Operation Allied Force: Unintended Consequences of the NATO Bombing on Children's Outcomes[*]

Lara Lebedinski

Institute of Economic Sciences[†]

Giuseppe Migali

Lancaster University and Magna Graecia University[‡]

Miloš Popović

Leiden University[§]

Sunčica Vujić

University of Antwerp, University of Bath, VU Amsterdam and IZA, Bonn[¶]

October 13, 2021

[*]Acknowledgments: We are grateful to Wiji Arulampalam, Janet Currie, Michele Pellizzari, Jonathan James, Anja Grujović and Jaroslaw Kantorowicz for their constructive and insightful comments. The participants at the Annual conference of the Institute of Economic Sciences (IES) in Belgrade in October 2019, the EALE-SOLE-AASLE virtual conference in June 2020, the ASSA virtual annual meeting in January 2021, the ESPE, the IWAEE and the IAAE virtual conferences in June 2021, the EALE virtual conferences in September 2021, as well as the seminar participants at the University of Antwerp, University of Alicante, and the Bath Applied Brown Bag (BABB) research seminar series provided useful comments and suggestions. We thank the Statistical Office of the Republic of Serbia and the Serbian Ministry of Education, Science and Technology for making the datasets available to us. Miloš Popović acknowledges that this project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 798392.

[†]Corresponding author: Institute of Economic Sciences, Zmaj Jovina 12, 11000 Belgrade, Serbia. Email: lara.lebedinski@gmail.com

[‡]Lancaster University Management School, United Kingdom. Email: g.migali@lancaster.ac.uk

[§]Institute of Security and Global Affairs, Leiden University, the Netherlands. Email: m.popovic@fgga.leidenuniv.nl

[¶]Department of Economics, University of Antwerp, Belgium. Email: suncica.vujic@uantwerpen.be

Abstract

This is the first paper that estimates the causal effect of the NATO's Operation Allied Force in Serbia in 1999, on children who were in the womb during the bombing. We investigate the *in utero* effect in terms of short-term outcomes, such as birthweight, as well as medium-term outcomes measured by grades of 15-year-old pupils at the end of primary school. Using the birth records of the Serbian Statistical Office, we estimate difference-in-differences models, combined with propensity score matching. We compare the birthweight of children born in the same year (1999) and in the months just before and after the bombing, and children born in the same months of the previous year (1998). We then exploit the data on educational achievement at the end of primary school, provided by the Ministry of Education, to estimate matching models of the effect of the bombing on individual grades. Our findings suggest that children *in utero* during the bombing were 2pp more likely to be born with a lower than average birthweight. In the medium-term, we find a statistically significant and negative effect (around -1%) of the bombing on maths grades and Serbian language at primary school, and a 1% increase in the probability to enrol on vocational secondary schools. Overall, our results confirm the importance of the negative effects on children in the aftermath of large-scale disasters, and the necessity of policy interventions to mitigate them.

Keywords: Human Capital Formation; Children; Armed Conflict; In-Utero Effect.
JEL classification codes: I15, J13, O15.

1 Introduction

On March 24, 1999, the North Atlantic Alliance (NATO) initiated air strikes against Yugoslavia (now Serbia),[1] under the name "Operation Allied Force." The military intervention consisted of an air campaign targeting not only military facilities, but also strategic targets such as factories, bridges and government buildings. Since the bombing of Britain and Germany in the Second World War, the NATO bombing of Serbia was the largest air campaign in Europe. The intervention lasted for 78 days, between March 24, 1999 until June 10, 1999, and hit 108 out of 160 Serbian municipalities, excluding Kosovo and Montenegro. We use this arguably exogenous variation to show that adverse shocks during the intrauterine period affected children outcomes in the short-run, such as birthweight, as well as later educational outcomes, such as primary schools grades and probability to enrol the non-vocational secondary schools.

It is not straightforward to estimate the causal effect of the early childhood circumstances on later outcomes. The results of this exercise may be confounded by the unobserved factors which affect the socio-economic and medical conditions of both mother and child. For example, both parents' income and children's health may be affected by the family circumstances and genetic makeup which are transmitted from one generation to another. To be able to detect causal effects, one needs independent (exogenous) variation in early-life conditions and relate this to the outcomes of interest later in life. Similar in spirit to Akbulut-Yuksel (2014), we look at the effect of NATO bombing on children's birthweight and later educational outcomes.

Our identification strategy for the short-term outcomes is based on a difference-in-differences (DD) estimation approach, combined with propensity score matching. We first compare children *in utero* during the whole period of bombing with children born few months before in the same year, 1999, and children born in the same months of the previous year. This approach should avoid the issue of selection into pregnancies, as bombing was arguably unforeseeable. For this analysis we use birth records from the Statistical Office of the Republic of Serbia. Our findings for the short-term outcomes suggest that children who were *in utero* during bombing had a higher likelihood of having lower than average birthweight of about 2 percentage points.

To investigate the medium-term outcomes, we use data on educational achievement at the

[1] In 1999, official name of the country was the Federal Republic of Yugoslavia. In 2003, the country was renamed to Serbia and Montenegro in order to reflect its constituent parts after a dissolution of former Yugoslavia in the early 1990s. After Montenegro's independence in 2006, Serbia became the legal successor of Serbia and Montenegro. In the remainder of the paper, Yugoslavia and Serbia are used interchangeably and they both refer to the territorial space of Serbia without Kosovo.

end of primary school, provided by the Serbian Ministry of Education. Due to the rule on the starting age in primary school, we cannot perform a DiD analysis, hence we estimate an inverse probability weighting regression-adjustment (IPWRA) model. We find that children who were *in utero* during bombing had statistically significant lower marks in mathematics and Serbian language of about 1% (or about 0.03 standard deviations); and also a 1 percentage point reduction in the probability to enrol a 4-year secondary school compared to a 3-year vocational school.

There are many ways for a lethal catastrophe such as bombing to affect the pregnant mothers. There are both direct channels, such as physical destruction, displacement and deteriorated socio-economic and health conditions, as well as indirect ones, such as contamination of air and soil, as well as malnutrition. Considering that the goal of bombing was to maximise material damage and limit collateral (civilian) damage,[2] mothers of the treated and the control children had the same socio-economic, health conditions and access to prenatal care. The mothers of the treated children didn't have significantly higher number of stillbirths or different behavioural and health outcomes during pregnancy in comparison to the mothers of the control children. Due to the United Nations (UN) sanctions against Yugoslavia, as well as tightened visa travel regime for its citizens, migratory movements out of the country were limited. Therefore, we believe that the main transmission mechanism is *in utero* environment of both mother and the child due to the prenatal maternal stress (Aizer et al., 2016; Black et al., 2016; Berthelon et al., 2018; Persson and Rossin-Slater, 2018).[3]

One policy implication of our findings would be that governments need to intervene and design policies to alleviate the negative *in utero* effects on children in the aftermath of large-scale disasters. Another policy implication questions bombing as a legitimate tool of intervention in the international conflicts – this type of interventions should be re-evaluated, taking all possible consequences into account.

1.1 The NATO Intervention

The North Atlantic Treaty Organisation (NATO) "Operation Allied Force" was the codename of the aerial bombing campaign against the Federal Republic of Yugoslavia during the Kosovo

[2]Less than 7% of damaged civilian objects were hospitals, houses or private farms. According to the Humanitarian Law Centre (HLC), the NATO bombing killed a total of 754 people: 454 civilians and 300 members of the armed forces. There were 260 casualties in Serbia alone.

[3]Contamination of the soil could be a potential channel for the more long-term outcomes and for a different treated group than the one considered in the paper.

War.[4] As a result of the failed peace talks in Rambouillet, NATO initiated punitive aerial strikes on March 24, 1999. The military intervention used modern precision weaponry, such as aerial bombing and surface-to-air missiles, against Yugoslav strategic military targets (military barracks, industrial facilities, transportation networks and communication lines, as well as governmental buildings). It was a precision aerial bombing similar to bombings of Iraq, Libya, Syria and Afghanistan (Sardoschau, 2018; Oskorouchi, 2019), with the aim of maximising material damage and limiting collateral damage (Fenrick, 2001). The NATO's operation lasted 78 days and hit 108 out of 160 Serbian municipalities at the time, excluding Kosovo and Montenegro. It ended on June 10, 1999, when an agreement was reached that led to the withdrawal of Yugoslav armed forces from Kosovo. The bombing was the largest aerial bombing campaign in Europe since the bombing of Britain and Germany in the Second World War.

In our work we use a novel and unique dataset of the NATO bombing of Serbia, which covers the whole period of bombing from March 24, 1999 until June 10, 1999. The dataset was manually coded and includes information on the location of bombings as reported in the media.[5] The data at our disposal, collected at the level of settlements (4,721) in 160 municipalities, are the most comprehensive and precise data of the NATO bombing of Yugoslavia (present Serbia). For example, we have information on which settlement was bombed and for how long, the number of fatalities per settlement and the distance to the nearest strike/fatality in kilometres.

The Serbian case is especially useful for examining the effects of bombing on child development in a quasi-experimental framework. First, the NATO intervention was arguably unanticipated and provides a source of exogenous variation. Second, apart from the NATO bombing, there was no other armed conflict on the territory of Serbia, which enables us to isolate the effect of the intervention. Third, the magnitude of this event exceeds the average terrorist bombings, offering a unique opportunity to empirically investigate the effect of prolonged prenatal exposure to "disaster" conditions on child development and human capital outcomes. Finally, this conflict has not been previously studied in the economics literature.

Figure A1 summarises the main features of the bombing data. Figure A1a shows that the bombing was dispersed across the country with the highest concentration of attacks in large towns such as Belgrade, Niš, Novi Sad, and Kraljevo. Figure A1b captures the intensity of the NATO bombing of Serbia, showing the number of days a settlement was bombed. It ranges

[4] The Kosovo conflict originates from the collapse of Yugoslavia, which broke up through a series of armed conflicts on the territories of Slovenia, Bosnia, Croatia and Kosovo during the 1990s.

[5] More information on the data collection process is provided in the Appendix.

between 0 and 35, with the majority of settlements experiencing less than ten days of bombing.

1.2 Literature

The 'Fetal origin hypothesis' (FOH) or 'Barker's hypothesis' goes back to David Barker, a British physician and epidemiologist, who proposed a direct link between prenatal nutrition and adult coronary heart disease, including hypertension, adult-onset diabetes and stroke. The idea is that adverse shocks while *in utero* "tend to have permanent effects on the body's structure and function" (Barker, 2001), which may lead to increased vulnerability and chronic conditions later in life. Besides this direct effect, a shock early in life may trigger adverse outcomes, such as worse health and educational outcomes in childhood and subsequently worse labor-market and health outcomes (Van den Berg and Lindeboom, 2018; Aizer et al., 2016; Ben-Shlomo and Kuh, 2002).

Almond and Currie (2011) and Almond et al. (2018) provide an overview of the epidemiological literature on the 'fetal origin hypothesis' and contributions from economics.[6] They further summarise studies in economics exploiting natural variation of *in utero* environment of both mother and the child due to lethal catastrophes, such as famines, pandemics, wars, and hurricanes as natural experiments' ('disaster literature'), as well as more 'mild shocks', such as malnutrition, infectious diseases, macroeconomic conditions, pollution and toxic exposure, weather and climate changes during pregnancy. Our paper directly contributes to this literature.

In Table 1 we provide a summary of recent studies which demonstrate that experiencing negative events during pregnancy leads to worse birthweight and human capital outcomes of the affected children. We compare these findings to our estimates of the impact of the NATO bombing. In Panel A, we provide an overview of papers which use exposure to terrorist attacks while pregnant, such as 9/11, ETA or Jihadi terrorist attacks, as well as the violent clashes between the Palestinians and Israel (the al-Aqsa Intifada) (Eccleston, 2011; Mansour and Rees, 2012; Quintana-Domeque and Ródenas-Serrano, 2017; Duque, 2017; Brown, 2020; Armijos Bravo and Vall Castelló, 2021). In Panel B we summarise the effects of natural disasters such as earthquakes (Torche, 2011; Kim et al., 2017; Menclova and Stillman, 2020), hurricanes, tropical storms, floods and temperature shocks (Currie and Rossin-Slater, 2013; Andalón et al., 2016;

[6]Since economists joined this line of research, they have contributed in terms of plausible strategies for identification of causal effects, they have contributed to the nurture versus nature debate in this context, they focussed on whether some types of shocks are more detrimental than others, as well as the timing and cost-effectiveness of different remedial interventions (income transfers or more targeted interventions) designed to mitigate the harms generated by the *in utero* shocks.

