User Fee Abolition and the Demand for Public Health Care[*]

Steven F. Koch[†]

May 2016

Abstract

This research examines the effect of the abolition of user fees in South Africa, a policy implemented in 1994 for uninsured children under the age of six and the elderly uninsured, as well as pregnant and nursing mothers. The analysis focuses on the implementation of the policy and the use of curative public healthcare services by children following strict and fuzzy regression discontinuity designs. The estimates point to statistically insignificant average and local average policy effects, even though the policy appears to have been implemented reasonably effectively, albeit imperfectly. In other words, the policy did not, on average, affect the use of curative public healthcare, at least for those children who should have benefited from the policy.

Keywords: User Fee Abolition, Regression Discontinuity

JEL: I18, I11

[†]Department of Economics, University of Pretoria, Pretoria, Republic of South Africa; (O) 27-12-420-5285, (F) 27-86-691-2749. *steve.koch@up.ac.za*

1 Introduction

The imposition and abolition of health care user fees has been a feature of public health care delivery in Africa for a number of decades. Spurred on by the goal of raising additional funds for health budgets and increasing the efficiency of health care delivery, a number of countries on the continent imposed user fees for public health services. However, experience with those fees was not particularly positive.[1] A number of African countries have, since, reversed tack, abolishing their user fee programs. The literature examining these policies and changes has spawned a number of reviews, such as that by Ridde and Morestin (2011) and Lagarde and Palmer (2008).

Largely, early studies focused on registrations at the clinic level, typically for one to four clinics – see McCoy and Khosa (1996), Mwabu and Wang'ombe (1997), Wilkinson et al. (1997), Wilkinson et al. (2001). Although these studies find relatively large increases in registrations associated with user fee abolition, Lagarde and Palmer (2008) worry that these studies cover too small a sample of clinics to be meaningful to policy-makers, and that the studies are unlikely to have uncovered policy impacts. In South Africa, the focus of this analysis, the most extensive sample is contained in Bayat and Cleaton-Jones (2003), who consider 10 facilities. There are, however, more recent facility-level studies in other countries that include a broader sample. For example, Nabyonga et al.'s (2005) and Burnham et al.'s (2004) samples include six to ten health districts in Uganda, respectively, while Masiye et al. (2008) and Chama-Chiliba and Koch (2014) include data from most of the 72 districts in Zambia. With the exception of Chama-Chiliba and Koch (2014), all of the aforementioned studies uncover large user fee abolition associations, ranging from 17% to 54%; those increases are shown to coincide with user fee abolition, but are not causal (Lagarde and Palmer, 2008). In addition to increases in usage, Burnham et al. (2004), Walker and Gilson (2004) and Masiye et al. (2008) uncover worker disgruntlement associated with the increased workload.

Although the literature suggests large and beneficial impacts are to be had, following

---

[1]In Kenya, Willis and Leighton (1995) find user fees associated with a 27% decrease in utilization at provincial hospitals, a 46% decrease at district hospitals and a 33% decrease at health centres. In Zambia, Blas and Limbambala (2001) find outpatient attendance reductions of 35%. Biritwum (1994) finds a 40% decrease in outpatient attendance in Ghana.

the reduction in healthcare access/use charges, Lagarde and Palmer's (2008) review is sceptical. Supported by their interrupted time series analysis, conceptually similar to regression discontinuityn (RD), they argue that the published clinic-level literature, including that in South Africa, leaves much to be desired. With respect to South Africa, they are especially concerned by the willingness to imply national conclusions, based on data from such a small set of clinics or hospitals, and they worry about possible concurrent changes that were also likely to have affected the impact.[2]

One approach to addressing the causality limitations within the literature is through randomised controlled trials (RCTs). For example, in the case of bednet provision, there is strong evidence from Kenya that free distribution increased take-up (Cohen and Dupas, 2010). Similarly, Miguel and Kremer (2004) find that the introduction of cost-sharing for deworming drugs reduced take-up. More generally, Ashraf et al. (2010) present convincing evidence that lower prices are more likely to lead to purchases, at least for chlorine disinfectant, while post-acceptance discounts do not further influence purchase decisions. However, large-scale randomised controlled trials of public health care user fee abolition have not been undertaken. One large-scale analysis, McKelvey et al. (2012), suggests that there are rather limited price effects associated with contraceptive use in Indonesia. Their analysis is reasonably representative of 'complete scale', because the financial crisis impacted the entire country, even if not uniformly. Although one would expect user fee abolition to increase take-up, the user fee is not the only cost associated with accessing health facilities. Furthermore, as implied by Burnham et al. (2004), Walker and Gilson (2004) and Masiye et al. (2008) the health system may not be in a position to scale-up in response. Therefore, the equilibrium impact of user fee abolition could be much smaller than policy makers expected, or as has been previously reported.

Because there are no RCTs allowing for the isolation of user fee abolition effects, researchers have either assumed that user fee abolition was an exogenous policy shock (Deininger and Mpuga, 2005) or applied quasi-experimental methods. Most common amongst these quasi-experimental methods are differences-in-differences (DD), and the

---

[2]For example, Benatar (1997) highlights the construction of almost 100 new primary care clinics that were opened by the end of 1996.

recent results available are, for the most part, supporting the earlier findings, but at smaller magnitudes. McKinnon et al. (2015) find small but statistically significant effects on facility-based deliveries, an approximate 3% increase, resulting from user fee abolition in a range of countries in Africa. Chama-Chiliba and Koch (2014) find an increase of approximately one-half of one percent per quarter increase in facility-based deliveries in Zambia, although there is no evidence of an immediate increase and their post-abolition period contains a limited number of quarters. In South Africa, Brink and Koch (2015), on the other hand, find statistically insignificant effects; plausibly, the assumptions required for DD methods are not met in their study, although their results are robust to a set of sensitivity checks.

In the following analysis, we set out to re-examine the effects of the policy change in South Africa, which we outline in the following section. The policy that was adopted allows for the application of Regression Discontinuity (RD) to estimate the causal effects of user fee abolition on a subset of the affected population, specifically children under the age of six (referred to as eligible young children). The analysis begins with a check on policy implementation, and finds that the policy was implemented, albeit imperfectly. Despite some evidence of implementation, although imprecisely estimated and economically small, the subsequent analysis finds no convincing evidence suggesting that public healthcare-seeking behaviour undertaken by (caregivers of) eligible young children was affected.

