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

Lara Lebedinski

University of Vienna and Institute of Economic Sciences†

Giuseppe Migali

Lancaster University and Magna Graecia University‡

Miloš Popović

Booking.com§

Sunčica Vujić

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

February 26, 2025

*Acknowledgments: We are grateful to Wiji Arulampalam, Janet Currie, Anja Grujović, Jonathan James, Dave Marcotte, Emily Nix, Michele Pellizzari and Jaroslaw Kantorowicz for their constructive and insightful comments. The participants at the Annual conferences of the Institute of Economic Sciences (IES) in Belgrade in October 2019 and October 2021, the European Association of Labour Economists (EALE) – Society of Labor Economists (SOLE) – Asian and Australasian Society of Labour Economics (AASLE) virtual conference in June 2020, the Allied Social Science Associations (ASSA) virtual annual meeting in January 2021, the European Society for Population Economics (ESPE), the International Workshop on Applied Economics of Education (IWAEE) and the International Association for Applied Econometrics (IAAE) virtual conferences all held in June 2021, the EALE virtual conference in September 2021, the Royal Economic Society (RES) virtual conference in April 2022, and the SOLE annual meeting in May 2022 provided useful comments and suggestions. The seminar participants at the University of Antwerp, University of Alicante, University of Bath (BABB), University of Pompeu Fabra (UPF), University of Wageningen, University of Essen (Health Economics Seminar), University of Göttingen, University of Hagen and University of Minho in Braga seminar series, gave insightful 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. This paper is accompanied by Supplementary Material available here.

†Corresponding author: University of Vienna and 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

§Booking.com, Amsterdam, the Netherlands. Email: milos.agathon@gmail.com

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

Abstract

This paper estimates the causal effect of 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 birth weight and stillbirth outcomes, as well as long-term outcomes measured by grades and high school preferences for/enrolment in non-vocational secondary schools. Using the entire birth records of the Serbian Statistical Office we compare the birth weight outcomes 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 and secondary school preferences/enrolment. Our findings suggest that children who were *in utero* during the bombing were 2pp more likely to be born with a lower than average (< 3500g) birth weight and 1pp less likely to be born with high birth weight (≥ 4000g). We find no effects for low birth weight (< 2500g) and stillbirth outcomes. In the long-term, we find a statistically significant negative effect of the bombing on grades in mathematics (around -0.9%) and Serbian language (around -0.6%) at the end of primary school, and a 1pp decrease in the preference for/enrolment in academically oriented secondary school. Our results emphasize that war-related bombing has devastating consequences for pregnant women and affected children, and the necessity of policy intervention to prevent conflicts and mitigate their consequences for the most vulnerable part of the population.

Keywords: Human Capital Formation; Children; War-Related Bombing; In-Utero Effect.

JEL classification codes: I15, J13, O15.

1 Introduction

Motivation — War-related bombing has devastating consequences for the affected population and is essentially a 'war on public health' (Ashford and Gottstein, 2000), where the first victims are the most vulnerable populations – pregnant women and their unborn children. Recent conflicts in Ukraine and Palestine have only reignited the importance of this topic and drew public attention to the severe consequences of wars and wartime bombing on affected children (Akbulut-Yuksel, 2022; Chupilkin and Kóczán, 2022). Birth weight of a child, as one of the main health indicators, has consequences on subsequent health, education, and labor market outcomes (Black et al., 2005; Oreopoulos et al., 2008) and is determined by what happens during the pregnancy. The negative effect of prenatal shocks on children's birth weight and later human capital outcomes can be explained by maternal stress (Aizer et al., 2016; Black et al., 2016; Berthelon et al., 2021; Persson and Rossin-Slater, 2018) due to both direct effects, such as physical destruction, malnutrition, displacement and deteriorated socio-economic and health conditions, and indirect effects such as contamination of air and soil. There is limited quantitative evidence on the prenatal consequences of war-related bombing. This paper aims to fill this gap.

It is not straightforward to estimate the causal effect of early childhood circumstances on later outcomes. The result of this exercise may be confounded by the unobserved factors that 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. To this end and similar in spirit to Akbulut-Yuksel (2014), we look at the effect of NATO bombing on children's birth weight and later educational outcomes.

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

[1] In 1999, the official name of the country was the Federal Republic of Yugoslavia. In 2003, the country was renamed Serbia and Montenegro to reflect its constituent parts after the 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.

as factories, bridges, and governmental 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, and 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 have a negative effect on affected children, in the short-run, in terms of reduced birth weight, and in the long-run, in terms of negative educational outcomes, as measured by primary schools grades and preferences for/enrolment in the non-vocational secondary schools.

What we do – identification and results preview — The aim of this paper is to establish a causal link between the NATO bombing of Serbia on the health and educational outcomes of children who were *in utero* during the bombing and were born between the months of June and October 1999. Our identification strategy for the short-term outcomes is based on a conditional before/after estimation approach, combined with propensity score matching. The purpose of the latter is to find a control group of newborns more similar to the treated ones in all relevant pre-bombing characteristics. We first compare children who were *in utero* during the whole period of bombing in 1999, with children born a few months before in the same year, and children born in the same months in the previous year, 1998. This approach, based on the variation across time of the birth weight of 4 cohorts of children, avoids 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 suggest that children who were *in utero* during the bombing were 2 percentage points (pp) more likely to be born with a lower than average (< 3500g) birth weight and 1pp less likely to be born with high birth weight (≥ 4000g). We find no effects on low birth weight (< 2500g) and stillbirths. When examining the spatial intensity of bombing, we find that there was a common war effect and all pregnant mothers need to be considered as treated. We provide evidence that children born in more bombed settlements had a more negative impact of bombing compared to children born in less affected settlements.

To investigate the long-term outcomes, we use data on educational achievement at the 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 the same conditional before/after estimation approach, because the whole cohort finishing primary school in 2014 was exposed to bombing to some degree, hence we estimate an inverse probability weighting regression-adjustment (IPWRA) model. We find that children who were *in utero* during bombing have statistically significant lower grades in mathematics (around −0.9%) and Serbian language (around −0.6%) at the end of primary school (both effects are equivalent to about 0.03 standard deviations (SD)), and a 1pp increase in the preference for/enrolment in 3-year vocational secondary schools in comparison to 4-year grammar (more academically-oriented) secondary schools.

The main transmission mechanism was *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., 2021; Persson and Rossin-Slater, 2018), a channel also suggested by the works of Krstić et al. (2006) and Krstić et al. (2007) which look at the effect of the NATO bombing on the pregnancy outcomes of the affected women.

Our contribution — This paper is the first to rigorously examine the effect of the NATO bombing on a specific population subgroup (unborn children) affected by this event, using the National Registry of Birth Records from the Statistical Office of the Republic of Serbia (SoRS), covering the whole territory of Serbia without Kosovo. In addition, we have also compiled a novel and unique dataset of the NATO bombing of Serbia – this is the most comprehensive and precise dataset of the NATO bombing of Serbia. As such, we contribute to the literature on short- and long-term effects of conflicts on future generations by shedding light on an important conflict that has not previously received much attention in the economics literature.[2]

The Serbian case is especially useful for examining the effects of bombing on child development in a quasi-experimental framework. First, the NATO intervention is arguably unantici-

[2]One notable exception is a recent paper by Tkalec and Žilić (2021) which identifies the effect of NATO bombing in Kosovo on tourism outcomes in Croatian Adriatic counties. There are also a few papers in medical science literature that examine the effect of the NATO bombing of Serbia: Marić et al. (2010), Krstić et al. (2006) and Krstić et al. (2007). Marić et al. (2010), henceforth METAL10, examines the effect of the 1999 NATO bombing on birth outcomes of the affected children, using the data of one hospital in Belgrade (The Institute of Gynaecology and Obstetrics) and focusing on lower birth weight as the main outcome. In comparison to METAL10, the current paper (a) is the first paper that looks at the effect of the 1999 NATO bombing on birth outcomes of the affected children using the SoRS National Registry of Birth Records for the whole territory of Serbia without Kosovo. The paper of METAL10 examines the effect of the 1999 NATO bombing on birth outcomes of the affected children, using the data of one hospital in Belgrade (The Institute of Gynaecology and Obstetrics), without discussing whether this group of children is representative of the rest of the country. This can further be seen in the difference in the sample size of the two papers – the sample size of the treatment group exposed to prenatal stress in METAL10 is 1198, while in our paper the treatment group is 27,154 (23,536) without (with) father data. Therefore, METAL10 analysed less than 5% of the children affected by the 1999 NATO bombing. (b) When it comes to the main outcome of interest, METAL10 looks at the "lower birth weight", without providing a further definition of what this lower birth weight is. The current paper provides estimation results for the probability of low birth weight P(LBW) (<2500g), probability of below-average birth weight P(BABW) (<3500g), probability of high birth weight P(HBW) (≥4000g) and stillbirths. (c) The current paper also uses very detailed data on bombing and explores the effect of the intensity of bombing on the birth outcomes of the affected children, not considered before. (d) The current paper further focuses on the long-term educational outcomes, not considered by METAL10. (e) Last but not least, the current paper uses econometrics techniques such as difference-in-differences (with/without matching) and a battery of robustness checks, not considered by METAL10.

pated 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 enabled us to isolate the effect of the war-related bombing alone. Third, the magnitude of this event exceeds the average number of terrorist bombings, offering a unique opportunity to empirically investigate the effect of prolonged prenatal exposure to "disaster" conditions on child development and later human capital outcomes. Finally, we base our analysis on individual registers of birth weight and educational outcomes, which is an advantage in comparison to the underlying literature which often uses survey data and/or cannot trace the exact place of birth of a child.

The paper contributes to the discussion that 'birth weight does matter' (Black et al., 2005), and that being born with below-average birth weight has detrimental consequences in both the short- and long-run. Oreopoulos et al. (2008) show that "even for infants born between 2,500 and 3,500 grams... there is about a one percentage point higher risk of death within one year". The negative long-term human capital outcomes provide further evidence for the lack of ambition and achievement and could be explained through the reduced ability and cognition of the affected cohorts.

Related 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 have consequences on individual outcomes throughout the life cycle, as measured by worse health and educational outcomes in childhood and subsequently worse labor-market, adult health outcomes and other indicators of socioeconomic status (Atella et al., 2024; Van den Berg and Lindeboom, 2018; Aizer et al., 2016; Currie and Rossin-Slater, 2015; 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.[3] 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

[3]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.

