Destruction from Above: Long-Term Impacts of WWII Tokyo Air Raids*

Masataka Harada†, Gaku Ito‡, and Daniel M. Smith§

October 16, 2019

Abstract

What are the long-term socioeconomic impacts of wartime violence? We use historical aerial imagery of the aftermath of the United States' indiscriminate firebombing of Tokyo in 1945 to generate detailed neighborhood-level data on damages—helping us to overcome the methodological challenges of nonrandom assignment and coarse measurement. Decades after the air raids, the most heavily bombed neighborhoods continued to suffer socioeconomically—with higher crime, lower educational attainment, and higher unemployment. Causal mediation analysis reveals that these patterns cannot be explained by the ratio of new residents or the construction of high-rise buildings, and a geo-coded survey featuring behavioral experiments indicates less altruism in affected neighborhoods. In contrast to previous studies that stress how violence might affect social cohesion (positively or negatively) through individual or family-level trauma, our findings suggest that community-level exposure to violence might create persistent legacies by displacing victims and altering the urban landscape, thereby fragmenting local communities. (150 words)

Word Count

11,817 words (excluding Online Appendix)

Keywords

political violence, World War II, Tokyo firebombing, Japan, historical legacies, causal inference, geographic information systems (GIS), remote-sensing

*We thank Ryotaro Inumaru for research assistance. For helpful feedback, we thank Sanford Gordon and participants at the 1st Annual Meeting of the Japanese Society for Quantitative Political Science, the microeconomics workshop at the University of Tokyo, and the CAES seminar at Fukuoka University. This work was supported by JSPS KAKENHI Grant Numbers JP15K16977, JP18KT0054, JP18KK0039, and JP19K21681; and financial support (No. 187102) from the Central Research Institute of Fukuoka University.

†Associate Professor, Fukuoka University. Email: masatakaharada@gmail.com. Corresponding author.

‡Assistant Professor, Hiroshima University. Email: gkit@hiroshima-u.ac.jp

§Associate Professor, Harvard University. Email: danielmsmith@fas.harvard.edu

A growing body of literature highlights the long-term impact of historical events on contemporary social, economic, and political outcomes (e.g., Acemoglu, Johnson, and Robinson, 2001; Davis and Weinstein, 2002; Nunn and Wantchekon, 2011; Michalopoulos and Papaioannou, 2016; Acharya, Blackwell, and Sen, 2016b). The long-term legacies of war are particularly important to understand, since wartime violence is believed to affect the entire lives of those involved, as well as the lives of their descendants, regardless of whether the impact is positive or negative (e.g., Wood, 2008; Bellows and Miguel, 2009; Miguel and Roland, 2011; Grosjean, 2014). Previous studies have also revealed that the transmission of these historical legacies across generations occurs not only through individuals, but also through entire communities (Kocher, Pepinsky, and Kalyvas, 2011; Dell and Querubin, 2018; Lowes, Nunn, Robinson et al., 2017; Rozenas, Schutte, and Zhukov, 2017).

Nevertheless, much of the existing research does not measure the impact of war destruction *per se* but rather individuals' responses to suffering, and how this suffering is spatiotemporally translated to their families and descendants (Miguel, Saiegh, and Satyanath, 2011; Lupu and Peisakhin, 2017). Surprisingly few studies evaluate the *community-level* impact of wartime destruction, and causal investigation of these effects, in particular, has been limited due to the methodological challenges of nonrandom assignment of bombings or other forms of violence, or coarse measurement of the destruction at higher levels of aggregation (e.g., Davis and Weinstein, 2002; Brakman, Garretsen, and Schramm, 2004; Miguel and Roland, 2011; Zhukov and Talibova, 2018). In this study, we contribute to this emergent literature by examining the long-term legacies of the United States' air raids over Tokyo, Japan, in the final stages of World War II. More specifically, we examine the long-term impacts of indiscriminate firebombing on contemporary socioeconomic outcomes, making use of an original data set of georeferenced damages at the neighborhood level.

