Co-payment exemption and healthcare consumption: quasi-experimental evidence from Italy

Vanessa Cirulli[1] · **Giuliano Resce[2]** · **Marco Ventura[1]**

Abstract

This paper investigates the causal effect of co-payment exemption on the number of specialist visits in the Italian National Health System. Exploiting a discontinuity in the multiple eligibility criteria, we apply multiple regression discontinuity in a quasi-experimental setting, considering both age and income requirements. Differently from the standard regression discontinuity, this twofold discontinuity allows to identify the effect of co-payment on a particularly needy sub-population of less wealthy people and how it changes according to the eligibility criteria. We find positive effects of co-payment exemption and the effects are stronger for less wealthy and older individuals. The result may be useful to the policy maker to tailor ad-hoc policies aimed at disadvantaged sub-populations.

Keywords Multiple regression discontinuity (MRD) · Healthcare · Co-payment · Demand effects · National Health System

JEL Classification C31 · I11 · I14 · I18

1 Introduction

Given the constraints of limited resources and increasing demand for healthcare, particularly due to higher life expectancy and an aging population, governments are compelled to balance effectiveness, equity, and financial stability in their healthcare strategies (Grassetti and Rizzi 2019). As the growth of public health spending has been widely documented (Baltagi et al. 2017; Vincenzino 1995; White 2007; WHO et al. 2018), the discussion on co-payment healthcare services is unsettled, demonstrated by

✉ Marco Ventura
  marco.ventura@uniroma1.it

1 Department of Legal and Economic Studies, Sapienza University of Rome, Via del Castro Laurenziano 9, 00161 Rome, Italy

2 Department of Economics, University of Molise, Campobasso, Italy

the existence of arguments both supporting and opposing its implementation or facilitation and has consistently remained on the agenda of governments in recent decades (Atella et al. 2006; Färdow et al. 2019). Co-payment, intended as a financial contribution for healthcare services and goods consumed, works as a lever to enhance consumers' responsibility in demanding only the optimal quantities of healthcare goods and services. Co-payments are often used to reduce moral hazard in healthcare. Indeed, moral hazard can arise when individuals have insurance coverage that shields them from the full cost of medical services, leading to over-utilization or unnecessary medical expenses.

However, the effectiveness of this approach is subject to debate, as co-payments can also act as a barrier to healthcare access, particularly for individuals with lower income or limited resources. Co-payment can impose a financial burden on low-income individuals, as it may represent a significant portion of their disposable income, making it difficult to afford essential healthcare. Higher co-payments may discourage individuals from seeking care for minor or preventive health issues, such as screenings, leading to delayed diagnosis and treatment. Discouraging individuals from accessing prevention services, co-payments may lead to missed opportunities for early detection and prevention of diseases (Rezayatmand et al. 2013). These are the reasons why many healthcare systems have implemented measures to reduce or eliminate co-payments for certain needy populations. This paper tests whether a co-payment exception for needy populations increases the use of healthcare services. We focus on specialist physician services, which play a crucial role in preventive healthcare, as they enable early detection, provide specialized knowledge and guidance, assess individual risks, can develop personalized treatment plans, and it has the potential of offering valuable health education (Jusot et al. 2012; Lueckmann et al. 2021). Increasing the use of these services for needy populations can significantly improve health outcomes and quality of life (Solanki and Schauffler 1999).

Estimating the effect of co-payment on healthcare utilization is empirically challenging because, to a certain extent, the amount of co-payment depends on individual's characteristics, e.g., income, that also determine the demand for health services. This paper addresses this issue by exploiting a multidimensional discontinuity in age and income, required in order to benefit from co-payment exemption in the Italian National Health System (NHS). Notably, people aged over 65 and with a household income less than 36,165.98 euro are eligible for co-payment exemption, defined as *co-payment exemption for income*. Both requirements are not exogenous with respect to health conditions and the ensuing healthcare consumption. It has been largely shown that age is one of the main drivers of healthcare utilization (Oliveira and de la Maisonneuve 2006), and it is well known that individuals' income is a fundamental factor of individual health (Subramanian and Kawachi 2004; O'Donnell et al. 2007). The (multidimensional) discontinuity in age and income provides us with a natural experiment that allows us to study the phenomenon in a quasi-experimental setting where treatment, i.e., co-payment exemption, is "as if" randomly assigned in proximity of the multidimensional discontinuity. Considering that the discontinuity is determined by two variables, we use a recent methodological development of Regression Discontinuity (RD), the Multiple Regression Discontinuity (MRD). Furthermore, this complex discontinuity offers another advantage in that it allows us to target the estimates on

the sub-population of less wealthy people, considered by the policy makers as one of the most needy and, for this reason, worth protecting and therefore studying. In this regard, our paper provides the policy makers and practitioners with useful knowledge not yet available, that may help them to tailor ad-hoc interventions. Our findings provide convincing evidence of a positive and significant effect of co-payment exemption on the utilization of specialist visits.

Empirical evidence for similar frameworks began with the 1974 RAND Health Insurance Experiment that provided the first opportunity to estimate the effects of health insurance and plan characteristics on medical spending (Aron-Dine et al. 2013). Between 1974 and 1981, the RAND experiment randomly assigned health insurance to a representative sample of families in the United States (Newhouse 1974). Results show that more generous insurance coverage leads to increases in healthcare utilization. This finding has been confirmed more recently by the Oregon Health Insurance Experiment (Finkelstein et al. 2012). A substantial number of works based on observational data followed the seminal works analyzing the consequences and effectiveness of co-payment scheme adoption in terms of demand for health services, health outcomes and distributional effects, according to the type of health service (see Kiil and Houlberg (2014), Lin et al. (2020) and the literature therein for a survey).

Even though healthcare co-payment may rationalize the demand for public health services and potentially enhance the efficiency of the public health sector, it can also create or increase inequality of access and utilization among users' (De Matteis et al. 2019). Thus, the introduction of co-payment may have a negative impact on equality in healthcare, i.e., it is a sort of regressive form of healthcare finance (Atella et al. 2006; Gemmill et al. 2008; Costa et al. 2012; Inoue and Kachi 2017). In these cases, a sort of protection for more needy patients can be adopted by introducing exemptions from healthcare co-payment. Such a policy can be seen as a Social Health Insurance (SHI) targeted to improve health and to provide more vulnerable and needy people with protection against the financial consequences of health shocks. The research question of how a SHI targeted to the most vulnerable and needy people may affect healthcare consumption is now increasingly explored in low and middle income countries (Bernal et al. 2017; Serna 2021), but evidence about the effect of such a policy in an advanced economy is still limited. As there is ample evidence of increasing worldwide within-countries inequality (Arestis et al. 2011; Liberati 2015; Goodhart 2017), it is becoming crucial to understand how healthcare consumption could react to such a policy in a relatively advanced economy. With this aim, we empirically test the relationship between co-payment exemption and the utilization of specialist visits (outpatient services) in Italy. In particular, we focus on the effect of *co-payment exemption for income* as one of the several forms of co-payment exemptions introduced by the Italian Legislator to guarantee healthcare services to needy and destitute people.[1]

Italy is characterized by a substantial presence of heterogeneity in income (Lagravinese 2015), multidimensional well-being (Greco et al. 2018) and health conditions (Lagravinese et al. 2019) and for these reasons the Government establishes and regulates co-payment exemption for people with specific features. Evidence about the effects of co-payment on healthcare services in the Italian NHS are provided by

---

[1] Art. 32 of the Italian Constitution.

Fiorio and Siciliani (2010), Atella and Kopinska (2014), and Ponzo and Scoppa (2021). Fiorio and Siciliani (2010) estimate the relationship between variation in co-payment and demand for pharmaceuticals. Atella and Kopinska (2014) analyze the causal effect of co-payment on compliance with statins in a sample of Italian hypercholesterolemic patients. By using a quantile regression model, they find that the introduction of co-payment reduces the compliance in regions with low healthcare organization and performance (compared to the national average). In particular, by interacting geographic area and town size, they show that the compliance increases with the municipality size of northwest and northeast, especially among low compliers; conversely, it diminishes with the city size in the center and in the south. In addition, they find that the exemptions from co-payments are associated with higher compliance, and this association is three to four times greater for poor compliers considering all exemption categories. Moreover, females are significantly less compliant than males, especially among poor compliers. Also the clinical history (case of acute events) are important determinants of compliance among poor compliers as well as the treatment length and type of statin therapy reduce the compliance mainly among poor compliers. Ponzo and Scoppa (2021) investigate the effect of co-payment exemption on the demand for different health services. In particular, the authors do not disentangle the effects of the exemption due to chronic diseases from the one granted for economic conditions. This limitation is essentially due to the lack of information about respondents' income in their database. Currently, the European healthcare systems, and particularly the Italian system, have been strongly put under pressure, due to the pandemic. By the same token, the US private system has raised considerable doubts.[2] In this scenario, disentangling the effect for less wealthy individuals is of great importance for a correct and informed health policy conduction. Our paper moves a step further, exactly in this direction. Indeed, our dataset is similar to that of Ponzo and Scoppa (2021) but we complement it with respondents' income as we are interested in the effect of *co-payment excemption for income*. Thus, we consider all people aged more 65[3] and we enrich previous analyses introducing also income requirement for co-payment exemption eligibility[4] This allows us to refine the analysis on co-payment exemption due to economic status, obtaining revealing results. Interestingly, we find that the effect of exemption on the number of specialist visits decreases as income increases, and it increases as individuals age. This piece of information would have not been attainable by following a standard RD. From a methodological point of view, this novelty comes at the cost of higher econometric complexity, which at the same time makes the problem more challenging and stimulating. The enhanced database allows us to consider eligibility for exemption according to the twofold requirement in terms of age and income. To tackle this empirical challenge we rely on one of

---

[2] As a support to this claim we simply recall the controversy raised by the high cost of a COVID-19 test during the first wave of the outbreak, in March 2020.