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. Currie and Rossin-Slater (2015) further discuss that early-life conditions can have consequences on individual outcomes throughout the life cycle. Our paper directly contributes to this literature.

In Table 1 we provide a summary of recent studies that demonstrate that experiencing negative events during pregnancy leads to worse birth weight and human capital outcomes for the affected children. In Panel A, we provide an overview of papers that 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). In Panel B we summarise the effects of natural disasters such as earthquakes, hurricanes, tropical storms, floods and temperature shocks, and Asian flu. The largest effects in terms of children's reduced birth weight are observed when the mothers are exposed to natural disasters (reduction of 45-50 grams in De Oliveira et al. (2025) and Torche (2011)). The effect is comparable in magnitude to the effect of risky behaviors during pregnancy, such as tobacco consumption (Lien and Evans, 2005). If we focus on the effects on the higher probability of low birth weight P(LBW), the size of the estimated effects is small and similar across different literature – most papers find less than 2pp. 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.

If we compare these findings to our estimates of the impact of the NATO bombing on children's lower birth weight and educational outcomes, we conclude that the findings of this paper are comparable in magnitude to the existing literature summarised in Table 1. Similar to our paper, Currie and Rossin-Slater (2013) do not find a significant effect of hurricane exposure during pregnancy on low birth weight. Their explanation is that the incidence of low birth weight in children in their sample is only 6%. In our paper, only 5% of the children affected by the NATO bombing have low birth weight.

Paper structure — The remainder of the paper is organised as follows. Section 2 discusses the NATO intervention and potential mechanisms, section 3 discusses the methodology and the results of the short-term outcomes, such as birth weight. Section 4 looks at the long-term outcome in terms of educational achievement. Section 5 concludes the paper.

2 The NATO Intervention and Potential Mechanisms

2.1 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 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, 2024; Oskorouchi, 2019), with the aim of maximising material damage and limiting collateral damage (Fenrick, 2001). The NATO 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.[5]

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 included information on the location of bombings as reported in the media.[6] 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 kilometers.

Figure 1 summarises the main features of the bombing data. Figure 1, panel (a), shows that the bombing was dispersed across the country with the highest concentration of attacks in large cities such as Belgrade, Niš, Novi Sad, and Kraljevo. Figure 1, panel (b), captures the intensity of the NATO bombing of Serbia, showing the number of days a settlement was bombed. It

[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] Serbia's economy was largely left in ruins in the aftermath of the 1999 NATO bombing. Overall, industrial production went down by 21% compared to 1998, and by 40% compared to 1989 (Teodorović, 2000). Dozens of factories were either severely damaged or destroyed (Hosmer, 2001). The destruction of key factories dealt the strongest blow to its employees and their families, followed by suppliers and dealers located in other parts of the country. The destruction of the industry left 230,000 workers jobless, with a further 2 million affected by this loss of employment (Teodorović, 2000). A group of 17 independent Yugoslav economists estimated a direct damage to the economy excluding Kosovo of about $3.8 billion (Vreme, 2000).

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

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

2.2 Potential Mechanisms

There are many ways for a lethal catastrophe such as bombing to affect pregnant mothers. There are both direct channels, such as physical destruction, malnutrition, displacement, and deteriorated socio-economic and health conditions, as well as indirect ones, such as contamination of air and soil. (i) According to the data which we collected, about 90% of targets fell into the 'strategic military targets' and only about 10% were 'civilian targets' (such as a farm, hotel, house, hospital or school). The number of casualties was limited; the Humanitarian Law Center (HLC) in Belgrade reported that NATO attacks killed a total of 754 people: 454 civilians and 300 members of the armed forces. There were 260 casualties in Serbia alone while the remaining 494 casualties happened in Kosovo and Montenegro.[7] Therefore, considering that the goal of the bombing was to maximise material damage and limit collateral (civilian) damage, the potential "civilian destruction channel" could not have been the main driver of the estimated effects. (ii) As shown in the paper, the mothers of the treated children didn't have a significantly higher number of stillbirths, hence we can also rule out this potentially confounding mechanism. (iii) Using the Multiple Cluster Indicator Survey (MICS) data in 2000 conducted just one year after the bombing,[8] Table A1 in the Online Appendix compares the mothers of the treated children to the mothers of the control children on a range of socio-economic, health and behavioral outcomes. We see that the two groups didn't have different behavioral 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. (iv) Due to the United Nations (UN) sanctions against Yugoslavia, which at the moment of the bombing lasted already for nine years, as well as tightened visa travel regime for its citizens, migratory movements out of the country were limited. Within-country mobility was possible, but since it was difficult to predict the next-day target location, it was not clear where and

[7] Source: http://www.hlc-rdc.org/?p=34890&lang=de

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

when to move. Krstić et al. (2007) write "the bombardment of the whole territory lasted three months without any possibility to evacuate the population into safety zones..."[9]

Summing up, 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., 2021; Persson and Rossin-Slater, 2018) The prenatal maternal stress channel has also been suggested by the papers of Krstić et al. (2006) and Krstić et al. (2007) which look at the effect of the NATO bombing on the pregnancy outcomes of the affected women.[10]

3 Short-Term Outcome at Birth

3.1 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 birth weight outcomes of children who were *in utero* during the 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, birth weight in 11 categories, and whether a child was born in a hospital or elsewhere. The dataset features socio-demographic information about 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 (2020)), we exclude the following observations: births from mothers who were younger than 15 and older than 49, and multiple

[9]We don't have exact statistics on internal/external migration. Throughout the bombing period, Serbia was under general UN sanctions, and travelling abroad was very difficult. The following article summarises the situation well (`https://www.motherjones.com/politics/1999/09/serbias-lost-generation/`): "According to the UN High Commission for Human Rights, 50,000 people fled from Serbia over the three months of the NATO bombing campaign. Many of the evacuees were mothers who left with their children to avoid the bombs. But the Budapest-based Refugee Action Project estimates that 15,000 to 20,000 were draft evaders, many of whom crossed into Hungary, the sole NATO country bordering Yugoslavia." In 1999, the population of Serbia was 7.54 million. Therefore, about 0.4% of mothers with children temporarily or permanently emigrated from Serbia in 1999. "Serbs who wish to move to the West are likely out of luck: They are finding it practically impossible to get visas to travel anywhere, even for a preliminary immigration office interview. "The situation of those who tried to leave Yugoslavia during the previous wars was much better than for those who left during the Kosovo crisis," says Koszeg. "I think they [Western governments] are just fed up with refugees from the Yugoslav wars."

[10]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.

births. The exclusion of twins is based on the findings of Bhalotra and Clarke (2020) that exposure to bomb casualties in the second and third trimesters decreases the likelihood of multiple births. Stillbirths are only recorded if they occur after the 27th gestational week.[11]

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 (the last day of bombing) and October 31, 1999. 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 the possibility of premature births and to exclude the possibility that they are driving our results, we restrict the sample to children born up to October 31, 1999. As part of robustness checks (see Table 6), we include in the sample children born until December 15, 1999, and check whether their inclusion changes the main findings. The decision to focus on children who were *in utero* throughout the bombing is motivated by our intention and goal to isolate the impact of bombing on infants' birth outcomes. Therefore, we exclude children born during the bombing, to avoid the confounding effects of the exposure to bombs both *in utero* and in the early days after birth. While the impact of bombing in the early days after birth does not have any effect on birth outcomes, it could have an impact on later educational outcomes, and we prefer to isolate the *in utero* impact for both birth and educational outcomes. Children conceived during the 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 on children who were conceived during the bombing.

In Figure A1(a) in the Online Appendix, we graphically show that in the period between 1990 and 2010, the annual number of total births in Serbia had a declining trend, despite the positive improvements in the period between 2000 and 2005. In Figure A1(b) in the Online Appendix, we show the number of total births per month in the period between 1996 and 2003, and we use two vertical lines to indicate the months June to October in 1999. We observe that the monthly trend of births in these months for the year 1999 is similar to the previous years.

In Table 2, in column (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 as the treated children, but in the year prior to the bombing (June 10 to October 31, 1998). In columns (4) and (5) we show the statistics of infants born in the calendar months just before the bombing

[11] We do not have information on miscarriages.

(January to mid-March), observed in the year of the treatment (1999) and in the previous one (1998). All the control groups are obtained considering children that we think are the most similar to the treated in terms of background characteristics. A preliminary inspection of our data is reported in Table 2, panel A, for measures of birth weight: categories of low (< 2500g), below average (< 3500g), high birth weight (≥ 4000g) and stillbirths.

The differences 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 birth weight compared to children born in the same period in 1998. In particular, affected infants were more likely to be born with below average birth weight and they were less likely to be born with high birth weight. These descriptive statistics suggest that the right tail of the birth weight distribution was affected, i.e. children with a higher birth weight and not children at risk of lower birth weight. 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 difference 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 standard measures of birth weight: low (< 2500g), below average (< 3500g), high birth weight (≥ 4000g), and stillbirths. According to the World Health Organisation (WHO), the normal birth weight of an infant (term delivery) is between 2500-4200g, and above and below this range, infants have low and high birth weights, respectively. Small baby indicators are low birth weight (< 2500g), very low birth weight (< 1500g), or small-for-gestational-age newborns. Among these three outcomes, the former two can be measured with our data and we focus on low birth weight (< 2500g). Moreover, motivated by the descriptive statics and visual inspection of the birth weight data, which suggests a shift in the upper tail of the distribution, we look at below average ($< 3500g$), and high birth weight (≥ 4000g). The negative long-term effects of reduced birth weight (BW) are well documented. Oreopoulos et al. (2008) show that "even for infants born between 2,500 and 3,500 grams... there is about a one percentage point higher risk of death within one year".

The stillbirth outcome (delivery of a dead foetus at more than 27 weeks of gestation) is relevant for the present analysis for two main reasons. First, 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.[12] Our preliminary analysis based on Table 2 puts forward that most of the variation is observed around and above the mean of the birth weight distribution.

The distribution of the birth weight of children born from January to March 1999 and June to October 1999 is shown in Figure A2(a), while the birth weight of children born in the period June to October in 1998 and 1999 is shown in Figure A2(b) (both available in the Online Appendix). 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 instance, the average age of the mother at birth was 26.111 in 1999 while it was 25.949 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 in the Online Appendix.

