The Legacy of the Tehrik-e-Taliban Pakistan (TTP) on Girls' Access to Education

IDS 701: Final Project Report

Clarissa Ache, Preet Khowaja, John Owusu Duah, & Cameron Ratliff

April 5, 2022

Summary

The Tehrik-e-Taliban Pakistan (TTP) have found safe haven in north-western Pakistan since 2007, and have been at war with the Pakistani state and army. Part of their agenda is implementing *sharia* law in Pakistan, which has severely restricted girls' education in areas where they are most active. This study seeks to investigate the impact that the Taliban has had on girls' education in rural areas.

To achieve this objective, we compared primary school enrollment of girls in the 20 districts with strong TTP influence, to the girls who are in districts which are not controlled by the Taliban. Using a *difference in difference* approach, we estimate the causal treatment effect on the treated group (ATT) by comparing the girl's enrollment rate in the two groups both before and after the Taliban incursion that started in 2007. We assume here that both groups would trend similarly over time if Taliban intervention had not occurred. Pakistan's education data has been sourced from Pakistan Living Standards Measurement (PSLM) survey. The survey covers rural and urban areas of the four main provinces of Pakistan and collects information on a wide range of topics such as education, immunization, reproductive health, and access to basic services.

Our analysis did not reveal a significant average effect in the enrollment rate in school for girls TTP influenced cities. A thorough discussion of the challenges faced using the PSLM Survey has been outlined in the report. Some of these limitation include missing and varied data across years, which could explain the absence of the treatment effect that we expect to see in our study.

Introduction

The origins of Malala Yousufzai's story lies in the valley of Swat, in northern Pakistan. As the youngest Nobel Peace Prize laureate in the world, she says:

> Without education, girls are more likely to marry young, suffer from preventable diseases and experience poverty throughout their lives. Without educated girls, communities, countries and our world suffer from slower economic growth, more conflict, poor public health and increased risk from the effects of climate change. [10]

Pakistan's women have long been deprived of access to education under the pretext of cultural and religious justifications. As an Islamic Republic, from its inception in 1947, Pakistan has harboured a strong sensitivity to religion in its political landscape. In this arena, the emergence of Islamic militant groups in Pakistan, due to various geo-political factors, has made it considerably harder for girls to attend schools. The most prominent of these militant organizations is the Tehrik-e-Taliban (TTP), which has launched violent attacks across the country, with the mission of bringing *sharia* law to Pakistan. According to the TTP, women must not leave the house unaccompanied by a male guardian and need not obtain any education, particularly that which follows a Western format.

The emergence of the TTP is closely associated with that of the Afghan Taliban in the 1990's, but it was not until 2007 that it established itself as an official organization as it started a violent terrorist campaign against the Pakistan armed forces and the state in areas of north-western Pakistan that fell

under the Khyber Pakhtoonkhwa province and Federally Administered Tribal Area (FATA). Over time, the TTP launched attacks in other major cities of Pakistan (see Figure 1), causing thousands of civilian casualties. Malala was a victim of one such attack in Swat Valley, in northern Pakistan, as she travelled on a school bus that took her and her classmates to school. The TTP used various methods to incite violence among local populations, including threatening parents who sent their daughters to school and suicide attacks in educational institutions.

TTP violence peaked in 2010 and after a steady decline in their activities from 2014-2018, there has been a recent resurgence of attacks and a revitalized propaganda campaign in 2019, which could be a serious cause of concern for women's education in Pakistan. In the first two months of 2021, alone, 32 attacks have been claimed by the TTP across the country, and 2021 overall saw a 42% rise in attacks from the previous year. The Pakistani government has made numerous attempts of peace deals TTP leaders and umbrella organizations, both officially and unofficially. However, negotiating with a militant and fragmented organization has proved futile, with most peace deals lasting only a few months. The effect of TTP incursions has affected life for civilians in many ways, but the threat to women's rights in these regions is an enduring one, endangering lives of girls who choose to empower themselves through education.