[3] The Italian Law 214/2011 established two requirements for full retirement benefits: minimum contribution level of at least 20 years and age. According to the age, for employees (private and public), for self-employed and female employees in the public sectors, the retirement age is fix at 66 years old in the 2012 and it increases of around 6 months every three years. Today, the retirement age is at 67 years old.)

[4] We exclude from our analysis other forms of co-payment exemption, such as co-payment exemption for rare diseases, for disability, for chronic diseases, for pregnant women.

the most recent developments of Regression Discontinuity (RD), namely the Multiple Regression Discontinuity (MRD) and we consistently adapt some tests developed for the single score case to this more complex setup. Our results show positive effects of co-payment exemption on the number of specialist visits.

The remainder of the paper is organized as follows: Sects. 2 and 3 introduce the institutional setting and the dataset, respectively. Section 4 explains the identification strategy which is tested in Sect. 5. The results are presented in Sect. 6, while Sect. 7 includes the robustness checks and Sect. 8 assesses to what extent the results hold away from the frontier. Finally, Sect. 9 concludes the study.

2 Institutional setting

The Italian National Health System (NHS), introduced in 1978, is a universal healthcare system providing comprehensive health insurance coverage and uniform health benefits to the whole population.

During the nineties, the Constitutional reform has introduced a decentralization process, called *devolution process* to contain the public expenditure, transferring greater formal responsibility and power to make decisions from Central Government to Regions in public finance issues and in the healthcare sector. Thus, each region introduced specific taxes (especially on activities and on personal income) and healthcare sector was rethought at regional level. Different Regional Health Systems (RHS) were established acquiring more autonomy in programming, funding, organization and delivery of healthcare.

In this context, regions adopted specific healthcare co-payment schemes in terms of application and structures to contain their public healthcare expenditure.[5] A co-payment was introduced for specialistic visits, for diagnostic care utilization and on the pharmautical expenditure,[6] where regions can independently apply these forms of healthcare cost-sharing to all these services or only for someones, varying costs in a range established by the Central Governement.[7] These regional heterogneity increase inequality in healthcare access and utilizations over time. However, in order to guarantee equal access and utilization of healthcare services and to safeguard more fragile and vulnerable patients, and also, for needy and destitute people, the Government establishes and regulates the co-payment exemption for people with specific features. Thus, the Italian NHS establishes different types of co-payment exemptions based on age and income level, health status (chronic diseases and rare diseases) and disability.

To our purpose, we focus on a case of total co-payment exemption strictly pertaining economic issues: the one granted according to income status. This type of exemption was introduced in 1994[8] and still applies substantially unaltered. We analyze cross-sectional data over the period 2011–2012 with exempted individuals according to the

---

[5] Decree Law No. 382/1989, Law no. 537/1993, Legislative Decree No. 124/1998.

[6] Hopital co-payment is required when the illness/sickness level cannot be classified as emergency.

[7] All regional cost-sharing information are provided by the Italian National Agency for Regional Healthcare Services (AGENAS): https://www.quotidianosanita.it/allegati/allegato573939.pdf.

[8] Law 724/1994.

aforementioned Law. In spite of the regional and local autonomy, the Italian NHS guarantees free healthcare services to patients having household income less than 36,151.98 euro and who are older than 65 years. Despite the double requirement this type of exemption is normally referred to as "exemption for income" and we will maintain this wording throughout. To the best of our knowledge there are no waivers at regional/local level for this benefit and at these thresholds.[9] Other exemptions of partial or total exemptions can be granted according to disabilities, chronic diseases, such as cancer, war-related experience, unemployment, social pension and other issues. We will confine our attention to the "exemption for income", being the most related to economic issues.

3 Data and summary statistics

The estimation sample we use is obtained by merging two different sources. The first one, the so called multipurpose survey on "Health conditions and use of health services", conducted by the Italian National Institute of Statistics in 2011–2012 (ISTAT 2013), is a cross section containing around 119,000 interviews. This survey is conducted every five years, to monitor the use of healthcare services in the Italian population and the prevalence of chronic health conditions. Each individual is randomly selected from the municipal registry lists, creating a statistically representative sample of resident population[10] Unfortunately, this data source lacks information about income. In line with recent contributions in machine learning (Feigenbaum 2016; Athey and Wager 2021) and the widespread practice in epidemiology (Murray 2018; Schenker and Welsh 1988) we have implemented a statistical matching to take the household income data for the same year (2012) from the Italian version of the EU's Statistics on Income and Living Conditions—EU-SILC (ISTAT 2021).[11] Focusing on income exemption we have dropped from the sample all the units benefiting from other types of exemptions, both partial and total, in general exempted individuals aged less than 65.[12]

---

[9] See https://www.salute.gov.it/portale/esenzioni/dettaglioFaqEsenzioni.jsp?lingua=italiano&id=206.

[10] It is a two-stage sample with stratification of first-stage units (municipalities). Information was collected by direct interview for a portion of the questions. In cases where the individual was not available for interview for particular reasons, information was provided by another household member. Self-completion was provided for another part of the questions. In data processing, probabilistic imputation techniques and donor imputation were used for missing partial responses (More details about Survey methodology at the Istat webpage (only in Italian) https://www.istat.it/it/archivio/7740.

[11] For analogous implementations see Resce et al. (2019); Carnazza et al. (2021) who match individuals on the basis of all observable common characteristics, i.e., education, job condition, location (NUTS2), gender, and age. In line with these studies, we have implemented the widely used procedure of statistical matching (D'Orazio 2017). Our function searches in EU-SILC the nearest neighbor of each individual in the survey "Health conditions and use of health services", according to a distance computed on five key variables: sex, age, region, education, and working condition. The purpose of this search is to find individuals within the EU-SILC dataset who closely resemble each participant in terms of the specified health conditions and health service utilization. By considering factors such as sex, age, region, education, and working condition, we aim to identify individuals who are similar in these aspects.

[12] Among the dropped units there are: pregnant women, disabled individuals, individuals with chronic diseases, unemployed, individuals earning social pension, etc. We use around the 32% of the full sample.

Information on the outcome variable of our interest, i.e., the total number of specialist visits, are provided through a detailed questionnaire in which respondents are asked about their utilization of one of these healthcare services in the last four weeks before the interview. In particular, the total number of specialist visits was built by summing the number of utilization of a set of specialist visits, such as geriatric, cardiology, gynecology, oculist, orthopedic, neurological, psychiatric and so on.

For our purpose, treatment is captured by a dummy variable, *exemption*, that is equal to 1 when the patient benefits from (total) co-payment exemption for income reasons. Table 1 presents descriptive statistics on the resulting merged sample. The sample span is 37,586 individuals and almost 4% benefit from total co-payment exemption. In the four weeks preceding the interview, on average, less than 1 respondent had a specialist visit. The average age is 41 years old, while the average income is about 24,000 euro. Furthermore, the *generic visits* are the number of visits made by the general practitioner in the four week preceding the interview, which are completely free in Italy for everybody.

Table 2 reports summary statistics by co-payment exemption. The average differences between the two sub-samples are significant for most of the variables. In general, treated individuals experience a greater number of specialist visits, whereas it is not possible to find a statistically significant difference when generic visits are concerned. This descriptive evidence raises the fundamental question we want to answer in this paper. Do exempted individuals consume more healthcare services because of the exemption? Descriptive evidence seems to suggest "YES", but the statistical evidence in Table 2 does not provide any clue in terms of causality. To unveil the possible causal relationship behind the difference in mean we resort to a more sophisticated tool described in the following sections. Finally, and consistent with the eligibility criteria, the table reports that on average treated individuals are older and less wealthy than non-treated individuals.

