0.020)	-0.029*(0.017)	-0.001 (0.007)
55-64	-0.018*(0.011)	-0.0006 (0.015)	-0.039**(0.017)	-0.021*(0.012)	-0.004 (0.023)	-0.034*(0.020)	-0.002 (0.014)
65-74	-0.021*(0.013)	-0.0008 (0.018)	-0.045**(0.020)	-0.025*(0.015)	-0.005 (0.026)	-0.039*(0.023)	-0.002 (0.013)
75+	-0.025*(0.015)	-0.0009 (0.021)	-0.053**(0.023)	-0.030*(0.018)	-0.005 (0.029)	-0.042*(0.024)	-0.003 (0.022)
AME by gender							
Male	-0.010*(0.006)	-	-	-0.012*(0.007)	-0.002 (0.012)	-0.021*(0.012)	-0.0005 (0.002)
Female	-0.011*(0.006)	-	-	-0.012*(0.007)	-0.002 (0.012)	-0.021*(0.012)	-0.0004 (0.002)
AME by location							
Rural	-0.009*(0.006)	0.0003 (0.007)	-0.023**(0.009)	-	-	-0.022*(0.012)	-0.0004 (0.003)
Urban	-0.009*(0.006)	0.0003 (0.006)	-0.023**(0.009)	-	-	-0.020*(0.012)	-0.0004 (0.003)
AME by chronicity							
Chronic	-0.014*(0.009)	0.0004 (0.011)	-0.031**(0.013)	-0.018*(0.010)	-0.003 (0.016)	-	-
Non-chronic	-0.004*(0.002)	0.001 (0.002)	-0.009**(0.004)	-0.004*(0.002)	-0.001 (0.006)	-	-
Pseudo R^2	0.2521	0.2647	0.2549	0.2498	0.2587	0.1433	0.1804
Number of observations	24022	11477	12545	18129	5893	11993	10711

Notes: The outcome variable is SRH (binary). AME: Average marginal effect. Robust standard errors in parentheses. All specifications use sampling weights provided by TurkStat. *, ** and *** denote statistical significance at 10, 5 and 1 percent level respectively.