Did Removing User Fees Improve Access To Maternal And Neonatal Health Care In Zambia? A Difference-In-Difference Study.

Author: Laura Sochas

Word count: 4,365

DO NOT CITE WITHOUT PERMISSION

Introduction

The removal of user fees charged for accessing health services is a popular intervention to improve access. Introduced in the 1980s and 1990s across many developing countries in the context of structural adjustment programmes to supplement scarce health financing resources, some countries subsequently dismantled user fees regimes starting in the 2000s.

The effect of removing user fees on access to healthcare has been widely debated because of the fact that the policy might result in fewer financial resources available at the facility level, as well as simultaneous increases in utilisation. This, in turn, might lead to decreasing quality of care, which may then reverse gains in utilisation. Furthermore, other barriers to care, such as distance, costs of access beyond user fees, or disrespectful care, remain unaffected by the removal of user fees and it is uncertain whether the removal of one barrier (the cost of care itself) is sufficient to improve access.

Out of all health areas where user fees may have an impact, maternal and neonatal health is of particular interest given that it is the health-related Millennium Development Goal where the world made the least amount of progress over the past 15 years. Furthermore, it is arguable that maternal health care suffers from a greater range of concurrent barriers than other health services: the need for care in childbirth is highly time-sensitive, lending more importance to distance barriers; the social and cultural significance of childbirth may result in a greater number of decision-makers around whether or not to seek care; childbirth is often seen by women and their families as a condition which might require medical assistance rather than as an illness which systematically requires it; and women seeking maternal health care may be at greater risk of receiving disrespectful care from health workers. While the removal of user fees addresses some of the financial barriers to access, women and children in need of maternal and neonatal health care may be facing a whole host of other barriers that remain unaddressed by this policy.

While there are many studies on the effect of user fee removal, recent reviews on this topic note that very few use quasi-experimental designs. Very recently, two studies have used a difference-in-difference design at the multi-country level to address these gaps in the user fees literature for maternal and neonatal health. McKinnon et al use three treatment countries (Ghana, Kenya and Senegal) and seven control countries in Sub-Saharan Africa, while Leone et al use two treatment countries (Burkina Faso and Ghana) and three control countries. Both find a positive and significant effect of the removal of user fees on levels of facility delivery. Leone et al find a positive effect on the use of C-sections while McKinnon et al do not.

---

*Mphil/PhD Candidate in Demography, London School of Economics and Political Science.*
Difference-in-difference designs at the multi-country level, however, suffer from the problematic assumption that trends in facility delivery in the treatment and control countries would have remained similar in the absence of the policy change. Countries in different parts of Sub-Saharan Africa are subject to highly differentiated socio-economic shocks that might have affected both the coverage of facility delivery and the government’s ability to afford the removal of user fees.

This study therefore focuses on evaluating the removal of health user fees in a single country, Zambia, which removed health user fees in 2006. Today, Zambia enjoys better maternal and neonatal health access and outcomes than the Sub-Saharan African average, but its situation in 2006 (e.g.: neonatal mortality rate of 27.1) was very similar to the Sub-Saharan African average today (28.6). Zambia is therefore a useful test case for other countries in the region considering such a policy change.

Using routine health systems data from the Zambian Health Management Information System, Masiye, Chitah, & McIntyre (2010) and Lagarde, Barroy, & Palmer (2012) respectively show that there was a 55% and 32% increase in the utilisation of health services in rural areas after the policy change, for adults and children over five. However these studies use a before-after comparison and a segmented linear regression respectively, which are not able to identify a robust causal effect. For example, a before-after comparison would be unable to distinguish between a statistically unusual improvement and an upward trend that started before the policy change, while the segmented linear regression is vulnerable to interpreting any positive shock, affecting both treated and control groups, as a causal effect. Furthermore, they both use Health Management Information Systems data, which is routinely reported by clinics themselves and is of lower quality than household surveys.

More recently, Lepine, Lagarde, & LeNestour (2015) used data from a series of household surveys on general living standards and a synthetic control method, which does permit causal inference. According to their results, the policy had no impact on general healthcare utilisation for adults. The literature is therefore split over whether the policy was successful in improving general health service utilisation, with more robust studies finding no effect. There is also a dearth of research about the impact of the policy on access to maternal and neonatal health care specifically.