In light of this evidence, we propose the hypothesis that women's educational rights are weakened in areas of strong TTP influence, which constitutes regions where the TTP have established their influence. For this reason we are interested in studying the legacy the Taliban groups have left for Pakistani girls. More specifically, we are interested in finding the **causal effect of TTP terrorist incursions that started in 2007 on girls' access to education in rural areas of Pakistan.**

Model

In order to measure the effect of the Taliban influence on women's access to education in Pakistan, we first define which areas of the country have been under Taliban control. In 2007, after its official formation, the TTP and sympathising militants established safe havens in what was known as the Federally Administered Tribal Areas (FATA) and the province of Khyber Pakhtoonkhwa (KPK, previously known as the North West Frontier Province, or NWFP). They controlled 20 tribal regions and districts as follows: South Waziristan, North Waziristan, Orakzai, Kurram, Khyber, Mohmand, Bajaur, and Darra Adamkhel, and districts of Swat, Upper Dir, Lower Dir, Bannu, Lakki Marwat, Tank, Peshawar, Dera Ismail Khan, Mardan, Charsadda, and Kohat [5]. In Swat Valley, the TTP had begun its influence years before the official incursion launch in 2007. The Pakistan Army was able to retake Swat Valley from TTP stronghold by the end of 2007 [3]. Therefore, the region of Swat was not considered occupied by TTP after the incursion in our study.

Since TTP presence and activity in the aforementioned regions was fragmented and varied in nature, there is no comprehensive official documentation of their exact geographical presence and those of its allies over the years. As a result, we assume that areas that were affected any time after 2007, endured militant activity for the duration of our study from 2007-2020.

We define the beginning of the TTP incursion from 2007 to 2009 as the intervention or "treatment" we want to measure the effect of. Our approach is to compare the difference across time in girls' school enrollment rates in the 20 cities that were controlled by Taliban after 2007 (our treated group), to the enrollment rate in cities that were not controlled by Taliban (our control group). We propose this *difference in difference* design to approximate an experimental model because there are permanent baseline differences between the populations that were under Taliban influence and those that weren't. The difference in difference model allows us to mitigate the bias that were to arise from those permanent differences if we looked simply at enrollment trends after the Taliban took over these regions. Using the difference in difference model, we can estimate the causal treatment effect on the treated group (ATT) by comparing the girl's enrollment rate in the two groups both before and after the intervention, assuming they would trend similarly over time if Taliban intervention had not occurred. This model, however, assumes that Taliban occupation (our treatment variable) must not be correlated to enrollment rates of girls. In the next section it is demonstrated that this assumption is plausible under our model.

The model proposed also assumes measured enrollment rates in both groups of districts (the Taliban-occupied and the rest) will follow parallel trends in the period of time before the Taliban intervention.