The key to measuring and assessing community destruction is to exploit exogenous variation of the treatment and identify the treatment assignment mechanisms.[1] The indiscrim-

[1]Here, the treatment is the bombing, and the treatment assignment mechanisms refer to the way in which the bombings were implemented.

inate nature of the firebombing of densely populated areas of Tokyo is a rare historical example of war destruction that meets these conditions for causal identification. Under Major General Curtis E. LeMay, the US military shifted its tactics in the strategic bombing of Japan in March 1945: instead of using its B-29 bombers in high-altitude precision bombing, the strategy shifted to low-altitude indiscriminate deployment of incendiary bombs designed to burn down civilian housing (primarily made of wood and paper) in densely populated cities such as Tokyo, Osaka, Nagoya, and Kobe. The Tokyo bombings were conducted over several air raid campaigns, burning about 2.5 million houses, killing about 100,000 people (mostly civilians), and injuring roughly the same number (Saotome, 1981b). The bombing of the *shitamachi* (downtown) areas of Tokyo on March 10, 1945, inflicted the worst damage by ordinary weapons ever recorded in past warfare.

The Tokyo firebombing left varying degrees of damage within the affected areas, due mainly to differences in locations of natural and artificial fire lanes ("green spaces"), poor weather, and the indiscriminate bombing strategy (Saotome, 1981b; Searle, 2002). The Kanto Plain in which Tokyo resides features small hills and valleys that contributed to natural variation in damages. The gardens of residences of members of the Imperial family, large parks, rivers, canals, wide roads, and other large structures also served as natural fire lanes. Moreover, the bombing on March 10 was the first large-scale bombing campaign under a new strategy. Bombers approached at night, with no radio communication, and in the face of a strong jet-stream, which resulted in near zero visibility of targets on the ground. These conditions made the damages of the air raids largely unpredictable, and generated plausibly exogenous variation in damages within a relatively small geographic area.

To measure the micro-geography of raid damages, we utilize historical aerial imagery and historical GIS and remote-sensing techniques to develop an original data set measuring the neighborhood-level damage in Tokyo. The data set relies mainly on aerial images taken by the US Army Air Force (USAAF) and the US Air Force (USAF) between 1946 and 1948. We first georeferenced the imagery relying on predetermined geographic control points, and then coded the neighborhood-level damage with a human-coding scheme. We measured damages at the *cho-chomoku* level, which is the lowest administrative units in Japan, and roughly corresponds to the popular notion of a neighborhood. A noteworthy aspect of our geographic unit of analysis is its disaggregated nature. Specifically, the geographic size of *cho-chomoku* is considerably finer than counterparts in previous studies, with an average size in geographic area of roughly 1.6% that of a typical hamlet or village-level unit of observation (e.g., Kocher, Pepinsky, and Kalyvas, 2011; Dell and Querubin, 2018).

A second important contribution of our extensive data development process is that our data set is *doubly* remote-sensed. It is *spatially* remote-sensed, as it primarily relies on aerial photographic images—or the "view from above" (Donaldson and Storeygard, 2016). The image-based, remote-sensing approach has been widely utilized in the natural sciences, such as forestry science and engineering, yet applications in the social sciences are still uncommon. Our data is also *temporally* remote-sensed, as we utilize archives of historical images. A growing literature in political economy now utilizes historical data, but the vast majority of studies exploit text-based records. To our knowledge, this study is among the first attempts to estimate the impacts of geographically distributed historical events using an image-based data set constructed from both spatially and temporally remote-sensed imagery.

We regress a variety of neighborhood-level socioeconomic outcome variables from the geocoded census on the extent of damage inflicted on the neighborhood in the bombing campaign, controlling for *all known treatment assignment mechanisms of the bombing.* Although the firebombing of Tokyo was broadly characterized by its indiscriminate targeting (the Japanese described it as "blind bombing" at the time), the treatment assignment was not completely random. Specifically, USAAF avoided bombing the Imperial Palace and set several aiming points before each operation (Saotome, 1981b; Williams, 2007). We compile the geocoordinates of these aiming points from declassified historical archives and secondary sources (Okuzumi, 2007). To account for these treatment assignment mechanisms, we carefully control for the proximity of each *cho-chomoku* to the Imperial Palace and aiming points, as well as an extensive set of additional geographic covariates. We also pay careful attention to concerns of nonrandom treatment assignment of the raid damages—as we show in balance checks, the difference in flammability between types of residences, or underlying income heterogeneities, does not explain difference in damages.