Three previous papers have considered different aspects of this and a related South African policy. Using DD, Tanaka (2014) presents evidence that user fee abolition increased the level of child health in the country within a few years. Brink and Koch (2015), who also use DD, examine whether or not the 1996 policy change, which is also described below, affected healthcare-seeking behaviour. Finally, Koch and Racine (forthcoming) use the fact that the policy change has a multinomial set of outcomes – in that patients can choose not to be treated, seek care at a public facility or seek care in a private facility, and that these choices are interrelated – to develop a nonparametric estimate of the effect across these three outcomes. Their estimates suggest economically small, but plausible effects of the policy on both home care and public healthcare seeking,

but are only sharp RD estimates. Thus, the research presented here is most closely related to Koch and Racine (forthcoming), since we consider the same policy and include both parametric and nonparametric analyses. However, we are able to differentiate our research in three important ways. The first is that we use a subset of the data focussing only on those who are expected to benefit from the policy change. Our estimated magnitudes are in line with Koch and Racine's (forthcoming) analysis, which gives further support to the nonparametric approach that they develop. The second is that we examine the effect of facility location on the policy effect. The third is that we consider the potential impact of imperfect policy implementation. Thus, we are able to undertake a local analysis of the policy, focussing on those expected to benefit, while controlling for less than perfect roll-out and examining the importance of physical access.

2   Background and Methodology

2.1   Policy Context

In 1994, the late Nelson Mandela – the newly elected president of South Africa at that time – announced the beginning of free access to primary healthcare. The policy was explicitly designed for young uninsured children (under the age of six) and the uninsured elderly, as well as pregnant and nursing mothers; free primary healthcare was extended to everyone in 1996. The 1994 announcement, affecting health care in the public sector, was strongly influenced by the Declaration of Alma-Ata, made in September 1978.[3] Underpinned by the declaration's tenets, the change in policy was expected to improve healthcare access for South Africans and help alleviate health inequalities, especially those associated with access.

Economically, Fuchs (1968) and Grossman (1999) argue that health demand is derived demand, having the expected properties, while Heller (1982) suggests that a decrease in access costs should positively impact welfare, although the impact could be minimal if healthcare demand is price inelastic. Given that backdrop and Gertler et al.'s

---

[3]The declaration proclaims that health is a fundamental human right, and that to achieve the "promotion and protection of the health of the people", primary healthcare – "essential healthcare based on practical, scientifically sound and socially acceptable methods and technology made universally accessible to individuals" – should provide "promotive, preventative, curative and rehabilitative services".

(1987) higher healthcare demand price responsiveness amongst the poor implies that user fee abolition in South Africa should have increased access to the public health sector and benefited the poor. The limited South African literature, which focuses on a small number of clinics, suggests that usage increased after user fee abolition – see Figure 1 – McCoy and Khosa (1996), Wilkinson et al. (1997), Wilkinson et al. (2001) and Bayat and Cleaton-Jones (2003). Although Walker and Gilson's (2004) nurses also felt the policy positively impacted the poor, a notion supported by Koch and Racine (forthcoming).

Figure 1 about here.

In 1994, South Africa was throwing off the yoke of racial oppression, which also included separate (by race) health systems in need of integration, while poverty was rife; see Coovadia et al. (2009) for a detailed review of the segregation and integration of the South African health care system. According to World Bank data, the $1.25 per day headcount poverty ratio stood at 24.3% in 1993 (21.4% in 1995). User fees at that time, outlined in McIntyre et al. (1995) were relatively high. For the poorest patients, fees ranged from South African Rand (ZAR) 8.00 for outpatient care at a community hospital to ZAR 26.00 per admission for inpatient care at a regional or academic hospital (in 1995, the exchange rate averaged ZAR 3.50/$). Despite a $1.25 per day poverty rate of at least 20%, the minimum user fee exceeded $2.25.

2.2 Empirical Identification

The revised user fee policy was designed for a myriad of different groups; however, the following analysis will focus on a subset of the expected direct effects of the policy: (i) whether or not payment was made for care in the public sector, when eligible children received treatment, and (ii) the demand for curative care services in the public sector for eligible children. These foci arose from data limitations. The data available does not make it possible to directly consider preventative care, antenatal care or effects related to nursing mothers. However, it should also be noted that the data, collected during 1995, was from a survey that was not designed specifically for the purposes for which it

6

is being used, and, therefore, some caution must remain in the interpretation.[4]

Since the components of the policy examined relate to young children, and there is an age threshold within the policy, the analysis follows an RD Design. The policy is expected to result in an increase in the proportion of young uninsured ill children treated by the public health care sector. However, positive impacts are not likely to have materialized, at least to the extent possible, if the policy was not properly implemented. Therefore, it is necessary to establish if payment was less common amongst the eligible being treated in the public sector. Unfortunately, data for this analysis is based on a survey question referring to payment either for the service or for medicine, and therefore, is not as clear as would be preferred for the analysis. Furthermore, the policy is meant to apply to a specific group of children, those under six and uninsured, while the analysis focuses only on the ill, because it is not possible to access meaningful information about preventive care in the survey. Thus, the analysis sample is selective, and, therefore, it is necessary to check if the selection influences other healthcare related reporting. In our analysis, not reported, we find no evidence that selection influenced other healthcare related behaviours; thus, we infer that it has not influenced what we report, below.[5] Finally, although the focus is on the uninsured, the implied expectation of the policy was that it would improve access; thus, we also controlled for access (based on reported distance), finding some evidence that access mattered.

The expected policy impacts are examined via RD. In order to specify the RD design, a binary indicator of policy eligibility, D, is defined. It is determined by an age threshold; age is denoted by a, and the threshold, six years, is denoted by a_0; note that in the subsequent analysis, we translate the axis for easier interpretation. Therefore, $D = \mathbb{1}\left[a < a_0\right]$. Further, denote the outcome of interest if the child is eligible for free public healthcare as $Y_1 \in \{0, 1\}$, while $Y_0 \in \{0, 1\}$ is the outcome of interest if the child

---

[4]The data, outlined in Section 3, is further limited to 1995 for two reasons. Firstly, the 1994 data does not contain useful information separating the private and public sectors, when it comes to health. Secondly, the 1996 survey was undertaken after a second fee abolition extension in July, following which nearly all uninsured individuals were able to access the sector *sans* user fees; see Brink and Koch (2015).

[5]We have considered, specifically, whether or not medical aid coverage was affected by the policy, or whether reports of illness were affected by the policy. We find no meaningful differences there. We also undertook McCrary's (2008) test of running variable manipulation, and looked for discontinuities in other variables that are determinants of health care demand. We find no evidence of such manipulation or other discontinuities. We are happy to share our analysis upon request.

does not qualify for free primary healthcare. Defining $\tau \equiv E[Y_1 - Y_0 | a = a_0]$ as the average treatment effect, the goal of this RD analysis is to estimate τ, which is the effect of free access to primary healthcare at the age cut-off on the use of public health facilities.