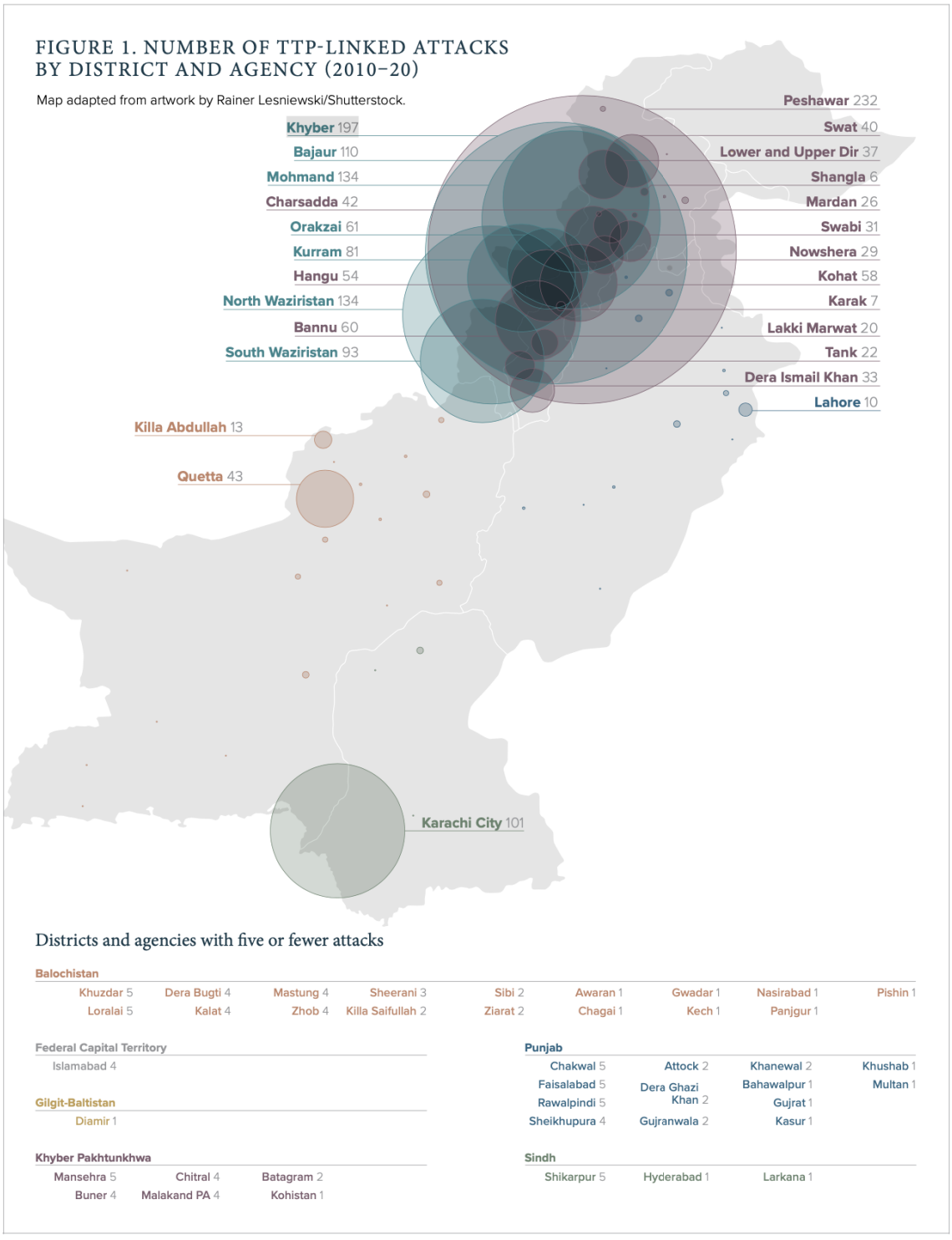

Figure 1: TTP Attacks in Pakistani Cities 2010-20

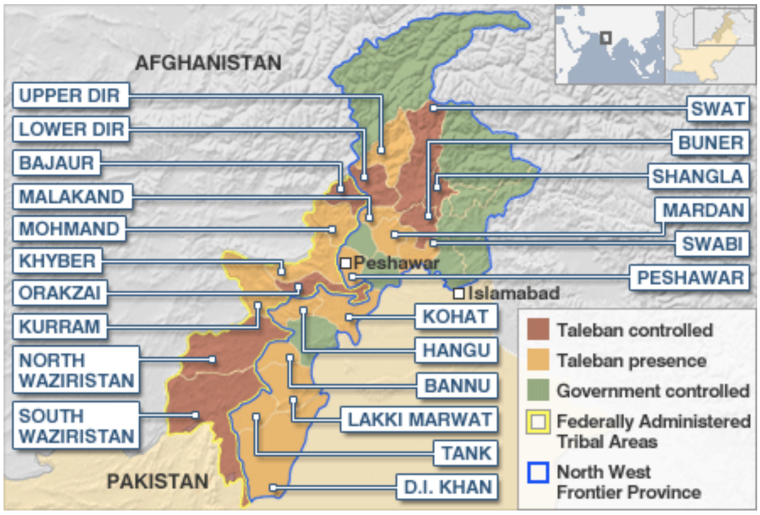

Figure 2: Taliban Occupation in Pakistan by 2009

Given the 20 cities that were occupied by Taliban militia were all considered rural, we decide to only assign rural areas to our control group. Rural areas in Pakistan have more similarities in socioeconomic and demographic indicators, for this reason we would expect education indicators to trend in similar rates across years if none had suffered Taliban occupations. The validity of this assumption is also explored in the next section.

Finally, in order to infer causal effects from our model we must validate the Stable Unit Treatment Value Assumption (SUTVA) holds. SUTVA indicates there are no different treatments applied to our treated group, which could be violated given some Taliban occupations are stronger, and thus, more influential than others. SUTVA also holds if there are no 'spillover effects', which could also be violated given refugee migrations happened in rural areas and many of the "treated" districts are geographically connected to rural areas in the control group. Taliban's message was often broadcast through media outlets such as radio stations, which has been proved to be influential in Swat Valley, and with the power of reaching any rural community in Pakistan (even the ones that were never physically occupied by the TTP, our control group). We keep these potential violations in consideration as we generate conclusions of our study.

Data