Our results indicate that the more heavily bombed neighborhoods experienced lower socioeconomic performance, even several decades later, including higher crime rates (apart from burglary), lower levels of educational attainment, higher unemployment, and fewer professional and executive workers. Notably, these findings deviate from previous studies using less fine-grained data to measure the long-term effects of US-led bombings of Japan (Davis and Weinstein, 2002), Germany (Brakman, Garretsen, and Schramm, 2004), and Vietnam (Miguel and Roland, 2011), which find little evidence of persistent impacts of the bombings on local population outcomes or economic performance. Our results are robust to several tests, including a subsample re-estimation. Moreover, mediation analyses using sequential g-estimation show that the effects do *not* appear to be mediated through the construction of new high-rise buildings or through higher proportions of current residents who moved to the neighborhood from elsewhere.

A geo-coded internet survey with two behavioral experiments provides some insight into the mechanisms behind our empirical findings. In the existing literature, the question of *how* violence affects socioeconomic outcomes in local communities is unclear. Some studies find that exposure to violence *weakens* local social cohesion or trust (e.g., Nunn and Wantchekon, 2011; Grosjean, 2014), putting communities at risk of falling into "conflict traps" or "poverty traps" that have long-term consequences (e.g., Collier, Elliott, Hegre et al., 2003). Other studies find that exposure to violence *strengthens* pro-social ties through collective action and "post-traumatic growth" (e.g., Wood, 2003; Bellows and Miguel, 2006; Blattman, 2009; Gilligan, Pasquale, and Samii, 2014; Bauer, Blattman, Chytilová et al., 2016).[2] Our survey,

[2] A challenge is that both processes might occur simultaneously (Gilligan, Pasquale, and Samii, 2014), making the total effects of wartime violence ambiguous and hard to disentangle. Moreover, the interpretation of any observed effects of violence must be conceptualized as the *combined effect* of death, injuries, and destruction, as well as human responses to this violence; and the long-term effects of violence on socioe-

which includes classic behavioral games designed to measure pro-social attitudes, shows that residents of more severely damaged neighborhoods are less likely to come from families that were personally affected by the bombings—suggesting displacement of the local populations who experienced the violence—and tend to display less altruistic behavior. Together, our findings thus suggest a causal pathway in which the firebombing left a negative legacy in parts of modern Tokyo by destroying local informal institutions and weakening social cohesion.

1 Legacies of Community Destruction

Large-scale wartime violence can have both immediate and long-lasting impacts on subsequent society through two distinct channels: individuals or families; and communities. At the individual or family level, violence might alter the preferences, attitudes, and behaviors of victims by imposing physical and/or psychological damages, and undermining subsequent sociopolitical status (Dell and Querubin, 2018). Previous studies focusing on immediate survivors generally suggest that exposure to wartime violence *increases* prosocial behaviors among those affected (e.g., Wood, 2003; Bellows and Miguel, 2006; Blattman, 2009; Bauer, Blattman, Chytilová et al., 2016), due to the stimulation of collective action in response to violence and post-traumatic growth in individuals' psychological development or outlooks.

Individual-level legacies can also be transmitted across generations as the descendants of victims would, at least partly, inherit the deteriorated socioeconomic status and psychological responses. For example, Lupu and Peisakhin (2017) investigate how parental suffering from political violence is passed down to descendants through a survey of Crimean Tatars, a minority Muslim population that was subject to deportation by the ethnic Russian majority of the Soviet state in 1944, and suffered starvation and disease: descendants of individuals who suffered intensely tend to have stronger ethnic identities, and tend to exhibit increased

conomic outcomes may be further conditioned by variables such as the nature (civil war versus interstate conflict) and outcome (victory versus defeat) of the conflict, and the ethnic or cultural heterogeneity of the affected populations. We cannot estimate the impact of war in general, nor a "typical" war, since there is no such thing.

support for the Crimean Tatar political leadership and stronger hostility toward Russia. Descendants of victims are also found to be more politically active, while family-level exposure does not systematically shape religious radicalization.

