# Patients are different!
# Explaining the relation between pills and cholesterol

Domenico Depalo *

**Abstract**

This paper investigates the relation between drugs and cholesterol, by jointly taking into account the possible sorting of patients into treatment and the heterogeneity in the effect, as suggested by a simple proposed theoretical model. To this aim, Marginal Treatment Effect (MTE) is estimated. Consistent with the predictions of the model, not only do drugs successfully decreases the level of cholesterol, but also patients who benefit most from the treatment are more likely to adhere to prescribed regime. These results are used in a second step to study the effects of thoughts policies aiming at increasing patients' compliance (and thus lower cholesterol level) that affect either the ultimate target (patients) or the intermediate target (doctors). A contribution of the paper is the description of characteristics of patients switched into treatment. Whilst targeting the patients provide the largest reduction of cholesterol, the policies differ substantially with respect to the population affected, therefore one of the two may be preferred to the other, depending on which population the policy maker wishes to target.

JEL classification:

Keywords: Marginal Treatment Effect, Cholesterol, Drug

*I would like to thank David Card, Alessandra Fenizia, Magne Mogstad, Edward Norton, Jeffrey Wooldridge .... Special thanks to Vincenzo Atella, who gave me the permission to use the data. Replication files and additional results will be available at the webpage: http://sites.google.com/site/domdepalo/ The views expressed in this paper are those of the author and do not imply any responsibility of the Bank of Italy. Corresponding address: Domenico Depalo, Banca d'Italia, Economics and Statistics Department, Via Nazionale, 91 - 00184 Rome (Italy), Tel.: +39-06-4792 5989, e-mail: domenico.depalo@bancaditalia.it

# 1 Introduction

Paul and John are homozygous twins with a very similar way of life. Recently, they were diagnosed by high blood cholesterol. Whilst John closely follows the prescribed treatment, Paul does not. He is not convinced by his doctor that it is worth following it.

The main motivation of this paper is about the precise meaning of the expression 'not worth following' the prescribed regime, using a simple theoretical framework. As one may easily predict, the worthiness depends on the comparison between benefits and opportunity costs. However, to the extent that they are heterogeneous in the population, standard quantities like the Average Treatment Effect (ATE), maybe for the treated (ATT) or untreated (ATU) patients, may not address the original question. For this reason, in this paper I focus on the marginal individual who is induced to follow the prescribed treatment regime (Heckman and Vytlacil, 2007a). To this aim, I identify the Marginal Treatment Effect (MTE). This quantity was introduced by Bjorklund and Moffitt (1987) and further developed in a series of contributions by Heckman and Vytlacil (2001, 2005, 2007a,b). MTE is not new in health economics: examples include Basu et al. (2007); Doyle (2007); Maestas et al. (2013); French and Song (2014); Kowalski (2016). However, I have not seen applications of MTE to health outcomes of patients.

The health outcomes analyzed in this paper is cholesterol. This pathology is widespread in the population: during the period 2009-12, the incidence of high 'bad' cholesterol in the Italian population between 35–75 years old was almost 70%, without large differences between men and women. The condition is common to several countries: for example, between 2009 and 2012 in the US more than 130 million adults had high cholesterol-level (American Heart Association, 2015). Besides biostatistical considerations, an higher adherence to the prescribed treatment regime is associated with higher patient welfare and would save money (Lamiraud and Geoffard, 2007). Therefore policy makers should enhance the compliance to prescribed regime. A key question is about which policy may better serves the scope. Carneiro et al. (2003, 2010, 2011) introduce and develop the Policy Relevant Treatment Effect (PRTE), that estimates the average treatment effect for individuals who are not treated under the current policy but that would be treated under

a different regime. To my knowledge, this is new contribution in health economics. Of course, several policies may be introduced. In this paper I propose three different scenarios: one affects the individual 'baseline' compliance of patients, one affects the communication skills of the doctors, and one affects the overall probability of following the prescribed treatment (for unspecified reasons). The effect of these hypothetical policies is maximized in the first policy. However, looking only at the magnitude of the effect might miss important differences.

As an extension to the literature employing MTE, I propose to look at which population is actually targeted by each policy. This question is the natural counterpart to what is nowadays standard for the Local Average Treatment Effect (LATE; Imbens and Angrist, 1994). The empirical content of this exercise is huge, because it may help tailor most appropriate policies, that targets the desired population.

This paper employs data from Italy. Although no attempt to generalize the results beyond Italy will be made, it should be clear at the outset that non-adherence is a generalized problem around the world. According to estimates provided in a report by the New England Healthcare Institute (2009), the annual cost of non adherence in the United States is approximately $290 billion (2.3% in terms of GDP), therefore improving compliance to medical prescriptions is key to improve health outcomes and lowering health care costs.

Using the standard Ordinary Least Square (OLS) compliance to prescribed treatment regime reduces cholesterol by 2.9 – 3.4 %, depending on the set of covariates. This estimate is biased if the sample of patients who follow the rule is not random, for example because they do not think it is worth following the regime, exactly like Paul does. The standard cure for this issue is an Instrumental Variable (IV) estimator, that suggests a reduction of cholesterol by 4.2 – 5.1 %. Even this consistent estimate is not fully informative. If the effect in the population is heterogeneous, it is unlikely (but not impossible) that a single number summarize the entire heterogeneity. By using MTE, I can draw better defined conclusions. Patients who benefit most from drugs 'need less argument' to be convinced to comply with the treatment; in contrast, those who benefit the least are harder to convince. This behaviour is clearly reflected in the MTE estimates: the former group of patients -to which John belongs to- enjoy the largest reduction in cholesterol (5%), while the

reduction for the latter group -to which Paul belongs to- is remarkably smaller. Among the policies proposed to enhance compliance, the most effective in terms of magnitude targets patients healthier than those targeted by the second most effective. This is a perfect setup where policy maker may maximize her/his objective function by targeting specific subpopulations not necessarily coinciding with those benefiting most from it.

The paper is organized as follows. In Section 2 I present a prototypical theoretical model that will be used to interpret the results. Section 4 describes the data that are used in the empirical application presented in Section 5, whose robustness is checked in Section 6. Section 7 offer some conclusions.

## 2 An economic framework

This paper first explains what the expression 'not worth following the prescribed regime' means. To this aim, a prototypical model of choice and associated outcomes is used to predict the patients' behviour (Roy, 1951). The model will be the background to interpret the results of the empirical section. See Heckman and Vytlacil (2007a,b); Carneiro et al. (2003) for a similar approach (from which I borrowed). A patient may be compliant with the therapy (or treated; subscript 1) or not (or untreated; subscript 0), but not simultaneously compliant *and* non compliant (more formally, the treatment states are mutually exclusive). Define $y_d$ the potential outcome under treatment $d$, the benefit from treatment is $-[y_1 - y_0]$.[1] Associated to treatment are opportunity costs ($c$). Following a recent but increasingly relevant strand of the literature, the cost I focus on will be directed towards understanding the relation between patients and doctors (Fichera et al., 2017; Atella et al., 2017). The relation is proxied by the communication skills of the doctor, who may convince individuals at the margin of indifference between treatment and non-treatment. A rational patient is compliant if

$$-[y_1 - y_0] - c \geq 0, \tag{1}$$

[1] More precisely, the benefit is *proportional* to $y_1 - y_0$. I normalized the monetary unit of benefits to unity; similar reasoning applies to opportunity cost.

i.e. if the benefit from a reduction of cholesterol induced by compliance is larger than the corresponding cost. This simple mechanism has a large empirical content. Using a growth rate $g$, define $y_1 = y_0 (1+g)^{-1}$, such that the higher $g$ the lower the cholesterol level for compliant patients with respect to non compliance status. The compliance will be endogenously preferred to non-compliance until

$$\begin{aligned} -y_0 \left[(1+g)^{-1} - 1\right] - c &\geq 0 \qquad (2) \\ -(1+g)^{-1} + 1 - \frac{c}{y_0} &\geq 0 \\ 1 - \frac{c}{y_0} &\geq (1+g)^{-1} \qquad (3) \end{aligned}$$

As simple as it is, this model provides a number of useful insights. First of all, 'it is worth following the prescribed regime' until the point where the cholesterol under treatment (indexed by $(1+g)^{-1}$) is smaller than the associated cost $(1 - \frac{c}{y_0})$. Second, focusing on the form of the equilibrium, the higher the cost to be compliant (i.e., high $c$), the larger the minimum gain required in terms of lower cholesterol and the lower the likelihood that the patient is treated. Third, the better the health conditions without compliance (i.e., low $y_0$), the larger the minimum gain required to comply with the prescribed treatment and the lower the likelihood that the patient is treated.

To better see these predictions at work, in Figure 1 I plot the threshold condition $1 - \frac{c}{y_0}$, which is function of costs and conditions without compliance. In order to have a 2-dimensions plot, I fixed $C \in \{C_1, C_2\}$, with $C_1 < C_2$ and let health conditions under non-compliance ($y_0$) free to vary. For any fixed health condition under non-compliance, the required gain increases with cost (the curve defined by $C_2$ is below the curve defined by $C_1$). For any fixed cost, as going from from $y^1$ to $y^2$ (i.e., from better to worse conditions under non compliance), the required gain becomes smaller for all the curves identified by cost $C$.