Assuming that the correlation between policy eligibility and unobserved determinants of the outcome of interest, which we represent by ν_i, is constant across the age threshold, as in (1), the treatment effect can be uncovered.

$$\lim_{\Delta \uparrow 0} E[D_i \nu_i | X_i, a_i = a_0 + \Delta] - \lim_{\Delta \downarrow 0} E[D_i \nu_i | X_i, a_i = a_0 + \Delta] = 0 \qquad (1)$$

The constant correlation assumption outlined in (1) allows the correlation to be differenced out, such that the treatment effect is identified as the difference between outcomes for the treatment and control groups.

$$\tau = \lim_{\Delta \uparrow 0} E[Y_i | X_i, a_i = a_0 + \Delta] - \lim_{\Delta \downarrow 0} E[Y_i | X_i, a_i = a_0 + \Delta] \qquad (2)$$

The preceding represents a sharp RD, which can be estimated via a parametric regression, which includes controls for age, as implied by the conditioning factors in (2).

$$Y_i = \beta_0 + \tau D_i + g(\tilde{a}_i) + D_i h(\tilde{a}_i) + \nu_i \qquad (3)$$

In (3), Y_i is the binary indicator of the outcome being analysed for ill child i, $\tilde{a}_i = a_i - 72$ is the age of the child in months, net of the threshold. D_i is the treatment indicator, and the treatment effect to be estimated is τ. Due to the dependence of the policy on age, different functions of age, represented by g and h, on either side of the threshold, are included in the regression. We also consider a nonparametric counterpart to (3), where $j = \mathbb{1}[D_i = 1]$ represents age eligibility and, across the threshold, D_i and ω_i are assumed to have constant covariance.

$$Y_{ij} = m_j(\tilde{a}_i) + \omega_{ij} \qquad (4)$$

The nonparametric regression is estimated separately on either side of the threshold and and $\hat{\tau}$ is estimated from the sample analog of (2), which is also bootstrapped.

Furthermore, it is possible that policy is not implemented perfectly, resulting in a fuzzy estimate of the policy effect. The Wald estimate of that effect is described in (5).

$$\tau_w = \frac{\lim_{\Delta\uparrow 0} E[Y_i|X_i, a_i = a_0 + \Delta] - \lim_{\Delta\downarrow 0} E[Y_i|X_i, a_i = a_0 + \Delta]}{\lim_{\Delta\uparrow 0} E[D_i|X_i, a_i = a_0 + \Delta] - \lim_{\Delta\downarrow 0} E[D_i|X_i, a_i = a_0 + \Delta]} \tag{5}$$

We estimated $\hat{\tau}_w$ nonparametrically by estimating both the numerator and the denominator, separately, calculating the ratio, and bootstrapping the process. As we will see, below, a sizeable proportion of those eligible for free public health care are reported to have paid for the care that they received in a public health facility. In order to interpret LATE (Imbens and Angrist, 1994) in this setting, we assume that actual reported payment status – whether or not the health care received required payment – is representative (even if noisy) of what caregivers expected, when they made their decision to seek public healthcare for the child. It is, thus, the increase in utilisation for those who expect to benefit from user fee abolition (*ie* from not paying). In other words, the estimate is expected to be negative, because fewer young children should be paying, while more children should be accessing public care.

The majority of RD examples, such as Hahn et al. (2001), Lee (2008) and Card et al. (2008) use data that is nearly continuous with respect to the running variable. However, in this design, as with the design considered by Duflo (2003), the running variable, age, is discrete, and was initially available only in years. Fortunately, birth date data is also available, allowing for an analysis in days, weeks and months, as well as years. Unfortunately, there are two problems. The first is that there are too few observations to take advantage of the days or weeks data, and, therefore, the RD results presented here focus on the months data. The second, and more concerning problem, is that the birth dates are only available for children living with mothers who are still alive. In an earlier version of this research, available from the author, an analysis based on the full sample of children, but only estimated across age in years was undertaken. Although some precision is lost from the slightly reduced sample used here, the results are not qualitatively different from what is reported here. Therefore, the analysis continues with its focus on monthly measures of age.

The structure of the analysis in the next few sections is quite similar. Due to Gel-

9

man and Imbens (2014), we consider only linear and quadratic polynomials for g and h. However, we also undertake nonparametric regression following Li and Racine (2007), and illustrate the estimates for reference. Thus, we esimate and present nonparametric RD effects, including an estimate of the fuzzy policy effect, linear RD effects and quadratic RD effects. All estimates are undertaken using using R (R Core Team, 2016), while nonparametric estimates use Hayfield and Racine's (2008) np package in R.[6] Even though cross-validation estimates an optimal bandwidth for the nonparametric regression model, we report our results for various age windows. Empirically, nonparametric methods are becoming more common in the RD literature, partly due to Imbens and Kalyanaraman's (2012) bandwidth selection approach. Finally, since the assumption contained in (2) requires the effect of control variables, other than age, to be constant across the threshold, Lee and Lemieux (2010), we undertook a test of running variable manipulation suggested by McCrary (2008), as well as examining other household variables for discontinuities.

3 The Data

Data for the analysis was sourced from the South African October Household Survey (OHS) of 1995. The main purpose of the OHS, Statistics South Africa (1995), was to collect information on households and individuals across the nine provinces of South Africa. The survey included questions related to dwellings and dwelling services, perceived quality of life, socio-demographic information, employment, informal labour market activity and formal labour market activity, as well as births and deaths in the household. Along with this information, there is a short series of questions related to illness, injury, healthcare-seeking behaviour and access to medical aid. The survey, which follows a stratified random sampling method, explicitly stratified by province, magisterial district, urban or rural locale and population group, included responses from 121 538 individuals living in 29 700 households. These enumeration areas were selected systematically based on probabilities proportional to their size, where the size was estimated

---

[6]We use least squares cross-validation to determine the optimal bandwidth in each (local linear) nonparametric regression model assuming an Epanechnikov kernel (see Li and Racine, 2007, for further details).

from the 1991 population census. Within a selected enumeration area, ten households were drawn for interview. Post-stratified weights are available, but are not used in the analysis, due to the fact that the analysis generally focuses on children aged 14 and under who have been reported as ill in the last 30 days. The weights are not calibrated for a sub-sample of this nature, and, therefore, the weights are not likely to lead to true population estimates.