An important implication from family-level studies of intergenerational legacies of violence is that individuals do not have to be directly exposed to the damages to be affected by them. This insight leads to the second causal channel linking exposure to violence to subsequent socioeconomic outcomes, one that focuses on the role of local communities and informal institutions rather than individual-level experiences.[3] Rozenas, Schutte, and Zhukov (2017) and Zhukov and Talibova (2018), for example, highlight how the community-level exposure to Stalin-era mass repression shapes contemporary political participation and voting behavior in Ukraine and Russia, respectively. Specifically, community-level exposure to indiscriminate violence is thought to generate a shared belief in a common threat and a collective identity, thereby resulting in a lasting negative attitude against the long-gone perpetrator.

Gilligan, Pasquale, and Samii (2014) also find that community-level exposure to violence increases social cohesion, but posit an alternative set of mechanisms. Communities that were more affected by Nepal's long civil war (1996-2006) exhibit significantly higher levels of altruistic giving and social trust—a pattern that the authors attribute to both a "purging mechanism" (less pro-social individuals flee, leaving a concentration of pro-social individuals) and a "coping mechanism" (those members of the community who remain "band together" to collectively survive and overcome the trauma).

We stress, in comparison, that the community-level exposure to violence might cause persistent legacies of lower social cohesion by *displacing immediate victims and altering the urban landscape, thereby fragmenting local communities.* Rather than generating strong collective identities among community members, extensive war damages might in some instances annihilate the community members who would otherwise pass down those memories to successive generations, creating a vacuum of informal institutions, norms, and social ties.

[3] By local community, we mean geographically small-sized regions such as neighborhoods, where physical proximity of residents facilitates face-to-face communication and interactions.

In other words, in contrast to a process of purging and coping among affected communities, extreme violence may result in a dramatic turnover of residents and neighbors, erasing whatever social capital had been built prior to the violence, and purging even the pro-social community members who might serve to rebuild the post-conflict community in positive ways.

In the context of post-WWII Japan, for example, war orphans and individuals who suffered from the air raids were often displaced to non-originating localities where their caregivers (usually relatives) resided. The vast land lots cleared by the air raids also stimulated land redevelopment, resulting in a reconfiguration of the affected neighborhood's space, environment, and residential composition. These dynamics in turn might have resulted in physical and psychological reconstructions of local communities in ways that weakened social cohesion and trust relative to unaffected communities.

This argument is in line with insights from institutional economics and historical institutionalist accounts in political science that provide theoretical and empirical illustrations for how institutions (or their removal) bridge historical treatments and contemporary socioeconomic outcomes (e.g., Acemoglu, Johnson, and Robinson, 2001; Michalopoulos and Papaioannou, 2016; Acharya, Blackwell, and Sen, 2016b; Lowes, Nunn, Robinson et al., 2017). The destruction of local communities impacts subsequent social, economic, and political realities because local communities are the foundation of the social ties and norms that serve as the informal institutions shaping local residents' beliefs, attitudes, and behaviors. Local communities in the severely bombed areas were devastated, but variation in raid damages generated different postwar institutional contexts across neighborhoods. This, in turn, shapes contemporary socioeconomic outcomes in the longer term by imposing varying costs accompanying specific behaviors (e.g., unemployment and crime) and altering residents' incentives for inculcating children with norms and values (e.g., education and training for higher skills).

2 The Firebombing of Tokyo

This section provides readers with information regarding the exogenous nature of the raid damages resulting from multiple uncontrollable factors, as well as the identified treatment assignment mechanisms that, without our data collection effort, would violate the conditional independence assumption.

2.1 From Hansell to LeMay

The strategic bombing of Japan during WWII targeted 138 cities over the course of several hundred campaigns, but can be classified into two major phases. The first phase began after the fall of Saipan in July 1944. These campaigns, commanded by General Haywood S. Hansell, were conducted at high-altitude (more than 30,000 feet) and targeted military facilities and affiliated industries.