3.2 Estimation Strategy

Our main identification strategy is based on a conditional before/after approach, where we compare 4 cohorts of pupils and exploit the fact that the bombing is a random event. We then repeat the estimation using a propensity score matching approach, in order to have more similar comparison groups and to take into account of potential additional confounding factors.

[12] We looked at very low birth weight and male birth as outcomes, but these were not significant and are not reported in the paper. They are available upon request.

Conditional Before-After Approach

We estimate the following regression:

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

where Y_{itmdl} is a binary variable for birth weight 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 on four outcomes: low birth weight (< 2500g), below average birth weight (< 3500g), high birth weight (≥ 4000g), and stillbirths. $Y1999_t$ is a dummy variable equal to one if the year is 1999, and zero if 1998. $I(treated)_{dm}$ is an indicator taking the value one if the child is born between June 10 and October 31 and zero if the child is born from January 1 to mid-March.

The vector X'_{itmdl} contains the following individual-level characteristics: gender of a baby, 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;[13] γ_m is a calendar-month fixed effect and τ_l is the municipality fixed effect. Standard errors are clustered at the municipality level.

Since we cannot observe children born during the bombing period and, at the same time, unaffected by it, we use as a control group children born just before the bombing in the same year, and we compare those two groups to children born in the same months the year before. Specifically, the first difference is given by the comparison of the birth weight of children born in the year of the 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. We consider all children in Serbia to be treated independently of their location at the time of the bombing. We also assume that in the absence of the bombing shock, the birth weight of babies born between June 10 and October 31, 1999 would have followed a similar trend as the birth weight of babies born between January and mid-March 1999. For our main outcomes, we examine this

[13] We follow the multiple imputation approach by Rubin (2018) to impute missing values for the missing father characteristics (years of education, employment status and age of father). We create 50 imputed datasets by imputing missing values using the multivariate normal regression procedure. Father characteristics are estimated using the marriage status of parents at the child's birth and the following mother's characteristics: age, years of education, and employment status of the mother.

assumption in Figure 2. Overall, the graph displays that there were no significant differences before the bombing in the years 1995 to 1998 and suggests that our empirical approach is valid. As robustness checks, we run an event study and placebo tests for the pre- and post-bombing periods. In this way, we can check whether there were any diverging trends prior to or after the adverse event.

Although this is not a standard difference-in-differences (DD) setting since we do not have a proper spatial variation in 1999, we interpret the coefficient of the interaction $I(treated)_{dm}$ and $Y1999_t$, as the impact of the NATO bombing on birth weight. Note that we do not include the variable $I(treated)_{dm}$ as it is collinear with the month fixed effects.

Propensity Score Matching

The second approach we employ is conditional before/after propensity score matching. The purpose is to find a control group of newborns more similar to those treated in all relevant pre-bombing characteristics. This implies the satisfaction of the conditional independence assumption (CIA), which in our context means that the treatment should be conditionally independent of the before-and-after difference in the outcome. The second assumption of the matching is common support, which ensures that newborns with the same characteristics have a positive probability of being treated. Therefore, in the 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, using solely attributes observed in 1998. We are also assuming, that the difference in observables characteristics in 1998 between the groups born in June/October and January/March of 1998 should have been the same as the difference between the groups born in June/October and January/March in 1999 in the absence of any treatment. Furthermore, the variables included in the propensity score model span a relatively brief timeframe—merely a few months within the same year (1998)— and were carefully chosen based on their relevance to the outcomes and their expected stability within the short timeframe under examination.[14] Table A4 in the Online Appendix provides crucial evidence supporting the stability of these attributes across the specified period. This table shows the balance in these variables before matching, indicating their similarity across the groups being compared. The initial balance suggests that these variables do not exhibit time variation within the period from January to October 1998. Therefore, their inclusion in the propensity score model is unlikely to compromise the validity

[14] We use 'Parents married at birth,' 'Mother's years of education,' 'Mother employed,' and 'Mother's age.' We have also re-estimated our models, excluding the time-varying covariates, and the results remain unchanged.

of the matching process.

Following, Heckman et al. (1997) and Smith and Todd (2005), we estimate a conditional before/after 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 is a non-linear and weighted version of equation (1),

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

where Y_1 and Y_0 are the birth weight (as defined above) of the newborns in the treated and control groups; t' and t are the years before (1998) and during the bombing (1999), respectively. More 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).

3.3 Estimation Results

Main Results

Table 3 shows the main results of the effect of bombing on birth weight outcomes and stillbirth. In panel A, we use a conditional before/after model (see equation 1) and in panel B we add father controls. We repeat the analysis in panel (C) using a propensity score matching model (see equation 2). In the Online Appendix Table A4 we report, for each model of Table 3 panel C, 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.

In column (1) of Table 3, the outcome is the low birth weight ($<$ 2500g), in column (2) it is below average birth weight ($<$ 3500g), in column (3) the outcome is the high birth weight

($\geq$ 4000g) and in column (4) it is stillbirth. The results of our main model from panel A are shown in Figure 3. The figure shows that there is no change in the probability of low birth weight and stillbirths, but we find that being in utero during bombing increases the likelihood of being born with below average birth weight (<3500g) by 2.1pp and reduces the likelihood to be born with a high birth weight ($\geq$4000g) by 1.0pp. These results hold when we include father controls in panel B and when we use matching difference-in-differences in panel C. We have examined heterogeneous effects by month of birth (June to October) and we could not find statistically significant differences in any outcome.[15]

The results suggest that there was no impact on the likelihood of being born with low birth weight (<2500g), which is considered to be the main adverse outcome of prenatal exogenous shocks (De Oliveira et al., 2025; Torche, 2011; Currie and Rossin-Slater, 2013). Similar to our paper, Currie and Rossin-Slater (2013) do not find a significant effect of hurricane exposure during pregnancy on low birth weight. They explain that the incidence of low birth weight of children in their sample is only 6%. In our paper, only 5% of the children affected by the NATO bombing have low birth weight. However, we do find some reduction in terms of weight especially around the average and for high birth weight. In terms of the size of the estimated effects, existing literature (see Table 1) finds an about 1.6-1.7pp increase in the probability of low birth weight, using natural disasters such as hurricanes, tropical storms, and earthquakes as natural experiments. Our finding that being in utero during bombing increases the likelihood of being born below average birth weight (<3500g) by 2.1pp is somewhat bigger, but not that far off from the existing literature, albeit for an outcome further up the birth weight distribution.

Heterogeneity of Results

In this part we aim to explore whether some groups were more affected by the bombing than others, considering the gender of the newborns, maternal parity, parental socio-economic status, and urban/rural divide. There is literature showing that male foetuses are more delicate than female ones (Catalano and Bruckner, 2006; Catalano et al., 2006). For example, in populations exposed to exogenous stressors such as earthquakes or political and social disruptions, there is reduced conception and increased foetal death of males. The impact of bombing could vary by socio-economic status as wealthier households have more resources to mitigate the negative consequences of the exogenous shocks. Having unmarried parents, a mother with lower education and a higher birth order are all characteristics of, on average, a lower socio-economic

[15]Results are available upon request.

status (SES) of a newborn. For example, Cozzani et al. (2022) show that children exposed to the Madrid bombing had a higher risk of prematurity and low birth weight and that this detrimental effect is consistently concentrated among low-SES offspring. Maternal parity (i.e. the number of previous births) can be viewed as a proxy for the costs of reproduction. Literature shows that reproductive value increases with the age of a child and that younger children are less valued than older ones. Since maternal investment begins already *in utero*, lower birth weight of higher-order children can be viewed as the main indicator of lower maternal investment during pregnancy (Merklinger-Gruchala et al., 2019). Figure 1, panel (a), in the paper shows that the bombing was dispersed across the country with the highest concentration of attacks in large urban areas such as the cities of Belgrade, Niš, Novi Sad, and Kraljevo. Based on these arguments and findings in the previous literature, we focus on the following heterogeneous groups: male/female, unmarried/married, low education/high education, birth order higher than 1/birth order equal to 1, and urban/rural. In this section, we look only at below average birth weight and high birth weight as these two outcomes were the ones impacted by the bombing.

Figure 4 and Table 4 show the heterogeneity results for the outcomes below average birth weight and high birth. We find that the male infants, infants with married parents, infants of a higher birth order, and infants in urban settings were more adversely affected than their counterparts. Our findings confirm that male foetuses are more vulnerable to external shocks and that socio-economic status can play a role, i.e. children of higher birth order are more negatively affected. The latter could also be due to the age of the mother. Shorter labor time for higher-order births might also create extra stress when expecting to deliver during an extreme event period such as a bombing. The results for urban settings support the hypothesis that the more intensive bombing of these areas is reflected in more negative outcomes for infants in urban settings. The result that exposure to bombing increases the probability of below-average birth weight among children born into families with married parents is contrary to the findings in the literature which stresses the importance of two-parent investments for the well-being of offspring (Merklinger-Gruchala et al., 2019). One potential explanation could be that some fathers were mobilised due to the Kosovo war and sent to the frontline, which induced additional stress on pregnant mothers in married and cohabitating relationships and was reflected in a higher probability of below-average birth weight. An additional explanation could also be the age of the mother.

Spatial Variation between the Municipalities

An alternative way to define the treatment and control group would be to exploit spatial variation, in which we define bombed municipalities as treated and not bombed municipalities as a control group. In this specification, we restrict the sample to the years 1998 and 1999 and compare children in bombed versus those in not bombed municipalities. The coefficient plot for the four main outcomes is reported in Figure 5 and in the Online Appendix in Table A2. The results suggest that there are no significant differences between children born in bombed and not bombed municipalities. This specification can be interpreted as a test of our primary design and the findings suggest that the spatial control group is contaminated by spillover effects from the treatment group, thereby substantiating the main design in our paper.

Spatial Variation between the Settlements

A further investigation has been performed exploiting the spatial variation of the settlements. This approach is more similar to a traditional difference-in-differences (DD). We include in the treatment group children born between June-October 1999, who are born to pregnant mothers during the whole bombing period and living in the bombed settlements. Ideally, as mentioned above, we would want in the control group pupils born in the same period June-October 1999, but unaffected by the bombing. Since virtually all children in Serbia are affected by the war, we try to minimise the effect of the bombing, exploiting the distance of each settlement not directly bombed from the nearest bombed settlement. To properly select the control group, we have disaggregated in deciles the distance from the nearest bombed settlement and included in our analytical sample only those settlements ranked in the top decile (i.e. top 10% of the distance ranking). Our 'weak' assumption is that pupils born in very distant and not bombed settlements are relatively less affected. We estimate equation (1) by adding a further interaction that takes into account the spatial variation of the settlements.