Finally, and most important for the following analysis, the sample of compliant patients is not random, but defined by inequality 1. This simple prototypical model delivers who switches into treatment and who not: Paul does not comply with the regime prescribed by the doctor because the (marginal) opportunity cost of the therapy is larger than the marginal benefit; at the opposite

sight, for John the benefit from adherence consumption overcomes its cost. Both costs and health conditions under non-compliance vary at patient level, therefore the minimum required gain is individual specific, which leads to essential heterogeneity in the treatment effect. The rest of the paper is about an appropriate estimator for the heterogeneous effect under non random sampling.

# 3 Treatment Effect(s)

In this section I review methods that identify the treatment effect under essential heterogeneity, a feature that emerges directly from the prototypical model of Section 2. Throughout the section I will focus only on the main issues for the present analysis; the reader interested in further details and/or technical details should refer to the original papers.

Define the (potential) outcomes as:

$$\begin{aligned} \ln y_0 &= \alpha + U_0 \\ \ln y_1 &= \alpha - \overline{\beta} + U_1 \end{aligned} \tag{4}$$

where $\alpha$ is an intercept, $\overline{\beta}$ is a parameter equal to $\ln y_1 - \ln y_0$ and $U_0$ and $U_1$ are random errors under treatment states equal to 0 and 1, respectively. The observed outcome is $\ln y = D \ln y_1 + (1 - D) \ln y_0 = D(\alpha - \overline{\beta} + U_1) + (1 - D)(\alpha + U_0) = \alpha + \{-\overline{\beta} + (U_1 - U_0)\} D + U_0$; upon substitution of $\eta = (U_1 - U_0)$, $\beta = -\overline{\beta} + \eta$ and $\epsilon = U_0$, the model for the outcome can be written $\ln y = \alpha + \beta D + \epsilon$. This specification fits perfectly into the economic framework of Section 2, with $\beta = \ln(1 + g)$.[2] It also shows that $D$ may be correlated with $\epsilon$, for example because of genetic endowment, and with $(U_1 - U_0)$, for example because compliant patients anticipate their (expected) gain from treatment. The Roy (1951) model predicts that $D = 1$ if $\ln(y_1) \leq \ln(y_0)$, or equivalently $\beta \leq 0$. Therefore, patients are compliant with prescribed treatment regime because they enjoy a sufficiently high reduction of cholesterol, namely higher than a patient chosen at random, i.e $ATT < ATE$ (under essential heterogeneity); viceversa, for non-treated individuals $ATE < ATU$. These are

[2] From eq. 4, $\ln y_1 - \ln y_0 = -\beta$. By definition of $y_1 = y_0(1 + g)^{-1}$, it follows that $\ln y_1 = \ln y_0 - \ln(1 + g)$. Therefore, $-\beta = \ln y_1 - \ln y_0 = \ln y_0 - \ln(1 + g) - \ln y_0 = -\ln(1 + g)$.

necessary conditions for concordance between the model specification and the data, under essential heterogeneity. If the effect in the population is homogeneous, $ATE = ATT = ATU$ (notice that if we knew all these quantities, this definition may be used as a null hypothesis of non-essential heterogeneity, as I do in Section 5). An ideal estimator should jointly grasp all these properties.

The main imlpication of the Roy model is that the sample of compliant patients is not a (completely) random sample of the population (Little, 1995), and the OLS is inconsistent (Heckman, 1978, 1979, 2010). To formalize the selection into treatment, I define

$$D^* = \gamma Z - V \tag{5}$$

$$D = 1 \text{ if } D^* \geq 0 \tag{6}$$

where $Z$ includes an indicator, called instrument, that is relevant to explain the compliance mechanism but not the outcome, e.g. the communication skill of the doctor. Separability of $D^*$ is maintained here only for simplicity of exposition, but is not required to derive what follows. For later reference, by taking the cumulative distribution function (CDF) of $V$, eq. 5 implies that $D = 1$ if $F_V(\gamma Z) \geq F_V(V)$, where $F_V(\gamma Z) \equiv P(Z)$ is a propensity score and $F_V(V) \equiv U_D \sim [0, 1]$ are quantiles of the distribution of the unobserved component.

A common approach to cure for the non-random selection of the sample is based on an Instrumental Variable (IV). If the treatment effects are heterogeneous, IV identifies a *local* average treatment effect (LATE):[3]

$$\frac{E[y|Z = z'] - E[y|Z = z]}{E[D|Z = z'] - E[D|Z = z]} = E[y_{z'} - y_z | P(z) < U_D < P(z')], \tag{7}$$

the treatment effect for *compliers* (Angrist et al., 1996), i.e. patients that are induced to comply with the treatment by a change in the instrument. It has been emphasized that not only do different instruments target two different subpopulations of compliers (i.e., estimate two different,

[3] The following assumptions are needed: *1)* the potential outcomes for each patient are unrelated to the treatment status of other patient (Stable Unit Treatment Value Assumption); *2)* the instrument is randomly assigned; *3)* exclusion restriction (i.e. $y_d \equiv Y(0, d) = Y(1, d)$); *4)* nonzero average causal effect of $Z$ on $D$ (i.e. $E[D_1 - D_0] \neq 0$); *5)* monotonicity (i.e. $\forall z' > z$, $D_{z'} \geq D_z$ for all patients, such that an increase in the level of the instrument does not decrease the level of the treatment, or vice-versa).

consistent, parameters; Card, 1999), but also that when the instrument takes more than 2 values, LATE using $(z_2, z_1)$ is different from LATE using $(z_4, z_3)$ because different marginal individuals sort into treatment (Heckman, 2010; for this reason, sometime I prefer the notation LATE$(z', z)$ instead of the more compact LATE). Doyle (2007) provides a nice graphical intuition about compliers, who is (marginally) induced into treatment, and the corresponding treatment effect.

The coefficient attached to treatment effect may thus be seen as a random variable with a distribution that can be summarized in different ways. To this aim, in this paper I exploit the Marginal Treatment Effect (MTE), introduced by Bjorklund and Moffitt (1987) and further developed in various papers by Heckman and Vytlacil (2000, 2001, 2005, 2007a,b). It identifies the treatment effect as $z' \to z$. Let $p \equiv P(Z)$ (only for notational convenience),

$$\begin{aligned} \lim_{z' \to z} \text{LATE}(z, z', X) &= E[y_{z'} - y_z | p < U_D < p'] \\ &= X\,\beta + \lim_{z' \to z} E[U_1 - U_0 | p < U_D < p'] \\ MTE(U_D = p) &= E[\Delta | U_D = p], \end{aligned} \tag{8}$$

i.e. the average reduction of cholesterol from adherence for patients just indifferent between treatment and non treatment at level of unobservables $U_D = p$; it may also be interpreted as a willingness to pay for patients at the margin of indifference for compliance given characteristics $(X, p)$. See Carneiro et al. (2003, 2017); Cornelissen et al. (2016). Patients with low realizations of $U_D$ are those whose unobservable characteristics make more likely to be compliant (because they would prefer treatment even if $\gamma Z$ is low); at the opposite side, for realizations of $U_D$ approaching 1 patients need high values of $\gamma Z$ to be induced to participate. This reasoning and the model of Section 2 together imply that the MTE should be 1) negative and 2) an *increasing* function of $U_D$: patients who enjoy a *larger reduction* of cholesterol ($MTE(U_D = p_{\text{low}}) << 0$) are convinced to be treated even though the communication skills of the doctor are low (i.e., low value of $U_D$) and therefore enter *first* into treatment; patients who enjoy a *smaller reduction* of cholesterol ($MTE(U_D = p_{\text{high}}) \to 0$) need more convincing arguments to be treated, i.e. the communication skills of the doctor must be high (i.e., high value of $U_D$), and therefore enter *last* into treatment (Carneiro et al., 2003, 2010;

Basu et al., 2007; Cornelissen et al., 2016; Carneiro et al., 2017). The ability to jointly grasp these properties makes of the MTE the 'ideal estimator' we were looking for. Also, a particularly attractive feature of the MTE is the unifying approach to the treatment effects, as one can summarize the distribution of $\overline{\beta}$ that best addresses the economic question of interest. Heckman and Vytlacil (2005) provide a set of weights that allow one to go from MTE to other relevant quantities, like ATE, ATT, ATU and LATE.

Applications of MTE are increasing over time. Cornelissen et al. (2016) provide several references, mostly in labour economics, where a much explored research question involves the return to education. Related to health, Basu et al. (2007) investigate the effects on costs of breast conserving surgery as opposed to mastectomy after 5-years. Doyle (2007) identifies the causal effects of foster care on long-term outcomes, like juvenile delinquency, teen motherhood, and employment, finding that children on the margin of placement tend to have better outcomes when they remain at home. Maestas et al. (2013); French and Song (2014) investigate the effect of the program Social Security Disability Insurance benefit receipt on labor supply finding that employment would have been higher by more than 25% had individuals not received benefits, with substantial heterogeneity of the effect, ranging from no effect for those with more severe impairments to about 50 percentage points for entrants with relatively less severe impairments. Brinch et al. (2012) assess the interaction between the quantity and quality of children to find that the family size effects vary in magnitude and even sign, and that families act as if they possess some knowledge of the idiosyncratic effects in the fertility decision. Kowalski (2016) investigate the effect of the program Oregon Health Insurance Experiment on the emergency room utilization and finds that it decreases from always takers to compliers to never takers. I have not seen paper that use MTE to investigate the health outcomes of patients.