The Pakistan Federal Bureau of Statistics (PBS) conducts the Pakistan Living Standards Measurement (PSLM) survey as part of a program started in 2004 with the objective of monitoring of key social indicators. These indicators would serve to design poverty reduction strategies **??**. The PSLM survey is geographically representative, and the data collected is publicly available in the PBS website for every year since 2004 to 2020 except for years 2009, 2016, and 2017. In this study, we collected all the survey responses for each of the available years.

PSLM and HEIS Survey Structure and Sample Design

The PSLM is collected in a two-stage stratified sample design. Districts (also called "Stratum") are classified as urban and rural areas. Each urban or rural district is divided into enumeration blocks consisting of 200-250 households. A two-stage stratified sample design was adopted by PSLM and HIES, districts and enumeration blocks in urban and rural areas respectively have been taken as Primary Sampling Units (PSUs). The sample size for the survey has been fixed by using average household consumption (in Rupees) as key indicator at 95% level of confidence, 5% margin of error (MOE) for three provinces namely, KPK, Punjab and Balochistan while 4% MOE has been used for province Sindh. Keeping in view the variability that exists within the population for the characteristics for which estimates are to be prepared, population distribution, level of estimates, and field resources available; the sample size varied marginally

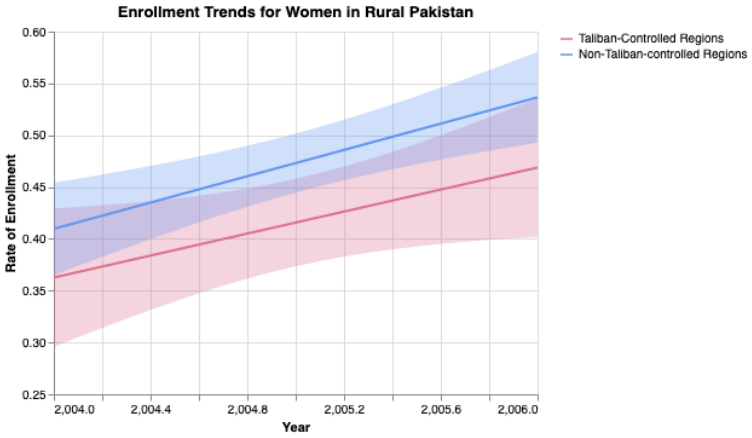

Figure 3: Our proposed model assumes parallel trends in absence of Taliban intervention. We will assume the difference in the trends is negligible.

across the years.