A series of different sections in the survey cover a variety of different topics, and it is possible to merge the relevant information to create data at the child level. For this analysis, data for each child was taken from the individual questionnaire, including information on household head, which was merged back into the child data set. Finally, specific child birth dates were also merged into the data. This last source of information was the only way to capture age more finely than in years.

For the analysis, a series of binary indicators have been developed. One indicates whether or not the caregiver of an ill uninsured child is reported to have had to pay for either the service or for medicine, if the child was treated in the public sector. This indicator underpins the local estimate of the policy effect, because it provides some evidence regarding policy implementation. If there is no evidence of implementation, there is little need for additional analysis. Another indicates whether or not an ill child received their healthcare from the public sector, and is our primary outcome of interest. We focus our attention on uninsured children (aged up to 14 years), who have been reported as either ill or injured in the 30 days prior to the survey. Focussing our efforts on this group provides the evidence-base for policy efficacy.[7]

In addition to the aforementioned outcome variables, a number of other variables are available in or can be created from the survey. A number of these were used to further check the adequacy of the RD assumptions. Because health, healthcare-seeking and insurance coverage are likely to be affected by household income, the size of the household, the education level of household members, employment status of household members, and the age of household members (the latter three proxied through the head

---

[7]As part of the analysis, the exogeneity of this indicator, relative to the policy, was examined. Further, since the policy change was meant only to help the uninsured, an indicator of insurance (medical aid) coverage is created, and its exogeneity with respect to policy was also examined. The results, not reported, but available upon request, are indicitive of exogeneity.

of the household), it is important for those measures to also be exogenous to the policy. In other words, it must be the case that different types of households are more likely to be found either above or below the policy threshold. Similarly, since distance to the health facility is an important proxy of the opportunity cost of health-seeking behaviour, and, more importantly, there is evidence that additional public clinics came on-line between 1994 and 1996, Benatar (1997), we want to examine whether there are differences in access, as measured by distance.

The primary results of those tests are easily summarized in Table 1, which also includes the number of observations associated with each analysis sample, the mean for younger children (those aged up to but not including six years), the mean for older children (aged from six years to 14 years), the mean difference, and the probability this difference is zero (derived from Welch's t-test). As can be seen from the table, with the exception of household head age, there are no statistically significant mean differences. However, to be certain that there are no disctontinuities, all of the variables were subjected to further tests. First, we performed a running variable manipulation test, as suggested by McCrary (2008), which finds no discontinuity in the running variable. Further, we looked for breaks in the variables reported in Table 1 using (3) and (4), where the outcome of interest was replaced by one of the variables from Table 1. All of those tests, available upon request, support the hypothesis that our main results are robust; we do not find any discontinuities in any other variable that could be affecting our main policy estimate.

Table 1 about here.

4 Policy Impacts Amongst Uninsured Children

The analysis begins with policy implementation, i.e., were younger (uninsured) children less likely to pay for public care than older children? After establishing that payment was less likely amongst the eligible, as it should be if the policy was implemented (even if imperfectly), the focus turns to public healthcare usage by uninsured ill children, and differences in behaviour based on location.

4.1 Paying for Services

The survey response underpinning the payment analysis is a binary indication of payment for either the service or for medicine (=1, if payment was made). A separate question for medical service and medicine would have been preferred, since individuals may have had to buy additional medicines, later, which could have influenced the answer. Similarly, it is possible that responses could include informal payments, such as bribes, or indirect payments, such as travel costs. Furthermore, the payment data that is available is *ex post*, rather than an *ex ante* expectation, i.e., respondents were not asked if they expected to be required to pay for care. However, finding a lower proportion of the eligible paying, relative to the ineligible, even in the face of a less than perfect survey instrument, does support the hypothesis that user fee abolition was implemented and enforced, although the resulting estimate itself is not likely to be a true reflection of implementation. As noted previously, in order for the local estimates to be interpretable in this context, we must assume that the *ex post* data is a fair reflection of *ex ante* expectations.

Figure 2 about here

Figure 2 contains the nonparametic fits (on either side of the age threshold) of the (proportion) of ill children in each age-month, who were both uninsured and did not report making a payment when using the public sector for the care that they received. The illustration suggests a rather large impact, which can be viewed as a measure of implementation. A more careful analysis suggests relatively good implementation, although more precision in the estimates would be welcome. The effect of removing user fees is found to reduce the probability of payment amongst age-eligible (and uninsured) children by between 9% and 41% (ignoring confidence bands). Thus, there is evidence across the age windows that payment for treatment at public health facilities was less likely to have been made for the treatment of an age-eligible uninsured child than for the treatment of an age-ineligible uninsured child.

Table 2 about here

13

4.2   Effects on Public Health Care Usage Decisions

Although a reduction in payment is observed in the data, and since the policy was formulated to improve access to public facilities for uninsured young children, amongst others, RD is applied, as above, to examine the effect of fee abolition on healthcare-seeking in the public sector. The outcome of interest is based on a binary indication of treatment in a public facility (=1) or not; thus, the results presented in this subsection relate to public healthcare usage probabilities, especially their differences at the policy cut-off for uninsured children, as well as local differences.

<div align="center">Figure 3 about here</div>

Although Figure 1 suggests a near doubling of outpatient days in Soweto, following the removal of user fees, the fitted nonparametric regression illustrated in Figure 3 is more circumspect, suggesting that the policy change did not have an impact on curative public healthcare-seeking. Even though a discontinuity is visible, it may not be statistically significant. We investigate its statistical significance through the paramater, τ, in the linear and quadratic models and via bootstrapping wihtin each age window for the nonparametric models. Within each window, the nonparametric regressions are first estimated, and a difference at the cut-off is calculated. Then, the data is sampled, with replacement, the nonparametric regression is refit, and a new difference is calculated. Resampling is repeated until there are 299 values, from which we calculate a mean and standard deviation, which represent the nonparametric RD estimate and the confidence that can be placed on that estimate. Estimates across the different age bands are available in Table 3. Although statistical significance is not a feature of the results reported in the table, the estimates suggest an increase in public healthcare-seeking behaviour amongst age-eligible uninsured children.[8] Focusing only on the nonparametric estimates, for children in the five to six years age window, the average effect is an increase of about 11%.

<div align="center">Table 3 about here</div>

---

[8]It should also be noted that an attempt was made to uncover whether or not insured children were able to take advantage of the policy. The RD estimates, available upon request, find no support for that hypothesis.