Rosales-Rueda, 2018; De Oliveira et al., 2021), and Asian flu (Kelly, 2011). In Panel C, we provide an overview of studies which look at the exposure to violence and domestic violence while pregnant (Aizer, 2011; Koppensteiner and Manacorda, 2016; Brown, 2018; Currie et al., 2018; Guantai and Kijima, 2020). In Panel D, we list other, more 'mild shocks': Ramadan fast (Almond and Mazumder, 2011; Almond et al., 2015a), macroeconomic shocks (Bozzoli and Quintana-Domeque, 2014; Olafsson, 2016; Akbulut-Yuksel et al., 2020), death of a family member (Black et al., 2016; Persson and Rossin-Slater, 2018), air pollution and radioactive fallout (Almond et al., 2009; Currie and Schwandt, 2015; Calzada et al., 2021).

All this literature shows that the causal impact of negative *in utero* shocks, on children birthweight and human capital outcomes, can be explained by maternal stress (Aizer et al., 2016; Black et al., 2016; Berthelon et al., 2018; Persson and Rossin-Slater, 2018), due to both direct effects, such as physical destruction, displacement and deteriorated socio-economic and health conditions, and indirect effects such as contamination of air and soil, as well as malnutrition.

The largest effects in terms of children reduced birthweight are observed when the mothers are exposed to domestic violence (reduction of 163 grams in Aizer (2011)) and violence as measured by homicide rates (reduction of 344 grams in Brown (2018)), macroeconomic shocks (reduction of 50-60 grams in Olafsson (2016) and Akbulut-Yuksel et al. (2020)), and natural disasters (reduction of 45-50 grams in De Oliveira et al. (2021) and Torche (2011)). The effects of all these factors on birthweight are comparable in magnitude to the effect of risky behaviours during pregnancy, such as tobacco consumption (Lien and Evans, 2005). If we focus on the effects on the higher probability of low birthweight (LBW), the size of the estimated effects is small and similar across different literatures – most papers find less than 2 percentage points.[7] Negative shocks *in utero* can also affect later human capital outcomes – estimated effects are up to 0.3 SD reduction in attained grades and test scores.

This paper is the first to rigorously examine the effect of the NATO bombing on a specific population subgroup affected by this event. The aim of this paper is to establish a causal link between the NATO bombing of Serbia on health and educational outcomes of children who were *in utero* during the bombing and were born between the months of June and October 1999. As such, we contribute to the literature on short- and medium-term effects of conflicts on future generations by shedding light on a conflict which has not been studied previously in

[7] An exception is the paper by Currie and Schwandt (2015), who find that exposure to the 9/11 dust cloud increased the probability of being born with low birthweight by 5.7 percentage points.

the literature. Our findings are comparable in size to the literature reviewed in Table 1.

The remainder of the paper is organised as follows. Section 2 discusses the methodology and the results of the short-term outcomes, such as birthweight. Section 3 looks at medium-term outcome in terms of educational achievement. Section 4 concludes the paper.

2 Short-Term Outcome: Birthweight

2.1 Data and Descriptive Statistics

We use the national registry of birth records from the Statistical Office of the Republic of Serbia (SoRS) to examine the impact of the NATO bombing on birthweight outcomes of children who were *in utero* during bombing. The birth records cover the whole population of births in Serbia and they include individual level information for each birth, such as whether a child was born alive, date of birth, gender, birthweight in 11 categories, and whether a child was born in a hospital or elsewhere. The dataset features socio-demographic information of mothers, including their place of residence, age, parity history (number of births that she had), marital status, educational background, and occupational status. Where possible, information on the father such as age, educational background and occupational status are also used.

Our main analysis is conducted for the years 1998 and 1999, and our robustness checks extend the pretreatment period up to 1996. In line with the previous literature (Quintana-Domeque and Ródenas-Serrano (2017), Bhalotra and Clarke (2014)), we exclude the following observations: births from mothers who were younger than 15 and older than 49, multiple births and newborns with birthweight below 500g. We also drop twins and stillbirths. The exclusion of twins is based on the findings of Bhalotra and Clarke (2014) that exposure to bomb casualties in the second and third trimester decreases the likelihood of multiple births. Stillbirths are only recorded if they occurred after the 27th gestational week, and we perform a separate analysis on the effect of bombing on the probability of stillbirths.

In our final sample, we define as treated those children who were *in utero* during the whole bombing period (78 days) and were born between June 10 (last day of bombing) and October 31, 1999.[8] Hence, children born during the bombing are excluded, to avoid the confounding effects of the exposure to bombs both *in utero* and in the early days after birth. Children conceived

[8] If we assume that the average pregnancy lasts 40 weeks or 280 days, the last children that could be included in our sample are those born on December 15, 1999. However, to take into account of the possibility of premature births, we restrict to children born up to October 31, 1999.

during bombing are excluded because there is evidence in the literature of postponing fertility during periods of war (Caldwell, 2006), and this, in turn, might have compositional effects.

In Figure A2a we graphically show that in the period between 1990 and 2010, the annual number of live births in Serbia had a declining trend, despite the positive improvements in the period between 2000 and 2005. In Figure A2b, we show the number of live births per month in the period between 1996 and 2003, and we shade the months June to October in each year. We observe that the monthly trend of births in the these months for the year 1999 is similar to the previous years.

In Table 2, in columns (1) we show the outcomes and background characteristics of babies in the treated group, and in columns (2), (4) and (5) the newborns in the control groups. Specifically, in column (2) we include infants born in the same calendar months of the treated children, but in the year prior to the bombing (June 10 to October 31, 1998). In columns (3) we show the statistics of infants born in the calendar months just before the bombing (January – mid March), observed in the year of the treatment (1999) and in the previous one. All the control groups are obtained considering children that we think are the most similar to the treated in terms of background characteristics. According to the World Health Organization a normal birthweight of an infant (term delivery) is between 2500-4200g, and above and below this range infants have low and high birthweight, respectively. A preliminary inspection of our data is reported in Table 2, panel A, for several measures of birthweight: categories of 500g, below the average ($< 3500g$), low ($< 2500g$) and high birthweight ($> 4500g$).

The tests of the difference in means, for newborns in columns (1) and (2), are reported in column (3) and are statistically significant. Overall, this suggests that children born from June to October 1999 had a lower weight compared to children born in the same period in 1998. We additionally observe that stillbirths among treated children were surprisingly lower than among children born in the same period in the previous year. We will explore this finding further, but it should be noted that the number of stillbirths is very low in both periods (143 in the observed period in 1998 and 108 in the same period of 1999). We repeat the tests for infants in the comparison groups in columns (4) and (5) and we show the results in column (6). We notice that there are no differences in outcomes between children born between January and mid-March in the year of the bombing and in the previous one.

In our analysis, we use as main outcome the measure of birthweight below average ($\leq$ 3500g) and we compare it to high birthweight ($\geq$ 3500g), because, as it is clear from Table 2, most of the variation is observed around the mean of the birthweight distribution.[9] We also re-define our dependent variable excluding newborns with low birthweight, to verify whether the estimated effect of the bombing is not driven by children in the tails of the distribution.

The distribution of birthweight of children born from January to March, 1999 and June to October, 1999 is shown in Figure A3a, while the birthweight of children born in the period June to October in 1998 and 1999 is shown in Figure A3b. Our treated group are children born from June to October, 1999. In the upper figure we observe a reduction in the 3500g-3999g and 4000-4499g category and a shift towards the 3000g-3499g category in the period June to October with respect to January to mid March. Similarly, in the lower figure where we compare treated children with children born in the same period in the previous year (1998), we observe a reduction in birth weight and a shift towards the 3000g-3499g category.

In Table 2, panel B, we show the individual background characteristics for the four groups. Among babies born from June to October, there are some statistically significant differences in characteristics between the years 1999 and 1998, however they are very small. For instances, the average age of the mother at birth was 11.009 in 1999 while it was 10.931 in 1998, hence this difference is smaller than one month overall. For the birth period January to mid-March, the statistically significant differences in some background characteristics are again very small. These differences disappear in the matching sample, as seen in Table A2.

2.2 Estimation Strategy

Our main identification strategy is based on a difference-in-differences (DD) estimation, that we first perform using a parametric approach, and then we combine it with a nonparametric propensity score matching methodology.

Difference-in-Differences (DD)

In our parametric model, the first difference is given by the comparison of birthweight of children born in the year of bombing, from June 10 to October 31, 1999, to those born from January 1 to mid-March, 1999. The second difference considers children born in the same months of the previous year, 1998. Our treated children are *in utero* for the whole 78 days of bombing.

[9]Negative long-term effects of reduced birthweight (LBW) are well documented. Small baby indicators are low birthweight (<2500g), very low birthweight (<1500g) or small-for-gestational age newborns.

We estimate the following regression:

$$Y_{itmdl} = \beta_0 + \beta_{DiD} I(June\_Oct)_{dm} \times Y1999_t + \beta_1 I(June\_Oct)_{dm} + \phi X'_{itmdl} + Y1999_t + \gamma_m + \tau_l + \epsilon_{itmdl} \quad (1)$$

where Y_{itmdl} is a binary variable for birthweight of newborn i in year t, in month m, in day d, in the municipality l. As mentioned in the previous section, we estimate the impact of NATO bombing using three definitions of dependent variable: below average birthweight; excluding low birthweight; stilbirths.[10]

The variable $I(June\_Oct)_{dm}$ takes the value one if the infant is born between June 10 and October 31 and it takes the value zero if the child is born from January 1 to mid-March. $Y1999_t$ is a dummy variable equal to 1 if the year is 1999, and zero if 1998. The coefficient of the interaction $I(June\_Oct)_{dm}$ and $Y1999_t$, is the DiD estimator which captures the impact of the NATO bombing on birthweight

The vector X'_{itmdl} contains the following individual level characteristics: gender of a baby, whether a baby was born in a hospital, a dummy variable if the parents are married, age of the mother, number of years of education of the mother, and a dummy variable indicating whether the mother is employed. In an extended model, we add the following father's characteristics: age of the father, number of years of education of the father, and a dummy variable indicating whether the father is employed. γ_m is a calendar-month fixed effect and τ_l is the municipality fixed effect. Standard errors are clustered at the municipality level.

The DD methodology relies on the parallel trends assumption. Therefore, we assume that in the absence of the bombing shock, the birthweight of babies born between June 10 and October 31, 1999 would have followed a similar trend as the birthweight of babies born between January and mid-March 1999. For our main outcome, we show the DD graph in Figure A4. Overall, the graph displays that there were no significant differences before the bombing event in the years 1995 to 1998 and suggests that a DD methodology is valid in this case.

Difference-in-Differences Matching

The second approach combines difference-in-differences and a nonparametric propensity score matching. The purpose is to find a control group of newborns more similar to the treated in all relevant pre-bombing characteristics. This implies the satisfaction of the conditional

[10] In the appendix in table A3 we provide results for two other outcomes (low birthweight ($< 2500g$) and high birthweight ($> 4500g$)), which were not impacted by the NATO intervention.

independence assumption (CIA): the selection into treatment is based solely on observable characteristics. The second assumption of the matching is the common support, which ensures that newborns with the same characteristics have a positive probability to be treated. Therefore, in a first stage we estimate the propensity score using a probit model of being born between June 10 and October 31 versus January and mid-March.

Following, Heckman et al. (1997) and Smith and Todd (2005), we estimate a DD matching regression, which allows for temporally invariant differences in outcomes between newborns in utero during the bombing and not. Indeed, the hidden bias due to the effect of unobserved heterogeneity is not required to vanish for any covariates but just to be the same before and after treatment. The estimator for repeated cross-section data is given by

$$\tau_{ATT} = \sum_{i \in T_t} \left(Y_{1ti} - \sum_{j \in C_t} W_{ij} \, Y_{0t} \right) w_{it} - \sum_{i \in T'_t} \left(Y_{0t'i} - \sum_{j \in C'_t} W_{ij} \, Y_{0t'j} \right) w_{it'} \qquad (2)$$

where Y_1 and Y_0 are the birthweight (as defined above) of the newborns in the treated and control groups; t' and t are the years before (1998) and after the bombing (1999). Precisely, $T_{t'}$ includes newborns between June 10 and October 31, 1998; $C_{t'}$ includes newborns between January and mid-March, 1998; T_t includes newborns between June 10 and October 31, 1999; C_t includes newborns between January and mid-March, 1999. W_{ij} is the weight obtained employing the nearest neighbour algorithm, and used to construct the counterfactual for the ith treated observation. w_{it} is the reweighing to reconstruct the outcome distribution for the treated sample. In our analysis, we only consider observations that are on the common support and we provide analytical standard errors (Abadie and Imbens, 2008).

2.3 Estimation Results

Main Results

Table 3 shows the main results of the effect of bombing on birthweight. In panel A, we use a parametric difference-in-differences model (see equation 1) and consider as dependent variable the birthweight below average. In column (1), including only individual controls and fixed effects, we find that being in utero during bombing increases the probability to be born below average birthweight by 2.2pp. When adding father controls the effect is slightly lower (1.9pp) but still highly statistically significant. In columns (3) and (4) we use a different definition of the dependent variable, obtained by excluding newborns with low birthweight. We do this because we want to verify whether the effect is driven by the reduction in birthweight around the mean of the distribution, as suggested in the descriptive statistics. This is, indeed, the case, because the effects of the bombing are unchanged.