4 Identification strategy

Some features of RD Design, such as its transparency and conceptual simplicity, have made it one of the most reliable techniques to identify the effects of policy interventions. Over recent years, we are witnessing advances in many directions, such as optimal bandwidth and robust inference (Calonico et al. 2014, 2017, 2020), manipulation tests (McCrary 2008; Cattaneo et al. 2018), local randomization (Cattaneo et al. 2015, 2016), power tests (Cattaneo et al. 2019), multiple cutoffs (Cattaneo et al. 2016, 2020), inference away from the cutoff (Angrist and Rokkanen 2015; Dong and Lewbel 2015; Battistin and Rettore 2008) just to cite a few. For our purpose, a relevant extension pertains to cases in which eligibility depends upon more than one running variable, or simply scores, the so called Multiple Regression Discontinuity (MRD), in which typically, but not necessarily, the scores are two. Examples are given by borders identified by latitude and longitude (Keele and Titiunik 2015; Dell 2010; Gerber et al. 2011) or test scores at school (Jacob and Lefgren 2004; Papay et al. 2011; Zajonc 2012; Clark and Martorell 2014) and multiple requirements for eligibility in an income support program (Frey 2019). A common practice to deal with such

Table 1 Summary statistics—full sample

Variable	N	Mean	SD	Min.	Max.
Specialist visits	37, 589	0.132	0.512	0	15
Generic visits	3731	1.293	0.748	1	20
Exemption	37, 589	0.039	0.194	0	1
Age	37, 589	41.04	16.467	15	90
Income	33, 845	24218.28	12816.22	2000	58, 930
Male	37, 589	0.578	0.494	0	1
Married	37, 589	0.448	0.497	0	1
Italian	37, 589	0.92	0.271	0	1
Smoker	37, 589	0.394	0.489	0	1
College	37, 589	0.133	0.34	0	1
High school	37, 589	0.423	0.494	0	1
Worker	37, 589	0.538	0.499	0	1
Employee	37, 589	0.404	0.491	0	1
Self-employed	37, 589	0.133	0.339	0	1
House	37, 589	0.744	0.436	0	1
North-West	37, 589	0.227	0.419	0	1
North-East	37, 589	0.193	0.394	0	1
Center	37, 589	0.176	0.38	0	1
South	37, 589	0.405	0.491	0	1

Specialist visits = no. of specialist visits; generic visits = no. of generic visits; exemption = 1 if patients have co-payment exemption; age = age; income = household's income; male = 1 if male, married = 1 if married, Italian = 1 if Italian, smoker = 1 if smoker, college = 1 if max educational achievement is college degree, high school = 1 if max educational achievement is high school, worker = 1 if individual works, employee = 1 if employee, self-employed = 1 if self-employed, house = 1 if house owner, North-West = 1 if located in the North-West, North-East = 1 if located in the North-Est, Center = 1 if located in the Center, South = 1 if located in the South. The number of visits refers to four weeks before the interview. N = number of observations

occurrences consists in reducing the MRD problem to a single score problem. This can be done in two ways. The simplest consists in using a single score on the units with the other score already crossing the cutoff. It is worth noting that in this case the estimated effect is not fully informative of the policy effect, neither at the vertex, i.e., the joint cutoffs defining eligibility, so that the local information content RD provides is made even narrower by this technique. Another slightly more complex practice consists of collapsing the two scores into one by computing a distance measure, e.g., Euclidean norm, Mahalanobis distance, and applying the standard single score RD on the new running variable. While this strategy seems perfectly sensible for geographic boundaries and when the scores are of the same scale, such as test scores, it casts doubts when the scores pertain to very different phenomena. Examples can be given by population and Human Development Index (Frey 2019) or age and income, as in our analysis. To be more clear, think of a situation in which the resulting collapsed score returns a certain value, let us say h, meaning that individual i is distant h to the treatment frontier. The same value h can apply to both "too young" individuals

**Table 2** Summary statistics by co-payment exemption

Variable	Exemption = 1		Exemption = 0		Difference
	N	Mean	N	Mean	
Specialist visits	1478	0.155	36, 108	0.132	0.023*
Generic visits	314	1.347	3417	1.288	0.059
Age	1478	74.118	36, 108	39.689	34.429***
Income	1421	20118.662	32, 424	24397.951	− 4279.288***
Male	1478	0.52	36, 108	0.580	− 0.060***
Married	1478	0.622	36, 108	0.441	0.182***
Italian	1478	0.989	36, 108	0.918	0.072***
Smoker	1478	0.370	36, 108	0.395	− 0.025*
College	1478	0.024	36, 108	0.138	− 0.113***
High school	1478	0.113	36, 108	0.435	− 0.322***
Worker	1478	0.023	36, 108	0.559	− 0.536***
Employee	1478	0.004	36, 108	0.42	− 0.416***
Self-employed	1478	0.019	36, 108	0.137	− 0.118***
House	1478	0.834	36, 108	0.740	0.093***
North-west	1478	0.286	36, 108	0.224	0.062***
North-East	1478	0.236	36, 108	0.191	0.045***
Center	1478	0.179	36, 108	0.175	0.003
South	1478	0.299	36, 108	0.409	− 0.110***

Specialist visits = no. of specialist visits; exemption = 1 if patients have co-payment exemption; age = age; income = household's income; male = 1 if male, married = 1 if married, Italian = 1 if Italian, smoker = 1 if smoker, college = 1 if max educational achievement is college degree, high school = 1 if max educational achievement is high school, worker = 1 if individual works, employee = 1 if employee, self-employed = 1 if self-employed, house = 1 if house owner, North-West = 1 if located in the North-West, North-East = 1 if located in the North-Est, Center = 1 if located in the Center, South = 1 if located in the South. The number of visits refers to four weeks before the interview. N = number of observations
***$p < 0.01$; **$p < 0.05$, *$p < 0.1$

or "too wealthy" individuals, bearing very different implications. For cases like this Frey (2019) suggests leaving the scores separate and applying a RD experiment in a multidimensional setting. Indeed, the distinct nature of the two scores make likely that the sub-populations being compared along the frontier differ considerably. Under the one-score approach this heterogeneity might defeat the spirit of the RD Design. Differently, keeping the two scores separate one can exploit heterogeneity along the frontier, i.e., along the combinations of the scores allowing for eligibility. In other words, when the scores have a distinct nature the sub-populations along the frontier may be heterogeneous and this heterogeneity can help unveiling the heterogeneous effects of the policy.

In our specific case, individual i aged x_{1i}, with family income x_{2i} is eligible for treatment, i.e., co-payment exemption, if $x_{1i} \geq 65$ AND $x_{2i} < 36,165.98$ euro. The combination of the two scores defines two regions in a plane, one containing eligible
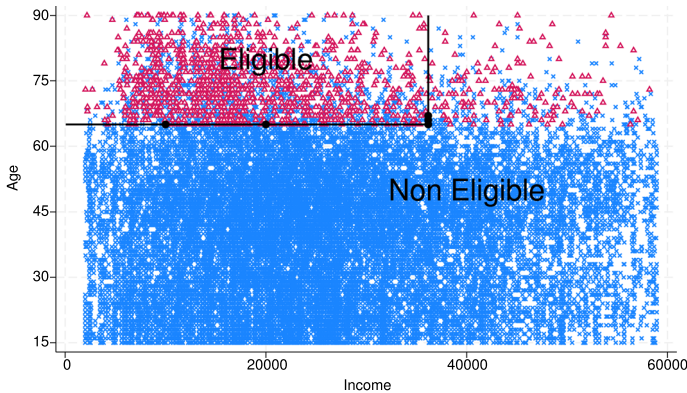
Fig. 1 Eligibility boundary. Note: the black solid line represents the eligibility boundary. Individuals located in the upper left quadrant are eligible for treatment. Red triangles represent actual treated and blue X's non-treated individuals. (Color figure online)

individuals and one non-eligible, the boundary between the two can be represented by $\mathbb{B} = (x_{1i} \geq 65 \cap x_{2i} < 36,165.98)$. The situation is depicted in Fig. 1.

The black line represents the boundary delimiting the eligibility region which is contained in the upper left, UL, quadrant. By far, this region is populated by treated individuals (red triangles), while non-treated are located mostly in the remaining region. As it is possible to see from the picture, we are in a typical situation of non-perfect compliance as there are (few) non-treated eligible individuals, i.e., blue X's, in the UL quadrant, and treated individuals in the non-eligibility region. The non-compliance problem arises from the fact that being exempted is not automatic, but rather potential eligible individuals must apply following a complex procedure that, to a certain extent, may vary according to the regions.[13] Briefly, potential eligibles every year must self-declare their household's income uploading it in the Electronic Health Folder, EHF, (*fascicolo sanitario elettronico*) a special section of the Ministry of Health website. In turn, most of the times, household's income must be previously computed by specialists operating in this sector, such as business consultants or tax assistance centers. Yet, access to EHF is possible only through by advanced electronic means of personal identification that can discourage elderly people. Once applied for the benefit, self-declarations are transferred to fiscal authorities who can indict applicants in case of false declarations. Even once successfully concluded this complex procedure, exemption is not automatic. Indeed, there can be cases in which valid applications do not appear in the system due to mere computer errors and delays. A further reason for non-perfect compliance is that exemption is granted on the basis of past income and this may generate an inconsistency between eligibility and current income. Another interesting feature arising from the picture is that along the horizontal side of the boundary there are many points, above and below the black line, while the sample is not as 'thick' along the vertical side, where the highest concentration

---

[13] For a general description of the steps to follow see https://www.salute.gov.it/portale/esenzioni/dettaglioFaqEsenzioni.jsp?lingua=italiano&id=206.