In contrast, the second phase of strategic bombing began in January 1945 and continued until Japan's surrender on August 15, 1945. These campaigns were characterized by indiscriminate bombing of urban residential areas rather than targeted bombing of military and industrial facilities. The effectiveness of carpet bombing of densely populated areas had been recognized by the US from the early days of the war (Fedman and Karacas, 2012). However, it was not until LeMay replaced Hansell that the strategy was officially adopted. LeMay drastically changed the bombing strategy, realizing that it was important to weaken Japanese citizens' support for the war on the home front in order to accelerate Japan's surrender. LeMay therefore changed tactics from high-altitude precision bombing to low-altitude (about 6,000 feet) nighttime bombing of residential areas, despite this change putting pilots at higher risk.[4]

The USAAF also developed incendiary bombs, made of thickened gasoline, to burn Japanese houses, which were mostly made of wood and paper. This Napalm-like incendiary

[4]These contrasting targeting strategies in the two major phases of bombing roughly correspond to the "denial strategy" (first phase) aiming at reducing enemy forces' ability to sustain combat activities, and the "punishment strategy" (second phase), aiming at imposing intolerable costs on civilians (Pape, 1996).

bomb has two important features for our purposes. First, it was not good at destroying concrete structures. Thus, we cannot determine from aerial photos whether the insides of concrete buildings were burned. Second, unlike landmines and nuclear weapons, it is less likely that unexploded shells prevented later land developments or contaminated the surrounding areas, since incendiary bombs used only a small amount of explosives.

2.2 Indiscriminate bombing with two exceptions

The bombing of Tokyo, code-named *Operation Meetinghouse* by the USAAF, was the most extensive bombing with normal weapons in terms of civilian casualties (Saotome, 1981b). In total, 100 thousand people were killed and another 100 thousand left injured, and more than 1 million people were made homeless (Fedman and Karacas, 2012). The major low-altitude indiscriminate bombings to the Tokyo area were performed five times in total; the first, known as the *Shitamachi Bombing*, incurred the greatest damages. The bombing was performed on March 10 to the southeast lowland areas of Tokyo. The fact that it was a nighttime bombing under a new and unfamiliar strategy, combined with strong winds, worsened the raid damages.[5] In this bombing raid alone, the 325 B-29 bombers dropped 1,665 tons of incendiary bombs, and killed 83,793 people (Saotome, 1981a). The four other major low-altitude bombings occurred within a few months.[6]

Although the bombings were indiscriminately implemented, there were *two* factors that USAAF tried to control systematically. First, the crews of the bombers were ordered not to attack the Imperial Palace (Saotome, 1981b; Williams, 2007).[7] This complicates our analysis because the Imperial Palace is located in the very center of the city. That is, the Imperial Palace was not only an object to be protected, but also an object close to the targets. Given this somewhat paradoxical fact, we control for the distance from the palace by adding a spline

[5] The air defense law (*Bōkūhō*) was one of the major reasons for the numerous casualties (Watanabe, 2013). This law required all residents to remain at the scene of a fire as long as possible to extinguish it.

[6] See Table A.1 for the complete list of these five bombings.

[7] Indeed, Claude Robert Eatherly, who gave up bombing his initially assigned target and tried to bomb the Imperial Palace instead (but missed), was "chewed out" for his violation of military regulations (Polmar, 2004).