If one were to assume perfect foresight amongst child caregivers, i.e., that when they expected not to have to make payments for public care, they did not, in fact, make any payments for public healthcare, a Wald esimate of the policy impact is avaialble; however, given the imprecision of the two components, we do not expect it to be precisely estimated. However, the local effects are consistently negative, as expected, due to the inverse relationship expected between price and quantity. Returning to the five to six years age window, the local effect is estimated nearer 42%. In terms of interpretation, recalling the perfect foresight caveat, this Wald estimate is a Local Average Treatment Effect (Imbens and Angrist, 1994); thus, for those expecting the policy to be appropriately implemented, the policy increased the probability of using public health care facilities by nearly two-fifths.

4.3 Heterogeneous Effects?

Although the previous results suggest imprecise, along with possibly small, policy impacts, it is possible that averaging the impact across all children misses important differences. Thus, the analysis was extended to consider possible heterogeneous effects. There are a number of ways heterogeneities might arise; however, Tanaka (2014) suggests that policy impacts are differentiated by location. Therefore, we separate the data by those who live relatively close to a facility, within five kilometers, and those who live farther away than that.[9] The survey data is somewhat limited, having only three responses (within one km, between one and five kms and beyond five kms); however, this distance provides a useful measure of broader accessibility. We initially expected to find that those closest to facilities were most likely to use them; however, we do not find any support for that conclusion. Instead, the evidence leans the other direction.

Figure 4 about here

As before, we present an illustration of the nonparametric regression of public healthcare usage amongst uninsured ill children. Initially, we focus on those children living

---

[9]As one reviewer suggested, this could, instead, be framed as an urban/rural question. We explored that possibility, results available upon request, and found results and sample splits that were similar (although not identical) to what is reported here.

within five kms. We then look at similarly delimited children, except that they live more than five kms from the facility. See Figures 4 and 5. The difference between the figures is striking. The discontinuity is only visible for children living farther away from facilities.

Figure 5 about here

To be sure, a more formal empirical analysis was undertaken. The RD results for the children located closer to the facilities are presented in Table 4. As can be seen in the table, there is little in the way of consistent results across the specifications. However, if one is willing to limit their focus to the five to six year window, the nonparametric estimate suggests a 1% increase in public healthcare utilisation. However, the Wald estimate of 30% is in line with what we reported for all young children in the previous subsection.

Table 4 about here

For those living farther away from public healthcare facilities, the estimates are more consistent. Although not precise enough to conclude the the policy increased utilisation amongst those living more than five kilometers from a health facility, all estimates suggest that is the case, and they range from 11% to 62%.[10] If one is willing to make the assumption that the caregivers making healthcare-seeking decisions for these children are making them based on payment expectations, and that realisations are appropriate proxies for expectations, then, for those children expected to benefit (not pay), the increased utilisation exceeds 70%, using only the five-six year old window, although precision remains problematic.

Table 5 about here

---

[10]Quadratic polynomials, given so few observations, however, would also be subject to the criticism leveled by Gelman and Imbens (2014) and Gelman and Zelizer (2015), and therefore, should be treated sceptically.

5 Conclusion

Economic intuition suggests that the abolition of public healthcare user fees should lead to increased demand for public healthcare. Most of the research related to user fee policies in Africa supports that intuition. To some degree, our analysis also supports that conclusion, although our results are statistically insignificant. However, a number of previous studies of this or related effects, in Africa, have not applied quasi-experimental methods, or been able to take advantage of the quasi-experimental environment made possible by the policy revision (Lagarde and Palmer, 2008). Furthermore, many African studies have been limited to clinic registration data, and typically for a very limited number of clinics. In South Africa, that has been especially true. Bayat and Cleaton-Jones (2003) examine 10 dental clinics, and McCoy and Khosa (1996) examine three or four health clinics, while Wilkinson et al. (1997) and Wilkinson et al. (2001) consider only one clinic.

Given those limitations, the preceding research adds to the literature in a number of ways. It is one of very few African studies making use of household information to examine the effects of user fee abolition. The research applies appropriate empirical methods in this quasi-experimental setting, and those methods allow for the analysis of both average policy impacts and local average policy impacts, neither of which have received attention in the South African literature. The study also highlights the benefits of Koch and Racine's (forthcoming) nonparametric estimation method, which allowed them to take advantage of more data than we could, resulting in more precision, although they were not in a position to consider local estimates. As such, our analysis extends the information that is available with respect to the impact of the 1994 South African user fee abolition policy at a national level.

Although this analysis provides a number of insights into the limited effects derived from free primary healthcare on households, further questions remain. In particular, the policy may have affected household welfare; in fact, part of the reason for the abolition of user fees was to improve household welfare. Therefore, analysing whether or not there was a pro-poor bias in terms of health related expenditures deserves attention, but requires more detailed expenditure data. Furthermore, the policy was enacted in

17

1994, over 20 years ago. Although only imprecise and relatively small policy impacts were identified in this analysis, it is plausible that the benefits of user fee abolition are long-term, rather than immediate. In that regard, considering the effect of the policy on the health of the population – extending the work of Tanaka (2014) – or education completion, in future years would shed light on the broader benefits of user fee abolition.

References

Ashraf, N., Berry, J. and Shapiro, J. M. (2010), 'Can higher prices stimulate product use? evidence from a field experiment in Zambia', *American Economic Review* **100**(5), 2383–2413.

Bayat, A. and Cleaton-Jones, P. (2003), 'Dental clinic attendance in Soweto, South Africa, before and after the introduction of free primary dental services', *Community Dentistry and Oral Epidemiology* **31**(2), 105–110.

Benatar, S. R. (1997), 'Health reform in the new South Africa', *The New England Journal of Medicine* **336**(12), 891–895.

Biritwum, R. B. (1994), 'The cost of sustaining Ghana's "cash and carry" system of health care financing at a rural health center', *West African Journal of Medicine* **13**(2), 124–127.

Blas, E. and Limbambala, M. E. (2001), 'User-payment, decentralization and health service utilization in Zambia', *Health Policy and Planning* **16**(Supplement 2), 19–28.

Brink, A. S. and Koch, S. F. (2015), 'User fee abolition in South Africa in 1994 and 1996: Difference-in-differences', *Development Southern Africa* **32**(2), 170–192.

Burnham, G. M., Pariyo, G., Galwango, E. et al. (2004), 'Discontinuation of cost sharing in Uganda', *Bulletin of the World Health Organization* **82**(3), 187–195.

Card, D., Dobkin, C. and Maestas, N. (2008), 'The impact of nearly universal insurance coverage on health care: evidence from Medicare', *American Economic Review* **98**(5), 2242–2258.