### 3.1 Policy Relevant Treatment Effect

Beside the heterogeneity of the treatment effect, an interesting question is whether it is possible to implement a different policy ($D^*$) -instead of the current $D$- that makes Paul more likely to be compliant, i.e. increasing the treatment probability from $p$ to $p'$. If so, how large would the

gain be? In aging population societies, this type of questions will be increasingly asked by policy makers. The money at stake for the public finance is much: according to projections from European Commission (2015, p.128), the increase in public expenditure on health care over the period 2013-2060 with respect to GDP, would be about 15% for the European Union, the Euro Area, and Italy.

To answer these key questions, Heckman and Vytlacil (2005) introduce the Policy Relevant Treatment Effect (PRTE), which is the average effect of going from a baseline policy to an alternative policy per net person shifted, defined as $\frac{E[y|D^*]-E[y|D]}{E[D^*]-E[D]}$, where $D^*$ is a policy alternative to the current $D$ (so for example $E[y|D^*]$ is the shorthand for $E[y|\text{alternative policy}]$). Necessary for the interpretation of this quantity is the assumptions that the policy change does not affect the distribution of observable characteristics and of $U_D$: hence, an external manipulation of the policy does not affect anything in the model apart from the selection of the treatment. Necessary for the estimation of this quantity is the fact that $E[D^*] - E[D] \neq 0$: apart from the trivial mathematical content, the condition says that the policy needs to affect at least some individuals (and more precisely: at least one). As simple as it is, this definition allows for a general array of potential policies. In the application at hand, two types of applications are most interesting. Would it be better to implement policies capable of affecting the behaviour of the patients (externally influencing their compliance; this is 'direct policy') or the communication of the doctors (that affect the behaviour of the patients and ultimately their compliance; this is 'mediated policy')? In both cases, some components of the probability of treatment would be changed. In the former case, the *baseline* probability of being compliant with the prescribed policy regime would go from $p = [1, Z][\gamma_1, \gamma_Z]'$ to $p' = [1, Z][\gamma_1 + \alpha, \gamma_Z]'$; in the latter case, the policy instrument that affects the probability of being compliant would go from $p = [Z_j, Z_{\neq j}][\gamma_j, \gamma_{\neq j}]'$ to $p' = [Z_j, Z_{\neq j}][\gamma_j + \alpha, \gamma_{\neq j}]'$. Carneiro et al. (2010, 2011) consider the case of policy that increase the probability of being compliant by an amount $\alpha$ (from $p$ to $p' = p + \alpha$) and a policy that shifts each person's probability of being compliant by a proportion $(1 + \alpha)$ (from $p$ to $p' = p\,(1 + \alpha)$).

# 4 Data and descriptive statistics

In this section I briefly describe the data at hand. The reader interested in more details should refer to Atella et al. (2017). The data are taken from the Health Search Database (HSD), a longitudinal observational database collected by SIMG, an association of the Italian College of General Practitioners (GP).

Although the participation to HSD is on voluntary basis, important features of these data are common to administrative sources, namely the information provided is not affected by measurement errors or by systematic unit-/item- non-response. Accordingly, Fabiani et al. (2004) show that the data are suitable to carry out pharmacoepidemiological studies generalised to the whole Italian population. Lamiraud and Geoffard (2007) provide further motivations for preference of real-world-data (when available) over questionnaire, because they reveal preferences in real situations rather than in a hypothetical scenario. The information recorded by GP is rich and detailed with respect to health indicators. In contrast, individual socio-economic conditions are missing.

The model of Section 2 is based on the assumption that patients do not sort across GP. There are two reasons why this is not a concern, one on the econometric side and one institutional. If there exists a sorting mechanism, it will likely underestimate the effects of interest. There are two possible situations that induce bias. In one case, a non-compliant patient is bothered by the pressure of the doctor for becoming a complier, therefore he prefers to move to a negligent doctors with lower communication skills (in the limit: $U_D \to 0$): the Roy (1951) model predicts that these individuals benefit the most from the treatment, therefore these 'adversely sorted' individuals would attenuate the MTE at low values of $U_D$. In the second case, a concerned patient wishes to have pressures from the scrupulous doctor with better (probably more pressing) communication skills (in the limit: $U_D \to 1$): for the same reason exploited before, the treatment effect at high values of $U_D$ would now be upward biased (in absolute value). If patients switch for constant level of $U_D$, there will be no bias in the estimated effect. In Italy random sorting of patients across GP is guaranteed by the institutional reasons. All the individuals enrolled in the NHS are entitled to a GP, who is responsible for providing a list of services (among which are prescription of drugs,

requests for specialist visits and diagnostic tests). When an individual turns 14 she/he chooses a GP, from a list of available doctors that is administered by the so called ASL. The maximum number of patients per GP is 1,500. The relationship between GP and individual can be broken for two reasons: the GP has the power to refuse the assistance to an individual by communicating the motivations to ASL, that afterwards selects a new GP; the individual may choose to change a physician at any time without motivating the withdrawal. In practice, splits between patients and doctors are rare: GP would loose money from refusing patients; patients would not be guaranteed to have a GP in their neighborhood (actually this is almost sure because the supply of GP is scars: this increases the opportunity cost of patients to visit the GP) nor to have access to their preferred one (due to the cap to the number of patients).

In this paper, for the year 2004-08, I consider patients between 40 and 80 years old when they first entered the HSD, diagnosed of "pure hypercholesterolemia" (or familiar hypercholesterolemia) and with at least a prescription of statin medications.[4] The initial sample is then made of 643,346 observations, pertaining to 40,387 patients observed for 4.3 years (17.3 quarters), on average.

The empirical specification follows closely the economic model. The health indicator of interest is the (log-)ratio between serum levels of low-density lipoprotein (LDL) and total cholesterol, which is considered a better predictor of future heart attack risk than total cholesterol itself (Baigent et al., 2005). Also, taking a ratio provides a natural normalization for a better comparison across individuals with the same LDL, but different total cholesterol.

The model predicts that the likelihood to follow the treatment regime is a function of health conditions without compliance and the associated cost. Therefore three aspects are of interest: 1) how we define the treatment regime, 2) what observed characteristics affect $y_0$ and 3) what the associated cost is.

A crucial aspect for policy makers is the extent to which patients follow the medical prescriptions, or compliance/(non-)adherence.[5] Although in principle one may have access to this infor-

[4]Hypercholesterolemia is the presence of high levels of cholesterol in the blood. It is not a disease but a metabolic derangement that can contribute to many forms of disease, most notably coronary artery disease. The primary therapy to treat hypercholestorelemia is statin medications which limit the body's production of cholesterol.

[5]According to the International Society for Pharmacoeconomics and Outcomes Research (ISPOR), medication compliance refers to adherence and persistency to medical treatment: the former refers to the proportion of prescribed

mation, in practice almost always only proxies are available. Various proxies have been proposed (Hughes et al., 2001). In this paper I use the medication possession ratio (MPR) at patient level, defined as the number of days of medication supplied within the refill interval over the number of days in refill interval. This proxy is widely used in the literature, mainly thanks to its simplicity of calculation and interpretation (Cramer et al., 2008). Atella et al. (2017) provide details about the construction of the variable with these data. As in Lamiraud and Geoffard (2007), I work with a discrete indicator of compliance: if MPR is at least 0.75, the patient is defined as compliant/adherent. Nonetheless, I experimented the threshold in the range $[0.7, 0.8]$ and verified that the qualitative conclusions of the paper are confirmed. **da fare**

The second important aspect is related to the conditions that affect $y_0$. The data provide several details on this aspect, which are exploited as covariates. Beside the level of cholesterol when the individual first enter the sample, I can control for a general summary of health conditions through the Charlson et al. (1987) index as well as for specific pathologies, namely diabetes, hypertension, TIA, PTCA, IMA, ictus and angina pectoris. These are dummy variables that take value 1 if the condition is verified in that specific patient. Finally, I add some control variable that may help better explain the outcome. Related to cholesterol, it is widely known that some statins are more powerful than others, so that an equipotency index is used (Maron et al., 2000): for a given quantity of statins, the higher the equipotency the larger the reduction of cholesterol and the lower the opportunity cost of compliance. Related more general to the overall health conditions of the individuals, I control for the only demographic characteristics that I have, namely gender and age; a set of dummies for years and quarters completes the specification.