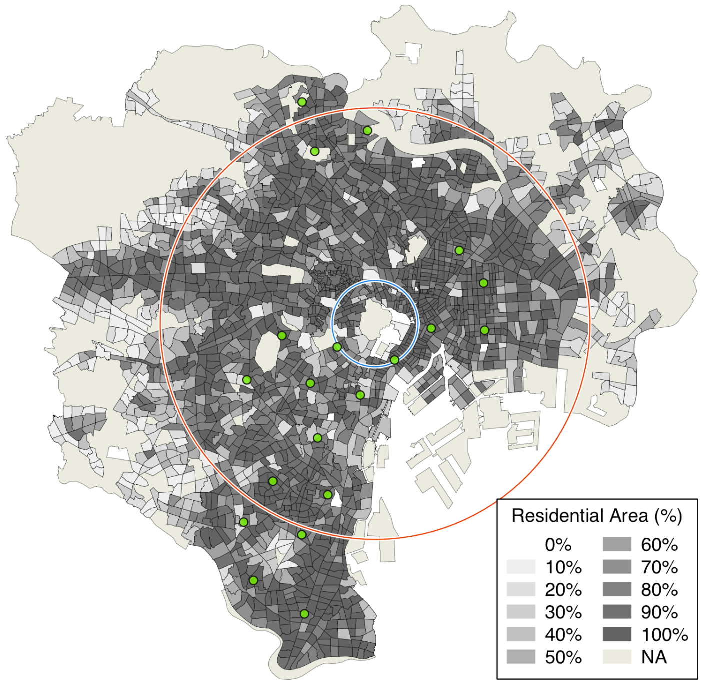


Figure 1: Primary Aiming Points and the Residential Area Ratio in Prewar Tokyo

Note: The inner circle centered at the Imperial Palace represents the size of the typical village-level unit with a 2 km radius, such as hamlets in Vietnam (Kocher, Pepinsky, and Kalyvas, 2011). The dots indicate the aiming points of the five major low-altitude bombings. The gray-shaded polygons represent the estimated ratios of residential development within each neighborhood before the air raids. Coding of the prewar residential area ratio and primary aiming points are described in the main text. The outer circle indicates a 10 km radius, which is used to define the subsample in the robustness check.

term as a control variable. Second, each bombing was accompanied by a bombing plan, with several aiming points (APs) for dropping the bombs. These APs were usually determined to destroy specific military targets in the first phase of bombings, but the APs in the low-altitude bombings of the second phase were designed to facilitate efficient destruction of the city. Information on the APs was not archived collectively in a single document; some APs were documented in one of the USAAF tactical mission reports, while others were printed on a stone board called *litho-mosaics* (USAAF, 10 March 1945; Okuzumi, 2007). Importantly, all of the APs for the major low-altitude bombings have been discovered by Okuzumi (2007). We geo-reference these APs and show them as green dots in Figure 1. The figure indicates that the APs were set all over the city, mainly at dense residential areas at equal distance. Thus, we also control for each neighborhood's distance to the nearest AP as a spline term.

Statistically speaking, these two factors are major potential confounders as they are key components of the treatment assignment mechanisms. The vast majority of past studies has either ignored treatment assignment mechanisms or used them as proxies rather than confounders (e.g., Miguel and Roland, 2011). In contrast, our study uses them as control variables measuring the actual damages separately from above. This research design contributes to the plausibility of the conditional independence assumptions.

2.3 Exogenous variation

As we have seen so far, the USAAF made the bombing plans and gave military orders to control raid damages. However, actual damages spread beyond expectations due to various natural factors. For example, winds of 40 knots (equivalently 46 mph or 20 m/s) on the night of March 10 immediately reduced visibility of the bomber crews to zero after the vanguard planes dropped the first bombs on the aiming-point targets (USAAF, 10 March 1945). This, combined with the radio interference by the Imperial Japanese Army, and the fact that it was *the very first risky low-altitude bombing*, made it highly difficult for the crews to follow the predetermined plans (Saotome, 1981b).[8]

Natural and artificial constructions, as well as geographical features, added other types of variation to the raid damages. The back side of a burning hill, for example, might have escaped from the fires. Moreover, canals, rivers, parks, large school buildings, roads, and rail tracks sometimes worked as fire lanes.

Finally, the disaggregation of the unit of analysis enables us to exploit the exogenous variation of the raid damage for our purposes. Our unit of analysis is the neighborhood within Tokyo's 23 large wards. The average size of a neighborhood is only 0.2 km^2. In comparison, the Vietnamese hamlets used in Kocher, Pepinsky, and Kalyvas (2011) are defined as equivalent to the 2 km radius circle ($\approx$ 12.57 km^2) indicated with the blue circle in Figure 1. If the raid damages were aggregated, we would lose much of the natural variation

[8] As another example of the uncontrolled nature of the air raids, the main part of the Imperial Palace, called *Kyūjō*, was also accidentally damaged by the bombing on May 25, 1945, despite the prohibition.

just mentioned.

3 Data

We rely on two primary data sources for our empirical analysis: (1) historical aerial imagery, topographic maps, and archival documents, and (2) georeferenced data on contemporary census measures, crime rate statistics, and the micro-geography of Tokyo. The following sections describe the data development and coding procedures.