Data Aggregation and Manipulations

To meet the objectives of the study, we obtained data from the education and demographic sections of the PSLM survey directly from its website, keeping only individuals with ages between 4 and 10. This age group has been selected to match the United Nations Millennium Development Goals that seek to ensure universal primary education. After subsetting for these individuals, we have 1,761,674 observations. That is the number of individuals surveyed in the data available.

Also, because the PSLM surveys a different random subset of the population each year. To ensure that the units observed were consistent across the years we aggregated enrollment rate at the district level as shown in Equation 1.

$$\text{Enrollment Rate }_{girls} = \frac{\text{Total enrolled girls in the district}}{\text{Total number of girls surveyed in the district}} \tag{1}$$

As mentioned, each district has a designation of whether it is considered a rural or urban area. However, some districts correspond to subdivisions of the same city (they are given the same name in the survey encoding). This happened for locations that have both rural and urban areas. Since many locations that had the two designations and were sampled (and encoded) twice in the PSLM, we defined them as a separate observations. That determines our unit of observation as district ("survey stratum") by year, by rural/urban designation, and by sex (girls only).

It is important to note that individual level data on other social and economic factors such as religion and ethnicity would have been very informative to our study, that is, if they had been collected in the survey. The regions of Taliban-stronghold as ethnically Pashtun, and considerably distinct from the ethnic composition of other rural areas in our survey. This variable could serve as a stand-in for culture and provide a more robust causal analysis.

Education and Demographic Data in the PSLM Survey

Since pre-Taliban intervention, enrollment trends for girls and boys were not parallel, we compare girls' enrollment rates in districts controlled by the Taliban to girls' enrollment rate in other rural districts, where the pre-2007 enrollment trends as similar as observed in Figure 3. This eliminates confounding information on whether girls live in rural or urban areas and also fulfills a key assumption of our difference-in-difference analysis.

In Figure 3, we can observe trends are almost parallel for enrollment rates for girls prior the interventions studied. In our model, we will assume that in absence of Taliban intervention, girls enrollment

in education would still move in parallel across years after 2007. Considering the limitations of the data used to produce this trends, we will assume the difference in the trends is negligible.

In order to establish causal inference, we want to ensure that the selection of districts in our treatment or control groups was not related to our response variable (rate of enrollment for girls in primary school). A literature review convinces us that the incursion of the Taliban into the given cities is mainly due to geo-political reasons that are not directly related to rate of enrollment [5].

Demographic data collected was on age, marital status, gender, province and districts where the household of the individual was located. We decided to select these variables because we estimate that the enrollment status of an individual can be influenced by the above-mentioned variables.

We notice that about 57% of children in the data set live in rural areas while 43% live in urban areas. Children between the ages of 4 and 10, in rural areas, are married three times more than individuals in urban areas. This an example of demographic differences between rural and urban areas and Pakistan, which is why only rural areas were used in the model for comparison.

For a difference in difference analysis to be valid, the demographic composition of treatment and control groups must be stable for repeated cross-sectional design. The trends in demographic indicators collected for the two groups shows that the composition is stable for the repeated panels (years) of the data.

Challenges Using the PSLM Survey Data

Over the course of our data cleaning, exploration and aggregation we flagged some issues with the dataset that could affect the outcome of our analysis. Primarily, there is missing data in various important ways:

- We do not have survey results from the surveys of years 2009, 2016 and 2017, and we rely on a regression fit in our final analysis estimate the trend in enrollment rate for these years. This is why in our results, confidence bands for the fitted linear model are larger for those years.

- The collection of districts surveyed is not the same each year. This is especially challenging when it comes to building paneled data that is comparable for each year. To mitigate this problem, we first ensure that we have enough districts with data across the all of the survey years. We also use fixed effects for the districts in estimating the difference in difference coefficient in a Panel linear regression. Table 1 summarizes the counts of districts that were included in the comparison each year. The discrepancy in the last two years' counts is caused by the regrouping and renaming of districts in the survey and an addition of new rural districts that had not been surveyed after 2018.