In Table 1 I report some descriptive statistics, for the pooled sample. The health outcomes of compliant patients are better than for compliant patients by 3.5% and the heterogeneity measured by the standard deviation is pretty similar between the two samples. The difference in cholesterol across the two groups is slightly decreasing in quantiles: it is about 4% at 10th quantile, 3.6 at the median and about 3 at 90th quantile. I take this sample statistics as suggesting that there might be heterogeneity in the treatment effect (although -at this stage- not necessarily 'essential'

doses taken in the prescribed time interval; the latter refers to the continued use of a prescribed therapy over time. With a little abuse of terminology, I use adherence and compliance as synonymous.

in the sense of Heckman et al., 2006). Non compliant patients are usually in better general-health conditions as measured by the Charlson et al. (1987) index and slightly younger than compliant. The differences for other observable covariates are usually small. Only by gender, the share of women among non-compliant is remarkably larger than the share of women among compliant (the share of women in the population is well above 55%, therefore this number simply reflects the demographic distribution in the population).

# 5 Empirical analysis

In this section I present the empirical analysis. I first show that the required hypothesis for the identification of MTE are verified and then I present the result.

### Preliminary analysis

The preliminary analysis tackles three different issues: 1) the existence of a valid instrument, 2) the common support assumption (for a meaningful comparison), 3) whether the 'essential heterogeneity' is a feature of the treatment effect. The first two aspects are common to any treatment evaluation; the last is somewhat specific to the approach taken in this paper.

Paul and John have a different attitude towards compliance. As increasingly emphasized, this attitude is attributable in large part to doctors: with these data, Atella et al. (2017) conclude that physicians have an important role in determining patient health status, through what they call 'physician endowment', like communication skills; similar conclusions were reached by Burroughs et al. (2002); Ginsburg et al. (2005); Mahlknecht and Voelter-Mahlknecht (2005); Stewart (1996); Roter et al. (2002). Fichera et al. (2017) successfully tested a formal theoretical model where doctors' effects exerts a positive effect on patients behaviour. Based on this literature, I proxy the communication skill of th doctors with the average compliance of their patients. Thus, for each patient $i$ I use as instrument the average compliance of the other patients within the same GP (excluding $i$), i.e. $Z_i = \sum_{j \neq i} \text{compliance}_j$. Similar instruments have been used extensively in other fields (Kling, 2006; Doyle, 2007; Maestas et al., 2013). It is valid if the communication

ability is relevant to explain the compliance of individual $i$, without impact on the cholesterol level. The need to break any dependence (even due to unobservable characteristics) affecting both the treatment status and the outcome is the reason why I exclude the $i$-th patient from his/her instrument. A possible critique to this instrument is that patients $j \neq i$ of doctor $A$ are more compliant to prescribed treatment regime than patients of doctor $B$, because the former doctor has better clinical knowledge than the latter and, as a consequence, the communication skill is in fact an ability channel with direct, unobservable, effect on the level of cholesterol. While this may certainly be true for some doctors, I think it does not jeopardize the results: 1) the pathology under analysis is widespread (with an incidence of high LDL values as large as 70% for basically the same ages considered in this analysis); 2) the related medical research is well developed; 3) the guidelines are available to assist the GP, when she/he needs it; 4) if guidelines are still not enough, GP can prescribe a specialist visit and it still would be recorded in HS database. Overall, this 'quality' channel seems of little relevance. In order to confirm these arguments, the first step of my analysis is to check on a formal statistical basis the assumptions related to monotonicity, exogeneity and relevance of the instrument. The first two null hypotheses can be jointly tested using a method proposed by Mourifié and Wan (2016) (see also Kitagawa, 2015; Laffers and Mellace, 2016): evaluated at equally spaced values of the treatment (a step of 0.2), they are not rejected at standard confidence level (results available on the website). As for the relevance of the instruments, the F-statistics from first stage are much higher than 20, so finite sample bias is not a concern Bound et al. (1995); Stock and Yogo (2002).

With a valid instrument, a meaningful comparison is possible only if $0 < Pr(D = 1|X = x) < 1$, i.e. for treated individuals there is a comparison group of untreated patients. I check the probability of compliance spanned by the instrument and covariates (Figure 2): in the tails (for very large or very small estimated probabilities), there are fewer observations and the common support may turn out to be restrictive. For this reason, the discussion of results will be focused on a smaller range (say between 0.1 and 0.9).

Having verified the two basic requirements above, it is interesting to see whether there is essential heterogeneity (Heckman et al., 2006). I address the issue by exploiting two different methods. One

uses the fact that $MTE(U_D = p)$ is a derivative with respect to $p$ (Heckman and Vytlacil, 1999), therefore if IV is linear in the propensity score $P(Z)$, the empirical content of LATE can be extended to the whole population (Carneiro et al., 2003): the null hypothesis that powers of $P(Z)$ are equal to zero is strongly rejected (Table 2), and MTE is needed. A second method is proposed by Kowalski (2016) and is based on the representation

$$y = \lambda_D D + \lambda_Z Z + \lambda_{DZ} DZ$$

where the joint null hypothesis $\lambda_D = \lambda_{DZ} = 0$ tests whether the treatment effect is globally externally valid and equal to zero. Coherent with the previous result, also this null is strongly rejected. These two tests suggest that 'a unique' IV does not jointly identify the LATE/ATE/ATT/ATU. For completeness, the Hausman (1978) test strongly rejects the null hypothesis of non systematic difference between IV and OLS: although in a heterogeneous treatment effect its empirical content is less clear than under homogeneity, I interpret this results as pointing towards the need for methods robust to endogeneity *and* heterogeneous effect.

This preliminary analysis provides large evidence in favour of a MTE approach, using as instrument the compliance by doctors (which is interpreted as a communication skill). The results delivered by this approach will be valid for a very large population (as identified by the common support). By discussing the results, it will be clear that the information provided by this approach greatly enrich what we know about the effects of pills on cholesterol.

## Results

There are three (families of) estimators for MTE: fully parametric (usually, but not necessarily, based on normality assumptions as in Bjorklund and Moffitt, 1987), semi-parametric (Heckman and Vytlacil, 2007b; Carneiro and Lee, 2009; Carneiro et al., 2017), and fully non-parametric (Heckman and Vytlacil, 2005). The empirical analysis is based on a semi-parametric method. Writing the observed outcome as $E[y|X,p] = E[\alpha + U_0 + D\,X\,\beta + \underbrace{D\,E[U_1 - U_0|p]}_{K(p)}]$, one can use the method of Local Instrumental Variable (Heckman and Vytlacil, 2000) and estimate $MTE = \frac{\partial E[y|P(Z)=p]}{\partial p} =$

$X\beta + K'(p)$. I approximate $K'(p)$ based on a polynomial of degree 3. Below I check the sensitivity of the results of this choice.

The model specification is delivered directly from the economic model of Section 2. The set of covariates includes clinical and socio-demographic characteristics. As for the former: on the clinical side, an indicator of general health condition (Charlson et al., 1987), dummies for occurrences of hospitalization (with a specific split for cardio-vascular diseases), diabetes, transitory ischemic attack (TIA), percutaneous transluminal coronary angioplasty (PTCA), heart attack, ictus, angina pectoris, dummies for consumption of specific statins (namely, simvastatin -possibly paired with other chemical compositions-, lovastatin, pravastatin, fluvastatin, atorvastatin, rosuvastatin). As for socio-demographic characteristics: on demographic side, an indicator for gender and age; on the environmental side, dummies for years, quarters and geographical area.

The MTE remains relatively flat up to $U_D \approx 0.4$, and then increases monotonically over the treatment probability (Figure 3). Coherent with the model's prediction, 1) there exists heterogeneity in return to pills, and 2) those whose reduction of cholesterol is larger (i.e., have a more negative treatment effect) enter first (i.e., at low values of $U_D$). Hence drugs may benefit more some patients than others and even being not very effective for a (small) part of the population (Heckman et al., 1997).[6] The huge heterogeneity in the distribution of treatment effects can be better appreciated by looking at Table 3, where I reported information that allows a quantitative assessment of the essential heterogeneity in the treatment effect: ATE, ATT, ATU, Sorting, Sorting$^*$, Selection.[7] These quantities are calculated using the procedure introduced by Carneiro et al. (2017). The sorting gain (i.e., $ATT - ATE = -0.023$) is negative and significant, confirming that treated patients have a larger benefit than those randomly chosen. Coherently, $ATE - ATU = -0.024$, because those non treated enjoy a lower benefit than those taken at random. This 'full decomposition' of the sorting allows me to introduce yet another test of the (non-) essential heterogeneity, based on

[6] Recall that from $U_D = 0.9$ the common support becomes a critical assumption, therefore from that point onwards results are not considered.

[7] Sorting is the difference between the ATT and the ATE, i.e. between $E(Y_1 - Y_0|D = 1)$ and $E(Y1 - Y0)$, a measure of the 'surplus' of the treatment effect enjoyed by patients who sort into compliance with respect to the treatment effect enjoyed by a randomly drawn patient. Selection is the difference between the ATT and the OLS, i.e. between $E(Y_1 - Y_0|D = 1)$ and $E(Y_1|D = 1) - E(Y_0|D = 0)$: in this application, statistically significant negative difference suggests a comparative advantage from drug compliance (i.e., positive selection).

the null hypotheses that $(ATE = ATT)\&(ATE = ATU)$. With respect to those run in Section 5, this test requires each quantity to be known (which is clearly a disadvantage; although notice that $(ATE \neq ATT) \Rightarrow (ATE \neq ATU)$ and viceversa); however, it provides more information and allows one to say which component of the heterogeneity is more important than the others (if any): in this full sample, heterogeneities on the positive and negative sides are equally (in clinical terms) important. Finally, the selection effect (i.e., $OLS - ATT = 0.024$), as one would expect from the Roy (1951) model.