Overall, we do not find any spatial effect, i.e. there is no difference between children born in bombed and not bombed settlements (Table A3) We have repeated the analysis at the municipality level, and we still do not find any effect (Table A5). All these results are reported in the Online Appendix in Tables A3 through A6. We argue that the absence of any spatial effect is because all pregnant mothers are affected by the bombing, whether they live in a bombed or not bombed settlement. There is a common war effect, and all pregnant mothers need to be considered as treated, thus we are only able to identify an effect when we compare four different cohorts of children across time.

Intensity of Treatment

Bombing had an impact on children below the average birth weight and high birth weight, but does this impact differ by the intensity of bombing? We directly test the intensity of the bombing in two ways. We first perform a within-bombed settlement analysis and then we extend this analysis looking for a spatial effect.

Initially, we restrict our sample to the bombed settlements only, which are disaggregated according to the number of days they have been bombed. Hence, we estimate equation (1) starting from at least 1 day of the bombing and incrementing the minimum number of days up to 10. The estimation is performed using as dependent variables the below average birth weight and the high birth weight, and always including fixed effects, individual and paternal controls (see Panel A and B of Table 5).

Secondly, we repeat the spatial analysis described above, restricting the treatment group to bombed settlements experiencing an increasing number of days of bombing. The control group has been left unchanged (settlements not bombed in the top decile of the distance from bombed settlements). The estimation is repeated for both below-average birth weight (Table 5 Panel C) and high birth weight (Table 5 Panel D).

The results in Table 5, Panel A, suggest that the impact of bombing on the likelihood of being born below average birth weight increases with the intensity of bombing. In column (1), restricting to settlements bombed at least one day, we find that the likelihood of being born below average birth weight increases by 1.5pp. When we restrict the sample to settlements experiencing at least two days of bombing, the effect is 1.7pp (column 2). When the days of bombing are at least five (column 3), there is a 70% increase in the effect which jumps to 3pp and reaches 3.8pp when the bombing lasts 10 days or more (column 4).

In Panel B, Table 5, we show the results for high birth weight. In columns (1) and (2) the results are similar with a decrease of around 1.2pp on the likelihood of being born with high birth weight. The effect reaches a peak of -1.7pp after 5 days of bombing and then disappears. The results of the spatial analysis in Panel C and D of Table 5, confirm our previous findings of no difference in birth weight when comparing children in bombed and not bombed settlements.

We have also defined a continuous variable including the number of days a settlement or municipality was bombed, and we use it as an interaction term in the triple difference setting. We do not find that being born in a settlement or municipality that was more intensively bombed has an impact on outcomes at birth (Tables A3 through A6 in the Online Appendix).

Robustness

In this section, we vary the treatment period and examine whether our results are sensitive to changes in the sample. Additionally, we estimate placebo tests to see whether there were some pre- or post-trends in the data in the years before and after the bombing. In our estimation, we have assumed that in the absence of the 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 event study and placebo tests.

The focus of our study are infants who were in utero all 78 days of the bombing. In the main sample, we consider children who were born from June 10, 1999, to October 31, 1999, to be treated. The average duration of a normal pregnancy is 40 weeks or 280 days, and if everyone in the sample had this duration of pregnancy, we should include children born up to December 15, 1999. However, the duration varies from person to person – out of precaution and to take into account the possibility of premature births, we restrict the main sample to children born up to October 31, 1999. In Panel A of Table 6, we now extend the sample up to December 15, 1999, and we find that our findings are unchanged. Similarly, in panel B we include a part of newborns who were born during the bombing – in this second case, our sample includes newborns born from May 1, 1999, to October 31, 1999. Again the size of the estimated coefficients remains very stable. Finally, in Table 6, Panel C, we include all infants who were born during the NATO intervention. The size of the coefficient for below average birth weight falls marginally, but it remains statistically significant. The minimal reduction in the coefficient is not surprising, because children born during the bombing were exposed fewer days to the shock and thus including them reduces the coefficient slightly.

We now turn to the placebo tests. We estimate the model in equation (1), but we extend the baseline analysis by including the years 1996 and 1997 before the bombing and the years 2000 to 2002 after the bombing. For each year we include in the analysis children born between January 1 to mid-March and those from June 10 to October 31. The omitted group are children born in 1998 from June 10 to October 31 (as in the main analysis), and the difference-in-differences estimates should be interpreted with respect to this group. The estimated impact of bombing is reported for all outcomes as the difference-in-differences estimate for 1999 in Figure 6. Coefficients for all outcomes are insignificant for the pre- and post-treatment periods. This result strengthens further our findings on the impact of bombing on the birth weight of infants.[16]

Overall, all these robustness tests provide additional evidence that we are in the right direction in identifying a causal impact of the NATO bombing on birth weight.

4 Long-Term Outcomes: Educational Achievement

4.1 Data and Descriptive Statistics

In order to capture the long-term effects of the NATO bombing, we use a dataset, provided by the Serbian Ministry of Education (SMoE), containing the educational achievement of the whole population of pupils finishing primary school.[17] In Serbia primary school lasts eight years, from around age 6 to age 14. The pupils can formally finish primary school only if they sit the examination, containing tests in mathematics, the Serbian language and a mix of different subjects (geography, chemistry, physics, history, and biology). The total score of these tests together with the average grades from P6 to P8 class count for 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.[18] After learning the results of that final test, the pupils finishing primary education 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 behavior in the P8 class as outcomes.[19] 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. Behavior is graded on the same 1 to 5 scale. We also consider two additional variables related

[16]While we do run the placebo tests for the post-bombing period, we note that there could be compositional effects due to delayed fertility in this period. In the presence of compositional effects, the placebo tests for the post-bombing period would not be valid.

[17]It is not possible to match the birth records with the educational data, because we could only access birth records at the Statistical Office of the Republic of Serbia.

[18]For more information, visit this link (*in Serbian*).

[19]While test scores, introduced in 2014, 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 have been our main control year, the contents of the tests of the final examination were illegally sold to pupils before the actual examination (some media sources referring to this event can be found following this link (*in Serbian*) and, consequently, the test scores are not reliable.

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 in 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 of the following year. Consequently, in the treated group we can only 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. Children born from November until the end of February were either *in utero* during the bombing (those born up to mid-December) or they were conceived during the bombing (from mid-December until the end of February). We know that there is selective conception in times of uncertain events and this group of infants cannot be used as a control group. The whole cohort finishing primary school in 2014 was exposed to bombing to some degree. 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 columns (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 behavior. 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.

4.2 Estimation Strategy

Our main identification strategy for the long-term educational 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 a different estimation strategy based on 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 estimates 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 uses those weights to fit weighted regression models of the outcome 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

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

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 behavior, secondary school track preferences, and secondary school actual enrolment. X_i includes gender and month of birth of a pupil, γ_l are municipality of birth fixed effects. We cluster the standard errors at the school level; T_t is equal to 1 if children were *in utero* during the NATO bombing (born between June 10, 1999, and October 31, 1999) and equal to 0 for children born in the same month of the 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 further investigate the main results, we also perform causal mediation analysis (Celli, 2022; Imai et al., 2008), aiming to separate components pertaining to pregnancy from other effects of the bombing. We want to evaluate whether the bombing affects educational outcomes through birth weight, or the latter is mainly a short-term effect of the bombing. To this end, we first merge the two datasets – the national registry of birth records from the Statistical Office of the Republic of Serbia (SoRS) and the Dataset provided by the Serbian Ministry of Education (SMoE), containing educational achievement of the whole population of pupils finishing primary school. Since we do not have a unique identifier that would allow us to merge the two datasets at the individual level, we aggregate the data at the daily level, i.e., we use mean outcomes and mean characteristics at the daily level for both birth and educational outcomes and merge the two datasets by date of birth. The sample size in the main analysis on the educational dataset and the sample size which allows us to perform the causal mediation analysis differ because the main analysis uses individual data and causal mediation analysis uses aggregate data. All regressions, estimating the direct and indirect effects of bombing, only contain the gender of the child as a covariate.

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.

4.3 Estimation Results

Main Results

We show in Table 8 the causal impact of being *in utero* during the NATO bombing on long-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 for mathematics, around -0.03 points, which corresponds to around -0.9% reduction of the average grade (for the Serbian language this amounts to around -0.6%). For pupils in the treated group, we also notice a negative and highly statistically significant reduction in the probability of choosing a 4-year high school track (column 4), which is confirmed by lower enrolment in the same type of school (column 5). The size of the effect is around 1pp. 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. Both variables provide further evidence for the lack of ambition and achievement and could be explained through the reduced ability and cognition of the affected cohorts.

Causal Mediation Analysis

The results of mediation analysis are presented in Table 9. The total effect (ATE) of the NATO bombing on educational outcomes stays robust and comparable in size and significance to the estimates presented in Table 8. Only the total effect of the NATO bombing on the behaviour of affected children becomes marginally significant, at a 10% level. In terms of mediation, we see that while the direct effect (ADTE) of the NATO bombing remains significant for almost all educational outcomes, the indirect effect via below average birthweight is only significant for behaviour with a negative sign which offsets the positive direct effect. We cannot exclude that the persistent and statistically significant long-term effect of the NATO bombing on educational outcomes may be due to other factors not related to birthweight, although they do not completely offset the birthweight effect. Indeed, when comparing the ATE and the ADTE, the latter is smaller in magnitude implying that below average birthweight might have a "residual" role. In our analysis, we cannot disentangle the factors behind the persistent direct effect of the bombing. We can only provide some speculative evidence, e.g. we do not observe variation in teacher composition, and the effect of migration is complex.[20]

Placebo Tests

In Table A7 in the Online Appendix, we show the results of a placebo test using 1997 as the 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 A8 in the Online Appendix, we report the results of the Oster test (Oster, 2019) 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 by setting the highest possible levels of $R^2_{\max} = 1$ and $\delta = 1$, and indeed we do not find any statistically significant effect. However, the coefficients for mathematics, Serbian language, and secondary school first choice remain statistically significant for a level of $R^2_{\max}$ up to 0.3, and

[20]We have additionally collected data from the Statistical Yearbook of the Republic of Serbia on the number of teaching staff in primary, secondary and tertiary education in the period between 1998/1999 and 2015/2016. The full-time equivalent (FTE) number of teachers is stable until 2009/2010. Starting from 2010/2011, there has been a decline in the FTE number of teachers. Due to a demographic decline of children of school age over the same period, the per-student number of teachers has remained relatively stable or has been decreasing over time, from 19 in 1998/1999 to 17 in 2015/2016 in primary education (from 17 in 1998/1999 to 15 in 2015/2016 in secondary education and from 23 in 1998/1999 to 18 in 2015/2016 in tertiary education). In terms of migration, the period in the aftermath of the bombing was characterised by emigration which mostly had economic motives. In her PhD dissertation, Despić (2015) discusses the emigration from Serbia over this period by age, education, and geographical dispersion. The emigration from Serbia since 1999 is characterised by a higher emigration of highly-qualified individuals ("brain drain"); emigration of young individuals immediately after completing a tertiary degree; more emigration of highly-educated men before the age of 40, while more educated women migrated after the age of 45, after the period of completed fertility; migration was also characterised by stays shorter than 1 and 5 years – one factor which contributed to shorter stays was the entry of Serbia to "white" Schengen list in 2010, which allowed Serbian citizens to travel to European Schengen countries without a visa and stay up to three months in six months period; there were also some geographical changes in emigration patterns, but they have remained stable in the period between 2002 and 2011.