Data and methods

Zambia introduced user fees for health services accessed by working-age adults and children over five years old in the early 1990s. Although pregnant women were supposedly exempted from user fees prior to the policy change, many sources report that fees were widely charged for facility delivery. A survey of 332 women in 1998-2000 in Kalabo district found that fees for facility delivery ranged between 3,000 Kz ($1) for hospitals and 800 Kz ($0.30) in rural health centres. The same study reported that 59% of respondents said that the delivery fees were not affordable. Ten years later, a study of 20 facilities across five districts in 2008 reported that in urban centres, where fees were not yet abolished by 2008, delivery bills ranged from 10,000 Kz ($3.6) to 30,000 Kz ($11).

All user fees were theoretically abolished in April 2006 in 54 districts administratively classified as “rural”, but not in the 18 districts administratively classified as “urban”. This administrative classification does not perfectly predict actual type of residence: there are towns in “treated/rural” districts and countryside areas in “control/urban” districts. The types of residence that respondents actually live in are divided into four categories by the Demographic Health Survey: large city, small
city, town and countryside. Geographic analysis confirms that “control/urban” and “treated/rural”
districts include DHS sampling clusters belonging to all four residence types (see Figure 2 and Table 1),
although the distribution of residence types varies across treated and control districts. In 2007,
user fees were also removed from districts administratively classified as urban, except for municipal
and city centres, where user fees were only removed in 2012.25

Figure 1: Map of Zambia, districts and DHS sampling clusters by type of residence

![Map of Zambia](image)

Table 1: Sample sizes by residence category and treatment status

<table>
<thead>
<tr>
<th>Type of residence</th>
<th>Control districts</th>
<th>Treated districts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capital city</td>
<td>103</td>
<td>39</td>
</tr>
<tr>
<td>Small city</td>
<td>128</td>
<td>12</td>
</tr>
<tr>
<td>Town</td>
<td>308</td>
<td>230</td>
</tr>
<tr>
<td>Countryside</td>
<td>549</td>
<td>1,757</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1,088</strong></td>
<td><strong>2,038</strong></td>
</tr>
</tbody>
</table>

*Sample: Births occurred where woman is living now, April to December, 2002-2006, information about facility delivery is not missing.*

This study is the first to use the Zambia 2007 Demographic Health Survey (DHS) to investigate the
impact of removing user fees. The DHS is a nationally and provincially representative survey that uses
stratified random sampling and allows the analysis of health care access in the field of maternal and
neonatal health specifically. Because female respondents are asked about where they gave birth and
about the baby’s survival for births occurring up to five years prior to the interview, it is possible to
construct yearly variables between 2002 and 2007 (see Figure 2).
Although the DHS does not indicate the respondent’s district of residence, each of the 319 DHS sampling clusters’ GPS location is provided in a separate dataset. I overlaid the GIS location of the clusters on a map of Zambian districts using the software ArcGIS, which allowed for the treatment assignment of each respondent (see Figure 2). The GPS location of sampling clusters is intentionally “displaced” in the dataset for confidentiality reasons, with 0-2 km of error for city and town clusters and 0-5 km of error for countryside clusters. Two districts, Hezhi-Tezhi (treated) and Kalulushi (control), did not contain any DHS sampling clusters and are therefore not included in our analysis.

A difference-in-difference (DiD) strategy was used to exploit the fact that user fees were removed from districts administratively classified as “rural” before they were removed from districts administratively classified as “urban”. This approach estimates the difference between the before-after change in the treated/rural group versus the before-after change in the control/urban group. The key identifying assumption is that treated and control districts would have continued to experience similar trends to each other in the absence of the policy change.

This design allows one to make stronger causal claims about the impact of removing user fees in the treated districts compared to before-after or segmented linear regression research designs, which are more common in the literature to date. It allows one to control for trends that affected treated and control areas equally over time (such as a general improvement in the quality and coverage of the health system), as well time-invariant district characteristics (for example, the fact that treated/“rural” areas have lower health care access than control/“urban” areas to begin with). Because of this, “urban” districts can serve as a control for treated/“rural” districts, despite the fact that they are not directly comparable. However it is important to keep in mind that the difference-in-difference regression does not control for asymmetric shocks that could have affected the outcome in treated and control districts differently (hence the importance of the parallel trends assumption).

Due to the fact that health facilities had some discretion about whether or not to implement the policy, and because the dataset contains no information about whether women actually paid to access care, this method only identifies the “Intention to Treat Effect” and not the “Average Treatment Effect on the Treated”. In other words, the study estimates the effect of encouraging
some districts but not others to apply the policy, for those districts being encouraged, but does not estimate the effect of actually making care free for all women giving birth in a health facility.