We repeat the same analysis using a matching difference-in-differences model (see equation 2). The purpose is to obtain a control group more similar to the treated group in a series of characteristics observed before the bombing. In panel B, Table 3, the results show that the negative effect of bombing on the below-average birthweight of newborns is confirmed. We only observe a slightly larger impact (2.2pp) on the likelihood of being born below average birthweight when adding father controls, in columns (2) and (4). The size of the estimated effect is comparable to similar findings in the literature (De Oliveira et al., 2021; Torche, 2011; Currie and Rossin-Slater, 2013), which use natural disasters such as hurricanes, tropical storms and earthquakes as a source of prenatal exogenous variation and look at its effect on the probability of low birthweight (see Table 1). In the Appendix Table A2, we report, for each model of Table 3 panel B, the differences in means, before and after matching, between the covariates included in the propensity score. The validity of the procedure is confirmed by the large reduction in the standardised bias, which implies the non-rejection of the null hypothesis of equality of means after matching.

Intensity of Treatment

Bombing had an impact on the likelihood of being born below the average birthweight, but does this impact differ by the intensity of bombing? To answer this question we restrict our sample to bombed settlements which experienced an increasing number of days of bombing. We start from at least 1 day and we repeat the estimation by incrementing the minimum number of days up to 10. The estimation is performed using as dependent variable only the birthweight below average, and always including fixed effects, individual and paternal controls. We adopt a parametric difference-in-differences methodology and then we combine it with matching. The four comparison groups are the same used in the analysis reported in Table 3, however, the main difference is the sample restriction to *bombed settlements* only. Hence, we are excluding infants *in utero* born in settlements which were not bombed.

The results in Table 4 suggest that the impact of bombing on the likelihood of being born below average birthweight increases with the intensity of bombing. In column (1), restricting to settlements bombed at least one day, we find that the likelihood to be born below average birthweight increases by 1.5pp. When we restrict the sample to settlements experiencing at least two days of bombing, the effect is 1.8pp (column 2). When the days of bombing are at least five (column 3), there is a 70% increase in the effect, which jumps to 3.1pp and reaches 3.8pp when the bombing lasts more 10 days or more (column 4).

In panel B, Table 4, we show the results of the matching DD. In columns (1) and (2) the results are virtually unchanged, however, we observe an increase of at least 1pp when the intensity of bombing is 5 or more days. In particular, when the minimum number of bombing days is 10, the probability to a have a birthweight below average increases by 5pp (column 4).

In the last two columns of Table 4 we report the results of the estimation of the effect of bombing in the least affected settlements. We first rank all settlements in deciles based on the distance from the closest bombed settlement, and then we select those in the top 10% and top 20% of the distance distribution, which we define as the least affected settlements. The coefficients shown in columns (5) and (6), in both panels A and B, are very small and statistically not significant. This confirms that the effect of the bombing is not obtained by chance, and the estimation in the least affected settlements should be interpreted as a placebo test.

Spatial Variation of the Settlements

A further investigation has been performed exploiting the spatial variation of the settlements. We define two groups of settlements for this exercise. In the first group, we include the bombed settlements and we disaggregate the number of days bombed similar to Table 4. In the second group we include the not bombed settlements in the top 10% of the distance distribution,[11] which we define as the least affected settlements. We estimate a modified version of equation 1, where the first difference is between the birthweight of newborns *in utero* during the bombing (i.e. born in 1999 between June 10 and October 31) resident either in the first group of settlements or in the second; whereas, the second difference is identical except for the year of birth, 1998. Hence, we want to isolate the effect of the bombing in the bombed settlements only.[12]

The results[13] show that the effect is positive but small and not statistically significant. We conclude that for children *in utero* during the period of the bombing, the location of the settlements does not affect their birthweight.

[11] We have perfomed the same exercise with the not bombed settlements in the top 20% of the distribution and the results are similar to the top 10% of the distribution.

[12] This analysis differs from the placebo test in Table 4 because we only focus on infants born between June and October.

[13] Shown in the Appendix in Table A5.

Stillbirths

The stillbirth outcomes (delivery of a dead foetus at more than 28 weeks of gestation) is relevant for the present analysis for two main reason. First, the exposure to high levels of stress could lead to an increase in stillbirths. Findings from both medical sciences (Wisborg et al., 2008) and economics (Eccleston, 2011) suggest that prenatal maternal stress is linked with the increased risk of foetal death and stillbirth outcomes. Second, a higher mortality of children resulting from the NATO bombing could invalidate our identification strategy because it would change the composition of children born in the treated cohort.

We estimate the parametric difference-in-differences model in equation 1, using an outcome indicator which takes the value of 1 for stillbirths and the value 0 for live births. In Table 5, the results show that mortality of newborns was not affected by the bombing.

Placebo Tests

Our identification strategy using the difference-indifferences models rests on the parallel trends assumption, which presumes that in the absence of NATO bombing the birth outcomes of babies not exposed to bombing and those exposed to bombing would have been the same. To assess this assumption, we perform two placebo tests. We estimate the DD model in equation 1, considering in the first difference the birthweight of children born from June 10 to October 31, 1997, and of those born from January 1 to mid-March, 1997. The second difference considers children born in the same months of the previous year, 1996. We repeat the same analysis using the years of birth 1998 and 1997.[14] The results are reported in Table 6, panel A and B, and show that all coefficients are not statistically significant.

In Table A4, we show other placebo tests corresponding to the intensity of treatment analysis presented in Table 3. All coefficients are still not statistically significant (with one exception, significant at 10%). Overall, these tests provide additional evidence that we are in the right direction in identifying a causal impact of bombing on less than average birthweight.

[14]We cannot not run placebo tests for the years of birth after 1999, because there could be compositional effects due to delayed fertility in the post-bombing period. In the presence of compositional effects the placebo tests would not be valid.

3 Medium-Term Outcomes: Educational Achievement

3.1 Data and Descriptive Statistics

In order to capture the medium-term effects of NATO bombing, we use a dataset, provided by the Serbian Ministry of Education (SMoE), containing educational achievement of the whole population of pupils finishing primary school. In Serbia primary school lasts eight years from around age 7 to age 15, and we can observe pupils' grades in all subjects from the class in year 6 (P6) to the last class in year 8 (P8), when they sit a final examination.[15] After knowing the results of that final test, they express their preference for secondary school, indicating up to 20 choices. The assignment to a school then depends on the available slots and preferences of other pupils.

In our analysis we use teacher assessments, i.e., marks in mathematics, Serbian language and behaviour in the P8 class as outcomes. While test scores would be a very relevant outcome in our setting, we cannot use them for two reasons. First, the whole cohort finishing school in 2014 was affected by the NATO intervention to some degree and the standardised test scores are only comparable within the same cohort. As a result we do not have comparison pupils within the same cohort. Second, in the year 2013, which should be our main control year, the contents of the tests of the final examination were illegally sold to pupils before the actual examination[16] and, consequently, the test scores are not reliable. Conversely, teachers do not change within the same school, in the period under investigation, and the comparison across cohorts is possible. The grades vary from a minimum of 1 which corresponds to a fail to a maximum of 5, whereas 2 is a pass. We also consider two additional variables related to the secondary school preferred choices. The first variable is a dummy equal to 1 if students prefer a 4-year secondary school track, which usually leads to university, and equal to zero if they prefer a 3-year vocational secondary school track. The second variable indicates the school track which they have actually enrolled.

The rules on the primary school starting age include in the same cohort pupils born between March 1 and February 28 in the following year. Consequently, in the treated group we can only

[15] The pupils can formally finish primary school only if they sit the examination, containing tests in mathematics, Serbian language and a mix of different subjects (geography, chemistry, physics, history and biology). The latter was only introduced in 2014. The total score of these tests together with the average grades from P6 to P8 class count for the admission to secondary school. Pupils are assigned to the secondary schools based on the results of the final examination, average grades from the P6 to the P8 class, as well as the results of the pupils' competitions in the P8 class of primary school, following an algorithm. For more information, visit this link (*in Serbian*).

[16] Some media sources referring to this event can be found following this link (*in Serbian*).

consider children born between June 10 and October 31, 1999. We exclude those born between March 1 and June 10 because they are not in utero for the whole period. Those born between January and February 1999 are not in the school cohort 1999, therefore they are not a valid control group. We therefore use the cohort before the bombing, those born between June 10 and October 31, 1998, and the cohort born in the same months of 2000, one year after the bombing.

Unfortunately, we have limited information on individual characteristics, except for the gender, but we know the date of birth, the municipality of birth and the school name and location (which corresponds to the residence of the pupil).

Table 7, Panel A, shows the average grades of pupils *in utero* in column (1), pupils born in the year before bombing (control group 1) in column (2), and pupils born in the year after bombing (control group 2) in column (3). In column (4) and (5) we report the test of the difference between the outcomes of treated and not treated pupils, in the two control groups, respectively. It is clear that pupils who were *in utero* during the intervention have statistically significant lower grades in Serbian language and mathematics, while no differences are found for behaviour. The preference for a four-year secondary school track is smaller for the treated pupils, indeed there is a negative and statistically significant difference when compared to the pupils in the control group (see columns 4 and 5). However, in terms of the corresponding enrolments the same differences are smaller and less precise. In Panel B, we report the only background characteristic available, and we do not observe any statistically significantly difference in the gender composition of the treated and control groups.

3.2 Estimation Strategy

Our main identification strategy for the long-term outcomes cannot replicate the same structure used for the short-term outcomes because of the rules on the starting age at primary school. Hence, we employ an inverse probability weighting regression-adjustment (IPWRA), a quasi-experimental approach (Imbens and Wooldridge, 2009; Cattaneo, 2010), which involves a two-stage estimation process. The first stage estimate a probit model to account for the effects of (pupils) observed variables on the probability to be in utero during the bombing, and computes inverse probability weights. The second stage use those weights to fit weighted regression models of the outcome (primary school grades) for pupils in the treated and the control group, and computes the difference of the corresponding predicted outcomes. Such difference provides an estimate of the average treatment effect, which is a consistent estimator if the conditional independence assumption, $(Y_1, Y_0 \perp D \,|\, p(X))$, and common support, $(0 < Pr(D = 1|X) < 1)$, hold. We estimate the following model

$$\tau_{ATE} = N^{-1} \sum_{i=1}^{N} (E[Y_{itmds}|\mathbf{X_i}, \gamma_m, T_t = 1] - E[Y_{itmds}|\mathbf{X_i}, \gamma_m, T_t = 0]) \quad (3)$$

where Y_{itmds} is the schooling outcome of child i, born in year t, month m, day d, in the school s. The outcomes of interest are P8 marks in mathematics, Serbian language and behaviour, high school track preferences and high school actual enrolment. X_i includes gender and month of birth of a pupil, γ_m are municipality of birth fixed effects. We cluster the standard errors at the school level. T_t is equal 1 if children were *in utero* during the NATO bombing (born between June 10, 1999 and October 31, 1999) and equal 0 for children born in the same month of previous year (control group 1), or children born from 10 June to end of October in the year after the bombing (control group 2). To test the robustness of our results to omitted variables we perform a placebo test, using as treated the cohort born in 1998 compared to the cohort born in 1997. We also run an Oster test (Oster, 2019), where we vary the value of the maximum R^2 and the level of the relative degree of selection on observed and unobserved variables, δ, up to the point that makes the average treatment effects (ATE) not statistically significant.

3.3 Estimation Results

Main Results

We show in Table 8 the causal impact of being *in utero* during the NATO bombing on medium-term schooling outcomes. Columns (1)-(3) show the results for the grade outcomes, and we observe that being in utero during the bombing has a negative and statistically significant effect on language and mathematics, using both control group 1 (Panel A) and control group 2 (Panel B). The magnitude of the effect is slightly higher in maths, around -0.03 points, which corresponds to a 1% reduction on the average grade. For pupils in the treated group we also notice a negative and highly statistically significant reduction in the probability to choose a 4-year high school track (column 4), which is confirmed by lower enrolment in the same type of school (column 5). Overall, the size of the reduction is a bit higher in Panel A, when we consider as control group the cohort born the year before the bombing.

Placebo Tests

In Table A6, we show the results of a placebo test using 1997 as year of treatment and 1996 as control, and we notice that there is no statistically significant effect on any outcome, except for enrolment in 4-year secondary school track, but the sign is actually positive.