$\delta = 1$. Keeping $R^2_{\max} = 1$, the effects on mathematics and first choice still remain 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 behavior has never been significant in the main model, so the results of the Oster test do not add any additional information.[21] Nevertheless, taking into account that in the analysis of the long-term outcomes we do not have many available covariates to estimate the IPWRA model, we can conclude that our estimated coefficients are robust.

5 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 birth weight and stillbirth, as well as long-term outcomes, such as primary school grades and preferences for/enrolment in non-vocational secondary schools. Our main identification strategy uses a conditional before/after 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.

Our findings suggest that children who were *in utero* during the bombing were 2pp more likely to be born with a lower than average (< 3500g) birth weight and 1pp less likely to be born with high birth weight (≥ 4000g). We find no effects on low birth weight (< 2500g) and stillbirths. When examining the spatial intensity of bombing, we find that there is a common war effect – all pregnant mothers need to be considered as treated. We provide evidence that children born in more bombed settlements had a more negative impact compared to children born in less affected settlements. We conclude that the findings of this paper are comparable in magnitude to the existing literature summarised in Table 1, albeit for an outcome further up the birth weight distribution. For example, the papers by De Oliveira et al. (2025), 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, find an effect of

[21] We acknowledge the sensitivity analysis method proposed by Cinelli and Hazlett (2019) as an alternative approach. While robustness values in their method are based on R-squared, reflecting the proportion of variation explained by unobservables, the very low robustness value ($RV_q = 2\%$) observed in our application raises interpretability challenges, likely due to the limited covariates available in our model. For this reason, we opted for the Oster test, which offers greater flexibility in setting assumptions about unobservables, and we do not include the Cinelli and Hazlett results beyond this mention.

less than 2pp on the probability of low birth weight. Similar to our paper, Currie and Rossin-Slater (2013) do not find a significant effect of hurricane exposure during pregnancy on low birth weight. They explain that the incidence of low birth weight in children in their sample is only 6%. In our paper, only 5% of the children affected by the NATO bombing have low birth weight.

For the analysis of the long-term outcomes, due to the primary school starting age rule in Serbia, we cannot separate pupils within the same cohort, because the whole cohort finishing primary school in 2014 was exposed to bombing to some degree. 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 the Serbian Ministry of Education on educational achievement at the of primary school. We find that children who were *in utero* during the bombing had statistically significant lower grades in mathematics (around −0.9%) and Serbian language (around −0.6%) at the end of primary school (both effects are equivalent to about 0.03 standard deviations), and a 1pp increase in the preference for/enrolment in 3-year vocational secondary schools in comparison to 4-year grammar (more academically-oriented) secondary schools. Both variables provide further evidence for the lack of ambition and achievement and could be explained through reduced ability and cognition of the affected cohorts.

Compared to the findings in the disaster literature, Almond et al. (2015) 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 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, policymakers should seriously consider these estimated negative effects, not only because pupil's performance in mathematics is a useful measure of cognitive skills, but also because it is a good indicator for future educational and labor market outcomes (Machin and McNally, 2008; Schrøter Joensen and Skyt Nielsen, 2009).

Recent conflicts in Ukraine and Palestine have only reignited the importance of this topic and drew public attention to the severe consequences of wars and wartime bombing on affected children (Akbulut-Yuksel, 2022; Chupilkin and Kóczán, 2022). Sadly, they are only confirming the broader relevance of our paper – comparable people are being bombed and violence is upon European children again, with long-lasting effects. "The destruction of civil infrastructure, whether by the imposition of widespread sanctions or by bombing, is essentially a war on public health. The first victims are infants..." (Ashford and Gottstein, 2000). One policy implication of our findings 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 international conflicts – this type of intervention should be re-evaluated, taking all possible consequences into account.

References

Abadie, A. and G. W. Imbens (2008). On the failure of the bootstrap for matching estimators. *Econometrica 76*(6), 1537–1557. 16

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. 3, 5, 6, 10

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. (2022). Unaccounted long-term health cost of wars on wartime children. VoxEU.org, 10 May. 3, 28

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. 28

Almond, D., B. Mazumder, and R. van Ewijk (2015). In Utero Ramadan Exposure and Children's Academic Performance. *The Economic Journal 125*(589), 1501–1533. 28

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. 42

Armijos Bravo, G. and J. Vall Castelló (2021). Terrorist attacks, islamophobia and newborns' health. *Journal of Health Economics 79*, 102510. 42

Ashford, M.-W. and U. Gottstein (2000). The impact on civilians of the bombing of Kosovo and Serbia. *Medicine, Conflict and Survival 16*(3), 267–280. 3, 29

Atella, V., E. Di Porto, J. Kopinska, and M. Lindeboom (2024). Traumatic experiences adversely affect life cycle labor market outcomes of the next generation—evidence from wwii nazi raids. *Journal of the European Economic Association 22*(2), 963–1009. 6

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

Ben-Shlomo, Y. and D. Kuh (2002). 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 (2021). Maternal Stress during Pregnancy and Early Childhood Development. *Economics & Human Biology 43*, 101047. 3, 5, 10

Bhalotra, S. and D. Clarke (2020). The twin instrument: Fertility and human capital investment. *Journal of the European Economic Association 18*(6), 3090–3139. 10, 11

Black, S. E., P. J. Devereux, and K. G. Salvanes (2005). Why the apple doesn't fall far: Understanding intergenerational transmission of human capital. *The American Economic Review 95*(1), 437–449. 3, 6

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. 3, 5, 10

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. 42

Caldwell, J. C. (2006). Social upheaval and fertility decline. In *Demographic Transition Theory*, pp. 273–299. Springer. 11

Camacho, A. (2008). Stress and birth weight: Evidence from terrorist attacks. *The American Economic Review 98*(2), 511–515. 42

Catalano, R. and T. Bruckner (2006). Secondary sex ratios and male lifespan: Damaged or culled cohorts. *Proceedings of the National Academy of Sciences of the United States of America (PNAS) 103*, 1639–1643. 17

Catalano, R., T. Bruckner, A. R. Marks, and B. Eskenazi (2006). Exogenous shocks to the human sex ratio: the case of September 11, 2001 in New York City. *Human Reproduction 21*(12), 3127–3131. 17

Cattaneo, M. D. (2010). Efficient semiparametric estimation of multi-valued treatment effects under ignorability. *Journal of Econometrics 155*(2), 138–154. 24

Celli, V. (2022). Causal mediation analysis in economics: Objectives, assumptions, models. *Journal of Economic Surveys 36*(1), 214–234. 24

Chupilkin, M. and Z. Kóczán (2022). The Economic Consequences of War: Estimates Using Synthetic Controls. EBRD Working Paper. 3, 28

Cinelli, C. and C. Hazlett (2019). Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society Series B: Statistical Methodology 82*(1), 39–67. 27

Cozzani, M., M. Triventi, and F. Bernardi (2022). Maternal Stress and Pregnancy Outcomes Evidence from a Natural Experiment: The 2004 Madrid Train Bombings. *European Sociological Review 38*(3), 390–407. 18

Currie, J. and M. Rossin-Slater (2013). Weathering the Storm: Hurricanes and Birth Outcomes. *Journal of Health Economics 32*(3), 487–503. 7, 10, 17, 27, 28, 42

Currie, J. and M. Rossin-Slater (2015). Early-life origins of life-cycle well-being: Research and policy implications. *Journal of Policy Analysis and Management 34*(1), 208–242. 6, 7

De Oliveira, V. H., I. Lee, and C. Quintana-Domeque (2025). Natural Disasters and Early Human Development: Hurricane Catarina and Infant Health in Brazil. *Journal of Human Resources 60*. 7, 17, 27, 42

Despić, J. (2015). Migracije visokoobrazovanih lica iz Srbije od 1991. godine u Kanadu i Sjedinjene Američke Države. PhD Disertation. Belgrade: Faculty of Economics, University of Belgrade. 26

Duque, V. (2017). Early-life conditions and child development: Evidence from a violent conflict. *SSM - Population Health 3*, 121 – 131. 42

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. 13, 42

Fenrick, W. (2001). Targeting and Proportionality during the NATO Bombing Campaign against Yugoslavia. *European Journal of International Law 12*(3), 489–502. 8

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. 16

Hosmer, S. T. (2001). *The Conflict Over Kosovo: Why Milosevic Decided to Settle When He Did*. Rand Corporation. 8

Imai, K., L. Keele, and D. Tingley (2008). A general approach to causal mediation analysis. *Psychological methods 15*(4), 309–334. 24

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. 24

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. 42

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. 42

Krstić, D., D. Marinković, L. Mirković, and J. Krstić (2006). Pregnancy outcome during the bombing of Yugoslavia from March 24 to June 9, 1999. *Vojnosanitetski Pregled 63*(4), 377–382. (in Serbian). 5, 10

Krstić, D., S. Pop-Trajković, L. Mirković, D. Marinković, and J. Krstić (2007). The influence of prenatal stress on delivery outcome. *Facta universitatis-series: Medicine and Biology 14*(1), 30–37. 5, 10

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. 28

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. 42

Marić, N. P., B. Dunjić, D. J. Stojiljković, D. Britvić, and M. Jasovic-Gasić (2010). Prenatal stress during the 1999 bombing associated with lower birth weight-a study of 3,815 births from Belgrade. *Archives of Women's Mental Health 13*(1), 83–89. 5, 42