of points is next to the vertex.[14] The discontinuity created by the administrative rule, coupled with non-perfect compliance, give us the possibility to estimate a treatment effect along the frontier by using a fuzzy MRD. We report the econometric details of this estimator in the Online Appendix, while offering the main intuition in what follows. The main peculiarity of MRD with respect to RD is that while the latter allows to estimate only one treatment effect at the cutoff, the former, in principle, allows to estimate as many treatment effects as the points along the frontier, the so called location-specific treatment effects (Cattaneo et al. 2023). This makes the MRD particularly attractive, because the location-specific effects are informative of how the effects change according to the specific values of the scores, i.e., along the boundary. Empirically, the number of computable location-specific effects depends upon the variability of the scores in the data along the frontier. The higher the variability, the higher the number of computable location-specific effects and the more informative the estimates will be. From a practical point of view, this technique consists of estimating as many two-stage least squares, i.e., instrumental variable regressions, as the points along the frontier one wants to consider, where the instrument is eligibility. In the potential outcome framework terminology, each estimate coincides with a Local Average Treatment Effects (LATE). At each boundary point compliers are those units who change treatment status thanks to the instrument, i.e., the eligibility rule. In particular, the first stage is:

$$W_i = \alpha_0 + \alpha_1 Z_i + \alpha_2 \tilde{x}_{1i} + \alpha_3 \tilde{x}_{2i} + \alpha_4 Z_i \tilde{x}_{1i} + \alpha_5 Z_i \tilde{x}_{2i} + u_i \tag{1}$$

where W_i is the binary treatment indicator, $W = 1\{if\ treated\}$, Z is the eligibility dummy, $Z = 1\{x_{1i} \geq 65 \cap x_{2i} < 36,165.98\}$ and $\tilde{x}_{ji} = x_{ji} - \bar{x}_i$ for $j = 1, 2$ are the centred scores, where $\bar{x}_1 = 65$, $\bar{x}_2 = 36,165.98$ represent the cutoffs. In plain English, the treatment indicator is regressed on the binary eligibility indicator, on the centred scores, and on the interactions of eligibility and (centred) scores, plus an error term, u_i. The α_1 coefficient is equivalent to the proportion of compliers at the cutoff in a one-score RD setting, i.e., the jump in the probability of being treated. In this case, it represents the proportion of compliers at the vertex point of the boundary. The interaction terms are included to make the relationship between W and Z to depend also on the distance to the boundary. Of course, other controls can be easily added to (1). In the second step, the fitted value of W_i, i.e., $\hat{W}_i$, is plugged into the second stage regression:

$$Y_i = \beta_0 + \beta_1 \hat{W}_i + \beta_2 \tilde{x}_{1i} + \beta_3 \tilde{x}_{2i} + \beta_4 Z_i \tilde{x}_{1i} + \beta_5 Z_i \tilde{x}_{2i} + \eta_i \tag{2}$$

where Y is the outcome variable. β_1 is the *conditional* average treatment effect, CATE, of the policy at the vertex of the boundary, in other words it represents the jump in the outcome at the vertex point. To make valid inference the standard errors must be corrected to take into account that $\hat{W}_i$ is a generated regressor. At this point, two important issues must be discussed. First, the treatment effect holds only at the vertex,

---

[14] As a consequence of this uneven distribution, the estimates along the horizontal side have better power properties, as we will see in Sect. 6.

therefore an optimal two dimension bandwidth, i.e., an optimal sphere, must be derived. Second, the same procedure must be repeated at different points along the frontier and an average effect can be computed by averaging over these estimates.[15] As far as the first issue is concerned, Zajonc (2012) in its application argues that the optimal sphere can be proxied by the Scott's rule:

$$\hat{h}_j = \hat{\sigma}_j \cdot N^{-1/6} \ \text{ for } \ j = 1, 2 \tag{3}$$

where $\hat{h}_j$ and $\hat{\sigma}_j$ are the estimated bandwidth and standard deviation of score x_j, respectively, and N is the sample size. As for the second issue, actually it stems from the first. Since the estimate as obtained from Eqs. (1) and (2) are locally valid, i.e., in a narrow neighborhood of the vertex, it is possible to repeat the same procedure to other points along the frontier, such as to exploit heterogeneous effects that different combinations of the scores may give rise to. To this purpose, Eqs. (1) and (2) must be re-estimated by re-centering the scores, one at a time at different cutoffs. For instance, to estimate the conditional effect at a given point along the vertical side of the frontier, let us say 75 years of age, one must center x_1 at 75, $\tilde{x}_{1i}^1 = x_{i1} - 75$, keep x_2 centred at the vertex, and re-estimate (1) and (2). That is:

$$W_i = \alpha_0 + \alpha_1 Z_i + \alpha_2 \tilde{x}_{1i}^1 + \alpha_3 \tilde{x}_{2i} + \alpha_4 Z_i \tilde{x}_{1i}^1 + \alpha_5 Z_i \tilde{x}_{2i} + u_i \tag{4}$$

$$Y_i = \beta_0 + \beta_1 \hat{W}_i + \beta_2 \tilde{x}_{1i}^1 + \beta_3 \tilde{x}_{2i} + \beta_4 Z_i \tilde{x}_{1i}^1 + \beta_5 Z_i \tilde{x}_{2i} + \eta_i \tag{5}$$

Similarly, to estimate a point along the horizontal side of the frontier x_2 must be re-centred and x_1 kept at the vertex. The CATE's can be aggregated to obtain a summary statistic of the treatment effect along the boundary, $\hat{\tau}_{avg}$, by taking a weighted average;

$$\hat{\tau}_{avg} = \frac{\sum_{k=1}^{K} \hat{\beta}_{1k}(\tilde{x}_1^k, \tilde{x}_2^k) \cdot \hat{\lambda}(\tilde{x}_1^k, \tilde{x}_2^k)}{\sum_{k=1}^{K} \hat{\lambda}(\tilde{x}_1^k, \tilde{x}_2^k)} \tag{6}$$

the weights $\hat{\lambda}(\tilde{x}_1^k, \tilde{x}_2^k)$ are the estimated density of compliers at each boundary point. They can be obtained by following Abadie (2003) who suggests how the proportion of compliers can be identified along the boundary and by applying a one score RD algorithm to a minimum distance measure at the vertex. The intuition behind this simpler way of calculating the average effect along the entire frontier is that the weighted average of the effects is no longer a function of two scores and one score RD is a weighted average, therefore the average of the (average) effects along the frontier can be conveniently obtained by applying a one score RD with an appropriate kernel function. For more details about the possibility to estimate $\lambda(x)$ as a combination of local averages and kernel density estimates with potential boundary corrections we refer the interested reader to Hjort and Jones (1996), Jones (1993); Loader (1996)

Unfortunately, one immediate drawback of this appealing procedure is presented by the limited number of points in the spheres along the boundary. The small sample

---

[15] To be precise, notice that being the set up a fuzzy experiment each estimate is a LATE and the average is taken over the conditional LATE's.
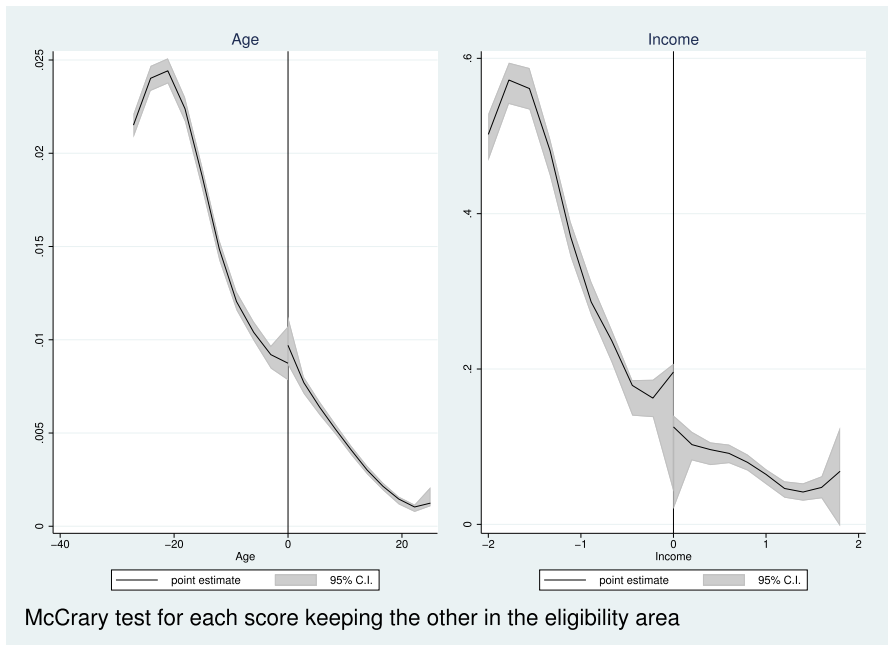
Fig. 2 McCrary manipulation test keeping the other score in the eligibility area. Note: the continuous black line represents the estimated density function, while the gray area represents a 95% CI. The left-hand-side of the figure reports the estimated density of age for units with income below the cutoff. The right-hand-side of the figure reports the estimated density of income for units with age above the cutoff

problem of one score RD is even exacerbated in the MRD because it is repeated as many times as the preferred points along the frontier, this may engender non statistically significant results. In order to lessen weak power problems it is advisable not to include covariates.

5 Testing the identification strategy

One of the central issues for valid identification of causal effects with RD experiments is *continuity*. Continuity of the score(s) density at the cutoff, continuity of the score(s) over the support and continuity of the regression function of the outcome over the score(s). In the case of MRD there is not a consensus in the literature about the tests that one can run in order to validate or reject the identification strategy. For this reason, we have taken a prudential standpoint by adapting the tests developed for single score RD to the MRD in different ways. Broadly speaking, the adaptation we have made consists in running a multiple number of tests. Specifically, for each null hypothesis to be put under empirical testing we run the test: (i) on each score keeping the other in the eligibility area; (ii) on each score keeping the other along the frontier. This procedure generates 2×2 tests for each null hypothesis.