3.1 Aerial Imagery, Topographic Maps, and Historical Archives

Aerial Imagery: The most noteworthy aspect of our data development efforts is the use of historical aerial imagery. Specifically, we employ a series of aerial photographs taken by the USAAF and USAF during the 1946–1948 period. These were developed as part of the US forces' occupation policy to evaluate the effectiveness and impact of the strategic bombing. We obtained 666 aerial images covering the residential areas of Tokyo's 23 wards, digitized and publicly distributed by the Geospatial Information Authority (GSI) of Japan.[9]

Figure 2 presents a sample of images. Remote-sensing imagery such as aerial photography needs to be, first, correctly georeferenced, and second, converted into meaningful information to facilitate empirical analysis. As the original digitized images are provided as scanned photos, we first originally georeferenced the imagery, relying on predetermined geographic control points (GCPs) that are time invariant. Examples of the GCPs include major roads, coast lines, and the Imperial Palace.[10] The resultant images are georeferenced with the cubic convolution method and assigned with the world geodetic system (WGS84) to facilitate the

[9] URL: https://mapps.gsi.go.jp/maplibSearch.do. Accessed March 21, 2019. See Appendix B for the selection process of the aerial photography.

[10] A reasonable question is whether the georeferencing of aerial photos of more than 70 years ago is feasible given that the geocoded map shows current-day Tokyo. Fortunately, this is feasible in most cases because Tokyo's road network was already developed at the time of WWII. For the same reason, the georeferencing was particularly hard and time-consuming where most of the area was farmland and where a new river or a new canal was to be built. As we discuss later, our primary results remain unchanged when we use the city center as a subsample.

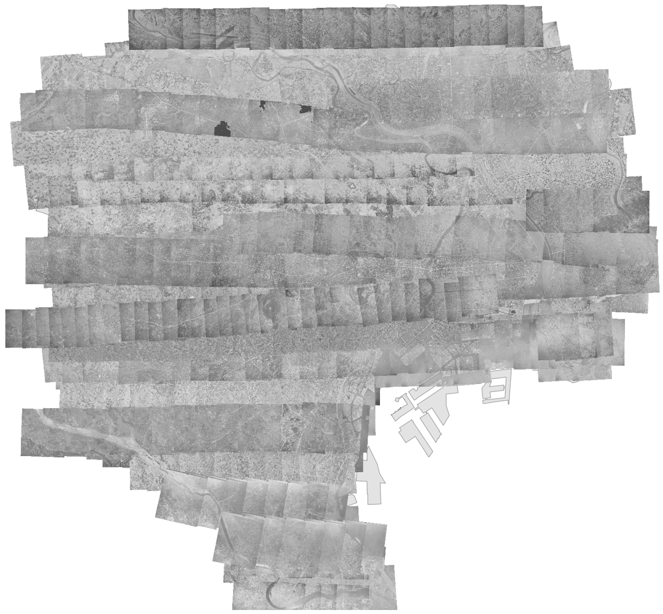
Figure 2: Sample Aerial Imagery of Tokyo, 1946–1948

Note: The neighborhoods in lighter colors broadly indicate heavily damaged areas, whereas the neighborhoods in darker colors represent the areas remaining half-burned or intact.

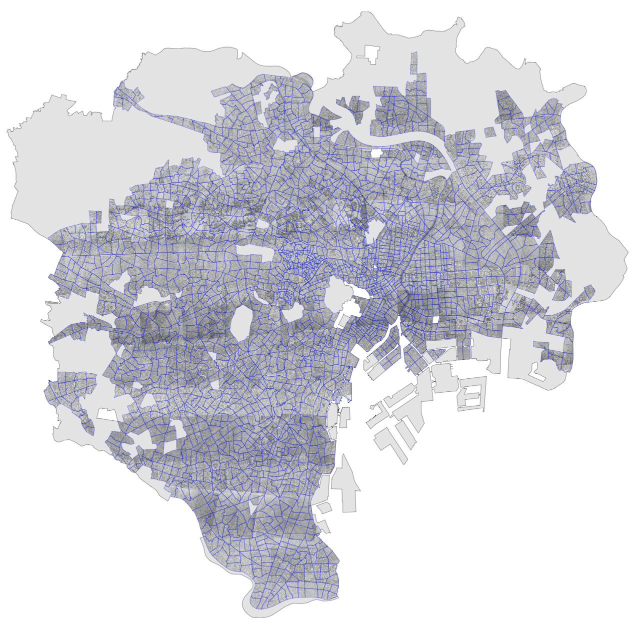
subsequent coding of raid damages. We then coded the neighborhood-level raid damages with these georeferenced images by a human-coding scheme which we discuss later. Panel (a) of Figure 3 plots the georeferenced aerial imagery overlaid onto the whole study region of Tokyo's 23 wards.