With respect to the ATE, the classical estimators provide either an estimate smaller (as in the case of the OLS, which is equal to -0.02942 ), or larger (as in the case of the IV, which is equal to -0.04194 ). These differences are not surprising: the case of the OLS may be seen as an omission of the variable indicating the seriousness of the disease, which is positively correlated to treatment and thus biases downward (in absolute value) the estimated effect. As for the IV, the result is empirically coherent with the fact that LATE is valid only for compliers and the MTE weights (see Table 1A in Heckman and Vytlacil, 2005) are maximized between $p = 0.4 - 0.5$, where the treatment effect is larger (in absolute value).

**What benefits if policies could make more patients compliant? For whom?**

Given the relevance of expenditures for Cardiovascular diseases, the policy maker may wish to increase the compliance of certain individuals and improve the general health conditions. Different policies can be implemented. Therefore the policy maker wishes to know: 1) which one maximizes the return in terms of lower cholesterol; 2) which one targets best the desired population.

In this paper I propose three policies. One affects the baseline of the individual compliance probability (e.g., through advertisements directed to patients), one affects the communication of the doctors (e.g., through courses that improve this skill), and one affects the overall probability maybe (e.g., through a mix of interventions). In all these cases, the amount of the shock is 10%. Therefore, from a strictly economic point of view, it is worth spending an euro as long as the PRTE is -0.0330 (i.e., a difference in ATE as large as $-0.0030$ in absolute values) . In Table 4 I report the PRTE from the three policies.

Increasing the baseline of the probability of treatment (i.e., the intercept of $P(Z)$) gives the highest return within the policies considered (Table 4). This happens because the goal of the policy involves the patient: if the policy targets the patients the effect is direct; on the contrary, if the policy targets the doctor, the effect on the patient is mediated and the treatment effect lower.

In many circumstances this is all the information the policy maker needs. In other circumstances more is needed. Two different policies may switch into treatment two different individuals that, from the viewpoint of the policy maker, are completely different. A contribution of this paper to the literature employing MTE is a description of observable characteristics of individuals who are switched into treatment. Invoking the results in Vytlacil (2002), I heavily borrow from existing results (Angrist, 2004; Angrist and Pischke, 2008). Thus the technical contribution to address this issue exploits only well known results; nonetheless, the exercise conveys important information that so far has never been provided to the policy maker, whilst they may be of key relevance.

Similar to LATE, a limitation of this exercise is that I can describe the average characteristics of the switchers, but I am not able to flag who is *actually* switched from non-treatment to treatment. Vytlacil and Heckman (2001) show that the size of those treated thanks to the new policy is $\Delta P(x) = P[D_{p'} = 1|X = x] - P[D_p = 1|X = x]$. In order to describe their characteristics, I exploit that $E[X|\Delta P, u_d] = \frac{E[\Delta P|X,u_d]\,E[X|u_d]}{E[\Delta P|u_d]}$, $\forall u_D$ (Angrist, 2004). Although one may be interested in measuring the quantity at specific $u_D$, I summarize it over all $u_D$, i.e. $\int E[X|\Delta P, u_d]d\,u_d$.[8] In order to compare the characteristics of the shifted patients to those of the entire population, I also evaluate $E[X]$: if $X$ is diabetes then a ratio $\frac{E[X|\Delta P,u_d]}{E[x]} < 1$ implies that the share of those induced into treatment by the policy that suffers from diabete is lower than in the overall population (and viceversa if the same ratio is greater than 1).

The results from this exercise are reported in the lower panel of Table 4. Even though improving the skill communication of the doctor provides a lower return than increasing the baseline of the patient, such a policy would target an healthier population as the share of those switched into treatment is in general lower than 1 when existing health conditions are considered. As age and

[8] The procedure by Carneiro et al. (2017), changes the intercept/instrument/probability of the selection equation and averages MTE for members that are switched into treatment by the simulated policy. Therefore, once I have this 'new' sample of individuals, I simply evaluate the sample averages of their characteristics obtaining $E[X|\Delta P, u_d]$.

geographical area (not shown), the two policies are only marginally different. If the scope is to switch into compliance the middle-age to old patients, then a mix of policies may work better (column 'Overall' which also provides the lowest PRTE).

What do these results suggest? If the policy maker wishes to reduce the current cholesterol as much as possible, because the associated costs are high, the most rewarding policy is to increase the baseline compliance of the individuals. If the policy maker wishes to reduce the cholesterol of the (relatively) healthier patients in an attempt to reduce the associated (indirect) costs for the NHS, the appropriate policy is increasing the communication skills of the GP. If the policy maker wishes to reduce the perspective cholesterol as much as possible (European Commission, 2015), the most rewarding policy would be a mix of the policies that targets the younger patients. Without the investigation of the characteristics of switchers, I could not understand the source of the difference between the two policies and would have missed a lot of important information. In my opinion, this piece of information should always be provided.

# 6 Robustness

To conclude the empirical analysis, I run two robustness checks: first I test the sensitivity of results of Section 5 to more technical aspects; second, I investigate whether relevant differences are estimated by specific, observable, characteristics.

The analysis is run using a semiparametric approach with third degree polynomial. This choice is made after checking that the differences as going from semi- to non-parametric model is negligible (a conclusion reached also in Carneiro, 2003; Carneiro et al., 2003; this can be seen by Figure 4, where -if anything- results from non-parametric approach are even more striking than those discussed above, as the treatment effect is monotonically increasing over $p$), and both are remarkably

different from the parametric model.[9] [10] I interpret the difference between non-/semi- parametric approach and fully parametric approach as an informal, though strong, evidence that the normality assumption is violated in this application. **figura con logit/probit & normal,semi,non** A second potentially critical technical assumption is about the estimation of the selection mechanism: in this case, using either a logit, probit or linear probability model, the estimated MTE are very similar. Overall, the results presented in Section 5 are robust to alternative empirical choices.

A second more interesting robustness check is about differences across relevant subsamples. I split the sample by age cohorts and by geographical area. As for age cohorts, I split the sample in 3 classes: $[49, 58]$ (the 'young' patients),$[59, 73]$ (the 'middle age' patients),[74+] (the 'old' patients). Various other possibilities are legitimate, but this seems a good compromise between ages relevant for the disease under investigation and (sub)sample sizes. Overall, the conclusions for all the subsamples are in accordance to those from the whole sample. Interesting differences emerge when we compare their magnitude. Those who benefit most from being compliant are the middle-age patients, whose benefit are almost as large as twice that of the patients who benefit least (the old patients). The reason why ATE for the young patients is larger than that for the old patients can be understood by looking closer at other components of ATE: the ATT for the old patients is the largest among these sub-samples, whilst the ATU is positive [and non significant] for the former group. Therefore, the gain for old patients who are actually treated is the highest in the population. To reinforce this claim, notice that also the sorting effect -however considered- is the largest for this sub-sample. From a clinical viewpoint, these results reinforce the conclusions in Atella et al. (2017) that drugs are part of the story to lower cholesterol, while in fact a substitution between drugs and physical activity is desirable: it refines the conclusions, to say that it is more so at younger ages (where this substitution is more likely to occur); it also says that for old patients, who on average

[9] The fully parametric approach has the disadvantage of fully specifying the Data Generating Process (DGP), thus being more efficient than the other methods when the assumptions are indeed verified, but at the cost that even minimal departures from them leads to inconsistency (Arabmazar and Schmidt, 1981, 1982). The fully non parametric approach requires that the support of the distribution $U_D$ conditional on the design matrix to span the full unit interval (Carneiro et al., 2011), and is slower than the semi-parametric approach. The semi-parametric approach does not impose distributional assumptions (in contrast to the fully parametric approach) and has weaker requirement on $U_D$ with respect to the non-parametric approach. See Carneiro et al. (2003) for further technical details. A non trivial aspect is also the computational time, whose difference is about 1:2.

[10] As for inference, I only *estimate* the probability of treatment, but do not *observe* it: following the entire existing literature, the inference is based bootstrap approximation (Carneiro et al., 2017).

have an higher Charlson et al. (1987) index, drugs are not effective at all in some cases (that is, for the untreated), probably because of the side effects. **davvero????**From a statistical viewpoint, this exercise is useful to appreciate the distinctive features of the test of essential heterogeneity introduced in Section 5.