Continuity of the regression funtion keeping one score in the eligibility area
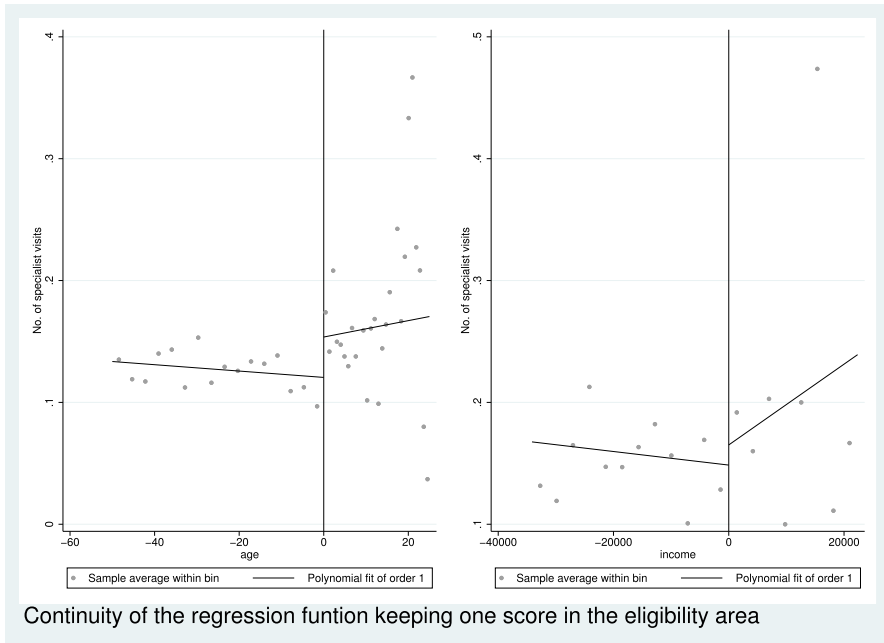
Fig. 3 Continuity of the regression function keeping the other score in the eligibility area. Note: the figure plots polynomial approximations of the conditional expectations of the outcome variable. The optimal number of bins is determined according to the evenly spaced optimal IMSE criterion. The left-hand-side reports the conditional expectation of the outcome with respect to the centred age, for units with income below the cutoff. The right-hand-side of the figure reports the conditional expectation of the outcome with respect to the centred income, for units with age above the cutoff

Identification strategy can break down if units can manipulate the score(s). The McCrary (2008) test is based on the idea that in the absence of manipulation of the score the density of the units does not exhibit a discontinuity around the cutoff. Presence of continuity is usually interpreted as a signal for absence of manipulation, namely random assignment into treatment or control group. The test is implemented as a Wald test and the null is continuity of the density. We implement an advanced version of the original McCrary test as suggested by Cattaneo et al. (2018) who allow us to select the bandwidth entering the test in a data-driven way, consistent with the estimation procedure. The formal test reports a P-value=0.464 and 0.398 for age and income, respectively, leading us not to reject the null hypothesis for the first version of the test, i.e., keeping the other score in the eligibility area, while visual evidence is provided in Fig. 2. Virtually identical results are obtained when repeating the test along the frontier, obtaining a P value of 0.674 and 0.32 for age and income, respectively, and a similar picture in Figure A.4.

As far as continuity of the running variables are concerned, an additional empirical test is reported in Figure A.1, A.2 and A.3 in the Online Appendix, in which the distributions of the scores over their entire support are reported, keeping the other score in the eligibility area (A.1), keeping the other score at its cutoff (A.2) and

Table 3 First stage results at preferred points along the frontier

	(1) Vertex	(2) H1	(3) H2	(4) V1	(5) V2
Z	0.324*	0.264***	0.226***	0.312**	0.349*
	(0.180)	(0.0847)	(0.0448)	(0.137)	(0.170)
N	91	156	229	87	85

Robust Std. Err. clustered at regional level in parentheses

Z = dummy variable, 1 if eligible. The roughly optimal BW according to Scott's rule in standard deviations of the centred running variables is $\tilde{x}_1, \tilde{x}_2 = \pm 0.176$; N = number of observations in the sphere. H1 = [age = 65, income = 10k]; H2 = [age = 65, income = 20k]; V1 = [age = 66, income = 36,151]; V2 = [age = 67, income = 36,151]. The coefficient represents the estimated first stage at the preferred frontier point

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$

keeping the score in a narrow interval along the boundary A.3 while the other is in the eligibility area. A substantial continuity can be visualized in all the cases.

A third check consists in ascertaining the continuity of the regression function. This can be done by dividing the support of the running variable(s) into bins and computing local means of the outcome within those bins. To avoid choosing an arbitrary number of bins Calonico et al. (2015) suggest a data-driven procedure based on the minimization of the integrated mean square error (IMSE). Interestingly, this test is a sort of placebo test in that it tests for discontinuity in the regression function away from the cutoffs. If we found discontinuities at some point different from the cutoffs, we would conclude against a correct identification. Visual evidence is reported in Fig. 3. The pictures are quite informative. On the one hand, they do not provide any evidence of potential discontinuities away from the cutoff in the underlying regression functions, thus acting as falsification tests and validating the key assumption of the design. On the other hand, the pictures show a clear jump in the conditional regression function at the (centred) cutoffs, providing evidence of a positive effect of treatment on outcome. Additional evidence is provided in the Online Appendix in Figure A.5 that repeats the exercise along the boundary and confirms the result.

6 Results

The first-stage results are reported in Table 3.

Each column reports the first stage repeated at different frontier points. Notably, column (1) is estimated at the vertex, as in Eq. (1), the models H1 and H2 in column (2) and (3) have been estimated along the horizontal side of the frontier, that is [age = 65, income = 10k]; [age = 65, income = 20k], respectively, whereas, models V1 and V2 in column (4) and (5) have been estimated along the vertical side [age=66, income=36,151]; [age=67, income=36,151], respectively. Put differently, columns (2) through (5) represent the empirical counterpart of Eq. (4). Each regression equation has been estimated using only the units in the two-dimensional bandwidth. Given the limited number of observations within the spheres no covariates have been added

Table 4 Balancing test

Variable	Effect	Pval
Male	− 1.171	0.143
Married	0.348	0.525
Italian	0.18	0.34
Smoker	0.27	0.689
College	− 0.881	0.166
High	0.324	0.665
Work	0.176	0.76
Employee	0.274	0.6
Selfempl	− 0.098	0.834
House	− 0.224	0.704
NW	1.138	0.101
NE	0.207	0.788
CE	− 0.852	0.17
South	− 0.493	0.507

The table reports the balancing test on covariates. The test has been performed by estimating Eq. 2 at the cutoffs $\bar{x}_1$ and $\bar{x}_2$. For each covariate the column "effect" reports the coefficient β_1 and "Pval" is the associated P value. male = 1 if male, married=1 if married, Italian = 1 if Italian, smoker=1 if smoker, college = 1 if max educational achievement is college degree, higher = 1 if max educational achievement is high school, work = 1 if individual works, employee = 1 if employee, selfempl = 1 if self-employed, house = 1 if house owner, NW = 1 if located in the North-West, NE = 1 if located in the North-Est, CE = 1 if located in the Center, South = 1 if located in the South

in order not to lose further degrees of freedom. Notwithstanding, in order to make identification as plausible as possible, following the suggestion by Cattaneo et al. (2015) and Cattaneo et al. (2016), we have tested whether pre-intervention covariates, such as demographic and socio-economic characteristics, are affected by treatment. Indeed, if treatment assignment is "as if random" within the sphere, the distribution of pre-intervention covariates should be the same for treated and control units. Table 4 reports the results showing that in no case we find evidence of treatment effects on covariates.[16]

The centred scores have been normalized to the same scale by dividing them by their standard deviations, so that the two dimensional bandwidth is the same across all the estimates, easing comparability and interpretability. The first stages are always statistically significant and compliers are between 22% and 35%. It follows that being eligible increases the probability of being exempted from co-payment by 22–35%. From another vantage point, these figures represent the take-up rate of the policy and compared to another Italian policy introduced in 2012, the same year our dataset refers to, such as a support for young innovation oriented firms, they are substantially

---

[16] The table can be also read in the sense that it reports the (bias corrected) difference in mean between exempted and not exempted individuals around the cutoffs showing similarity between the two groups. The authors are indebted with an anonymous referee for having made them notice this point.