In Table A7, we report the results of the Oster test applied to the model in Panel A of Table 8. We show how the coefficient of our treatment variable, being *in utero* during the bombing, changes for different levels of R^2_{max} and degree of selection on unobservables with respect to selection on observable, δ. The lower bound effect is obtained setting the highest possible levels of $R^2_{\text{max}} = 1$ and $\delta = 1$, and indeed we do not have any statistically significant effect. However, the coefficients for mathematics, Serbian language and high school first choice remain statistically significant for a level of R^2_{max} up to 0.3, and $\delta = 1$. Keeping $R^2_{\text{max}} = 1$, the effects on maths and first choice remain still statistically significant up to $\delta = 0.254$. The same happens, when we assume intermediate levels of R^2_{max} and δ, i.e., both at 0.5. The effect on behaviour has never been significant in the main model, so the results of the Oster test does not add any additional information. Nevertheless, taking into account that in the analysis of the medium-term outcomes we do not have many available covariates to estimate the IPWRA model, we can conclude that our estimated coefficients are robust.

4 Conclusions

This paper estimates the causal effect of NATO's Operation Allied Force in Serbia on children who were in the womb during the bombing. We examine the so-called *in utero* effect on children, both in terms of short-term outcomes, such as birthweight, as well as medium-term outcomes, such as primary school grades and preferences and secondary school enrolment. Our main identification strategy uses a difference-in-differences approach, combined with propensity score matching, to first compare children *in utero* during the bombing, born between June and October 1999, with children born between January and March of the same year, and second to children born in the same months of 1998. We use the birth records from the Statistical Office of the Republic of Serbia, and we find that children *in utero* during the bombing had a higher likelihood of having lower than average birthweight, compared to children in the control group. The magnitude of the effect is about 2pp and it is comparable to the effects found by De Oliveira et al. (2021), Torche (2011), and Currie and Rossin-Slater (2013), which use natural disasters such as hurricanes, tropical storms and earthquakes as a source of prenatal exogenous variation and look at its effect on the probability of low birthweight (see Table 1).

For the analysis on the medium-term outcomes, due to the primary school starting age rule in Serbia, we cannot separate pupils within the same cohort to perform a DD estimation. Therefore, we compare children born between June and October 1999 with those born in the same period of 1998, and we adopt an inverse probability weighting regression-adjustment (IPWRA) approach. We use administrative data from Serbian Ministry of Education on educational achievement at the of primary school, and we find that children in *in utero* during the bombing had statistically significant lower marks in mathematics and Serbian language of about 1% (or about 0.03 standard deviations); and a reduction of the same size on the preference and enrolment to non-vocational high schools. Compared to the findings in the disaster literature, Almond et al. (2015b) find that academic test scores are 0.05-0.08 standard deviations lower for students exposed to Ramadan in early pregnancy, while Almond et al. (2009) find that exposure to radioactive fallout from the 1986 Chernobyl disaster between weeks 8 and 25 of gestation reduces marks in mathematics by 3-6% (see Table 1). One explanation for the smaller estimated effect on primary school results in our paper might be due to using teacher assessment rather than test scores – it might be that teacher assessments positively favoured this cohort of students. However small, policy makers should seriously consider these estimated negative effects, not only because pupil's performance on mathematics is a useful measure of cognitive skills, but also because it is a good indicator for future educational and labour market outcomes (Machin and McNally, 2008; Schrøter Joensen and Skyt Nielsen, 2009).

There are many ways for a lethal catastrophe such as bombing to affect the pregnant mothers. There are both direct channels, such as physical destruction, displacement and deteriorated socio-economic and health conditions, as well as indirect ones, such as contamination of air and soil, as well as malnutrition. Considering that the goal of bombing was to maximise material damage and limit collateral (civilian) damage, we can rule out that the observed effect is due to the direct (physical) destruction. As shown in the paper, the mothers of the treated children didn't have significantly higher number of stillbirths, hence we can also rule out this potentially confounding mechanism. Using the Multiple Cluster Indicator Survey (MICS) data in 2000 conducted just one year after the bombing,[17] Table A1 compares the mothers of the

[17]The Multiple Cluster Indicator Survey (MICS), administered on a five-year basis, is a household survey implemented by countries under the programme developed by the United Nations Children's Fund (UNICEF). It is designed to provide internationally comparable, statistically rigorous data on key social indicators on the most sensitive part of the population such as mothers, children and vulnerable and marginalised groups. As such, the MICS survey aims to collect and analyse the data necessary to monitor the situation of women, children as well

treated children to the mothers of the control children on a range of socio-economic, health and behavioural outcomes. We see that the two groups didn't have different behavioural and health outcomes, as measured by the exposures to health and crime risks, as well as changes in alcohol, food and physical activity consumption. Therefore, we conclude that mothers of the treated and the control children had the same socio-economic, health conditions and access to prenatal care. Due to the United Nations (UN) sanctions against Yugoslavia, as well as tightened visa travel regime for its citizens, migratory movements out of the country were limited. Similar to Currie and Rossin-Slater (2013), prenatal maternal stress is the 'residual' transmission mechanism of *in utero* environment of both mother and the child (Aizer et al., 2016; Black et al., 2016; Berthelon et al., 2018; Persson and Rossin-Slater, 2018).[18]

One policy implication could be that governments need to intervene and design policies to alleviate the negative *in utero* effects on children in the aftermath of large-scale disasters. Another policy implication questions bombing as a legitimate tool of intervention in the international conflicts – this type of interventions should be re-evaluated, taking all possible consequences into account.

as vulnerable and marginalised groups in terms of education, health, child protection, HIV/AIDS, etc.

[18]Contamination of the soil could be a potential channel for the more long-term outcomes and for a different treated group than the one considered in the paper.

References

Abadie, A. and G. W. Imbens (2008). On the failure of the bootstrap for matching estimators. *Econometrica 76*(6), 1537–1557. 12

Aizer, A. (2011). Poverty, Violence, and Health: The Impact of Domestic Violence during Pregnancy on Newborn Health. *The Journal of Human Resources 46*(3), 518–38. 7

Aizer, A., L. Stroud, and S. Buka (2016). Maternal Stress and Child Outcomes: Evidence from Siblings. *The Journal of Human Resources 51*(3), 523–555. 4, 6, 7, 21

Akbulut-Yuksel, M. (2014). Children of War: The Long-Run Effects of Large-Scale Physical Destruction and Warfare on Children. *Journal of Human Resources 49*(3), 634–662. 3

Akbulut-Yuksel, M., S. Mumin Cilasun, and B. Turan (2020). Children of Crisis: The Effects of Economic Shocks on Newborns. IZA DP No. 12898. 7

Almond, D. and J. Currie (2011). Killing Me Softly: The Fetal Origins Hypothesis. *Journal of Economic Perspectives 25*(3), 153–172. 6

Almond, D., J. Currie, and V. Duque (2018). Childhood Circumstances and Adult Outcomes: Act II. *Journal of Economic Literature 56*(4), 1360–1446. 6

Almond, D., L. Edlund, and M. Palme (2009). Chernobyl's subclinical legacy: prenatal exposure to radioactive fallout and school outcomes in Sweden. *The Quarterly Journal of Economics 124*(4), 1729–1772. 7, 20

Almond, D. and B. Mazumder (2011). Health Capital and the Prenatal Environment: The Effect of Ramadan Observance During Pregnancy. *American Economic Journal: Applied Economics 3*(4), 56–85. 7

Almond, D., B. Mazumder, and R. van Ewijk (2015a). In Utero Ramadan Exposure and Children's Academic Performance. *The Economic Journal 125*(589), 1501–1533. 7

Almond, D., B. Mazumder, and R. van Ewijk (2015b). In Utero Ramadan Exposure and Children's Academic Performance. *The Economic Journal 125*(589), 1501–1533. 20

Andalón, M., J. P. Azevedo, C. Rodríguez-Castelán, V. Sanfelice, and D. Valderrama-González (2016). Weather Shocks and Health at Birth in Colombia. *World Development 82*, 69–82. 6

Arkin, W. M. (2000). *Civilian deaths in the NATO air campaign*, Volume 12. Human Rights Watch. 46

Armijos Bravo, G. and J. Vall Castelló (2021). Terrorist Attacks, Islamophobia and Newborns' Health. IEB Working Paper 2021/05. 6

Barker, D. J. P. (2001, Jun). Fetal and infant origins of adult disease. *Monatsschrift Kinderheilkunde 149*(1), S2–S6. 6

Ben-Shlomo, Y. and D. Kuh (2002, 04). A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. *International Journal of Epidemiology 31*(2), 285–293. 6

Berthelon, M., D. Kruger, and R. Sanchez (2018). Maternal Stress during Pregnancy and Early Childhood Development. IZA DP No. 11452. 4, 7, 21

Bhalotra, S. and D. Clarke (2014). The Twins Instrument and Family Size. University of Essex, mimeo. 8

Black, S. E., P. J. Devereux, and K. G. Salvanes (2016). Does Grief Transfer across Generations? Bereavements during Pregnancy and Child Outcomes. *American Economic Journal: Applied Economics 8*(1), 193–223. 4, 7, 21

Bozzoli, C. and C. Quintana-Domeque (2014, 07). The Weight of the Crisis: Evidence From Newborns in Argentina. *The Review of Economics and Statistics 96*(3), 550–562. 7

Brown, R. (2018). The Mexican Drug War and Early-Life Health: The Impact of Violent Crime on Birth Outcomes. *Demography 55*, 319–340. 7

Brown, R. (2020). The Intergenerational Impact of Terror: Did the 9/11 Tragedy Impact the Initial Human Capital of the Next Generation? *Demography 57*, 1459–1481. 6

Bulajić, M. (2000). NATO zločini u Jugoslaviji-kvalifikacija zločina genocida. In *NATO agresija na SR Jugoslaviju 99 - Zbornik radova*. Novi Sad: Udruženje za pravnu teoriju i praksu. 46

Caldwell, J. C. (2006). Social upheaval and fertility decline. In *Demographic Transition Theory*, pp. 273–299. Springer. 9

Calzada, J., M. Gisbert, and B. Moscoso (2021). The hidden cost of bananas: Pesticide effects on newborns' health. University of Barcelona Economics Working Paper No. 405. 7

Cattaneo, M. D. (2010). Efficient semiparametric estimation of multi-valued treatment effects under ignorability. *Journal of Econometrics 155*(2), 138–154. 17

Currie, J., M. Mueller-Smith, and M. Rossin-Slater (2018, July). Violence while *in Utero*: The impact of assaults during pregnancy on birth outcomes. Working Paper 24802, National Bureau of Economic Research. 7

Currie, J. and M. Rossin-Slater (2013). Weathering the Storm: Hurricanes and Birth Outcomes. *Journal of Health Economics 32*(3), 487–503. 6, 13, 19, 21

Currie, J. and H. Schwandt (2015). The 9/11 dust cloud and pregnancy outcomes: A reconsideration. *The Journal of Human Resources 51*(4), 805–831. 7

De Oliveira, V. H., I. Lee, and C. Quintana-Domeque (2021). Natural Disasters and Early Human Development: Hurricane Catarina and Infant Health in Brazil. *Journal of Human Resources*. Pre-print, March. 7, 13, 19

Duque, V. (2017). Early-life conditions and child development: Evidence from a violent conflict. *SSM - Population Health 3*, 121 – 131. 6

Eccleston, M. (2011). *In Utero* Exposure to Maternal Stress: Effects of the September 11th Terrorist Attacks in New York City on Birth and Early Schooling. Working Paper. 6, 15

Fenrick, W. (2001). Targeting and Proportionality during the NATO Bombing Campaign against Yugoslavia. *European Journal of International Law 12*(3), 489–502. 5

Guantai, F. and Y. Kijima (2020). Ethnic violence and birth outcomes: Evidence from exposure to the 1992 conflict in kenya. *Demography 57*, 423 – 444. 7

Heckman, J. J., H. Ichimura, and P. E. Todd (1997). Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *The Review of Economic Studies 64*(4), 605–654. 12

Imbens, G. W. and J. M. Wooldridge (2009, March). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature 47*(1), 5–86. 17

Kelly, E. (2011). The scourge of asian flu *in utero* exposure to pandemic influenza and the development of a cohort of British children. *Journal of Human Resources 46*(4), 669–694. 7

Kim, B., C. K. Carruthers, and M. C. Harris (2017). Maternal stress and birth outcomes: Evidence from the 1994 Northridge earthquake. *Journal of Economic Behavior & Organization 140*, 354–373. 6

Koppensteiner, M. F. and M. Manacorda (2016). Violence and birth outcomes: Evidence from homicides in brazil. *Journal of Development Economics 119*, 16 – 33. 7

Lien, D. S. and W. N. Evans (2005). Estimating the Impact of Large Cigarette Tax Hikes: The Case of Maternal Smoking and Infant Birth Weight. *The Journal of Human Resources 40*(2), 373–392. 7

Machin, S. and S. McNally (2008). The literacy hour. *Journal of Public Economics 92*(5), 1441–1462. 20

Mansour, H. and D. I. Rees (2012). Armed conflict and birth weight: Evidence from the al-Aqsa Intifada. *Journal of Development Economics 99*(1), 190–199. 6