The specification used in this study is a straightforward application of a difference-in-difference regression. Given that the probability of delivering in a health facility is close to 0.5, the model is estimated using Ordinary Least Squares, as the probabilities predicted by OLS are very similar to those estimated by a Binary Logistic Regression, and OLS coefficient aid interpretation.

\[ Y_{ijt} = \theta + \eta_j + \delta_t + aPolicy\ Change_{ijt} + \epsilon_{ijt} \]

The unit of analysis is the individual birth \( i \), occurring in district \( j \) and year \( t \). \( Y_{ijt} \) is a dummy equal to one if the birth occurred in a health facility, \( \theta \) is a constant, \( \eta_j \) is a district fixed effect (controlling for time-invariant district characteristics), \( \delta_t \) is a year fixed effect (controlling for common trends across districts), and \( Policy\ Change_{ijt} \) is an indicator variable equal to one if the birth occurred in 2006 in a treated district. Clustered standard errors are included in all specifications in order to account for serial correlation over time. The fact that births to the same mother are more similar than births to different mothers is accounted for by district-level fixed effects, given that the treatment is applied at the district level.

While individual-level controls do not aid causal inference, time-varying district-level controls could usefully account for measurable shocks that affected treated and control districts differently. Examples could be weather (which could affect the accessibility of care, e.g. because of impassable roads) or economic shocks (which could affect the affordability of care). Unfortunately, relevant district-level data were not available. To the extent that countryside areas in both treated and control districts share the same economy or the same weather-related accessibility constraints, however, this risk is mitigated by conducting sensitivity tests where the sample is restricted to respondents in specific residence categories.

Births to women who migrated since childbirth were excluded from the sample, as their location of delivery was unknown and therefore could not be assigned to a treated or control district. The resulting sample is therefore weighted towards more recent years and changes qualitatively, since women who have never migrated and will never do so may be different to women who may do so in future. I run regressions on the 2005 and 2006 data only, to test the sensitivity of the results to this problem. I also exclude births occurring after the second policy change, in 2007, which widened the scope of the policy to most of the control areas. Because there is some debate about the timing of this second policy change (January or June), I take a conservative approach by assuming that the policy change occurred at the earlier date, in January 2007, meaning that the period of exposure to treatment in this analysis is April to December 2006. All other years (2002-2005) also include births occurring between April to December, in order to avoid possible seasonal effects. I also conduct a sensitivity test that includes all 12 months of data in each year (April to March).

**Results**

As explained above, identifying a causal effect using difference-in-difference relies on the assumption that the trends in outcomes for treated and control groups would have remained parallel had neither group received treatment. While this assumption cannot be directly tested, confidence is improved by observing parallel trends before the policy change.
I first check for parallel trends in facility delivery across all types of residence areas at once (Figure 3a). This graph displays strongly parallel trends prior to 2006, the year of the policy change. I also investigate parallel trends within each of the residence categories separately as a sensitivity test. Although the trends for “all areas” are parallel over the time period, the different residence categories that compose treated and control groups are likely to be exposed to different kinds of shocks in any given year, and the treated and control groups have a different distribution of residence categories (see Table 1).

Because large and small city areas only contributed 51 births to the control group, I drop cities to examine parallel trends in towns and countryside together (Figure 3b). These trends are parallel prior to 2006. In the countryside sample, the trends are not highly parallel, but at least the direction of the trend is the same across treated and control samples prior to 2006 (Figure 3d). However, this is not the case for the smaller town sample in 2005 (Figure 3c). The town sample is also by far the smallest, which may have increased year on year variation.

Figure 3: Parallel trends

I first run a simple difference-in-difference Ordinary Least Squares (OLS) regression of the probability that the birth took place in any type of health facility, on an indicator variable equal to one if the birth occurred in a treated district in 2006, with district fixed effects and year dummies, for all types of residence (Model 1). The estimated effect is close to zero and not significant at any level. This is in line with the result found by Lepine et al (2015)’s synthetic control study for general adult health services.

While the outcome is binary, the mean probabilities are close to 0.5, implying that the linear probability model, estimated using OLS, is a good fit for the data. The predicted change in the probability of accessing a facility delivery under treatment vs. control in 2006 is very close to that
calculated by a binomial logistic regression (Model 2): a change of 0.49 percentage points (OLS) instead of 0.43 percentage points (Logit).

I also estimate a difference-in-difference regression for the years 2005 and 2006 only, in order to minimise the migration problem described above (Model 3). Again, the coefficient is very close to zero and not significant. Given the lack of clarity on the precise timing of the 2007 policy change and whether it occurred in January or June, I also run the regression while including months April to March in all years, instead of April to December, as in the basic specification. The estimated effect remains insignificant, close to zero, though now negative (not shown).