As for geographical area, I split the country in 4 areas: North-West (NW), North-East (NE), Center (C), and South (S). In all the areas, the qualitative conclusions that may be drawn from the analysis are identical to those discussed in Section 5. Even in this case, the differences in the magnitude of the effects are worth noticing. While the whole sample and North-West are very similar also with respect to the size of the effects, the other macro-regions are far apart: at one extreme there is Center, and on one opposite extreme North-East and South. In Center the ATE is remarkably smaller than in the overall sample, reflecting the lower than average initial level of cholesterol (for both treated and untreated individuals), which is coherent with the model outlined in Section 2. Similarly, for both NE and S the initial level of cholesterol is higher than the overall sample, overall for treated individuals; for untreated individuals, the initial level of cholesterol is slightly lower in NE than in NW or S and accordingly the ATU is remarkably low, driving the result.

# 7 Concluding remarks

TBA

This paper investigates the relation between drugs and cholesterol, by jointly taking into account the possible sorting of patients into treatment and the heterogeneity in the effect, that is suggested by a simple proposed theoretical model. Marginal Treatment Effect (MTE) is particularly convenient in this setup, because identifies the average reduction of cholesterol from adherence for patients just indifferent between treatment and non treatment at a specific level of unobservables. To the extent that there exists substantial heterogeneity (**?**), MTE will be different across the probability of treatment.

The results are coherent with the stylized model. Not only do drugs successfully decreases the

level of cholesterol, as estimated by all the estimators that I have employed, but also patients who benefit most from the treatment are more likely to adhere to prescribed regime. These results are used in a second step to study the effects of thoughts policies aiming at increasing patients' compliance (and thus lower cholesterol level) that affect either the ultimate target (patients) or the intermediate target (doctors). A contribution of the paper is the description of characteristics of patients switched into treatment. Whilst targeting the patients provide the largest reduction of cholesterol, the policies differ substantially with respect to the population affected, therefore one of the two may be preferred to the other, depending on which population the policy maker wishes to target.

## References


American Heart Association (2015). Heart disease and stroke statistics2015 update. Technical report.

Angrist, J. and J. Pischke (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Angrist, J. D. (2004). Treatment effect heterogeneity in theory and practice*. *The Economic Journal 114*(494), C52–C83.

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of causal effects using instrumental variables (Disc: p456-472). *Journal of the American Statistical Association 91*, 444–455.

Arabmazar, A. and P. Schmidt (1981). Further evidence on the robustness of the tobit estimator to heteroskedasticity. *Journal of Econometrics 17*(2), 253 – 258.

Arabmazar, A. and P. Schmidt (1982). An investigation of the robustness of the tobit estimator to non-normality. *Econometrica 50*(4), 1055–1063.

Atella, V., F. Belotti, and D. Depalo (2017, September). Drug therapy adherence and health outcomes in the presence of physician and patient unobserved heterogeneity. *Health Economics 26*, 106–126.

Baigent, C., A. Keech, P. Kearney, and L. Blackwell (2005). Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins. *Lancet 366*, 1267–1278.

Basu, A., J. J. Heckman, S. Navarro-Lozano, and S. Urzua (2007). Use of instrumental variables in the presence of heterogeneity and self-selection: an application to treatments of breast cancer patients. *Health Economics 16*(11), 1133–1157.

Bjorklund, A. and R. Moffitt (1987). The estimation of wage gains and welfare gains in self-selection models. *The Review of Economics and Statistics 69*(1), pp. 42–49.

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

Brinch, C. N., M. Mogstad, and M. Wiswall (2012, September). Beyond LATE with a discrete instrument. Heterogeneity in the quantity-quality interaction of children. Discussion Papers 703, Statistics Norway, Research Department.

Burroughs, V., R. Maxey, and R. Levy (2002, October). Racial and ethnic differences in response to medicines: towards individualized pharmaceutical treatment. *J Natl Med Assoc 94*(10), 1–26.

Card, D. (1999). The causal effect of education on earnings. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3 of *Handbook of Labor Economics*, Chapter 30, pp. 1801–1863. Elsevier.

Carneiro, P. (2003). Heterogeneity in the returns to schooling: Implications for policy evaluation. Thesis, The University of Chicago.

Carneiro, P., J. J. Heckman, and E. Vytlacil (2003). Understanding what instrumental variables estimate: Estimating marginal and average returns to education. *processed, University of Chicago, The American Bar Foundation and Stanford University, July 19*.

Carneiro, P., J. J. Heckman, and E. Vytlacil (2010, 01). Evaluating Marginal Policy Changes and the Average Effect of Treatment for Individuals at the Margin. *Econometrica 78*(1), 377–394.

Carneiro, P., J. J. Heckman, and E. J. Vytlacil (2011, October). Estimating marginal returns to education. *American Economic Review 101*(6), 2754–81.

Carneiro, P. and S. S. Lee (2009). Estimating distributions of potential outcomes using local instrumental variables with an application to changes in college enrollment and wage inequality. *Journal of Econometrics 149*(2), 191–208.

Carneiro, P., M. Lokshin, and N. Umapathi (2017). Average and marginal returns to upper secondary schooling in indonesia. *Journal of Applied Econometrics 32*(1), 16–36.

Charlson, M., P. Pompei, K. Ales, and C. R. MacKenzie (1987). A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chron Dis 40*(5), 373–383.

Cornelissen, T., C. Dustmann, A. Raute, and U. Schonberg (2016). From LATE to MTE: Alternative methods for the evaluation of policy interventions. *Labour Economics 41*, 47 – 60. SOLE/EALE conference issue 2015.

Cramer, J., A. Roy, and A. Burrell (2008). Medication compliance and persistence: Terminology and definitions. *Value in Health 11*.

Doyle, J. J. (2007, December). Child protection and child outcomes: Measuring the effects of foster care. *American Economic Review 97*(5), 1583–1610.

European Commission (2015). *The 2015 Ageing Report*. Belgium: European Commission.

Fabiani, L., M. Scatigna, K. Panopoulou, A. Sabatini, E. Sessa, F. Donato, M. Marchi, R. Nardi, C. Niccolai, F. Samani, and G. Ventriglia (2004). Health search: istituto di ricerca della societá italiana di medicina generale: la realizzazione di un database per la ricerca in medicina generale. *Epidemiol and Prev 28*, 156–162.

Fichera, E., J. Banks, L. Siciliani, and M. Sutton (2017, March). Does patient health behaviour respond to doctor's effort? Working paper.

French, E. and J. Song (2014, May). The effect of disability insurance receipt on labor supply. *American Economic Journal: Economic Policy 6*(2), 291–337.

Ginsburg, G., M. Donahue, and L. Newby (2005). Prospects for personalized cardiovascular medicine: the impact of genomics. *J Am Coll Cardiol 46*, 1615–1627.

Hausman, J. A. (1978, November). Specification Tests in Econometrics. *Econometrica 46*(6), 1251–71.

Heckman, J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica 46*(4), 931–59.

Heckman, J. (2010, June). Building bridges between structural and program evaluation approaches to evaluating policy. *Journal of economic literature 48*(2), 356–398.

Heckman, J. J. (1979, January). Sample Selection Bias as a Specification Error. *Econometrica, Econometric Society 47*(1), 153–61.

Heckman, J. J., J. Smith, and N. Clements (1997). Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *The Review of Economic Studies 64*(4), 487–535.

Heckman, J. J., S. Urzua, and E. Vytlacil (2006, August). Understanding Instrumental Variables in Models with Essential Heterogeneity. *The Review of Economics and Statistics 88*(3), 389–432.

Heckman, J. J. and E. Vytlacil (2005, 05). Structural Equations, Treatment Effects, and Econometric Policy Evaluation. *Econometrica 73*(3), 669–738.

Heckman, J. J. and E. J. Vytlacil (1999). Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences of the United States of America 96*(8), 4730–4734.

Heckman, J. J. and E. J. Vytlacil (2000). The relationship between treatment parameters within a latent variable framework. *Economics Letters 66*(1), 33 – 39.

Heckman, J. J. and E. J. Vytlacil (2001). Instrumental variables, selection models, and tight bounds on the average treatment effect. In *Econometric Evaluations of Active Labor Market Policies in Europe. Physica-Verlag.*

Heckman, J. J. and E. J. Vytlacil (2007a, January). Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6 of *Handbook of Econometrics*, Chapter 70. Elsevier.

Heckman, J. J. and E. J. Vytlacil (2007b, January). Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to

Evaluate Social Programs, and to Forecast their Effects in New. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6 of *Handbook of Econometrics*, Chapter 71. Elsevier.

Hughes, D., A. Bagust, A. Haycox, and T. Walley (2001). The impact of non-compliance on the cost-effectiveness of pharmaceuticals: a review of the literature. *Health Economics 10*, 601–615.

Imbens, G. W. and J. D. Angrist (1994, March). Identification and Estimation of Local Average Treatment Effects. *Econometrica 62*(2), 467–75.

Kitagawa, T. (2015). A test for instrument validity. *Econometrica 83*(5), 2043–2063.

Kling, J. R. (2006). Incarceration length, employment, and earnings. *The American Economic Review 96*(3), 863–876.

Kowalski, A. E. (2016, June). Doing more when you're running late: Applying marginal treatment effect methods to examine treatment effect heterogeneity in experiments. Working Paper 22363, National Bureau of Economic Research.

Laffers, L. and G. Mellace (2016). A note on testing instrument validity for the identification of late. *Empirical Economics*, 1–6.