Table 5 CATE's at preferred points along the frontier. Dependent variable: No. of specialist visits

	(1) Vertex	(2) H1	(3) H2	(4) V1	(5) V2
W	1.406*	1.265*	0.985*	1.296*	1.837**
	(0.837)	(0.766)	(0.596)	(0.728)	(0.906)
N	91	156	229	87	85

Robust Std. Err. clustered at regional level in parentheses. W = dummy variable, 1 if treated. The roughly optimal BW according to Scott's rule in standard deviations of the centred running variables is $\tilde{x}_1$, $\tilde{x}_2 = \pm 0.176$; N = number of observations in the sphere. H1 = [age = 65, income = 10k]; H2 = [age = 65, income = 20k]; V1 = [age = 66, income = 36,151]; V2 = [age = 67, income = 36,151]. The coefficient represents the average effect at the preferred frontier point
***$p < 0.01$; **$p < 0.05$; *$p < 0.1$

higher, 4–5 times, Mellace and Ventura (2023). This high take-up rate witnesses the relevance and the attractiveness of the policy under evaluation. Interestingly, the highest significance level, i.e., 1%, is achieved in correspondence with the largest number of observations, columns (2) and (3). Indeed, Fig. 1 shows a remarkable presence of points along the horizontal side and sparse points along the vertical side with concentration around the vertex. This explains why the estimates have been performed at age 66 and 67. Considering as outcome variable the number of specialist visits, we can now estimate the CATEs, namely the causal effects for complies along the boundary, which are obtained from Eqs. (2) and (5) for the vertex and the other points along the boundary, respectively. Angrist (2001) guarantees that linear two-stage least squares are still valid with limited or count dependent variables, as in our case. The results are reported in Table 5.

Of course, the points along the boundary considered in Table 3 are the same as Table 5, so that the estimate in the first column is at the vertex, H1 and H2 in columns (2) and (3) are along the horizontal side and the remaining two along the vertical side. Overall, the effect of co-payment exemption on the number of specialist visits is positive and significant at all the points considered. The limited number of observations at each point is likely to raise weak power problems, in this light a 10% significant level is encouraging and can be regarded as a sort of lower bound. The point estimates range between about 1 and 1.8 extra visits generated by the policy in the four weeks before the interview. Caution should be exercised when interpreting this figure since it represents only individuals who have recently received the exemption, not an average of all subjects. As a result, it is reasonable to consider that in the subsequent years, after having the exemption for an extended period, the number of visits might decrease. In essence, these results cannot be generalized to the entire sample of exempt individuals; instead, they should be regarded as "short-term" estimates applicable to those who have just received the exemption at around 65 years of age. These results are in line with the literature showing a negative demand effect of the co-payment (Cherkin et al. 1992; Cockx and Brasseur 2003; Layte et al. 2009; Nolan 2008; Van de Voorde et al. 2001; Winkelmann 2004), while they contrast with studies finding no (or positive) effects of co-payment on healthcare use (Layte et al. 2009; Rosen et al.

**Table 6** Average effect along the frontier, MSE optimal BW

	(1) (lin.)	(2) (quad.)	(3) (asy)	(4) (epa.)
$\hat{\tau}_{avg}$	0.376**	0.400**	0.289*	0.402**
Rob. bias-corr SE	(0.201)	(0.217)	(0.185)	(0.206)
Rob. bias-corr pval	[0.040]	[0.042]	[0.062]	[0.028]
Eff. number of obs left	7110	13,090	13,265	5764
Eff. number of obs right	1613	2376	2180	1428
Kernel type	Triangular	Triangular	Triangular	Epanechnikov
BW type	mserd	mserd	msetwo	mserd
Order Loc. Poly	1	2	1	1
BW	± 0.420	± 0.750	$[-0.766, 0.627]$	± 0.343

Robus bias-corrected Std. Err. clustered at regional level in parentheses, robust bias-corrected Pval in square brackets, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The table reports the average treatment effect averaged over the entire frontier. The optimal bandwidth (BW) has been obtained by minimization of the Mean Square Error (MSE) as in Imbens and Kalyanaraman (2012). lin = linear model; quad = quadratic model; asy = asymmetric BW around the cutoff; epan = Epanechnikof kernel

2011). Interestingly, considering column (2) and (3) the CATE decreases as income increases, while considering column (4) and (5) the CATE increases as individuals age, whereas, the effect at the vertex is remarkably close to the simple mean of the CATE's in the table. As just noted, the limited number of observations at each point prevents us from testing the statistical equality of the estimated effects. However, their dynamics with respect to income and age is encouraging, as fully sensible in economic terms and consistent with the literature.

The estimated average effect computed along the entire frontier, as in Eq. (6) is reported in Table 6.

The table reports the estimate of the average effect along the frontier with the robust bias-corrected standard error and the associated Pval, as suggested by Calonico et al. (2017) and Cattaneo et al. (2017). This effect is a weighted mean of all the effects computable along the frontier, therefore it is not a function of a bivariate score. One score RD can be regarded as a weighted mean within the bandwidth, where the weights are provided by a kernel function. For the score we have followed Cattaneo et al. (2020) and Cattaneo et al. (2023) implementing the closest perpendicular distance to the boundary defined by the treatment assignment, with positive values indicating treated units and negative values indicating control units. To avoid discretionary choices of the bandwidth we have applied the data driven optimal MSE bandwidth, according to Imbens and Kalyanaraman (2012). The effect is always positive and significant at the 5% level. In the baseline specification reported in column (1) the effect amounts to 0.376, meaning that on average among the treated individuals with combination of income and age such as to be along the frontier being exempt from co-payment increases the number of specialist visits by 0.376 every four weeks. Such an average value is below the interval estimated in Table 5, i.e., 1–1.8 for two reasons. First, for practical reasons Table 5 refers to five points while Table 6 to an infinite collection.
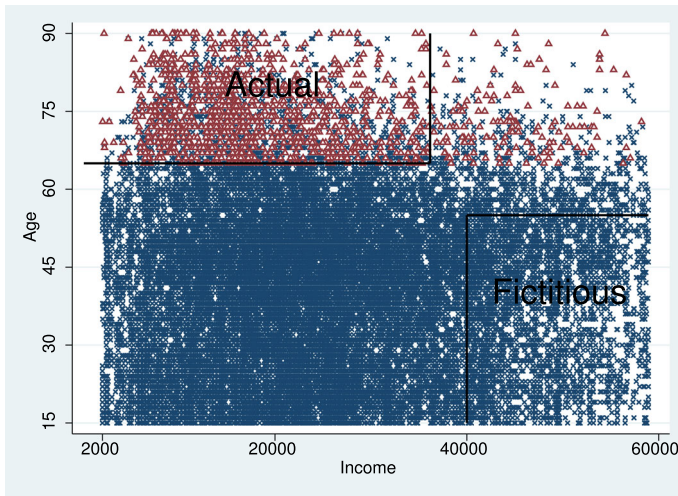
Fig. 4 Actual and fictitious treatment area. Note: The black solid lines delimit boundaries. The quadrant "Actual" represents the actual eligibility area for treatment. The quadrant "Fictitious" represents the placebo area for fictitious treatment

Second, the one score estimator puts less weights (because of a smaller density along the frontier) on the effect for lower income and older individuals, whose effect is greater. To check for the robustness of the result, we have carried out a number of perturbations of the baseline by including squared values of the score, column (2), allowing for asymmetric bandwidth on the left and on the right-hand-side of the cut-off, column (3), and using a different kernel function, column (4). The average effect is always positive and significant and does not show appreciable differences along the 4 different specifications. As a very final check we have repeated the same exercise by using a different data-driven optimal bandwidth, the Coverage Error Rate (CER), instead of the MSE, because Calonico et al. (2017) argue that CER has better inference properties. All the results are confirmed and evidence is provided in the Online Appendix in Table A.1.

7 Robustness

To check the robustness of our results we implement a placebo test in the spirit of Abadie et al. (2010) and the ensuing copious literature. For the placebo treatment we construct a fictitious treatment area based on the joint event ($x_{1i} \leq 55 \cap x_{2i} > 40,000$). The situation is depicted in Fig. 4.

The CATE's estimated at the vertex and along the fictitious boundary show no significant results, as reported in Table 7 as well as in Figure A.6.

As further proof, the average CATE along the entire fictitious frontier amounts to 0.0126 with a P value=0.67.

Another falsification test consists in keeping the actual eligibility area but changing the outcome. In particular, we know that generic and pediatric visits are free from

Table 7 Falsification test along the fake frontier

	(1) Vertex	(2) H1	(3) H2	(4) H3	(5) H4	(6) V1	(7) V2	(8) V3
W	-0.580	-0.427	-0.0204	0.0992	0.0485	0.0914	-0.0185	-0.191
	(0.375)	(0.334)	(0.405)	(0.220)	(0.536)	(0.0893)	(0.115)	(0.197)
N	203	121	125	71	19	233	281	220

Dependent variable: no. of specialist visits. Robust Std. Err. in parentheses clustered at regional level. $W = 1$ if treated. Roughly optimal BW according to Scott's rule in standard deviations of the centred running variables. The coefficient represents the average effect at the preferred frontier point

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

payment, for any patient both exempted and not. Thus, if we find a significant effect of exemption on this outcome that will signal something wrong in the identification strategy. In no case it is possible to find an effect at any conventional significance level. For sake of space evidence about the average effect along the entire frontier is reported in the Online Appendix, Table A.2 and Table A.3, respectively, for the MSE and CER optimal data driven bandwidth, and their robustness checks.