Chama-Chiliba, C. M. and Koch, S. F. (2014), 'Assessing regional variations in the effect of the removal of user fees on institutional deliveries in rural Zambia', Economic Research Southern Africa Working Paper No. 427.

Cohen, J. and Dupas, P. (2010), 'Free distribution or cost-sharing? evidence from a randomized malaria prevention experiment', *Quarterly Journal of Economics* **125**(1), 1–45.

Coovadia, H., Jewkes, R., Barron, P. et al. (2009), 'The health and health system of South Africa: historical roots of current public health challenges', *The Lancet* **374**(9692), 817–834.

Deininger, K. and Mpuga, P. (2005), 'Economic and welfare impact of the abolition of user fees: evidence from Uganda', *Journal of African Economies* **14**(1), 55–91.

Duflo, E. (2003), 'Grandmothers and granddaughters: old-age pension and intrahousehold allocation in South Africa', *World Bank Economic Review* **17**(7), 1–25.

Fuchs, V. (1968), 'The growing demand for medical care', *The New England Journal of Medicine* **279**(192), 190–195.

Gelman, A. and Imbens, G. W. (2014), 'Why higher-order polynomials should not be used in regression discontinuity designs', NBER Working Paper No 20405.

Gelman, A. and Zelizer, A. (2015), 'Evidence on the deleterious impact of sustained use of polynomial regression on causal inference', *Research & Politics* (January-March), 1–7.
URL: *http://dx.doi.org/10.1177/2053168015569830*

Gertler, P., Locay, L. and Sanderson, W. (1987), 'Are user fees regressive? The welfare implications of health care financing proposals in Peru', *Journal of Econometrics* **36**(1-2), 67–88.

Grossman, M. (1999), 'The human capital model and the demand for health', NBER Working Paper no. 7078.

Hahn, J., Todd, P. and van der Klaauw, W. (2001), 'Identification and estimation of treatment effects with a regression disontinuity design', *Econometrica* **69**(1), 201–209.

Hayfield, T. and Racine, J. S. (2008), 'Nonparametric econometrics: The np package', *Journal of Statistical Software* **27**(5).
URL: *http://www.jstatsoft.org/v27/i05/*

Heller, P. S. (1982), 'A model of the demand for medical and health services in peninsular Malaysia', *Social Science & Medicine* **16**(3), 267–284.

Imbens, G. W. and Angrist, J. D. (1994), 'Identification and estimation of local average treatment effects', *Econometrica* **62**, 467–476.

Imbens, G. W. and Kalyanaraman, K. (2012), 'Optimal bandwidth choice for the regression discontinuity estimator', *Review of Economic Studies* **79**, 933–959.

Koch, S. F. and Racine, J. S. (forthcoming), 'Healthcare facility choice and user fee abolition: Regression discontinuity in a multinomial choice setting', *Journal of the Royal Statistical Society, Series A* .
URL: *http://dx.doi.org/10.1111/rssa.12161*

Lagarde, M. and Palmer, N. (2008), 'The impact of user fees on health service utilization in low- and middle-income countries: how strong is the evidence?', *Bulletin of the World Health Organization* **86**(11), 839–848.

Lee, D. S. (2008), 'Randomized experiments from non-random selection in U.S. House elections', *Journal of Econometrics* **142**(2), 675–697.

Lee, D. S. and Lemieux, T. (2010), 'Regression discontinuity designs in Economics', *Journal of Economic Literature* **48**(2), 271–355.

Li, Q. and Racine, J. S. (2007), *Nonparametric Econometrics*, Princetone University Press, Princeton.

Masiye, F., Chitah, B. M., Chanda, P. et al. (2008), 'Removal of user fees at primary health care facilities in Zambia: a study of the effects on utilization and quality of care', Regional Network for Equity in Health in east and southern Africa Discussion Paper no. 57.

McCoy, D. and Khosa, S. (1996), "Free health care" policies, *in* D. Harrison, P. Barron and J. Edwards, eds, 'South African Health Review 1996', Health Systems Trust and Henry J. Kaiser Family Foundation, pp. 157–165.

McCrary, J. (2008), 'Manipulation of the running variable in the regression discontinuity design: A density test', *Journal of Econometrics* **142**, 698–714.

McIntyre, D., Bloom, G. H., Doherty, J. et al. (1995), *Health Expenditure and Finance in South Africa*, Health Systems Trust and the World Bank, Durban, South Africa and Washington, DC, USA.

McKelvey, C., Thomas, D. and Frankenberg, E. (2012), 'Fertility regulation in an economic crisis', *Economic Development and Cultural Change* **61**(1), 7–38.

McKinnon, B., Harper, S., Kaufman, J. S. and Bergevin, Y. (2015), 'Removing user fees for facility-based delivery services: A difference-in-difference evaluation from ten sub-Saharan countries', *Health Policy and Planning* **30**(4), 432–441.

Miguel, E. and Kremer, M. (2004), 'Worms: Identifying impacts on education and health in the presence of treatment externalities', *Econometrica* **72**(1), 159–217.

Mwabu, G. and Wang'ombe, J. (1997), 'Health service pricing reforms in Kenya', *International Journal of Social Economics* **24**(1), 282–293.

Nabyonga, J., Desmet, M., Karamagi, H. et al. (2005), 'Abolition of cost-sharing is pro-poor: evidence from Uganda', *Health Policy and Planning* **20**(2), 100–108.

R Core Team (2016), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL: *http://www.R-project.org/*

Ridde, V. and Morestin, F. (2011), 'A scoping review of the literature on the abolition of user fees in health care services in Africa', *Health Policy and Planning* **26**(1), 1–11.

Statistics South Africa (1995), *October Household Survey (South Africa), 1995*, Statistics South Africa (producer) and South African Data Archive (distributor), Pretoria.

Tanaka, S. (2014), 'Does abolishing user fees lead to improved health status? Evidence from post-apartheid South Africa', *American Economic Journal: Economic Policy* **6**(3), 282–312.

Walker, L. and Gilson, L. (2004), '"We are bitter, but we are satisfied": nurses as street-level bureaucrats in South Africa', *Social Science & Medicine* **59**(6), 1251–1261.

Wilkinson, D., Gouws, E., Sach, M. et al. (2001), 'Effect of removing user fees on attendance for curative and preventive primary health care services in rural South Africa', *Bulletin of the World Health Organization* **79**(7), 665–671.

Wilkinson, D., Sach, M. and Abdool Karim, S. (1997), 'Examination of attendance patterns before and after introduction of South Africa's policy of free health care for children aged under 6 years and pregnant women', *British Medical Journal* **314**(7085), 940–941.