Menclova, A. K. and S. Stillman (2020). Maternal stress and birth outcomes: Evidence from an unexpected earthquake swarm. *Health Economics 29*(12), 1705–1720. 42

Merklinger-Gruchala, A., G. Jasienska, and M. Kapiszewska (2019). Paternal investment and low birth weight – The mediating role of parity. *PLOS One 14*(1), e0210715. 18

Oreopoulos, P., M. Stabile, R. Walld, and L. L. Roos (2008). Short-, medium-, and long-term consequences of poor infant health: An analysis using siblings and twins. *The Journal Human Resources 43*(1), 88–138. 3, 6, 12

Oskorouchi, H. R. (2019). Learning to Fight: Afghan Child Health and In-utero Exposure to Conflict. *Population and Development Review 45*(2), 275–300. 8

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics 37*(2), 187–204. 25, 26

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. 3, 5, 10

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. 10, 42

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. 42

Rubin, D. B. (2018). Multiple imputation. In *Flexible Imputation of Missing Data, Second Edition*, pp. 29–62. Chapman and Hall/CRC. 14

Sardoschau, S. (2024). In utero exposure to violence and child health in Iraq. *Journal of Population Economics 37*(40). 8

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. 28

Smith, J. A. and P. E. Todd (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics 125*(1), 305–353. 16

Teodorović, M. (2000, April). Lokalni i savezni izbori – zajedno? *Glas Javnosti*. http://arhiva.glas-javnosti.rs/arhiva/2000/05/02/srpski/P00042808.shtm. 8

Tkalec, M. and I. Žilić (2021). Does proximity to conflict affect tourism: Evidence from NATO bombing. *PLoS ONE 16*(10), e0258195. 5

Torche, F. (2011). The Effect of Maternal Stress on Birth Outcomes: Exploiting a Natural Experiment. *Demography 48*(4), 1473–1491. 7, 17, 27, 42

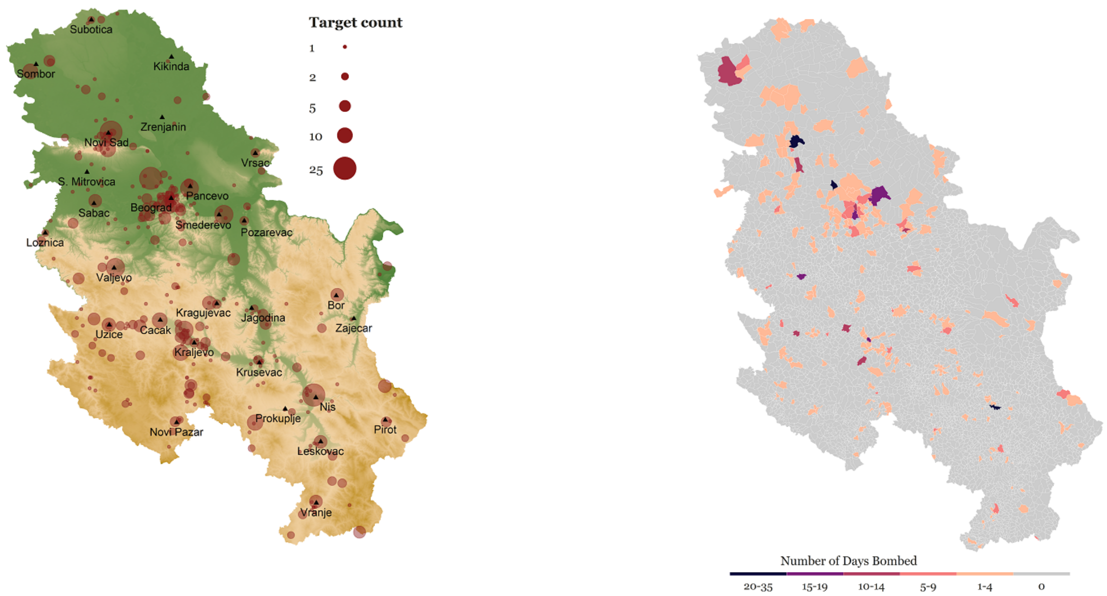
Van den Berg, G. J. and M. Lindeboom (2018). Famines, Hunger, and Later-Life Health. Oxford Research Encyclopaedia of Economics and Finance. Oxford University Press. 6

Vreme (2000, June). Studije: Srbija 2000 – godinu dana posle bombardovanja. *Vreme* (494). https://www.vreme.com/arhiva_html/494/08.ASP. 8

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. 13

Figures

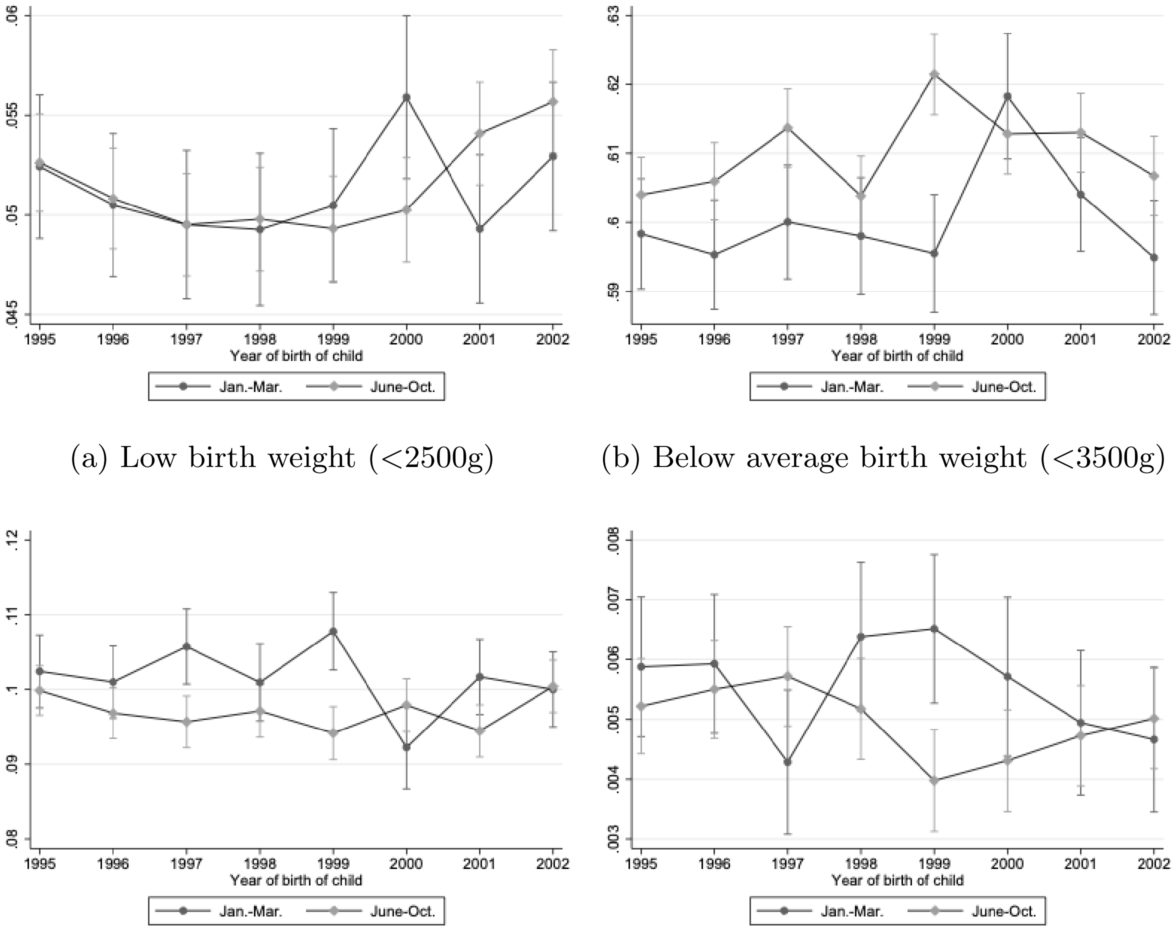
Figure 1: Spatial Distribution of 1999 NATO Bombing of Serbia



(a) Attacks by Target

(b) Number of Days Bombed (settlements)

Figure 2: Trend of Birth Outcomes



(a) Low birth weight (<2500g)

(b) Below average birth weight (<3500g)

(c) High birth weight (>4000g)

(d) Stillbirth

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

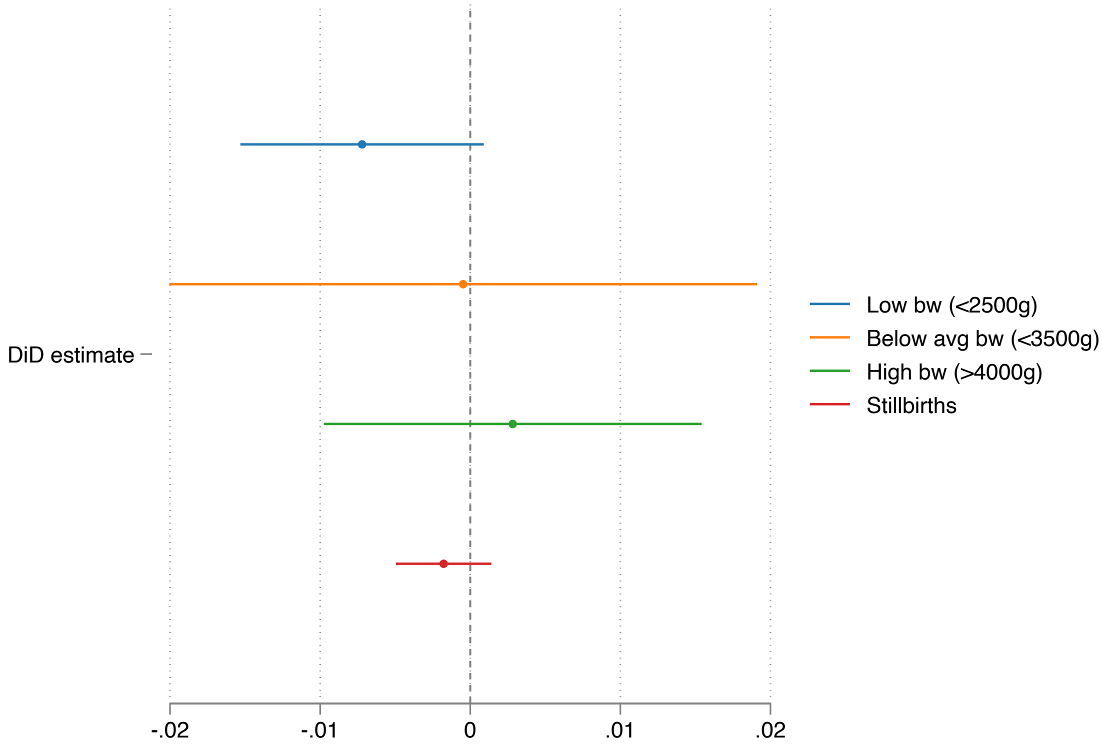


Figure 4: The Impact of NATO Bombing on Below Average and High Birth Weight– Heterogeneity