Because the policy would have changed the cost of delivering in a public health facility but not in a private one, I estimate the same specification as Model 1 for public facilities only. Similarly, because the cost of care prior to the policy change would have been higher in hospitals relative to health centres, I estimate the same specification for public hospitals. The policy had no significant impact on either of these two outcomes – the coefficients on the difference-in-difference indicator variable remain close to zero and insignificant (not shown).

Table 2: Impact on facility delivery and sensitivity tests

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1) OLS 2002-2006 All areas</th>
<th>(2) Logit 2002-2006 All areas</th>
<th>(3) OLS 2005 &amp; 2006 All areas</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006 &amp; Treated</td>
<td>0.00494 (0.0450)</td>
<td>0.0204 (0.2212)</td>
<td>0.00662 (0.0538)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.438*** (0.0249)</td>
<td>0.481*** (0.0146)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>3,126</td>
<td>3,074</td>
<td>1,569</td>
</tr>
<tr>
<td>Nb of districts</td>
<td>70</td>
<td>66</td>
<td>70</td>
</tr>
<tr>
<td>District FE</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Year FE</td>
<td>YES</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
</tr>
</tbody>
</table>

Clustered standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
Sample: births occurred where woman is living now, Apr-Dec, 2002-2006, information about facility delivery is not missing. Model (2): 4 districts and 52 obs dropped due to 100% facility attendance or 100% home births over the period

As discussed in the last section, we might expect the impact of the policy to vary according to the residence category, or for residence-specific shocks to interfere with the estimated causal effect of the policy. Using a linear probability model, I find that there is no significant effect of the policy on the probability of facility delivery, in any residence category (Models 4 to 6). The difference-in-difference coefficient in the “town & countryside” sample subset is still very close to zero, though it is now negative. The coefficients in the “town only” and “countryside only” samples are somewhat higher (around 0.5 percentage points), negative, and remain insignificant.

Table 3: Impact on facility delivery by residence type

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(4) Towns &amp; country</th>
<th>(5) Town only</th>
<th>(6) Countryside only</th>
</tr>
</thead>
</table>

As discussed in the last section, we might expect the impact of the policy to vary according to the residence category, or for residence-specific shocks to interfere with the estimated causal effect of the policy. Using a linear probability model, I find that there is no significant effect of the policy on the probability of facility delivery, in any residence category (Models 4 to 6). The difference-in-difference coefficient in the “town & countryside” sample subset is still very close to zero, though it is now negative. The coefficients in the “town only” and “countryside only” samples are somewhat higher (around 0.5 percentage points), negative, and remain insignificant.
<table>
<thead>
<tr>
<th>2006 &amp; Treated</th>
<th>0.00390</th>
<th>-0.0615</th>
<th>-0.0407</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.0508)</td>
<td>(0.0814)</td>
<td>(0.0580)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.395***</td>
<td>0.777***</td>
<td>0.313***</td>
</tr>
<tr>
<td>(0.0253)</td>
<td>(0.0658)</td>
<td>(0.0226)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>2,844</td>
<td>538</td>
<td>2,306</td>
</tr>
<tr>
<td>Nb of districts</td>
<td>68</td>
<td>40</td>
<td>66</td>
</tr>
<tr>
<td>District FE</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Year FE</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Controls</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
</tr>
</tbody>
</table>

Clustered standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
Sample: births occurred where woman is living now, Apr-Dec, 2002-2006, information about facility delivery is not missing.

Why did the policy have no impact on facility deliveries? There are many possible reasons, ranging from the fact that quality of care may have fallen after the policy change due to drug shortages and insufficient health workers, to the presence of other remaining barriers, to the fact that user fees were low or inconsistent to start out with, or that the policy was not evenly implemented across the treated districts. Disentangling these reasons is important to inform policies in other countries and to continue making progress on maternal health care access in Zambia itself.

An important potential explanation to explore further is the last one mentioned above: does this study find no “Intention to Treat Effect” because too few facilities promptly complied with the new policy in the treated districts? If so, this would imply that the appropriate policy response is to implement the policy more thoroughly. The data prevents us from investigating this question directly as we have no information about the proportion of women actually paying for care in each district. However, Lepine et al’s analysis (2015) suggest that this explanation is incorrect. Using a different dataset, the authors find that for general adult health care, the degree of implementation of the policy at district level was not associated with adult health care utilisation in that district. Secondly, they find that facilities offering better quality of care were more likely to implement the policy. This bias would tend to over-estimate the causal effect, thereby strengthening our confidence in this study’s finding of no effect.