Menclova, A. K. and S. Stillman (2020). Maternal stress and birth outcomes: Evidence from an unexpected earthquake swarm. *Health Economics 29*(12), 1705–1720. 6

Olafsson, A. (2016). Household Financial Distress and Initial Endowments: Evidence from the 2008 Financial Crisis. *Health Economics 25*(S2), 43–56. 7

Oskorouchi, H. R. (2019). Learning to Fight: Afghan Child Health and In-utero Exposure to Conflict. *Population and Development Review 45*(2), 275–300. 5

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics 37*(2), 187–204. 18

Pebesma, E. and R. S. Bivand (2005). Classes and Methods for Spatial Data: the sp Package. *R news 5*(2), 9–13. 46

Persson, P. and M. Rossin-Slater (2018). Family Ruptures, Stress, and the Mental Health of the Next Generation. *American Economic Review 108*(4–5), 1214–1252. 4, 7, 21

Quintana-Domeque, C. and P. Ródenas-Serrano (2017). The Hidden Costs of Terrorism: The Effects on Health at Birth. *Journal of Health Economics 56*, 47–60. 6, 8

R Core Team (2019). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. 46

Rosales-Rueda, M. (2018). The impact of early life shocks on human capital formation: evidence from El Niño floods in Ecuador. *Journal of Health Economics 62*, 13–44. 7

Sardoschau, S. (2018). Children of War: In-Utero Stress and Child Health in Iraq. Paris School of Economics and Université de Paris Working Paper. 5

Schrøter Joensen, J. and H. Skyt Nielsen (2009). Is There a Causal Effect of High School Math on Labor Market Outcomes? *The Journal of Human Resources 44*(1), 171–198. 20

Smiljanić, S. (2009). *Agresija NATO na SRJ, Vojska Jugoslavije u odbrani od agresije 1999. godine, knjiga prva.* Beograd: Svedočanstva. 46, 47

Smith, J. A. and P. E. Todd (2005). Does matching overcome LaLonde's critique of non-experimental estimators? *Journal of Econometrics 125*(1), 305–353. Experimental and non-experimental evaluation of economic policy and models. 12

Torche, F. (2011). The Effect of Maternal Stress on Birth Outcomes: Exploiting a Natural Experiment. *Demography 48*(4), 1473–1491. 6, 7, 13, 19

Van den Berg, G. J. and M. Lindeboom (2018). Famines, Hunger, and Later-Life Health. Oxford Research Encyclopaedia of Economics and Finance. Oxford University Press. 6

Wisborg, K., A. Barklin, M. Hedegaard, and T. Henriksen (2008). Psychological stress during pregnancy and stillbirth: prospective study. *BJOG: An International Journal of Obstetrics & Gynaecology 115*(7), 882–885. 15

Tables

Table 1: Comparison of the In Utero Results Across Literature

				Results		
Type exogenous variation	Event	Country	Year of the event	BW (grams)	P(LBW)	Grades and test scores
A. Wars, bombing, terrorist and violent attacks (‘disaster literature’)						
Eccleston (2011)	The 9/11	US (NYC)	2011	↓8-19g	-	-
Mansour and Rees (2012)	Fatalities from the al-Aqsa Intifada	Palestine (West Bank)	2004	↓2.1g	↑0.10-0.27pp	-
Quintana-Domequea and Ródenas-Serrano (2017)	ETA terrorism (Hipercor bombing)	Spain (Barcelona)	1987	↓3g	↑0.015pp	-
Duque (2017)	Terrorist attacks	Colombia	1999-2007	-	↑0.01pp	↓0.04-0.25SD
Brown (2020)	The 9/11	US (NYC/DC)	2011	↓15g	↑0.4-0.5pp	-
Armijos Bravo and Val Castello (2021)	Jihadi terrorist attacks (Muslim women)	Spain (Catalonia)	2017	↓12.89g	↑1.6pp -	
This paper (2021)	The NATO bombing	Serbia	1999	-	↑2pp	↓1% (0.03SD)
B. Natural disasters (famines, hurricanes, pandemics)						
Torche (2011)	Tarapaca earthquake	Chile	2005	↓51g	↑1.8pp	-
Kelly (2011)	Asian flu	Britain	1957	↓0.02-0.04SD	-	↓0.06-0.07SD
Currie and Rossin-Slater (2013)	Hurricane/tropical storm	US (Texas)	1996-2008	-	↑1.5pp	-
Andalon et al. (2016)	Hot and cold temperature shocks	Colombia	1999-2008	↓4.1g	-	-
Kim et al. (2017)	Northridge earthquake	US (LA)	1994	↓9-11g	↑0.2-0.5pp	-
Rosales-Rueda (2018)	El Niño floods	Ecuador	1997-98	-	-	↓0.10SD
Menclova and Stillman (2020)	Earthquake	New Zealand	2010	↓10g	-	-
De Oliveria et al. (2021)	Hurricane Catarina	Brazil	2004	↓44.4g	↑1.7pp	-
C. Violence and domestic violence						
Aizer (2010)	Domestic violence	US (California)	1991-2002	↓163g	-	-
Koppensteiner and Manacorda (2015)	Violence measured by homicide rates	Brazil	2000-2010	↓2.5g	↑0.13pp	
Brown (2018)	Violence measured by homicide rates	Mexico	2003-2009	↓344g	↑2.5-5pp	-
Currie et al. (2019)	Intimate partner violence	US (NYC)	2004-2012	-	↑1.7pp	-
Guantai and Kijima (2020)	Electoral violence	Kenya	1992	↓271g	↑2.1pp	-
D. Other ‘mild’ shocks						
Almond et al. (2009)	Chernobyl fallout	Sweden	1986	No effect	-	↓3-6%
Almond and Mazumder (2011)	Ramadan fast	US (Michigan)	1989-2006	↓18-25g	-	-
Bozzoli and Quintana-Domeque (2014)	Macroeconomic crisis	Argentina	2000-2005	↓34-35g	↑0.007pp	-
Almond et al. (2014)	Ramadan fast	England	1998-2007	-	-	↓0.05-0.08SD
Currie and Schwandt (2015)	The 9/11 dust cloud	US (NYC)	2011	-	↑5.7pp	-
Black et al. (2016)	Loss of a grandparent	Norway	1967-2009	↓21g	-	-
Olafsson (2016)	Financial crisis	Iceland	2008	↓66g	↑1.9pp	-
Persson and Rossin-Slater (2018)	Family death	Sweden	1973-2011	↓11g	↑0.39pp	-
Akbulut-Yuksle et al. (2020)	Economic crisis	Turkey	2001-2008	↓50g	-	-
Calzada et al. (2021)	Aerial fumigation of banana plantations	Ecuador	2015-2017	↓29-89g	↑0.35pp	-

Notes: This paper: P(below-average BW). Reduction in grades in mathematics and Serbian language. Brown (2020): Children exposed in utero to increased maternal stress due to the terrorist attacks of September 11. Intrauterine growth is restricted by the exposure in the first trimester. Quintana-Domequea and Ródenas-Serrano (2017): Results interpreted in terms of ten additional bomb casualties. Duque (2017): In utero exposure is most detrimental in the first trimester. She looks at a decline in math reasoning. Rosales-Rueda (2018): The effect on the Peabody Picture Vocabulary Test (PPVT) is measured as three months exposure to floods while in utero. Kelly (2011): The negative birthweight is found for short mothers and mothers who smoke. Test score results at ages 7 and 11. Koppensteiner and Manacorda (2015): Effects are concentrated in the first trimester. Brown (2018): The estimated effect on the P(LBW) refer to the bottom 50% of per capita expenditure. Almond et al. (2014): Test score outcomes at the age of 7. Almond et al. (2009): Grade average and test score in mathematics and Swedish at the end of compulsory school (at age sixteen). Also has info on qualifications for high school. Currie and Schwandt (2015): The effect is for the 1st trimester and boys. Akbulut-Yuksle et al. (2020): The effect is for the affected children in the poorest regions, compared to the newborns in the region with average GDP.

Table 2: Birth Records: Mean Comparison t-Tests of Outcomes and Individual Characteristics

Month of birth	Jun.-Oct.		Diff.	Jan.-mid Mar.		Diff.
Year of birth	1999	1998	p-value	1999	1998	p-value
	(1)	(2)	(3)	(4)	(5)	(6)
A.Birth weight						
Birth weight (categorical)			<0.001***			0.72
500g-999g	0.1%	0.1%		0.1%	0.1%	
1000g-1499g	0.5%	0.4%		0.4%	0.3%	
1500g-1999g	0.9%	1.1%		1.0%	1.0%	
2000g-2499g	3.1%	3.0%		3.0%	3.0%	
2500g-2999g	16.7%	16.4%		16.4%	16.3%	
3000g-3499g	40.8%	39.5%		38.6%	39.1%	
3500g-3999g	28.4%	29.8%		29.6%	30.1%	
4000g-4499g	8.5%	8.5%		9.5%	8.9%	
More than 4500g	1.0%	1.3%		1.3%	1.3%	
Below average birth weight (<3500g)	62.1%	60.5%	<0.001***	59.5%	59.8%	0.68
Low birth weight (<2500g)	4.6%	4.6%	0.93	4.5%	4.5%	0.79
High birth weight (>4500g)	1.0%	1.3%	<0.001***	1.3%	1.3%	0.68
Stillbirths	0.04%	0.5%	-0.1%**	0.06%	0.06%	0.0%
B. Individual charasteristics						
Female	47.9%	48.2%	0.38	48.3%	48.7%	0.52
Born in hospital	98.8%	98.9%	0.084*	98.8%	99.1%	0.006***
Parents married at birth	79.7%	79.8%	0.69	80.6%	81.0%	0.45
Mother's years of education	11.009	10.931	0.001***	10.967	10.964	0.93
Mother employed	40.0%	39.7%	0.34	39.9%	40.1%	0.75
Mother's age	26.098	25.926	<0.001***	26.034	25.927	0.11
Number of years married	3.423	3.355	0.048**	3.926	3.874	0.31
Has father data	86.0%	86.3%	0.26	86.2%	87.8%	<0.001***
Father's years of education	11.649	11.629	0.35	11.623	11.561	0.053**
Father employed	83.4%	85.7%	<0.001***	84.0%	84.8%	0.088*
Father's age	30.033	29.897	0.009***	29.925	29.851	0.34
Observations	27,016	29,348		12,640	12,749	

Notes: The children affected by bombing are born from June 10, 1999 to October 31, 1999 and their outcomes and characteristics are reported in column (1). Column (2) reports the outcomes and characteristics of children born from June 10, 1998 to October 31, 1998. Column (3) reports the p-value of the differences between 1999 and 1998 for the given period. Column (4) shows the characteristics of children born prior to bombing in the same year, in the period from January 1 to mid-March 1999. Column (5) shows the characteristics of children born from January 1 to mid-March 1998. Column (6) shows the p-value of the differences of children born from January 1 to mid-March, in years 1999 and 1998. p-values: * significant at 10%, ** significant at 5%, *** significant at 1%. Source: Birth records from the Statistical Office of the Republic of Serbia.

Table 3: The Impact of NATO Bombing on Birth weight Outcomes – Main Results

	Below average birthweight (<3500g)			
	(1)	(2)	(3)	(4)
A. Difference-in-Differences				
	0.022***	0.019**	0.022***	0.019**
	(0.007)	(0.008)	(0.008)	(0.008)
Dep. var. mean	0.601	0.582	0.583	0.566
Dep. var. SD	0.490	0.493	0.493	0.496
Observations	79,343	69,109	75,767	66,475
Adj. R-squared	0.046	0.037	0.044	0.037
B. Matching Difference-in-Differences				
	0.022***	0.021***	0.023***	0.0212**
	(0.008)	(0.008)	(0.008)	(0.008)
Dep. var. mean	0.600	0.582	0.582	0.565
Dep. var. SD	0.490	0.493	0.493	0.495
Observations	77,771	67,996	74,324	65,422
Controls	X	X	X	X
Father controls		X		X
Month of birth FE	X	X	X	X
Year of birth FE	X	X	X	X
Municipality FE	X	X	X	X

Notes: This table presents estimated baseline coefficients for the differences-in-differences model in panel A, and difference-in-differences matching model in panel B. Columns (1) and (2) include the whole sample, in columns (3) and (4) we exclude children with weight below 2500 grams. Controls: age, years of education and employment status. Additional father controls: age, years of education and employment status. All regressions include month, year and municipality fixed effects. Standard errors clustered at municipality level in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%. *Source:* Birth records from the Statistical Office of the Republic of Serbia.