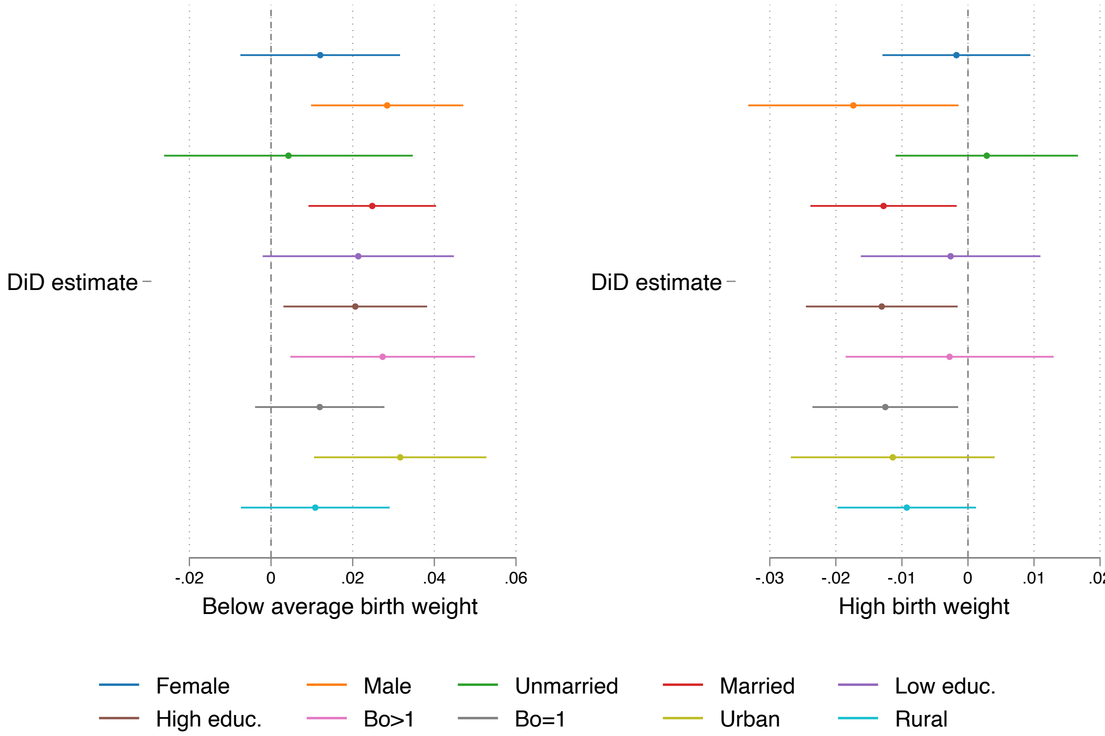


Figure 5: Impact of Bombing on Birth Outcomes: Spatial Difference-in-Difference Model

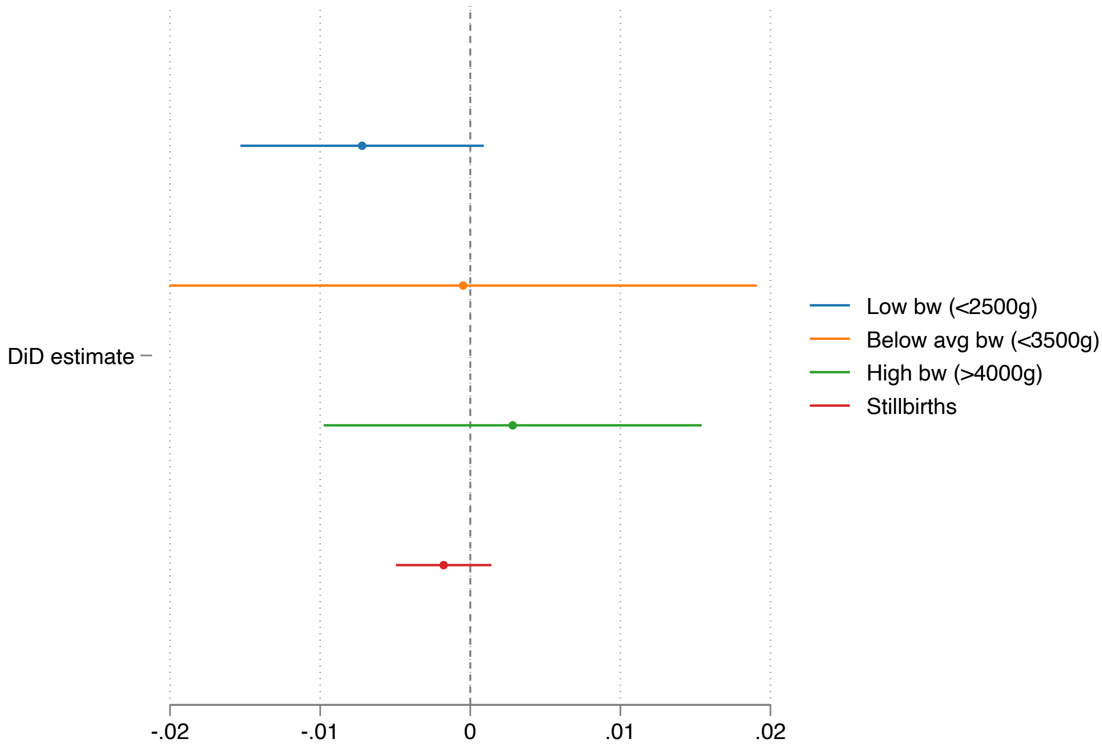
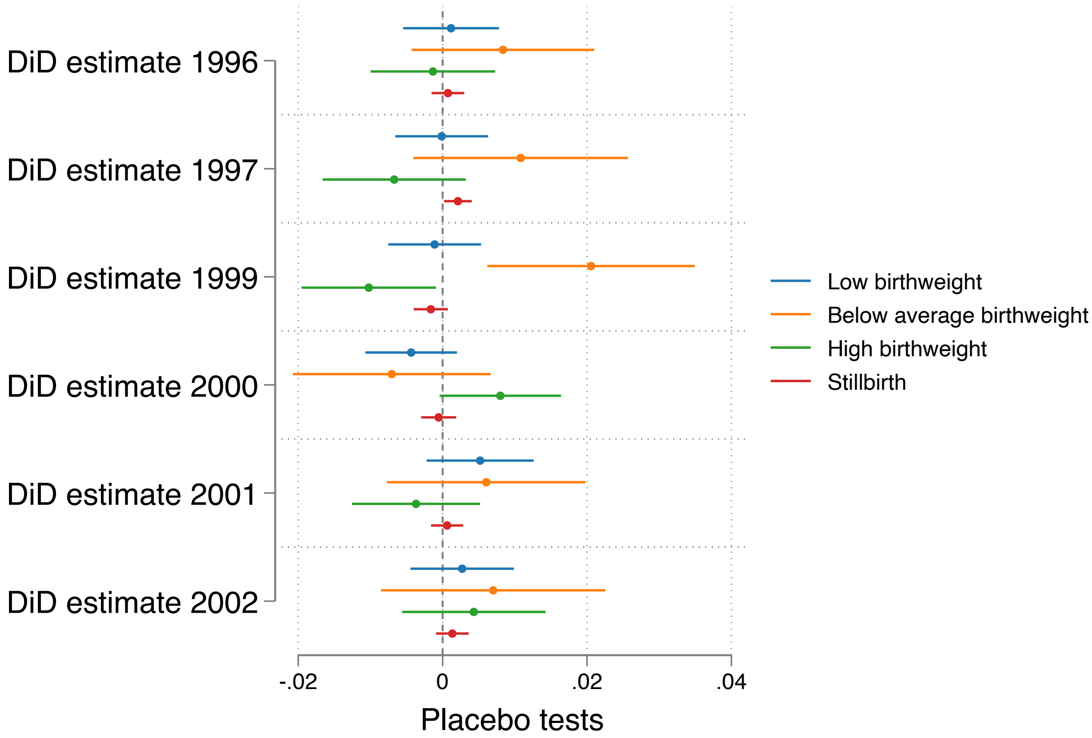


Figure 6: Placebo tests for pre- and post-bombing period



Tables

Table 1: Comparison of the In Utero Results Across Literature

Type exogenous variation	Event	Country	Year of the event	Results: BW (grams)	Results: P(LBW)	Results: 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-Domeque 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 Vall Castelló (2021)	Jihadi terrorist attacks (Muslim women)	Spain (Catalonia)	2017	↓12.89g	↑1.6pp -	
Marić et al. (2010)	The NATO bombing	Serbia (Belgrade)	1999	↓86g	-	-
This paper (2022)	The NATO bombing	Serbia	1999	-	↑2pp	↓1% (0.03SD)
B. Natural disasters (famines, hurricanes, pandemics)						
Camacho (2008)	Landmine explosions	Colombia	1998-2003	↓8-12g	-	-
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	-
Andalón 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 Oliveira et al. (2025)	Hurricane Catarina	Brazil	2004	↓44.4g	↑1.7pp	-

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-Domeque 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 birth weight is found for short mothers and mothers who smoke. Test score results at ages 7 and 11.