8 External validity

One of the major limitations one comes across when using RD is about the local validity of the estimates, in the sense that the results apply only at the cutoff. In the case of MRD this limitation is mitigated by the fact that the discontinuity consists of a border, rather than a point, so that the estimates take on a wider interpretation, with respect to the single score case. Notwithstanding, the main question still applies: do the results we have found apply away from the frontier? Put differently, if the government slightly changed the eligibility criteria, would we find the same effect we have estimated? To answer this question, Dong and Lewbel (2015) suggest that the treatment effect derivative, TED, is an appropriate statistic for the stability of the results around the cutoff. The intuition is very simple: if small changes in a score (computed at its cutoff) lead to significant changes in the treatment effect, there is evidence of instability of the results, namely they cannot be taken as valid away from the cutoff. Thus, in our specific case, the idea is to apply this technique twice, once for each score keeping the other passing the cutoff, i.e., in the eligibility area. Table 8 reports a full set of TED estimates computed on x_1 keeping $x_2 < \bar{x}_2$. In other words, in Table 8 we are checking whether by slightly shifting the vertical side of the frontier the results still apply. Similarly, in Table 9 we are checking whether by slightly shifting the horizontal side of the frontier the results still apply. The double check is performed under a variety of alternative specifications, just to make sure that the results are general enough and not driven by the specific functional assumption of the regression equation nor by other ad-hoc choices.

Column (1) of Tables 8 and 9 reports the estimated TED by using the same bandwidth we have used in MRD estimates, as suggested by Scott's rule. Columns (2) through (5) use the optimal data-driven MSE bandwidth (Imbens and Kalyanaraman

Table 8 TED for eligible units by income

	(1) $(\hat{h}_1)$	(2) (lin.)	(3) (quad.)	(4) (asy.)	(5) (epa.)
TED	0.943	− 1.033	− 0.992	− 0.905	− 1.079
	(4.122)	(0.641)	(1.418)	(0.482)	(0.737)
BW type	Scott	MSE	MSE	MSETWO	MSE

Robust Std. Err. in parentheses. The table reports the TED using x_1 as the running variable and keeping only units having $x_2 < \bar{x}_2$. $\hat{h}1$ = optimal BW according to Scott's rule; lin = linear model with optimal BW obtained by minimization of the Mean Square Error (MSE) as in Imbens and Kalyanaraman (2012); quad = quadratic model with optimal BW obtained by minimization of the MSE; asy = asymmetric BW around the cutoff; epan = Epanechnikof kernel. MSETWO= asymmetric BW obtained by MSE minimization
***$p < 0.01$; **$p < 0.05$; *$p < 0.1$

Table 9 TED for eligible units by age

	(1) $(\hat{h}_1)$	(2) (lin.)	(3) (quad.)	(4) (asy.)	(5) (epa.)
TED	1.441	− 1.413	− 2.395	− 1.434	− 1.108
	(12.456)	(2.097)	(3.284)	(2.097)	(2.374)
BW type	Scott	MSE	MSE	MSETWO	MSE

Robust Std. Err. in parentheses. The table reports the TED using x_2 as the running variable and keeping only units having $x_1 \geq \bar{x}_1$. $\hat{h}2$= optimal BW according to Scott's rule; lin= linear model with optimal BW obtained by minimization of the Mean Square Error (MSE) as in Imbens and Kalyanaraman (2012); quad= quadratic model with optimal BW obtained by minimization of MSE; asy= asymmetric BW around the cutoff; epan= Epanechnikof kernel. MSETWO=MSE optimal asymmetric BW
***$p < 0.01$; **$p < 0.05$, *$p < 0.1$

2012) which is re-calculated each time under different circumstances. In column (2) a linear model is adopted, in column (3) a quadratic model, column (3) adopts an asymmetric bandwidth allowing for different lengths on the right- and on the left-hand-side of the cutoff, and, finally a different kernel function is used in column (5). The results show a great regularity. In each column of both tables the estimated TED is always statistically non-significant at the conventional levels, indicating that our results might have higher external validity.

Finally, in order to rule out inference problems we have followed the procedure suggested by Calonico et al. (2017) replacing the MSE optimal data-driven bandwidth with the CER. Evidence is provided in Tables A.3 and A.4 in the Online Appendix and show exactly the same results as the ones in Table A.4 and A.5 lending further support to the hypothesis of external validity.

9 Conclusions

This paper investigates the causal effect of co-payment exemption for income on the utilization of healthcare services with a particular focus on the number of specialist visits in the Italian NHS. From a methodological perspective we implement the recently

developed multiple regression discontinuity (MRD) model exploiting the discontinuity in the two features connected to the eligibility criteria: household's income and age. Given our context, the technique allows us to estimate the causal effect in a quasi-experimental setting.

Results provide strong evidence in favor of a positive and significant effect of co-payment exemption for income on the number of specialist visits. Our evidence is in line with the literature supporting a negative healthcare demand effect of the co-payment, while it is in contrast with studies finding no (or positive) effects of co-payment on healthcare use.

From a policy perspective this evidence supports the hypothesis that a reduction in the use of specialist visits can be obtained introducing an appropriate co-payment framework. However, it is worth recalling that a reduction in the use of specific healthcare services may potentially decrease the total healthcare costs, but not necessarily. Indeed, it is possible that co-payment for some services may cause substitution effects with services that are free or subject to a lower co-payment. Furthermore, the reduction in some effective and necessary care may lead to a deterioration in public health, and in the long run the savings brought by co-payment may be canceled by increases in the use of other services. Finally, by eliminating co-payment exemption it is reasonable to expect that low income individuals will reduce their use of specialist visits more than wealthier individuals. Under the standard assumption of decreasing marginal utility of income, low-income individuals will suffer from a higher utility loss associated with the additional cost they must bear. Our study presents compelling proof that waiving co-payment for underprivileged populations leads to a notable rise in the utilization of specialist physician services. These services play a pivotal role in preventive healthcare by facilitating early detection, offering specialized knowledge and guidance, evaluating individual risks, creating personalized treatment plans, and providing valuable health education. Consequently, the co-payment exception has the potential to greatly enhance health outcomes and overall quality of life.

One of the main limitations of this study is that given the low number of observations at each point along the frontier, we cannot condition the estimates upon covariates because that would lead to a further loss of observations and neither we can test for equality of the estimated effects. In this respect, the new waves of the Multipurpose Survey are expected to be richer and more accurate, especially in terms of sample span and questions asked to respondents. We leave this issue to further research.

Declarations

References

Abadie A (2003) Semiparametric instrumental variable estimation of treatment response models. J Econometr 113(2):231–263

Abadie A, Diamond A, Hainmueller J (2010) Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program. J Am Stat Assoc 105(490):493–505

Angrist JD (2001) Estimation of limited dependent variable models with dummy endogenous regressors: simple strategies for empirical practice. J Bus Econ Stat 19(1):2–28

Angrist JD, Rokkanen M (2015) Wanna get away? Regression discontinuity estimation of exam school effects away from the cutoff. J Am Stat Assoc 110(512):1331–1344

Arestis P, Martin R, Tyler P (2011) The persistence of inequality? Camb J Reg Econ Soc 4(1):3–11

Aron-Dine A, Einav L, Finkelstein A (2013) The rand health insurance experiment, three decades later. J Econ Perspect 27(1):197–222

Atella V, Kopinska JA (2014) The impact of cost-sharing schemes on drug compliance in Italy: evidence based on quantile regression. Int J Public Health 59(2):329–339

Atella V, Peracchi F, Depalo D, Rossetti C (2006) Drug compliance, co-payment and health outcomes: evidence from a panel of Italian patients. Health Econ 15(9):875–892

Athey S, Wager S (2021) Policy learning with observational data. Econometrica 89(1):133–161

Baltagi BH, Lagravinese R, Moscone F, Tosetti E (2017) Health care expenditure and income: a global perspective. Health Econ 26(7):863–874

Battistin E, Rettore E (2008) Ineligibles and eligible non-participants as a double comparison group in regression-discontinuity designs. J Econometr 142(2):715–730

Bernal N, Carpio MA, Klein TJ (2017) The effects of access to health insurance: evidence from a regression discontinuity design in Peru. J Public Econ 154:122–136

Calonico S, Cattaneo MD, Titiunik R (2014) Robust nonparametric confidence intervals for regression-discontinuity designs. Econometrica 82(6):2295–2326

Calonico S, Cattaneo MD, Titiunik R (2015) Optimal data-driven regression discontinuity plots. J Am Stat Assoc 110(512):1753–1769

Calonico S, Cattaneo MD, Farrell MH, Titiunik R (2017) Rdrobust: software for regression-discontinuity designs. Stand Genom Sci 17(2):372–404

Calonico S, Cattaneo MD, Farrell MH (2020) Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. Econometr J 23(2):192–210

Carnazza G, Liberati P, Resce G, Molinaro S (2021) Smoking and income distribution: Inequalities in new and old products. Health Policy 125(2):261–268

Cattaneo MD, Idrobo N, Titiunik R (2023) A practical introduction to regression discontinuity designs: extensions. arXiv preprint arXiv:2301.08958

Cattaneo MD, Frandsen BR, Titiunik R (2015) Randomization inference in the regression discontinuity design: an application to party advantages in the us senate. J Causal Inference 3(1):1–24

Cattaneo MD, Titiunik R, Vazquez-Bare G (2016) Inference in regression discontinuity designs under local randomization. Stand Genom Sci 16(2):331–367