- For the year 2007, the rates of enrollment for both male and females significantly drop to 0.3, which is inconsistent with previous trends that we observe. The time period of 2007-2008 was a time of political turmoil in Pakistan, with the assassination of former Prime Minister, Benazir Bhutto, in late 2007, and rampant public protests and strikes, which can explain this steep dip in enrollment rates. However, 2007-08 is also the year of major Taliban incursions into cities. We use that year as our intervention period, excluding the year's information from our difference-in-difference analysis.

- In the survey of 2018, we only obtained data for 32 rural districts, whereas the count ranges between 90 and 110 for other year's data. This is an issue for modeling since it means the composition of districts in 2018 may be different from that in other years. In order to control for this problem, we ensure that the mean enrollment rate in 2018 is consistent with the trends we see before and after 2018. The overall mean of enrollment rate does not deviate much to the surrounding years, however, when subsetting for TTP-control districts in 2018, we obtained a much lower mean compare to previous and next years. The nuances of this particular year are potentially pulling down the fitted linear model and cause it to trend downwards and generate bias in our conclusion.

- As mentioned in the first section of this report, the areas of Pakistan that had potentially the strongest Taliban influence are the most remote rural communities close to the Afghan border. These are in FATA and NWFP (now called KPK). In early surveys, many districts in these areas were not included in the sample, only 2018 and 2019 use samples that are most geographically representative of FATA and KPK, the same years we observed lowest enrollment rates of girls in primary school in TTP occupied areas.

Table 1: Total number of districts in Treated (occupied by Taliban) and Control districts each year the PSLM survey data was published

Treated	**No**	**Yes**
Year		
2004	87	10
2005	71	10
2006	91	10
2007	70	10
2008	100	10
2010	104	10
2011	80	10
2012	104	10
2013	106	10
2014	96	10
2015	107	10
2018	15	17
2019	104	17

Analysis

This section subsets our analysis into two parts. In the first part, we evaluate the difference in trends of girl's enrollment rates in the two groups. In the second section, we estimate the difference in difference coefficient using a paneled linear regression model.

Observed Difference-in-Difference

We look at women's rates of enrollment in school across the years the PSLM survey was conducted except for the period between 2007 and 2008. In Figure 4, we compare girls enrollment rate in Taliban-dominated regions to other rural areas. The trend of enrollment rate is negative after 2007 for women living in regions of Taliban domination. This is in contrast to their trend of enrollment before 2007 and also to the trends for women in other rural regions after 2007. Importantly, the enrollment trends for women in other regions continues to increase (at a slower rate) whereas for women in Taliban-controlled regions it declines. Visually, the fitted lines show that Taliban-influence may have a negative effect on girl's enrollment rates.

Note that just after the discontinuity between 2007 and 2008, values of enrollment are flipped compared to the trend prior 2007. That means that enrollment rate in Taliban-controlled cities is higher compared to the control cities in 2008. This inconsistency with the trends in years before the terrorist campaign started (2004-2006) may indicate a violation of model assumptions regarding parallelism. Therefore, visually, we may not have a strong enough conclusion about negative impact to enrollment rate of girls caused by Taliban presence.

Quantitative Difference-in-Difference

An average difference in difference coefficient is calculated using a paneled linear (Ordinary Least Squares) regression controlling for entity (district) and time (year) baseline differences. The time and entity controls are included to account for any kind factors that may have influenced our response variable differently in a particular year or district. This way, we are removing level the differences in enrollment rates across districts and years, which allows to estimate how the Treatment is causing with changes in enrollment rate, our research question.