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		1999	1998	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A.Outcomes						
birth weight (categorical)			***			Not sign.
Up to 499g	0.1%	0.1%		0.1%	0.1%	
500g-999g	0.1%	0.2%		0.1%	0.2%	
1000g-1499g	0.5%	0.5%		0.5%	0.5%	
1500g-1999g	1.0%	1.2%		1.1%	1.1%	
2000g-2499g	3.2%	3.0%		3.1%	3.1%	
2500g-2999g	16.7%	16.3%		16.2%	16.4%	
3000g-3499g	40.7%	39.3%		38.9%	38.4%	
3500g-3999g	28.3%	29.7%		29.9%	29.4%	
4000g-4499g	8.4%	8.4%		8.8%	9.4%	
More than 4500g	1.0%	1.3%		1.3%	1.3%	
Low birth weight (<2500g)	4.9%	5.0%	-0.1%	4.9%	5.0%	-0.1%
Below average birth weight (<3500g)	62.3%	60.6%	1.7%***	60.0%	59.8%	0.2%
High birth weight ($\geq$ 4000g)	9.4%	9.7%	-0.3%	10.1%	10.8%	0.7%*
Stillbirth	0.4%	0.5%	-0.1%**	0.6%	0.7%	0.89
Panel B. Individual charasteristics						
Female	47.9%	48.2%	-0.03%	48.7%	48.3%	0.4%
Born in hospital	98.8%	98.9%	-0.1%*	99.1%	98.7%	0.4%***
Parents married at birth	79.6%	79.7%	-0.1%*	80.9%	80.6%	0.8%
Mother's years of education	11.008	10.924	0.084***	10.956	10.963	-0.007
Mother employed	40.1%	39.7%	0.4	40.1%	39.8%	0.3%
Mother's age	26.111	25.949	0.162***	25.939	26.050	-0.127
Number of years married	3.425	3.353	0.072**	3.883	3.930	-0.047
Has father data	85.9%	86.2%	-0.3%	87.7%	86.1%	1.6%***
Father's years of education	11.650	11.629	0.021	11.562	11.623	-0.061*
Father employed	83.3%	85.7%	-2.4%***	84.8%	84.0%	0.8%*
Father's age	30.042	29.915	0.127**	29.859	29.938	-0.079
Observations	27,154	27,820		12,849	12,740	
Observations with father data	23,536	24,125		11,328	11,088	

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 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 differences of children born from January 1 to mid-March, in years 1999 and 1998. * 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 Outcomes – Main Results

	Low bw (<2500g) (1)	Below avg bw (<3500g) (2)	High bw (≥ 4000g) (3)	Stillbirths (4)
Panel A. Difference-in-Differences: Main model				
	-0.001	0.021***	-0.010**	-0.002
	[0.003]	[0.007]	[0.005]	[0.001]
Observations	79,837	79,837	79,837	79,837
Adj. R-squared	0.016	0.045	0.022	0.002
Panel B. Difference-in-Differences: Main model with father controls				
	-0.001	0.020***	-0.010**	-0.002
	[0.003]	[0.007]	[0.005]	[0.001]
Observations	79,837	79,837	79,837	79,837
Panel C. Matching Difference-in-Differences				
	-0.001	0.020***	-0.010**	-0.002
	[0.003]	[0.008]	[0.005]	[0.001]
Observations	79,831	79,831	79,831	79,831
Dep. var. mean	0.049	0.604	0.099	0.005
Dep. var. SD	0.216	0.489	0.298	0.072
Controls	X	X	X	X
Municipality FE	X	X	X	X

Notes: This table presents estimated baseline coefficients for the differences-in-differences model without and with father data in panel A and B, and difference-in-differences matching model in panel C. For the model presented in Panel B, we use multiple imputations to impute missing father controls. Controls: female, 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 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 Birth Outcomes – Heterogeneity

	Female	Male	Unmarried	Married	Low educ.	High educ.	Bo >1	Bo =1	Urban	Rural
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Below average (<3500g)										
	0.012	0.028***	0.004	0.025***	0.027**	0.012	0.032***	0.011	0.027***	0.012
	[0.010]	[0.009]	[0.015]	[0.008]	[0.012]	[0.009]	[0.011]	[0.009]	[0.011]	[0.008]
Dep. var. mean	0.665	0.546	0.720	0.576	0.693	0.564	0.571	0.634	0.624	0.592
Dep. var. SD	0.472	0.498	0.449	0.494	0.461	0.496	0.495	0.482	0.484	0.491
Observations	38,455	41,382	15,608	64,229	24,075	46,128	39,79	39,841	55,762	31,801
Adj. R-squared	0.026	0.036	0.041	0.032	0.048	0.028	0.054	0.038	0.046	0.044
Panel B: High birth weight (≥4000g)										
	-0.002	-0.017**	0.003	-0.013**	-0.003	-0.013**	-0.011	-0.009*	-0.003	-0.013**
	[0.006]	[0.008]	[0.007]	[0.006]	[0.007]	[0.006]	[0.008]	[0.005]	[0.008]	[0.006]
Dep. var. mean	0.068	0.127	0.057	0.109	0.068	0.112	0.118	0.080	0.095	0.100
Dep. var. SD	0.252	0.333	0.232	0.311	0.251	0.316	0.322	0.271	0.293	0.300
Observations	38,455	41,382	15,608	64,229	24,075	55,762	39,79	39,841	31,801	46,128
Adj. R-squared	0.007	0.017	0.014	0.019	0.020	0.018	0.025	0.017	0.023	0.021

Notes: This table presents estimated baseline coefficients for the differences-in-differences model for different groups: male/female (columns (1) and (2)), unmarried/married (columns (3) and (4)), low education/high education (columns (5) and (6)), birth order higher than 1/birth order equal to 1 (columns (7) and (8)) and urban/rural (columns (9) and (10)). Controls: female (not in columns (1) and (2)), parents married and the characteristics of the mother: age, years of education (not in columns (5) and (6)) and employment status. All regressions include month 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 5: The Impact of NATO Bombing on birth weight Within Bombed Settlements by the Number of Days Experiencing Bombing

Panel A: Outcome below average birth weight <3500g (within bombed settlements)				
	≤ 1	≤ 2	≤ 5	≥ 10
	0.014*	0.017**	0.030**	0.038*
	[0.008]	[0.008]	[0.013]	[0.020]
Dep. var. mean	0.591	0.590	0.592	0.578
Dep. var. SD	0.492	0.492	0.491	0.494
Observations	39,567	32,555	17,725	7,018
Adj. R-squared	0.043	0.042	0.044	0.035
Panel B: Outcome high birth weight ≥ 4000g (within bombed settlements)				
	≤ 1	≤ 2	≤ 5	≥ 10
	-0.013**	-0.012*	-0.017**	-0.009
	[0.006]	[0.006]	[0.007]	[0.010]
Dep. var. mean	0.102	0.102	0.101	0.105
Dep. var. SD	0.302	0.303	0.302	0.307
Observations	39,567	32,555	17,725	7,018
Adj. R-squared	0.020	0.020	0.017	0.014
Panel C: Outcome below average birth weight <3500g (spatial variation)				
	≤ 1	≤ 2	≤ 5	≥ 10
	0.005	0.006	0.014	0.027
	[0.015]	[0.016]	[0.017]	[0.022]
Dep. var. mean	0.601	0.601	0.606	0.609
Dep. var. SD	0.490	0.490	0.489	0.488
Observations	32,227	27,456	17,334	10,010
Adj. R-squared	0.045	0.045	0.047	0.049
Panel D: Outcome high birth weight ≥ 4000g (spatial variation)				
	≤ 1	≤ 2	≤ 5	≥ 10
	-0.002	-0.003	0.006	-0.005
	[0.020]	[0.016]	[0.017]	[0.022]
Dep. var. mean	0.106	0.107	0.106	0.103
Dep. var. SD	0.307	0.309	0.307	0.304
Observations	32,227	27,456	17,334	10,010
Adj. R-squared	0.052	0.048	0.051	0.049

Notes: This table presents estimated coefficients for different levels of bombing intensity. In Panel A and 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)). In Panel C and D, the treatment group includes bombed settlements experiencing the same increasing number of days of bombing. The control group includes settlements not bombed located in the top decile of the distance from bombed settlements. Controls: female, parents married and the characteristics of the mother: age, years of education and employment status. All regressions include month 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 birth weight Outcomes – Robustness

	Below avg bw (<3500g)	High bw ($\geq$ 4000g)
Panel A: Period 10/6/1999 - 15/12/1999		
	0.021***	-0.010**
	[0.007]	[0.005]
Dep. var. mean	0.603	0.098
Dep. var. SD	0.489	0.298
Observations	94,749	94,749
Adj. R-squared	0.046	0.022
Panel B: Period 1/5/1999 - 31/10/1999		
	0.021***	-0.011**
	[0.007]	[0.005]
Dep. var. mean	0.606	0.099
Dep. var. SD	0.489	0.298
Observations	93,815	93,815
Adj. R-squared	0.046	0.022
Panel C: Period 24/3/1999 - 31/10/1999		
	0.017**	-0.012**
	[0.007]	[0.004]
Dep. var. mean	0.608	0.099
Dep. var. SD	0.488	0.298
Observations	106,065	106,065
Adj. R-squared	0.045	0.022

Notes: This table presents estimated baseline coefficients when we vary the treated period. Columns (1) and (2) shows estimates for below average birth weight and high birth weight. Controls: female, parents married and the characteristics of the mother: age, years of education and employment status. All regressions include month 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	0.917	0.914	-0.010***	-0.006*
	[0.290]	[0.276]	[0.281]	(-3.985)	(-2.569)
Enrolled 4y[c]	0.883	0.892	0.878	-0.008**	0.005
	[0.321]	[0.311]	[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 2013 to 2015.

[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[c]
	(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 month of birth and school fixed effects. 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 2013 to 2015.

[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 9: Causal Mediation Analysis: The Effect of NATO Bombing on Schooling Outcomes

	Marks at the end of P8 class			Secondary school	
	Language	Mathematics	Behaviour	First wish 4y	Enrolled 4y
	(1)	(2)	(3)	(4)	(5)
AITE					
	-0.003	-0.004	-0.002**	-0.001	-0.001
	(0.003)	(0.003)	(0.001)	(0.001)	(0.001)
ADTE					
	-0.017	-0.027**	0.008**	-0.007**	-0.006*
	(0.011)	(0.012)	(0.003)	(0.003)	(0.003)
ATE					
	-0.020*	-0.030***	0.006*	-0.008***	-0.007**
	(0.011)	(0.011)	(0.003)	(0.003)	(0.003)
Observations	288	288	288	288	288

Notes: This table presents estimated coefficients with causal mediation analysis model with subject marks as outcomes. The data are aggregated at the day-of-birth level. Each outcome is estimated using female as a control variable. Standard errors in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%. *Source:* Final examination dataset from the Serbian Ministry of Education for years 2013 to 2015.

[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.