Table 4: The Impact of NATO Bombing on Below Average Birthweight Within Bombed Settlements by the Number of Days Experiencing Bombing

	Below average birthweight (<3500g)					
	Number of days bombed in settlement				Top distance	
	$1 \leq$	$2 \leq$	$5 \leq$	$10 \leq$	10%	20%
	(1)	(2)	(3)	(4)	(5)	(6)
A. Difference-in-Differences						
	0.015*	0.018**	0.031**	0.038*	0.002	0.010
	(0.008)	(0.008)	(0.013)	(0.020)	(0.027)	(0.019)
Dep. var. mean	0.589	0.588	0.590	0.575	0.635	0.628
Dep. var. SD	0.492	0.492	0.492	0.494	0.481	0.483
Observations	39,315	32,345	17,593	6,974	7,685	15,121
Adj. R-squared	0.043	0.042	0.044	0.035	0.057	0.051
B. Matching Difference-in-Differences						
	0.013	0.0174*	0.041**	0.053**	0.000	0.011
	(0.009)	(0.010)	(0.015)	(0.023)	(0.028)	(0.020)
Dep. var. mean	0.589	0.588	0.590	0.575	0.635	0.628
Dep. var. SD	0.492	0.492	0.492	0.494	0.481	0.483
Observations	39,315	32,343	17,590	6,974	7,683	15,115
Controls	X	X	X	X	X	X
Father controls	X	X	X	X	X	X
Month of birth FE	X	X	X	X	X	X
Year of birth FE	X	X	X	X	X	X
Municipality FE	X	X	X	X	X	X

Notes: This table presents estimated coefficients for different levels of bombing intensity for the difference-in-differences model in panel A, and differences-in-differences matching model in panel B. In columns (1) through (4) we restrict our sample to settlements experiencing at least one day of bombing (col. (1)), at least two days of bombing (col(2)), at least 5 days of bombing (col. (3)), and at least 10 days of bombing (col. (4)). Controls: female, born in hospital, parents married and the characteristics of the mother: age, years of education and employment status. Additional father controls: age, years of education and employment status. All regressions include month, year and municipality fixed effects. Standard errors clustered at municipality level in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%. In column (5) we restrict the sample to the top 10% of settlements furthest away from the closest bombed settlement, and in column (6) to the top 20% of settlements furthest away from the closest bombed settlement. *Source:* Birth records from the Statistical Office of the Republic of Serbia.

Table 5: The Impact of NATO Bombing on Stillbirths

	Stillbirths	
	(1)	(2)
DiD	-0.001	-0.002
	(0.001)	(0.002)
Dep. var. mean	0.005	0.004
Dep. var. SD	0.072	0.066
Observations	79,738	69,393
Adj. R-squared	0.003	0.002
Controls	X	X
Father controls		X
Municipality FE	X	X

Notes: This table presents estimated baseline coefficients for the differences-in-differences model. Stillbirths include only births after the 28th week of pregnancy. Controls: female, born in hospital, parents married and the characteristics of the mother: age, years of education and employment status. Additional father controls: age, years of education and employment status. All regressions include month, year and municipality fixed effects. Standard errors clustered at municipality level in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%. *Source:* Birth records from the Statistical Office of the Republic of Serbia.

Table 6: The Impact of NATO Bombing on Birthweight Outcomes – Placebo Years

	Below average birthweight (<3500g)			
	(1)	(2)	(3)	(4)
A. Placebo years 1996 – 1997				
DiD	0.002	0.002	0.003	0.003
	(0.007)	(0.008)	(0.007)	(0.008)
Dep. var. mean	0.602	0.585	0.582	0.568
Dep. var. SD	0.490	0.493	0.493	0.495
Observations	84,844	75,810	81,006	72,876
Adj. R-squared	0.047	0.039	0.045	0.039
B. Placebo years 1997 – 1998				
	-0.011	-0.012	-0.011	-0.012
	(0.008)	(0.008)	(0.008)	(0.008)
Dep. var. mean	0.609	0.592	0.591	0.576
Dep. var. SD	0.488	0.492	0.492	0.494
Observations	81,385	71,808	77,752	69,099
Adj. R-squared	0.047	0.038	0.045	0.037
Controls	X	X	X	X
Father controls		X		X
Month of birth FE	X	X	X	X
Year of birth FE	X	X	X	X
Municipality FE	X	X	X	X

Notes: This table presents estimated baseline coefficients for placebo years for the difference-in-differences model. Columns (1) and (2) include the whole sample, in columns (3) and (4) we exclude children with weight below 2500 grams. Controls: age, years of education and employment status. Additional father controls: age, years of education and employment status. All regressions include month, year and municipality fixed effects. Standard errors clustered at municipality level in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%. *Source:* Birth records from the Statistical Office of the Republic of Serbia.

Table 7: Grade: Mean Comparison t-Test

Year	0	-1	+1		
	In utero	Control 1	Control 2	Diff. 1	Diff. 2
	(1)	(2)	(3)	(4)[(1) - (2)]	(5) [(1) - (3)]
A. Outcomes					
Marks at the end of P8 class[a]					
Language	3.739	3.765	3.768	-0.026**	-0.029**
	[1.138]	[1.131]	[1.135]	(-2.724)	(-3.022)
Mathematics	3.383	3.418	3.417	-0.035***	-0.034**
	[1.216]	[1.207]	[1.221]	(-3.433)	(-3.249)
Behaviour	4.930	4.925	4.935	0.005	-0.005
	[0.363]	[0.383]	[0.348]	(1.687)	(-1.602)
First wish 4y[b]	0.907		0.914	-0.010***	-0.006*
	[0.290]		[0.281]	(-3.985)	(-2.569)
Enrolled 4y[c]	0.883		0.878	-0.008**	0.005
	[0.321]		[0.327]	(-2.950)	(1.697)
B. Characteristics					
Female	0.486	0.491	0.488	-0.005	-0.003
	[0.500]	[0.500]	[0.500]	(-1.288)	(-0.629)
Observations	27,165	28,433	28,270	55,598	55,435

Notes: Standard deviations in parenthesis [ ] in columns (1) through (3). t-statistics in parenthesis () in columns (4) to (5). Standard errors clustered at municipality level in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%. *Source:* Final examination dataset from the Serbian Ministry of Education for years 2012 to 2016.

[a] Marks refer to last year of primary school (P8 class) at the end of ISCED 2 level. Marks range from 1 (lowest mark) to 5 (highest mark).

[b] Student recorded a 4-year secondary school track as his/her first choice.

[c] Student enrolled a 4-year secondary track profile.

Table 8: Main Results: The Effect of NATO Bombing on Schooling Outcomes

	Marks at the end of P8 class[a]			Secondary school	
	Language	Mathematics	Behaviour	First wish 4y[b]	Enrolled 4y[b]
	(1)	(2)	(3)	(4)	(5)
Panel A: Control 1: Year -1					
ATE In utero (=1)	-0.022*	-0.030**	0.006	-0.010 ***	-0.008 **
	(0.013)	(0.013)	(0.005)	(0.003)	(0.003)
Dep var mean	3.763	3.417	4.925	0.917	0.892
Dep var SD	1.131	1.206	0.386	0.276	0.311
Observations	53,989	53,989	53,989	51,460	51,289
Panel B: Control 2: Year +1					
ATE In utero (=1)	-0.023*	-0.029**	-0.005	-0.006**	0.005*
	(0.013)	(0.013)	(0.003)	(0.003)	(0.003)
Dep var mean	3.739	3.383	4.930	0.907	0.883
Dep var SD	1.138	1.216	0.363	0.290	0.321
Observations	53,910	53,910	53,910	51,858	51,646

Notes: This table presents estimated coefficients with an IPWRA model with subject marks as outcomes. Each outcome is estimated using female as individual level control and fixed effects, such as month of birth and school id. Standard errors are clustered at municipality level in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%. *Source:* Final examination dataset from the Serbian Ministry of Education for years 2012 to 2016.

[a] Marks refer to last year of primary school (P8 class) at the end of ISCED 2 level. Marks range from 1 (lowest mark) to 5 (highest mark).

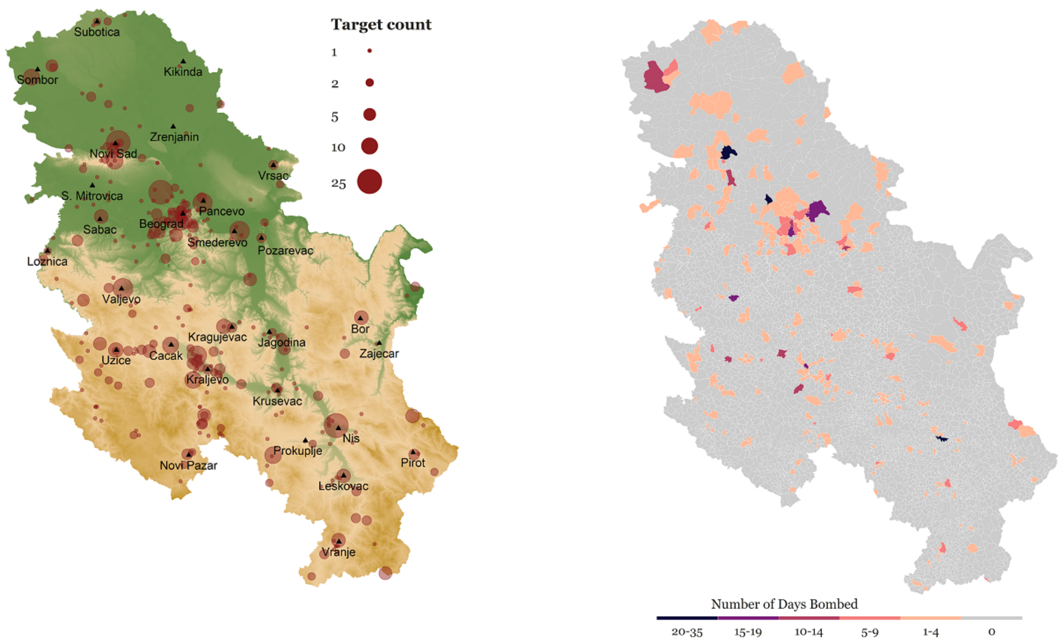
[b] Student recorded a 4year secondary school track as his/her first choice.

[c] Student enrolled a 4 year secondary track profile.

Online Appendix

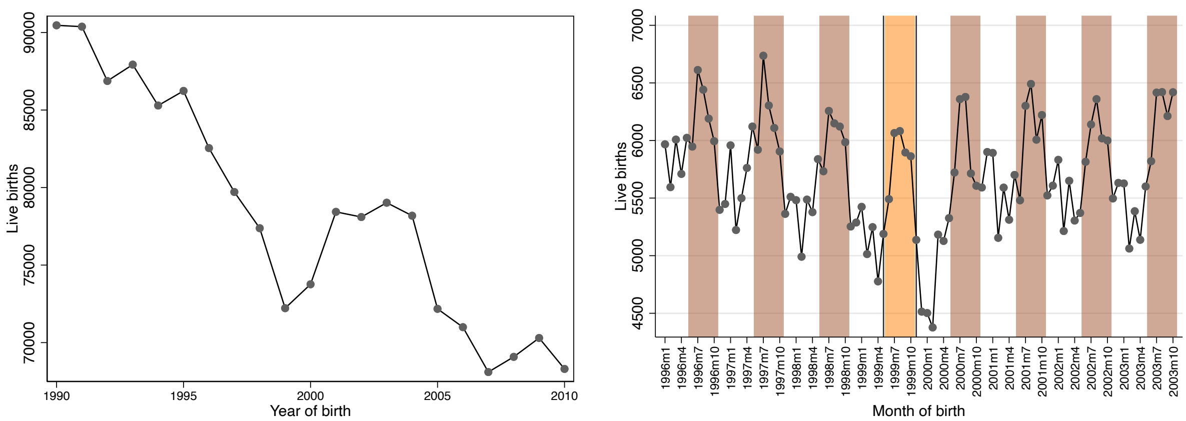
Figures

Figure A1: Spatial Distribution of 1999 NATO Bombing of Serbia



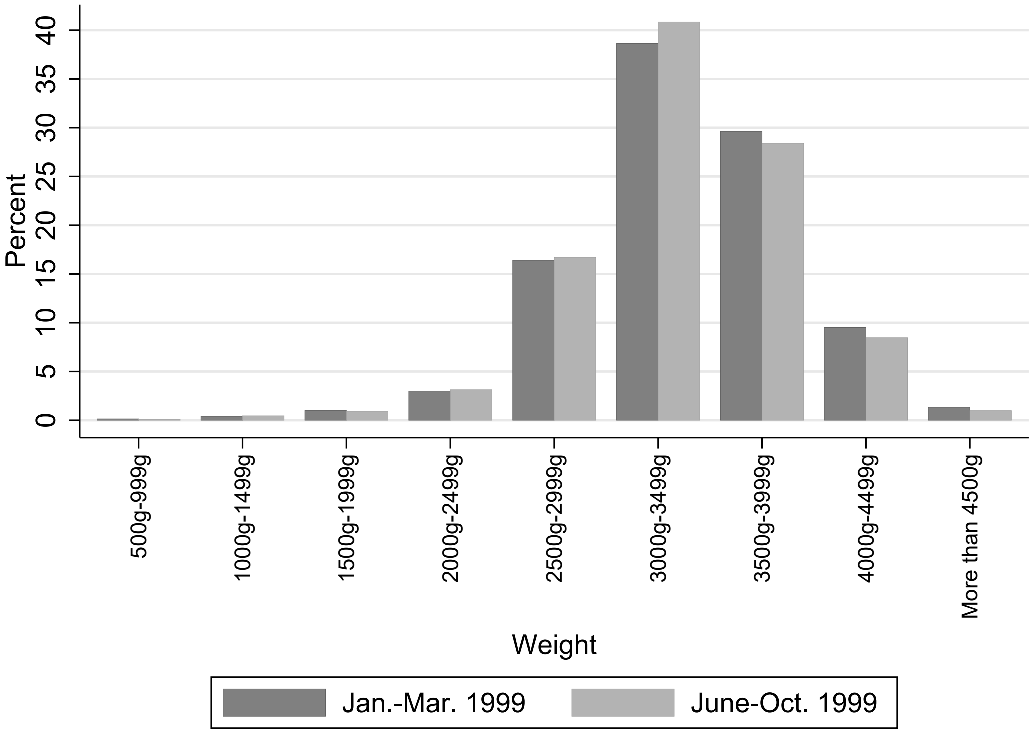
(a) Attacks by Target (b) Number of Days Bombed (settlements)