Cattaneo MD, Titiunik R, Vazquez-Bare G, Keele L (2016) Interpreting regression discontinuity designs with multiple cutoffs. J Polit 78(4):1229–1248

Cattaneo MD, Titiunik R, Vazquez-Bare G (2017) Comparing inference approaches for RD designs: a reexamination of the effect of head start on child mortality. J Policy Anal Manag 36(3):643–681

Cattaneo MD, Jansson M, Ma X (2018) Manipulation testing based on density discontinuity. Stand Genom Sci 18(1):234–261

Cattaneo MD, Titiunik R, Vazquez-Bare G (2019) Power calculations for regression-discontinuity designs. Stand Genom Sci 19(1):210–245

Cattaneo MD, Titiunik R, Vazquez-Bare G (2020) Analysis of regression-discontinuity designs with multiple cutoffs or multiple scores. Stand Genom Sci 20(4):866–891

Cherkin DC, Grothaus L, Wagner EH (1992) Is magnitude of co-payment effect related to income? Using census data for health services research. Soc Sci Med 34(1):33–41

Clark D, Martorell P (2014) The signaling value of a high school diploma. J Polit Econ 122(2):282–318

Cockx B, Brasseur C (2003) The demand for physician services: evidence from a natural experiment. J Health Econ 22(6):881–913

Costa G, Spadea T, Petrelli A (2012) Cost sharing, salute ed equità nella salute, il ruolo del copayment nella sanità italiana. Seminario di studio e di discussione promosso da Agenas e Aies, Roma

De Matteis D, Ishizaka A, Resce G (2019) The 'postcode lottery' of the Italian public health bill analysed with the hierarchy stochastic multiobjective acceptability analysis. Socioecon Plann Sci 68:100603

Dell M (2010) The persistent effects of Peru's mining mita. Econometrica 78(6):1863–1903

Dong Y, Lewbel A (2015) Identifying the effect of changing the policy threshold in regression discontinuity models. Rev Econ Stat 97(5):1081–1092

D'Orazio M (2017) Statistical matching and imputation of survey data with StatMatch. Italian National Institute of Statistics, Rome, Italy

Färdow J, Broström L, Johansson M (2019) Co-payment for unfunded additional care in publicly funded healthcare systems: ethical issues. J Bioethical Inq 16(4):515–524

Feigenbaum JJ (2016) A machine learning approach to census record linking. Retr March 28:2016

Finkelstein A, Taubman S, Wright B, Bernstein M, Gruber J, Newhouse JP, Allen H, Baicker K, Group OHS (2012) The Oregon health insurance experiment: evidence from the first year. Q J Econom 127(3):1057–1106

Fiorio CV, Siciliani L (2010) Co-payments and the demand for pharmaceuticals: evidence from Italy. Econ Model 27(4):835–841

Frey A (2019) Cash transfers, clientelism, and political enfranchisement: evidence from Brazil. J Public Econ 176:1–17

Gemmill MC, Thomson S, Mossialos E (2008) What impact do prescription drug charges have on efficiency and equity? Evidence from high-income countries. Int J Equity Health 7(1):1–22

Gerber AS, Kessler DP, Meredith M (2011) The persuasive effects of direct mail: a regression discontinuity based approach. J Polit 73(1):140–155

Goodhart D (2017) The road to somewhere: the populist revolt and the future of politics. Oxford University Press

Grassetti L, Rizzi L (2019) The determinants of individual health care expenditures in the Italian region of Friuli Venezia Giulia: evidence from a hierarchical spatial model estimation. Empir Econ 56(3):987–1009

Greco S, Ishizaka A, Matarazzo B, Torrisi G (2018) Stochastic multi-attribute acceptability analysis (SMAA): an application to the ranking of Italian regions. Reg Stud 52(4):585–600

Hjort NL, Jones MC (1996) Locally parametric nonparametric density estimation. Ann Stat 1619–1647

Imbens G, Kalyanaraman K (2012) Optimal bandwidth choice for the regression discontinuity estimator. Rev Econ Stud 79(3):933–959

Inoue M, Kachi Y (2017) Should co-payments for financially deprived patients be lowered? Primary care physicians' perspectives using a mixed-methods approach in a survey study in Tokyo. Int J Equity Health 16(1):1–8

ISTAT (2013) Multipurpose survey on health conditions and use of health services (2011–2012)

ISTAT (2021) Indagine su reddito e condizioni di vita (eu-silc)

Jacob BA, Lefgren L (2004) Remedial education and student achievement: a regression-discontinuity analysis. Rev Econ Stat 86(1):226–244

Jones MC (1993) Simple boundary correction for kernel density estimation. Stat Comput 3:135–146

Jusot F, Or Z, Sirven N (2012) Variations in preventive care utilisation in Europe. Eur J Ageing 9:15–25

Keele LJ, Titiunik R (2015) Geographic boundaries as regression discontinuities. Polit Anal 23(1):127–155

Kiil A, Houlberg K (2014) How does copayment for health care services affect demand, health and redistribution? A systematic review of the empirical evidence from 1990 to 2011. Eur J Health Econ 15(8):813–828

Lagravinese R (2015) Economic crisis and rising gaps north-south: evidence from the Italian regions. Camb J Reg Econ Soc 8(2):331–342

Lagravinese R, Liberati P, Resce G (2019) Exploring health outcomes by stochastic multicriteria acceptability analysis: an application to Italian regions. Eur J Oper Res 274(3):1168–1179

Layte R, Nolan A, McGee H, O'Hanlon A (2009) Do consultation charges deter general practitioner use among older people? A natural experiment. Soc Sci Med 68(8):1432–1438

Liberati P (2015) The world distribution of income and its inequality, 1970–2009. Rev Income Wealth 61(2):248–273

Lin Y-L, Chen W-Y, Shieh S-H (2020) Age structural transitions and copayment policy effectiveness: evidence from Taiwan's national health insurance system. Int J Environ Res Public Health 17(12):4183

Loader CR (1996) Local likelihood density estimation. Ann Stat 24(4):1602–1618

Lueckmann SL, Hoebel J, Roick J, Markert J, Spallek J, von dem Knesebeck O, Richter M (2021) Socioeconomic inequalities in primary-care and specialist physician visits: a systematic review. Int J Equity Health 20(1):1–19

McCrary J (2008) Manipulation of the running variable in the regression discontinuity design: a density test. J Econometr 142(2):698–714

Mellace G, Ventura M (2023) The short-run effects of public incentives for innovation in Italy. Econ Model 120:106178

Murray JS (2018) Multiple imputation: a review of practical and theoretical findings. Stat Sci 33(2):142–159

Newhouse JP (1974) A design for a health insurance experiment. Inquiry 11(1):5–27

Nolan A (2008) Evaluating the impact of eligibility for free care on the use of general practitioner (GP) services: a difference-in-difference matching approach. Soc Sci Med 67(7):1164–1172

O'Donnell O, Van Doorslaer E, Wagstaff A, Lindelow M (2007) Analyzing health equity using household survey data: a guide to techniques and their implementation. The World Bank

Oliveira Martins J, de la Maisonneuve C (2006) The drivers of public expenditure on health and long-term care: an integrated approach. SSRN 917782

Papay JP, Willett JB, Murnane RJ (2011) Extending the regression-discontinuity approach to multiple assignment variables. J Econometr 161(2):203–207

Ponzo M, Scoppa V (2021) Does demand for health services depend on cost-sharing? Evidence from Italy. Econ Model 103:105599

Resce G, Lagravinese R, Benedetti E, Molinaro S (2019) Income-related inequality in gambling: evidence from Italy. Rev Econ Househ 17(4):1107–1131

Rezayatmand R, Pavlova M, Groot W (2013) The impact of out-of-pocket payments on prevention and health-related lifestyle: a systematic literature review. Eur J Public Health 23(1):74–79

Rosen B, Brammli-Greenberg S, Gross R, Feldman R (2011) When co-payments for physician visits can affect supply as well as demand: findings from a natural experiment in israel's national health insurance system. Int J Health Plann Manag 26(2):e68–e84

Schenker N, Welsh AH (1988) Asymptotic results for multiple imputation. Ann Stat 16(4):1550–1566

Serna N (2021) Cost sharing and the demand for health services in a regulated market. Health Econ 30(6):1259–1275

Solanki G, Schauffler HH (1999) Cost-sharing and the utilization of clinical preventive services. Am J Prev Med 17(2):127–133

Subramanian SV, Kawachi I (2004) Income inequality and health: what have we learned so far? Epidemiol Rev 26(1):78–91

Van de Voorde C, Van Doorslaer E, Schokkaert E (2001) Effects of cost sharing on physician utilization under favourable conditions for supplier-induced demand. Health Econ 10(5):457–471

Vincenzino JV (1995) Health care costs: market forces and reform. Oncology (Williston Park) 9(5):367–8

White C (2007) Health care spending growth: how different is the united states from the rest of the OECD? Health Aff 26(1):154–161

WHO et al (2018) Public spending on health: a closer look at global trends. Technical report, World Health Organization

Winkelmann R (2004) Co-payments for prescription drugs and the demand for doctor visits-evidence from a natural experiment. Health Econ 13(11):1081–1089

Zajonc T (2012) Essays on causal inference for public policy. Ph.D. thesis, Harvard University

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.