In this panel linear regression, we are modeling our outcome variable as in the following equation:

$$y_{it} = \alpha_i + \gamma_t + \delta(x_{it}) + \epsilon_{it} \tag{2}$$

where:
y_{it} = Enrollment rate (the response variable)
α = Entity or district effects

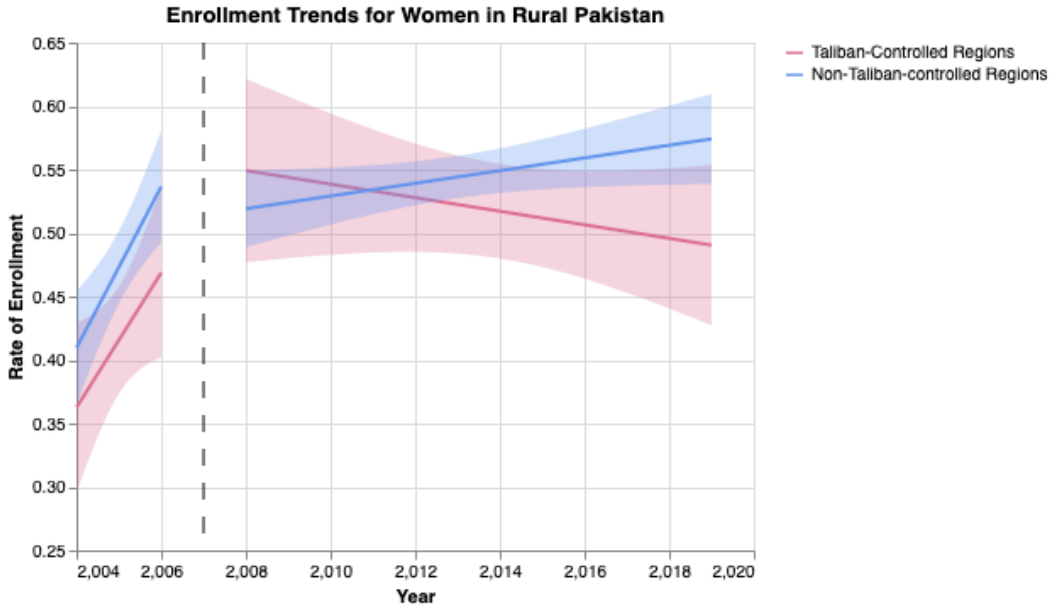

Figure 4: Difference in difference plot of women's enrollment rates in rural Pakistan

γ = Time or year effects
δ = average treatment effect of the treated (Difference in Difference coefficient)
ϵ_i = the random, unobserved "error" term which contains all determinants of y_{it}

The coefficient obtained from fitting the regression model to our data are shown in Appendix Figure 1. The estimated average effect of Taliban occupation (in the cities occupied) is 0.0409 , which means that the difference, on average, in enrollment rate of Taliban-occupied cities compared to the average of all other rural cities is 0.0409. This coefficient being small but positive suggests that the districts with Taliban presence, on average, did not see a decrease but a slight increase on average after the 2007 incursion. This coefficient, however, is not statistically significant, with a p-value of 0.080. Therefore, we will not make any conclusions about the causal effect of Taliban occupation in enrollment rate.

General Limitations of this Study

One important limitations of the model used in this study is that the average difference in difference coefficient does not test for parallelism of the trends because its merely a mean value.

Also, the regions that were likely the most influenced by Taliban are remote rural areas such as FATA and what was previously known as NWFP (now KPK) [7]. Many of these regions were not included in early PSLM surveys, which may be why average enrollment rate on the last two years of the PSLM data is much lower than all of the previous years.

There also may be other factors that influence the rural regions of Pakistan that have opposite effect on reported enrollment rates for girls in schools. People in rural regions may have been reported in the PSLM survey that girls in their household are officially enrolled in schools while they were not actually being allowed to attend school by Taliban rules. This potential bias in our data is possible since school enrollment in primary education is compulsory in Pakistan for children between 5 and 16 years old [11]. If this was true, our data could be biased and not a real representation of the education indicators we are interested in studying.