Figure A2: Live Births

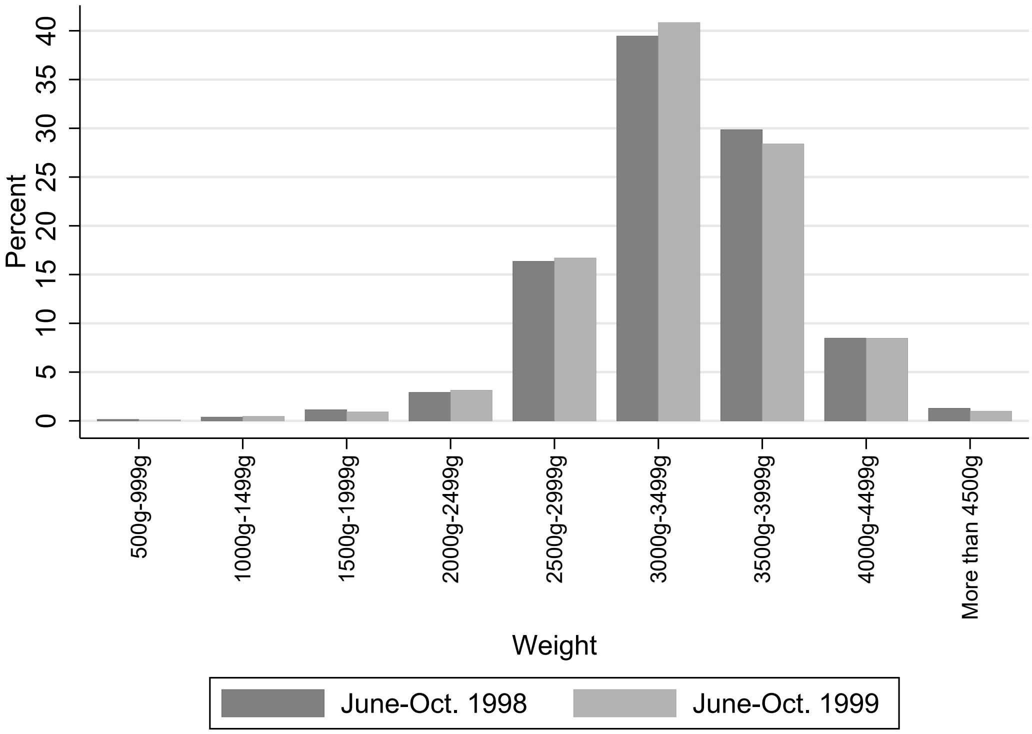


(a) Live Births by Year (b) Live Births by Month

Figure A3: Distribution of Birthweight

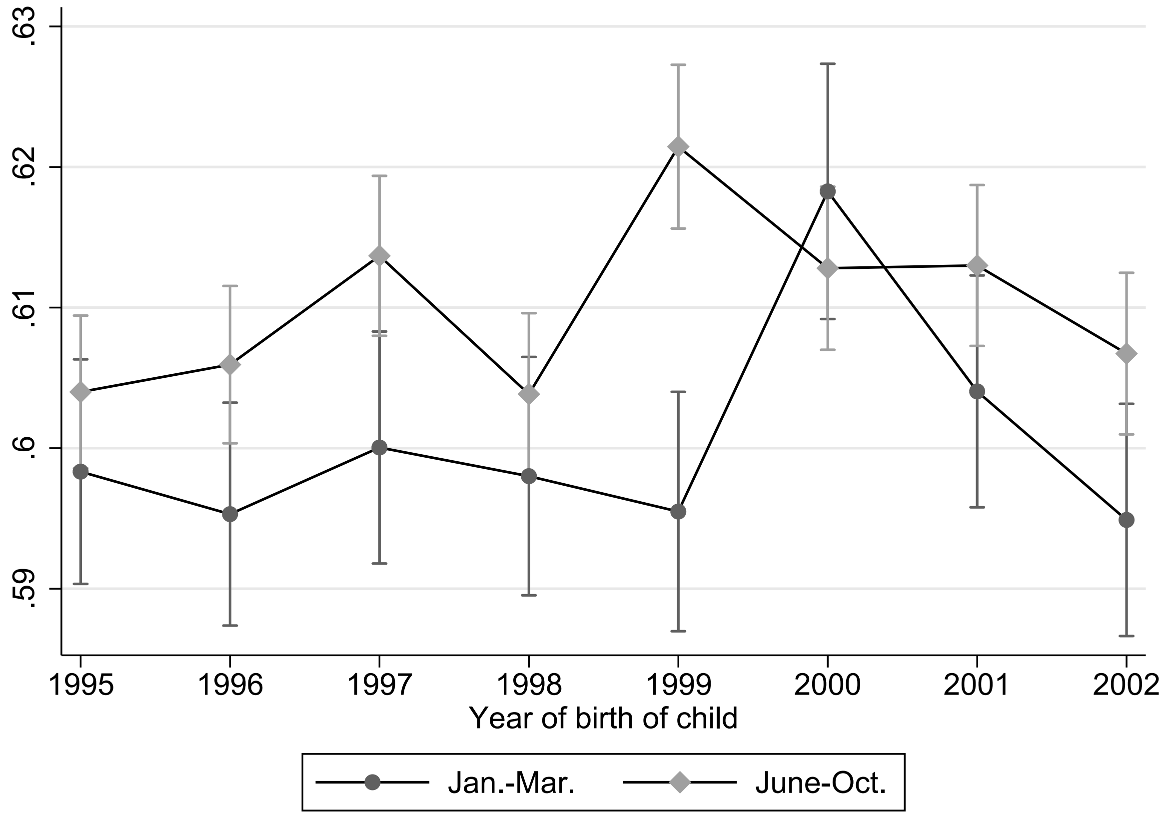


(a) Birthweight of children born in 1999



(b) Birthweight of children born from June to October in 1998 and 1999

Figure A4: Below Average Birthweight ($< 3500g$)



Tables

Table A1: Multiple Cluster Indicator Survey (MICS) Data:
Maternal Behaviour and Health Outcomes of the Treated and Control Children

	Jan-mid-Mar. 1999 N=100 (1)	Jun-Oct. 1999 N=181 (2)	Diff. *p*-value (3)
Mother's education			0.11
Primary or less	18.6%	10.3%	
Secondary	67.4%	60.9%	
Higher education	14.0%	28.7%	
Wealth index quintiles			0.15
Poorest	9.5%	13.8%	
Second	33.3%	16.1%	
Middle	47.6%	50.6%	
Fourth	9.5%	16.1%	
Richest	0.0%	3.4%	
Marital status: 0 Not married, 1 Married	95.3%	93.1%	0.62
Number of children younger than 5	1.535 (±0.631)	1.690 (±0.782)	0.26
Did you have an injury in the last 12 months? (yes/no)	4.7%	9.2%	0.36
Did you have poisoning in the last 12 months? (yes/no)	4.7%	3.4%	0.74
Risk of noise (yes/no)	30.6%	32.0%	0.88
Risk of air pollution (yes/no)	67.6%	67.1%	0.96
Risk of water pollution (yes/no)	45.7%	44.7%	0.92
Risk of waste materials (yes/no)	63.3%	51.4%	0.27
Risk of radioactive materials (yes/no)	55%	52%	0.76
Risk of crime and violence (yes/no)	34.3%	32.3%	0.84
Risk of venereal diseases and aids (yes/no)	2.6%	3.6%	0.78
Risk of heart diseases (yes/no)	15.6%	15.0%	0.93
Risk of diabetes (yes/no)	5.7%	11.1%	0.36
Risk of lung diseases (yes/no)	21.1%	9.9%	0.096
Risk of sickness due to injury (yes/no)	5.9%	2.6%	0.38
Risk of high blood pressure (yes/no)	17.6%	11.1%	0.34
Risk of liver cirrhosis (yes/no)	2.9%	2.5%	0.92
Risk of obesity (yes/no)	13.5%	17.7%	0.57
Reduced weight in the last 12 months (yes/no)	39.5%	36.8%	0.76
Reduced salt consumption in the last 12 months (yes/no)	11.6%	12.6%	0.87
Reduced sugar consumption in the last 12 months (yes/no)	14.0%	6.9%	0.19
Reduced fat consumption in the last 12 months (yes/no)	16.3%	12.8%	0.59
Reduced alcohol consumption in the last 12 months (yes/no)	7.0%	2.3%	0.20
Increased fruit and vegetable consumption in the last 12 months (yes/no)	27.9%	31.4%	0.68
Increased physical activity in the last 12 months (yes/no)	23.3%	16.5%	0.35
Reason for changed behaviour in the last 12 months			0.64
Changed because of health and healthier lifestyle	23.3%	27.1%	
Other or not changed	76.7%	72.9%	
Reason for high rate of sickness in the country			0.78
Nutrition	32.6%	40.2%	
Stress	51.2%	47.1%	
Difficult life	11.6%	10.3%	
Other	4.7%	2.3%	

Notes: Column (1) shows mothers' outcomes of the control children born from January 1 to mid-March 1999. Column (2) shows mothers' outcomes of the treated children affected by bombing, born from June 10, 1999 to October 31, 1999. Column (3) reports the *p*-value of the differences between mothers' outcomes of the treated and control children, as reported in 2000. *p*-values: * significant at 10%, ** significant at 5%, *** significant at 1%. Source: Multiple Cluster Indicator Survey (MICS) in 2000.

Table A2: Balancing results for differences-in-differences matching in Table 3

	Initial balance			Final balance		
	Means		Std. diff.	Means		Std. diff.
	Treated	Untread		Treated	Untread	
	(1)	(2)	(3)	(4)	(5)	(6)
Table 3, Panel B, column (1)						
Female	0.48	0.48	-0.009	0.48	0.48	0.000
Born in hospital	0.99	0.99	-0.005	0.99	0.99	0.000
Parents married at birth	0.8	0.82	-0.03	0.81	0.81	0.000
Mother's years of educ.	11	11	-0.001	11	11	0.000
Mother employed	0.4	0.41	-0.009	0.4	0.4	0.000
Mother's age	26.3	26.3	0.008	26.3	26.3	0.000
Has father data	0.87	0.88	-0.033	0.87	0.87	0.000
Table 3, Panel B, column (2)						
Female	0.48	0.48	-0.003	0.48	0.48	0.000
Born in hospital	0.99	0.99	-0.005	0.99	0.99	0.000
Parents married at birth	0.92	0.92	-0.006	0.92	0.92	0.000
Mother's years of educ.	11.28	11.27	0.001	11.28	11.28	0.000
Mother employed	0.43	0.43	0.001	0.43	0.43	0.000
Mother's age	26.25	26.17	0.016	26.22	26.23	0.000
Father's years of educ.	11.65	11.6	0.021	11.64	11.64	0.000
Father employed	0.85	0.85	0.001	0.85	0.85	0.000
Father's age	29.96	29.88	0.014	29.93	29.93	0.000
Table 3, Panel B, column (3)						
Female	0.48	0.48	-0.008	0.48	0.48	0.000
Born in hospital	0.99	0.99	-0.007	0.99	0.99	0.000
Parents married at birth	0.81	0.82	-0.027	0.82	0.82	0.000
Mother's years of educ.	11.04	11.04	0.001	11.04	11.04	0.000
Mother employed	0.41	0.41	-0.01	0.41	0.41	0.000
Mother's age	26.01	25.97	0.009	26	26	0.000
Has father data	0.88	0.89	-0.031	0.88	0.88	0.000
Table 3, Panel B, column (4)						
Female	0.48	0.48	-0.002	0.48	0.48	0.000
Born in hospital	0.99	0.99	-0.007	0.99	0.99	0.000
Parents married at birth	0.92	0.92	-0.005	0.92	0.92	0.000
Mother's years of educ.	11.3	11.3	0.002	11.3	11.3	0.000
Mother employed	0.43	0.43	0	0.43	0.43	0.000
Mother's age	26.22	26.15	0.015	26.2	26.2	0.000
Father's years of educ.	11.67	11.62	0.022	11.66	11.66	0.000
Father employed	0.85	0.85	0.005	0.85	0.85	0.000
Father's age	29.93	29.86	0.013	29.91	29.91	0.000

Notes: This table reports the differences in means, before and after matching, between the covariates included in the propensity score of Table 3, panel B.