Willis, C. Y. and Leighton, C. (1995), 'Protecting the poor under cost recovery: the role of means testing', *Health Policy and Planning* **10**(3), 241–256.

Table 1: The Balance Between Younger and Older Children

	Younger Children	Older Children	Difference Estimate	Test[a] Pr(> |t|)
Facility Distance<1km[i]	0.224	0.209	0.015	0.458
1km<Facility Distance<5km[i]	0.356	0.357	-0.001	0.965
Household Income[j]	571.9	577.3	-5.4	0.917
Household Size[j]	6.281	6.267	0.014	0.908
Household Head Age[j]	44.258	45.789	-1.531	0.018
Household Head Education[j,k]	4.696	4.625	0.071	0.688
Household Head Non-LFP[i]	0.286	0.274	0.011	0.595
Observations	973	802		

Descriptive statistics of data related to uninsured children, taken from the 1995 South African October Household Survey. $a-$ Probability derived from Welch's t-test of the null hypothesis of zero mean difference. [i]– Variable is binary. [j]– Variable is discretely continuous or continuous. [k] - Education in years.
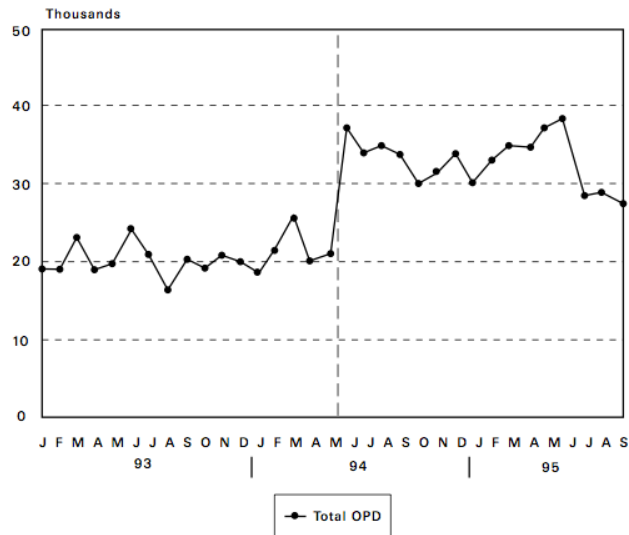

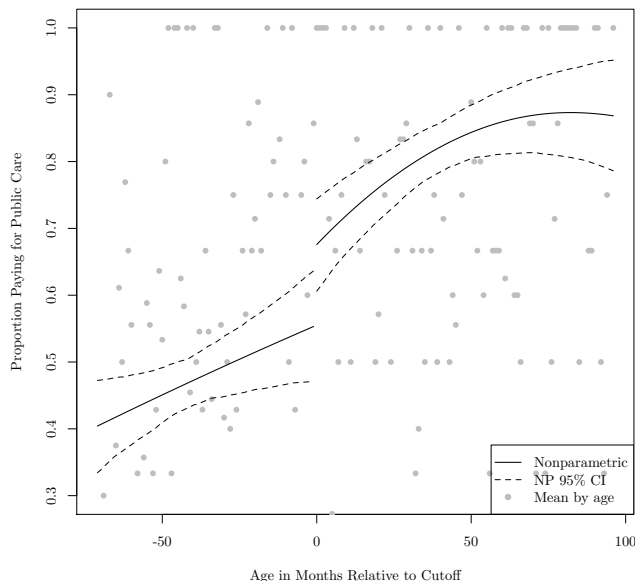Figure 1: Paediatric Attendance in Soweto Clinics (McCoy & Khosa, 1996).

Figure 2: Uninsured Ill Children Paying for Public Care. The figure illustrates estimates and 95% confidence intervals from a nonparametric regression fit of month of age on the proportion of uninsured ill children reported to have paid for public healthcare. Estimation performed separately on either side of the recentred cut-off (six year olds are illustrated at the age of 0 months). See (4) and discussion around it.

Table 2: Uninsured Children Paying for Public Care: Did the Policy Affect Payment at Point of Service?

Age Window	N	Linear	Quadratic	Nonparametric
0-14 Years	1003	-0.1698	-0.1703	-0.1582
		(0.059)	(0.088)	(0.057)
3-10 Years	527	-0.164	-0.1214	-0.2753
		(0.078)	(0.116)	(0.171)
4-7 Years	317	-0.0915	-0.2323	-0.2854
		(0.102)	(0.154)	(0.193)
5-6 Years	191	-0.1819	-0.406	-0.3856
		(0.141)	(0.228)	(0.23)

Regression discontinuity estimates from (3) and (4), reporting only $\hat{\tau}$ and either its standard error (in parentheses) for linear and quadratic models or its standard deviation (in parentheses) for the nonparametric model; the latter is based on 299 bootstrap replications using the sample analog of (2). All models estimated separately in each age window.
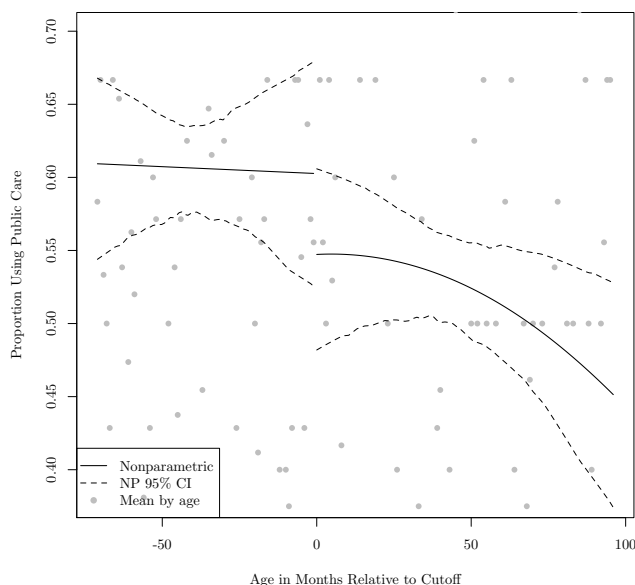
Figure 3: Uninsured Ill Children Making Use of Public Care. The figure illustrates estimates and 95% confidence intervals from a nonparametric regression fit of month of age on the proportion of uninsured ill children reported to have paid for public healthcare. Estimation performed separately on either side of the recentred cut-off (six year olds are illustrated at the age of 0 months). See (4) and discussion around it.