Using this study’s dataset, it is possible to analyse the reasons respondents gave for not delivering in a health facility. One of these reasons was “cost too much” – others included “too far/no transport”, “not necessary”, “not customary”, “don’t trust the facility”, etc. Using a simple tabulation, we see that among the treated group, the share of mothers who said “cost too much” fell in 2006 compared to 2005, whereas the proportion increased in the control group. This trend is in the expected direction, although because of the small sample (a dozen or less gave “cost too much” as a reason per treatment group per year), and the lack of parallel trends, it is not possible to draw causal conclusions.

The more interesting take-away is the very small percentage of mothers who gave “cost too much” as a reason in any year and across both treatment groups, compared to: “too far/no transport” (by the far the largest reason); “not necessary”; “caesarean section emergency”; and “other reasons”. The importance of distance as a barrier to healthcare access in Zambia has also been documented by other authors. This evidence would seem to indicate that the policy was addressing a barrier that was not highly influential to start out with, compared to other barriers.
Discussion

In this study, I sought to examine whether the removal of user fees in Zambia in April 2006 improved access to facility delivery. Using a method of analysis that permits stronger causal inference than in the literature to date, difference-in-difference, I conclude that the policy had no effect on access to facility delivery in the treated districts. I show that the cost of care prior to the policy change did not appear to be a major barrier among sampled respondents, especially when compared to barriers that were not addressed by the policy, such as distance to care and transport, or the perception that delivering in a health facility was not necessary. This finding contradicts other studies evaluating the effect of the policy in Zambia on general health care access14,15, as well as conclusions from multi-country studies in the field of maternal health10,11. In Zambia, this could be because the impact on maternal health care access differed from the impact on general adult and child health care access, although Lepine et al (2015) find no effect of the policy on general health service utilisation either16. Alternatively, the differences may be due to the nature of the methods used, particularly the stronger causal claims afforded by a natural experiment.

A number of limitations remain. Firstly, the short post-treatment period of nine months may have contributed to the fact that I did not find an overall effect of the policy. This short post-treatment period reduced sample size and may not have been long enough for implementation to be completed country-wide and for the population to be fully aware of the policy change. On the other hand, a longer period would be more likely to capture subsequent decreases in utilisation as a result of reduced quality of care, an effect documented in Zambia by Lagarde, Barroy, and Palmer 201214.

Secondly, the fact that the geographic location of DHS sampling clusters was scrambled would also have dampened the effect and may have resulted in allocating treated clusters to control districts and vice versa. However, the extent of the geographic displacement was small and should not have
affected a large number of clusters. A related issue is that of contamination: women living in a control district could have delivered in treated district, and the dataset contains no information on the place where care was accessed. Due to the fact that the data suggests that distance was a bigger deterrent to access than cost of care, we would expect the degree of contamination to be low.

Thirdly, it is possible that despite displaying parallel trends in the past, treated and control districts experienced differential shocks around the time of the policy change, which counteracted the positive impact of the policy in the treatment districts. This may have been an agricultural or industrial shock, for example, related to the weather or the price of copper, which would have affected families’ ability to afford a facility delivery. While the lack of weather, economic, or employment data at the district level did not allow me to rule out this possibility, I mitigated this challenge by comparing similar residence categories across treatment and control districts. This would have partly accounted for differential economic shocks, for example, if countryside areas in different parts of the country respond to similar economic stimuli.

A fourth limitation is the migration problem, whereby a greater share of the sample of births was excluded in earlier years due to the mother having moved residence since the birth. In 2006, the year of the policy change, a greater proportion of the sample is made up of “future movers”, who may have a different propensity to deliver in facilities relative to “never-movers”, with this propensity being different in treated/rural versus control/urban districts. This issue was addressed by conducting sensitivity tests that included only the year immediately before and the year of the policy change.

Conclusion

The evidence from this study suggests that removing user fees did not improve levels of access to facility delivery in Zambia. Other countries might be inclined to learn from Zambia’s experience with user fee removal: while the quality of any policy implementation is crucial, ambitious progress on facility delivery coverage may require a multi-pronged approach that addresses more than one barrier to access. In Zambia’s case, problems of geographic accessibility to maternal and neonatal health care have been repeatedly documented and remain a significant barrier to access despite the removal of user fees.

References


5. Gebrehiwot, T., Goicolea, I., Edin, K. & San Sebastian, M. Making pragmatic choices: women’s