Table A3: The Impact of NATO Bombing on Birth weight Outcomes – Low and High Birth Weight

	Low birth weight (<2500g)		**High birth weight (>4500g)**	
	(1)	(2)	(3)	(4)
Difference-in-Differences				
	0.001	-0.001	-0.003	-0.004
	(0.004)	(0.004)	(0.002)	(0.002)
Dep. var. mean	0.045	0.038	0.013	0.014
Dep. var. SD	0.207	0.191	0.113	0.116
Observations	79,343	69,109	79,343	69,109
Adj. R-squared	0.015	0.008	0.005	0.005
Controls	X	X	X	X
Father controls		X		X
Month of birth FE	X	X	X	X
Year of birth FE	X	X	X	X
Municipality FE	X	X	X	X

Notes: This table presents estimated baseline coefficients for the differences-in-differences model. Outcome in column (1) and (2) is an indicator for birth weight below 2500g, in columns (3) and (4) the outcome is an indicator for birth weight above 4500g. Controls: age, years of education and employment status. Additional father controls: age, years of education and employment status. All regressions include month, year and municipality fixed effects. Standard errors clustered at municipality level in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%. *Source:* Birth records from the Statistical Office of the Republic of Serbia.

Table A4: The Impact of NATO Bombing on Below Average Birthweight Within Bombed Settlements by the Number of Days Experiencing Bombing

	Number of days bombed in settlement				Top distance	
Treatment	1 ≤	2 ≤	5 ≤	10 ≤	10%	20%
	(1)	(2)	(3)	(4)	(5)	(6)
A. Placebo years 1996 - 1997						
DiD	-0.001	0.003	-0.007	-0.022	0.018	0.008
	(0.011)	(0.012)	(0.017)	(0.038)	(0.019)	(0.012)
Dep. var. mean	0.588	0.586	0.587	0.570	0.634	0.623
Dep. var. SD	0.492	0.492	0.492	0.495	0.482	0.485
Observations	41,092	33,330	17,928	7,171	8,622	16,877
Adj. R-squared	0.044	0.044	0.049	0.036	0.048	0.051
B. Placebo years 1997 - 1998						
DiD	-0.013	-0.018*	-0.017	-0.028	-0.015	-0.016
	(0.010)	(0.010)	(0.014)	(0.023)	(0.025)	(0.019)
Dep. var. mean	0.597	0.591	0.596	0.573	0.644	0.631
Dep. var. SD	0.491	0.492	0.491	0.495	0.479	0.483
Observations	39,868	32,680	17,745	7,039	8,154	15,931
Adj. R-squared	0.045	0.045	0.050	0.039	0.052	0.050
Controls	X	X	X	X	X	X
Father controls	X	X	X	X	X	X
Month of birth FE	X	X	X	X	X	X
Year of birth FE	X	X	X	X	X	X
Municipality FE	X	X	X	X	X	X

Notes: This table presents estimated coefficients for different levels of intensity for the placebo years for the difference-in-differences model. In columns (1) through (4) we restrict our sample to settlements experiencing at least one day of bombing (col. (1)), at least two days of bombing (col(2)), at least 5 days of bombing (col. (3)), and and at least 10 days of bombing (col. (4)). Controls: female, born in hospital, parents married and the characteristics of the mother: age, years of education and employment status. Additional father controls: age, years of education and employment status. All regressions include month, year and municipality fixed effects. Standard errors clustered at municipality level in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%. In column (5) we restrict the sample to the top 10% of settlements furthest away from the closest bombed settlement, and in column (6) to the top 20% of settlements furthest away from the closest bombed settlement. *Source:* Birth records from the Statistical Office of the Republic of Serbia.

Table A5: The Impact of NATO Bombing on Below Average Birthweight Within Bombed Settlements by the Number of Days Experiencing Bombing - Spatial Variation

	Below average birthweight (<3500g)			
	Number of days bombed in settlement			
	$1 \leq$	$2 \leq$	$5 \leq$	$10 \leq$
	(1)	(2)	(3)	(4)
A. Difference-in-Differences				
	0.004	0.006	0.014	0.027
	(0.015)	(0.016)	(0.017)	(0.022)
Dep var mean	0.599	0.598	0.604	0.606
Dep var SD	0.490	0.490	0.489	0.489
Observations	32,038	27,294	17,222	9,952
Adj. R-squared	0.045	0.045	0.047	0.049
Controls	X	X	X	X
Father controls	X	X	X	X
Month of birth FE	X	X	X	X
Year of birth FE	X	X	X	X
Municipality FE	X	X	X	X

Notes: This table presents estimated coefficients for different levels of bombing intensity for the difference-in-differences model. We compare children born from June to October 1999 (1) in bombed settlements to (2) children living in settlements most distant from the bombing (first difference). The second difference considers children born in the same months of the previous year, 1998 and living in the same settlements. Column (1) considers to be treated the settlements bombed at least one day as bombed, column (2) at least two days, column (3) at least five days and column (4) at least 10 days. Controls: female, born in hospital, parents married and the characteristics of the mother: age, years of education and employment status. Additional father controls: age, years of education and employment status. All regressions include month, year and municipality fixed effects. Standard errors clustered at municipality level in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%. *Source:* Birth records from the Statistical Office of the Republic of Serbia.

Table A6: The Effect of NATO Bombing on Schooling Outcomes – Placebo Years

	Marks at the end of P8 class[a]			Secondary school	
	Language	Mathematics	Behaviour	First choice 4y[b]	Enrolled 4y[b]
	(1)	(2)	(3)	(4)	(5)
Placebo years 1997 and 1996					
ATE In utero (=1)	0.001	-0.004	-0.039	-0.002	0.008 ***
	(0.013)	(0.012)	(0.004)	(0.003)	(0.003)
Dep var mean	3.728	3.380	4.969	0.905	0.860
Dep var SD	1.147	1.210	0.246	0.294	0.347
Observations	56,564	56,564	56,564	54,027	53,769

Notes: This table presents estimated coefficients with an IPWRA model with subject marks as outcomes. Each outcome is estimated using female as individual level control and fixed effects, such as month of birth and school id. Standard errors are clustered at municipality level in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%. *Source:* Final examination dataset from the Serbian Ministry of Education for years 2012 to 2016.

[a] Marks refer to last year of primary school (P8 class) at the end of ISCED 2 level. Marks range from 1 (lowest mark) to 5 (highest mark).

[b] Student recorded a 4year secondary school track as his/her first choice.

[c] Student enrolled a 4 year secondary track profile.

Table A7: Oster Test

	Marks at the end of P8 class[a]			Secondary school	
	Language	Mathematics	Behaviour	First choice 4y[b]	Enrolled 4y[c]
	(1)	(2)	(3)	(4)	(5)
Control 1: Year -1					
$\delta = 1, Rmax = 1$	0.002	-0.002	0.039	-0.004	0.005
	(0.029)	(0.043)	(0.025)	(0.016)	(0.017)
$\delta = 1, Rmax = 0.3$	-0.016*	-0.025*	0.015**	-0.008*	-0.005
	(0.011)	(0.014)	(0.007)	(0.005)	(0.005)
$\delta = 0.25, Rmax = 1$	-0.016	-0.025*	0.014**	-0.009*	-0.005
	(0.010)	(0.014)	(0.006)	(0.005)	(0.005)
$\delta = 0.5, Rmax = 0.5$	-0.016	-0.026*	0.014**	-0.008**	-0.005
	(0.010)	(0.013)	(0.006)	(0.004)	(0.005)

Notes: This table presents the coefficients of the in utero effect for varying levels of R^2 max and degree of selection on unobservables with respect to selection on observable, δ.

[a] Marks refer to last year of primary school (P8 class) at the end of ISCED 2 level. Marks range from 1 (lowest mark) to 5 (highest mark).

[b] Student recorded a 4 year secondary school track as his/her first choice.

[c] Student enrolled a 4 year secondary track profile.

Description of the Bombing Data Collection

The information on bombed localities, intensity, and duration comes from a novel dataset with information on over 1,000 targets in the Federal Republic of Yugoslavia, including the date, location, target type, and fatalities. The majority of targets were military objects and forces (63%) followed by the industry (13%), transport infrastructure (9%), civilian (7%), communications facilities (7%), and other targets (1%). This is, by far, one of the most comprehensive and precise datasets on the NATO bombing of Yugoslavia. Other essential datasets include the Human Rights Data Analysis Group's dataset on killings in Kosovo, and the Humanitarian Law Centre (HLC)'s database of NATO bombing victims. The dataset was manually coded and includes information on the location of bombings as reported in the media from March 24 until June 10, 1999. More specifically, the information on bombed municipalities mainly comes from then pro-opposition Serbian daily (Glas Javnosti), and pro-government daily (Večernje Novosti (Smiljanić, 2009), and two major Serbian weeklies (NIN and Vreme). Reports from the state-owned news agency Tanjug, the Human Rights Watch (Arkin, 2000), the Database on casualties of the Humanitarian Law Center (HLC) in Belgrade, and, sparingly, the White Book of the Yugoslav Government (Bulajić, 2000) were used for data triangulation as well as the identification of under-reported strikes against the army. NATO briefings were not used because they lack information on exact locations.

Cruise missile strikes and air raids were coded if the source entailed information on the exact location of incident.[19] For example, if the source reported an air raid on the Batajnica airport then the coordinates were coded for the airport using Google Maps. If, however, the source identified a strike on a 'wider area of Belgrade' or referred to a mountain range without a reference to a particular object or unit, then this attack was omitted. Fortunately, such occurrences were rare, less than 10 or equivalently less than 1% of all strikes, and in most cases it was possible to pin down a few unreported locations using the HLC database of casualties.

For each of the identified strikes, date of the incident as well as coordinates were coded. To determine whether an attack falls within the settlement boundaries, each point coordinate was intersected with the settlement polygon using a GIS intersect function from package `sp` in `R` programming language (Pebesma and Bivand, 2005; R Core Team, 2019). Next, the number of

[19]The data do not include reconnaissance flights. A reconnaissance aircraft or a 'spy plane' is a military surveillance aircraft designed or adapted to perform aerial reconnaissance with roles including collection of imagery intelligence (including using photography), signals intelligence, as well as measurement and signature intelligence.

strikes per settlement were identified by summing up those points that the function attributed to particular settlements. Finally, using information on location and date of strikes, the variable *Days* was constructed by counting unique dates of strikes that fall within settlement boundaries. In Figures A1a and A1b we present the number of attacks and the number of days bombed at the settlement level, respectively. The average duration of bombing by municipality is 3.4 days, and the mean number of strikes by municipality is 3.86.

Table A8: Codebook for the NATO bombing dataset

Variable	Description
Loc	Target location denoted by object name and/or settlement name
Long	Longitude of the target in EPSG:4326 (WGS84) coordinate
Lat	Latitude of the target in EPSG:4326 (WGS84) coordinate
Region	Denoting a wider area where the target was based, including Belgrade, Kosovo, Montenegro, Central Serbia and Vojvodina
Date	Date of the strike(s) on a target as reported in a source or sources.
Source	Online media sources used to determine a strike (or strikes) on a target on a specific date. The Glas Javnosti newspaper was used due to its links to the archived main page dedicated to the start of the bombing, because links to other pages sometimes failed to load or led back to this page. It is possible to access the whole archive on the dedicated page and search for specific dates in 1999 by clicking on the hyperlinked month name (e.g., April 1999). The resulting page opens a calendar with entries for every day of the month. News of the day are accessed by clicking on the specific date in the calendar. Other sources used are: Bela Knjiga or the White Book of the Yugoslav government, enlisting the targets and dates of strikes, the Humanitarian Law Center's database (FHP), and the Human Rights Watch (HRW) and their online reports. Additional sources include weekly newspapers NIN and Vreme, book "Agresija NATO" by (Smiljanić, 2009) as well as local online sources. While we massively relied on the Glas Javnosti reports, we also aimed to use multiple sources where possible.