Conclusion

We conclude we did not find a significant average effect in enrollment rate in school for girls in cities that were controlled by Taliban in 2007. The results we get can be attributed to the limitations in our data, its

incompleteness, and we propose using a complementary linear regression model matched at the individual level could help us understand better the factors affecting the likelihood of girls being enrolled in school. We could also include other economic and social controls such as income of the household and religion in this regression for a more informative study. Another analysis could look at the difference between boys and girls enrollment rates as the response variable. This difference-in-difference-in-difference (also known as a triple-difference) analysis would help provide reassurances that we are not capturing differences in schooling availability generally, but rather differences in schooling availability to young girls, in particular.

While this study doesn't sufficiently answer the causal question we are interested in, it lays the foundations of further investigations, such as those listed above. Such data-backed studies can serve as strong evidence that the TTP is a threat to local female populations and their right to education, pushing international organizations to care about the issue at hand and allow local activists to garner their support.

Appendix

0.1 Difference in Difference Regression

Dep. Variable:	rate_enrollment	**R-squared:**	0.0032
Estimator:	PanelOLS	**R-squared (Between):**	0.0125
No. Observations:	1089	**R-squared (Within):**	0.0098
Date:	Mon, Apr 25 2022	**R-squared (Overall):**	0.0104
Time:	16:45:44	**Log-likelihood**	1010.8
Cov. Estimator:	Unadjusted		
		F-statistic:	3.0586
Entities:	130	**P-value**	0.0806
Avg Obs:	8.3769	**Distribution:**	F(1,948)
Min Obs:	1.0000		
Max Obs:	19.000	**F-statistic (robust):**	3.0586
		P-value	0.0806
Time periods:	11	**Distribution:**	F(1,948)
Avg Obs:	99.000		
Min Obs:	32.000		
Max Obs:	121.00		

	Parameter	**Std. Err.**	**T-stat**	**P-value**	**Lower CI**	**Upper CI**
Treated:post_2009	0.0409	0.0234	1.7489	0.0806	-0.0050	0.0868

F-test for Poolability: 36.188
P-value: 0.0000
Distribution: F(139,948)

Included effects: Entity, Time

References

[1] Brumfield, B. Ng, N. (December 17, 2014) Who are the Pakistani Taliban? CNN
https://www.cnn.com/2014/12/17/world/asia/pakistan-taliban-explainer/index.html

[2] Jadoon, Amira. (May 2021). The Evolution and Potential Resurgence of the Tehrik-i-Taliban Pakistan. United States Institute of Peace. (p. 4)

[3] Pakistan army retakes Swat towns (December, 2007). BBC News.
http://news.bbc.co.uk

[4] PSLM Report (2015-2016). Pakistan Bureau of Statistics.
https://www.pbs.gov.pk

[5] Rahimullah Yusufzai (2009) A Who's Who of the Insurgency in Pakistan's North-West Frontier Province: Part Two – FATA excluding North and South Waziristan. Publication: Terrorism Monitor Volume: 7 Issue: 4

[6] Center for International Security and Cooperation (CISAC). Freeman Spogli Institute for International Studies: https://cisac.fsi.stanford.edu/mappingmilitants/profiles/tehrik-i-taliban-pakistan

[7] Article 130(4) in Chapter 3: The Provincial Governments, in Part IV: Provinces, of the Constitution of Pakistan

[8] Pakistan Conflict Map, (BBC 2009): http://news.bbc.co.uk/

[9] David Albouy. Berkeley Economics 131. Program Evaluation and the Difference in Difference Estimator https://eml.berkeley.edu/

[10] The Malala Fund. Malala's Story. https://malala.org/malalas-story

[11] Robert Hunter, World Education Services (2020). Education in Pakistan