Table 3: Regression Discontinuity Estimates for Usage of Public Facilities by Uninsured Ill Children

Age Window	N	Linear Polynomial	Quadratic Polynomial	Nonparametric Sharp	Fuzzy
0-14 Years	1775	0.0288	0.075	0.0438	-0.3903
		(0.049)	(0.072)	(0.047)	(0.893)
3-10 Years	914	0.0265	0.0603	0.0352	-0.0843
		(0.066)	(0.099)	(0.064)	(1.539)
4-7 Years	563	0.0345	0.1197	0.0342	-0.1178
		(0.084)	(0.129)	(0.075)	(1.033)
5-6 Years	340	0.0263	0.3661	0.1103	-0.4179
		(0.119)	(0.195)	(0.101)	(2.257)

Regression discontinuity estimates from (3) reporting $\hat{\tau}$ and either its standard error (in parentheses) for linear and quadratic models. Nonparametric estimates include both $\hat{\tau}$ and $\hat{\tau}_w$, and their standard deviation (in parentheses); the last are based on 299 bootstrap replications using the sample analog of (2) and (5), respectively. All models estimated separately in each age window.
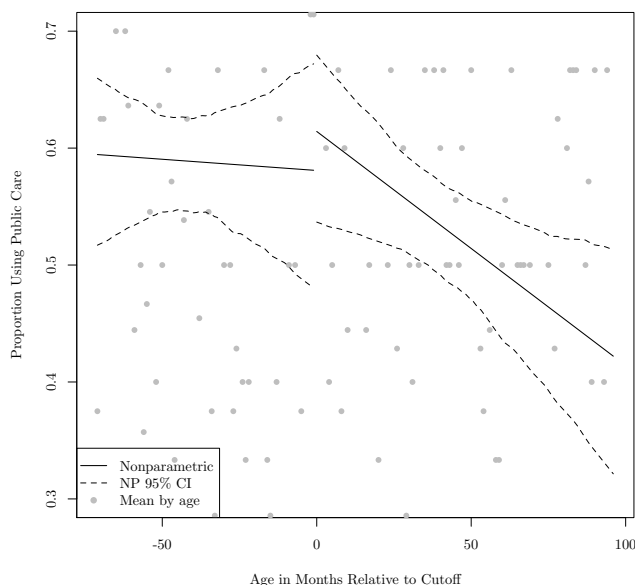
Figure 4: Uninsured Ill Children Living Within 5 Km of a Public Facility and Making Use of Public Care. The figure illustrates estimates and 95% confidence intervals from a nonparametric regression fit of month of age on the proportion of uninsured ill children reported to have paid for public healthcare, focusing on those living relatively close to facilities. Estimation performed separately on either side of the recentred cut-off (six year olds are illustrated at the age of 0 months). See (4) and discussion around it.

Table 4: Regression Discontinuity Estimates for Usage of Public Facilities by Uninsured Ill Children Living Within 5 Km of Facility

Age Window	N	Linear Polynomial	Quadratic Polynomial	Nonparametric Sharp	Fuzzy
0-14 Years	1018	-0.037	-0.0332	-0.0297	0.2077
		(0.066)	(0.097)	(0.062)	(1.017)
3-10 Years	500	-0.0435	0.0071	-0.009	0.1743
		(0.089)	(0.134)	(0.097)	(2.423)
4-7 Years	309	-0.0629	0.1188	0.0012	0.0144
		(0.113)	(0.17)	(0.114)	(2.888)
5-6 Years	185	0.0006	0.1968	0.0114	-0.3058
		(0.161)	(0.25)	(0.129)	(2.475)

Regression discontinuity estimates from (3) reporting $\hat{\tau}$ and either its standard error (in parentheses) for linear and quadratic models. Nonparametric estimates include both $\hat{\tau}$ and $\hat{\tau}_w$, and their standard deviation (in parentheses); the last are based on 299 bootstrap replications using the sample analog of (2) and (5), respectively. All models estimated separately in each age window.
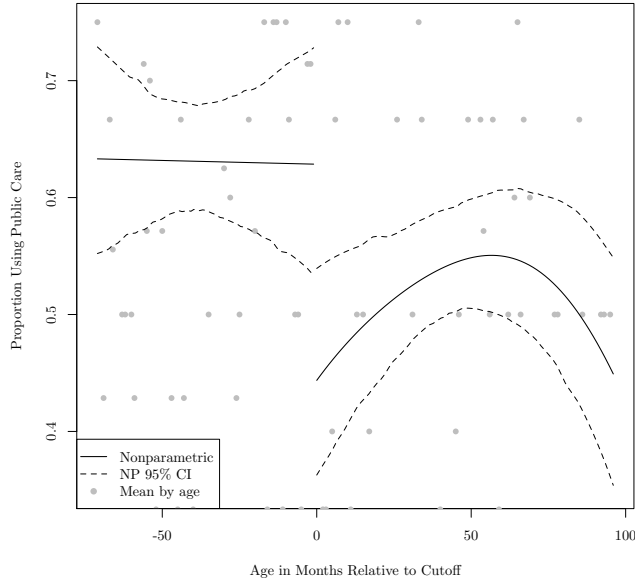
Figure 5: Uninsured Ill Children Living Farther than 5 Km of Public Facility and Making Use of Public Care. The figure illustrates estimates and 95% confidence intervals from a nonparametric regression fit of month of age on the proportion of uninsured ill children reported to have paid for public healthcare, focusing on those living farther away from facilities. Estimation performed separately on either side of the recentred cut-off (six year olds are illustrated at the age of 0 months). See (4) and discussion around it.

Table 5: Regression Discontinuity Estimates for Usage of Public Facilities by Uninsured Ill Children Living Outside 5 Km of Facility

Age Window	N	Linear Polynomial	Quadratic Polynomial	Nonparametric Sharp	Fuzzy
0-14 Years	757	0.1131	0.2097	0.15	-16.1424
		(0.073)	(0.108)	(0.072)	(255.943)
3-10 Years	414	0.1122	0.146	0.1179	-0.756
		(0.098)	(0.148)	(0.098)	(7.398)
4-7 Years	254	0.1571	0.112	0.1469	-0.3819
		(0.125)	(0.197)	(0.112)	(3.092)
5-6 Years	155	0.0757	0.6235	0.2497	-0.9218
		(0.179)	(0.319)	(0.19)	(8.593)

Regression discontinuity estimates from (3) reporting $\hat{\tau}$ and either its standard error (in parentheses) for linear and quadratic models. Nonparametric estimates include both $\hat{\tau}$ and $\hat{\tau}_w$, and their standard deviation (in parentheses); the last are based on 299 bootstrap replications using the sample analog of (2) and (5), respectively. All models estimated separately in each age window.