Lamiraud, K. and P. Geoffard (2007, 11). Therapeutic non adherence: a rational behavior revealing patient preferences? *Health Economics 16*(11), 1185–1204.

Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association 90*(431), 1112–1121.

Maestas, N., K. J. Mullen, and A. Strand (2013, August). Does disability insurance receipt discourage work? using examiner assignment to estimate causal effects of ssdi receipt. *American Economic Review 103*(5), 1797–1829.

Mahlknecht, U. and S. Voelter-Mahlknecht (2005). Pharmacogenomics: questions and concerns. *Current Medical Research and Opinion 21*(7), 1041–1047.

Maron, D., S. Fazio, and M. Linton (2000). Current perspectives on statins. *Circulation 101*, 207–213.

Mourifié, I. and Y. Wan (2016). Testing local average treatment effect assumptions. *Review of Economics and Statistics na*, n/a–n/a.

New England Healthcare Institute (2009). Thinking outside the pillbox a system-wide approach to improving patient medication adherence for chronic disease.

Roter, D., J. A. Hall, and Y. Aoki (2002). Physician gender effects in medical communication: A meta-analytic review. *JAMA 288*(6), 756–764.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers 3*(2), 135–146.

Stewart, M. (1996). Effective physician-patient communication and health outcomes: a review. *Canadian Medical Association Journal 152*, 1423–1433.

Stock, J. and M. Yogo (2002). Testing for weak instruments in linear iv regression. In D. Andrews and J. Stock (Eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas J. Rothenberg*. Cambridge University Press.

Vytlacil, E. (2002). Independence, monotonicity, and latent index models: An equivalence result. *Econometrica 70*(1), pp. 331–341.

Vytlacil, E. and J. J. Heckman (2001, May). Policy-Relevant Treatment Effects. *American Economic Review 91*(2), 107–111.

Table 1: Descriptive statistics

| | Non compliant (D=0) | | | | | Compliant (D=1) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Stat. | Chol. | Female | Charlson | Age | Equipot. | Chol. | Female | Charlson | Age | Equipot. |
| Mean | 0.621 | 0.611 | 0.556 | 66.0 | 4.195 | 0.586 | 0.520 | 0.695 | 66.8 | 4.421 |
| SD | 0.078 | 0.488 | 0.913 | 9.0 | 0.659 | 0.081 | 0.500 | 0.968 | 8.6 | 0.630 |
| 10th quant. | 0.522 | 0.000 | 0.000 | 53.0 | 3.000 | 0.484 | 0.000 | 0.000 | 55.0 | 4.000 |
| Median | 0.627 | 1.000 | 0.000 | 67.0 | 4.000 | 0.591 | 1.000 | 0.000 | 68.0 | 4.000 |
| 90th quant. | 0.708 | 1.000 | 2.000 | 77.0 | 5.000 | 0.678 | 1.000 | 2.000 | 77.0 | 5.000 |

Table 2: Test of effect homogeneity

| | Ability channel | | | |
|---|---|---|---|---|
| | With Cov. | | No cov. | |
| D | 1966.856 | *** | 16618.391 | *** |
| Z | 51.654 | *** | 1698.668 | *** |
| D × Z | 124.491 | *** | 145.355 | *** |
| Joint | 1012.743 | *** | 8409.882 | *** |
| Hausman (1978) | 179.294 | *** | 337.147 | *** |
| Carneiro et al. (2003) | 119.563 | *** | 176.380 | *** |

Table 3: Treatment effects; effect & S.E.

| Sample | ATE | ATT | ATU | Sorting | Sorting* | Selection |
|---|---|---|---|---|---|---|
| | | | Full sample | | | |
| All | -0.02997 | -0.05310 | -0.00594 | -0.02313 | -0.02403 | 0.02367 |
| | 0.00222 | 0.00036 | 0.00493 | 0.00258 | 0.00272 | 0.00041 |

Sorting=ATT-ATE; Sorting*=ATE-ATU; Selection=OLS-ATT

Table 4: Policy relevant treatment effect

| Parameter | Doctor | Patient | Overall |
|---|---|---|---|
| PRTE | -0.03631 | -0.04593 | -0.03186 |
| SE | 0.00063 | 0.00124 | 0.00338 |

| | | Description of characteristics | | | | | |
|---|---|---|---|---|---|---|---|
| Chracteristic | Overall | Affected | Ratio | Affected | Ratio | Affected | Ratio |
| Init | 0.64174 | 0.64245 | 1.00111 | 0.64062 | 0.99826 | 0.63636 | 0.99161 |
| Init sq | 0.41903 | 0.41977 | 1.00177 | 0.41756 | 0.99648 | 0.41235 | 0.98405 |
| Equip | 4.30883 | 4.28934 | 0.99548 | 4.33150 | 1.00526 | 4.41614 | 1.02491 |
| Equip sq | 18.99306 | 18.79108 | 0.98937 | 19.14827 | 1.00817 | 19.89580 | 1.04753 |
| Charlson 2 | 0.26160 | 0.25171 | 0.96219 | 0.26637 | 1.01825 | 0.29398 | 1.12377 |
| Charlson 3 | 0.09076 | 0.08816 | 0.97134 | 0.09159 | 1.00910 | 0.09708 | 1.06958 |
| Charlson 4 | 0.05335 | 0.05068 | 0.94991 | 0.05373 | 1.00713 | 0.05922 | 1.10999 |
| Ric cardio | 0.00677 | 0.00622 | 0.91820 | 0.00666 | 0.98286 | 0.00791 | 1.16812 |
| Ric | 0.03308 | 0.03282 | 0.99202 | 0.03320 | 1.00341 | 0.03516 | 1.06273 |
| Diab | 0.09794 | 0.09149 | 0.93413 | 0.09993 | 1.02034 | 0.11342 | 1.15812 |
| Hyper | 0.22891 | 0.22232 | 0.97119 | 0.23151 | 1.01135 | 0.24508 | 1.07064 |
| TIA | 0.01664 | 0.01545 | 0.92825 | 0.01681 | 1.01005 | 0.01967 | 1.18179 |
| PTCA | 0.00838 | 0.00517 | 0.61668 | 0.00657 | 0.78437 | 0.01167 | 1.39355 |
| IMA | 0.02115 | 0.01416 | 0.66939 | 0.01785 | 0.84395 | 0.03014 | 1.42505 |
| Ictus | 0.00499 | 0.00468 | 0.93842 | 0.00499 | 1.00057 | 0.00623 | 1.24881 |
| Angina | 0.01972 | 0.01695 | 0.85951 | 0.01914 | 0.97072 | 0.02492 | 1.26367 |
| Simvastatin | 0.31244 | 0.31227 | 0.99945 | 0.32410 | 1.03732 | 0.34270 | 1.09687 |
| Lovastatin | 0.01832 | 0.01283 | 0.70071 | 0.01012 | 0.55233 | 0.00702 | 0.38340 |
| Pravastatin | 0.11217 | 0.11185 | 0.99713 | 0.10709 | 0.95472 | 0.10003 | 0.89179 |
| Fluvastatin | 0.08398 | 0.08450 | 1.00617 | 0.08835 | 1.05195 | 0.08758 | 1.04282 |
| Atorvastatin | 0.31679 | 0.31440 | 0.99246 | 0.31590 | 0.99721 | 0.32705 | 1.03239 |
| Rosuvastatin | 0.15930 | 0.16676 | 1.04681 | 0.15780 | 0.99059 | 0.13871 | 0.87071 |
| Simva eze | 0.01352 | 0.01258 | 0.93067 | 0.01302 | 0.96290 | 0.01542 | 1.14060 |
| Simva acid acetyl | 0.00011 | 0.00010 | 0.87021 | 0.00011 | 0.94142 | 0.00014 | 1.19236 |
| Female | 0.56751 | 0.57693 | 1.01661 | 0.56033 | 0.98735 | 0.52388 | 0.92311 |
| Eta 2 | 0.05325 | 0.05301 | 0.99565 | 0.05117 | 0.96107 | 0.04751 | 0.89221 |
| Eta 3 | 0.10557 | 0.10583 | 1.00249 | 0.10363 | 0.98160 | 0.09932 | 0.94079 |
| Eta 4 | 0.15475 | 0.15477 | 1.00014 | 0.15463 | 0.99924 | 0.15382 | 0.99398 |
| Eta 5 | 0.20283 | 0.20402 | 1.00586 | 0.20418 | 1.00664 | 0.20543 | 1.01282 |
| Eta 6 | 0.20717 | 0.20593 | 0.99402 | 0.21007 | 1.01400 | 0.21702 | 1.04755 |
| Eta 7 | 0.17268 | 0.17241 | 0.99845 | 0.17459 | 1.01103 | 0.17801 | 1.03088 |
| Eta 8 | 0.06959 | 0.06991 | 1.00452 | 0.06970 | 1.00147 | 0.07017 | 1.00836 |
| Q 1 | 0.24818 | 0.24938 | 1.00481 | 0.24898 | 1.00321 | 0.24701 | 0.99528 |
| Q 2 | 0.25056 | 0.25174 | 1.00471 | 0.25057 | 1.00001 | 0.25015 | 0.99836 |
| Q 3 | 0.25123 | 0.25185 | 1.00246 | 0.25117 | 0.99977 | 0.25389 | 1.01060 |
| Y 2008 | 0.19219 | 0.19210 | 0.99954 | 0.19034 | 0.99038 | 0.19883 | 1.03455 |
| Y 2007 | 0.21753 | 0.21558 | 0.99102 | 0.21664 | 0.99592 | 0.22025 | 1.01251 |
| Y 2006 | 0.21661 | 0.21618 | 0.99798 | 0.21839 | 1.00822 | 0.22004 | 1.01583 |
| Y 2005 | 0.19979 | 0.19830 | 0.99254 | 0.19986 | 1.00035 | 0.19322 | 0.96714 |
| NW | 0.24007 | 0.24328 | 1.01339 | 0.24005 | 0.99994 | 0.24270 | 1.01097 |
| NE | 0.25733 | 0.27306 | 1.06112 | 0.26460 | 1.02822 | 0.27720 | 1.07719 |
| C | 0.14794 | 0.14706 | 0.99405 | 0.14648 | 0.99010 | 0.14632 | 0.98905 |

Table 5: Treatment effects by relevant samples; effect & S.E.

| Sample | ATE | ATT | ATU | Sorting | Sorting* | Selection |
|---|---|---|---|---|---|---|
| Split by age | | | | | | |
| [40-58] | -0.02353 | -0.04623 | -0.00395 | -0.02270 | -0.01958 | 0.01883 |
| | 0.00716 | 0.00580 | 0.00831 | 0.00136 | 0.00115 | 0.00600 |
| [59-73] | -0.03696 | -0.05497 | -0.01745 | -0.01801 | -0.01951 | 0.02522 |
| | 0.00112 | 0.00062 | 0.00159 | 0.00051 | 0.00047 | 0.00001 |
| [74+] | -0.02083 | -0.05592 | 0.01751 | -0.03509 | -0.03833 | 0.02509 |
| | 0.00326 | 0.00192 | 0.00463 | 0.00134 | 0.00137 | 0.00156 |
| Split by geaographical area | | | | | | |
| NW | -0.03090 | -0.05897 | -0.00106 | -0.02808 | -0.02984 | 0.03238 |
| | 0.00130 | 0.00238 | 0.00530 | 0.00367 | 0.00401 | 0.00260 |
| NE | -0.04407 | -0.07845 | -0.00216 | -0.03438 | -0.04190 | 0.04928 |
| | 0.00053 | 0.00026 | 0.00119 | 0.00079 | 0.00065 | 0.00050 |
| C | -0.01278 | -0.02300 | -0.00241 | -0.01022 | -0.01037 | -0.00592 |
| | 0.00092 | 0.00005 | 0.00172 | 0.00087 | 0.00080 | 0.00042 |
| S | -0.04183 | -0.06891 | -0.01867 | -0.02708 | -0.02316 | 0.03625 |
| | 0.00008 | 0.00918 | 0.00801 | 0.00927 | 0.00792 | 0.00871 |

Sorting=ATT-ATE; Sorting*=ATE-ATU; Selection=OLS-ATT

Figure 1: A graphical representation of the economic model

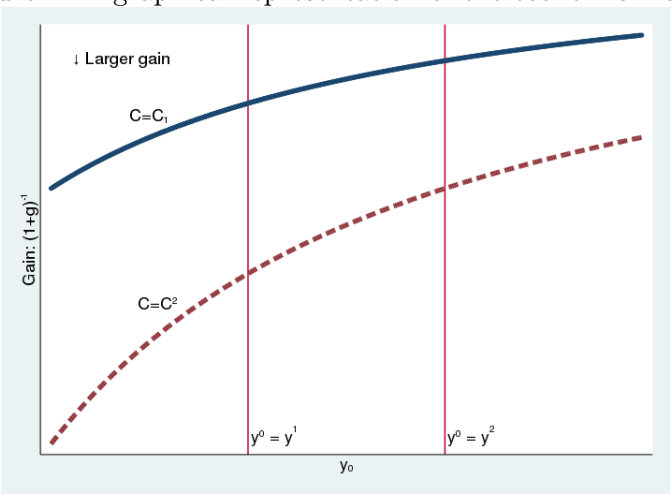



Figure 2: Histogram of the predicted probability of compliance

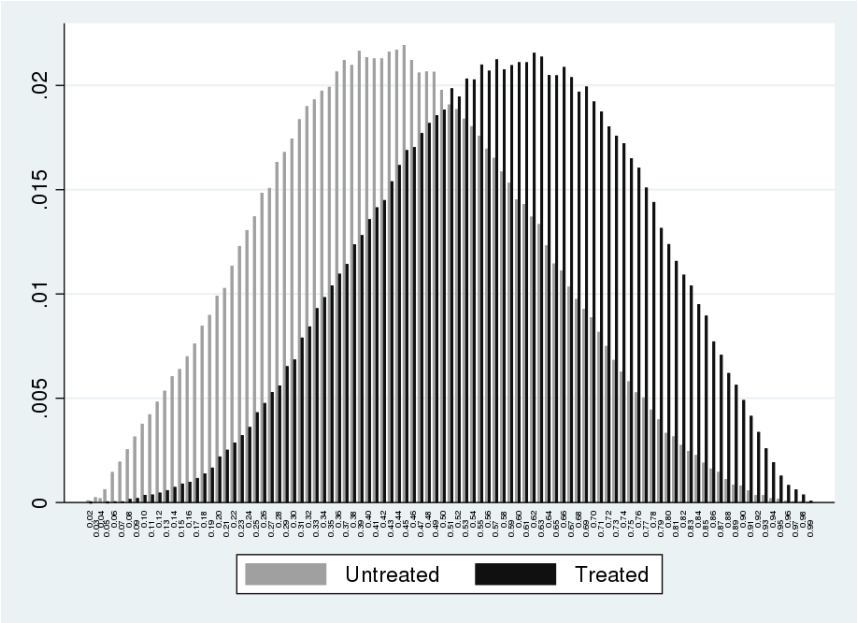


Figure 3: MTE of pills to cholesterol

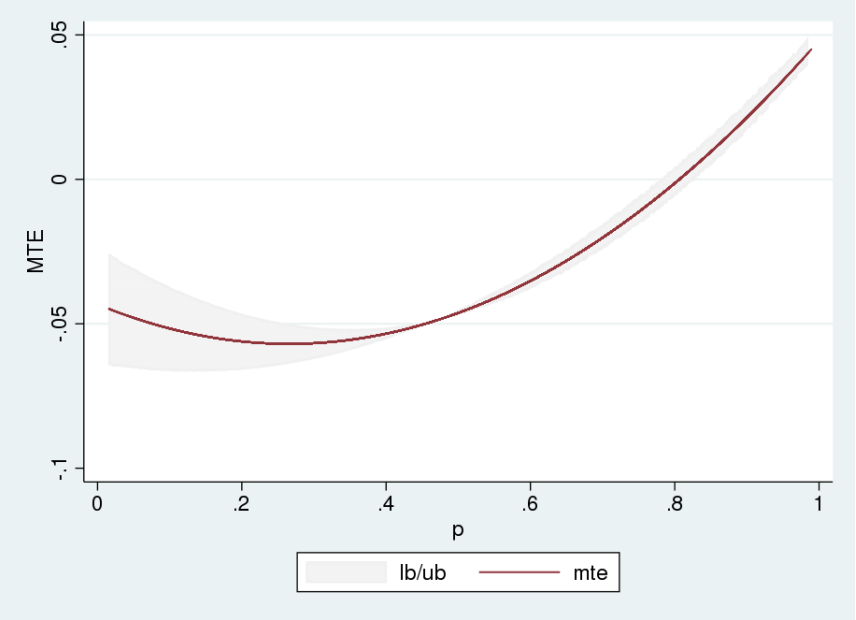
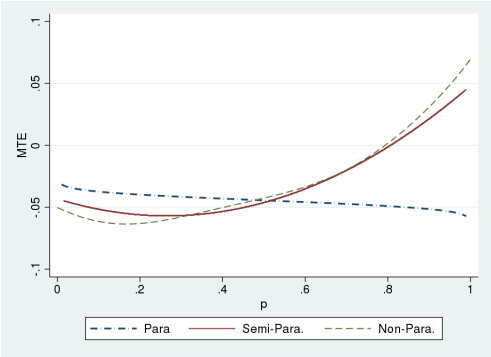
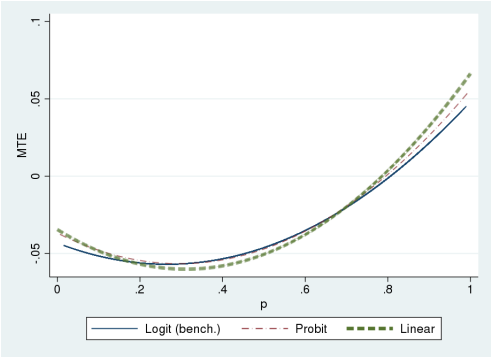



Figure 4: Sensitivity of MTE to technical aspects




(a) MTE approaches




(b) Sorting estimates