Military Topographic Maps and Archival Documents: To capture the prewar urban landscape of Tokyo and construct the sample neighborhoods, we employ the "Tokyo City Plans 1:12,500" maps, a series of military topographic maps developed by the US Army Map Service (AMS). These thematic maps were originally based on the city plan information compiled by the Imperial Japanese Army, and digitized and provided by the University of Texas Libraries.[11] As Figure A.1 shows, these maps effectively cover the urban area of Tokyo and mainly classify areas of high and low density construction. For the postwar imagery, we first georeferenced these maps using a series of the predetermined GCPs. We then overlaid them onto the contemporary census polygons to extract the neighborhood-level information on the prewar building-density proportions. As explained in detail below, we exclude neighborhoods that did not include any residential areas in the prewar period from

[11] URL: http://legacy.lib.utexas.edu/maps/ams/japan_city_plans/index_tokyo.html. Assessed March 21, 2019.



(a) Georeferenced Aerial Imagery



(b) Cropped Imagery and Sample Neighborhoods

Figure 3: Georeferenced Imagery of Tokyo

Note: (a) Georeferenced aerial imagery covering Tokyo's 23 wards. (b) Georeferenced imagery cropped by the sample neighborhoods. Blue segments represent the neighborhood borders in the 2010 Census. Our sample in the following analysis contains the neighborhoods in Panel (b).

our sample, as such areas were rarely targeted during the strategic bombing.

As described in the previous section, we also compiled the precise target locations of the strategic bombing campaigns from the declassified historical archive and secondary sources. Specifically, we rely on the "Records of the U.S. Strategic Bombing Survey: Security-Classified Tactical Mission Reports" developed by the USAAF Bomber Commands in 1945 and digitized and provided by the National Diet Library Digital Collections of Japan and *litho-mosaics* (USAAF, 10 March 1945; Okuzumi, 2007). The green dots in Figure 1 in the previous section represent the geographical distribution of the primary aiming points.

3.2 Georeferenced Data

Census Statistics: We also use georeferenced census data and a series of geocoded data sets provided by the GSI, the Statistics Bureau of the Ministry of Internal Affairs and Communications (MIC), and the National Spatial Planning and Regional Policy Bureau of the Ministry of Land, Infrastructure, Transport and Tourism (MLIT) to measure the neighborhood-level, contemporary demographic and socioeconomic characteristics and geographic conditions. The MIC conducts a simplified population census every five years and a full-scale census every ten years, and provides precisely georeferenced census statistics since the 1995 census. We employ the *Kokusei Chōsa Shōchiiki Shūkei Kekka* (the small-area aggregated census results), or the census results at the neighborhood level. The neighborhood-level census statistics provide disaggregated population composition including age and sex structure, education history, occupation, residence year and more.

As geographic borders of the neighborhood-level census units slightly vary across different census years, we employ the census units in the latest, 2010 full-scale census as our unit of analysis. The 2010 census polygons contain roughly 3,200 neighborhoods in Tokyo's 23 wards. Out of 3,200 neighborhoods, we restrict the sample neighborhoods for the following analysis by two criteria. First, we discard neighborhoods that had minimal residential construction during the WWII period. This preprocessing excludes neighborhoods that had little chance of being targeted for the air raids.[12] Second, we drop the census units with contemporary population counts smaller than 100 from our data set. We set the second criterion due to the limited availability of disaggregated statistics. The MIC does not disclose detailed census statistics for these sparsely-populated neighborhoods to maintain the anonymity of the census respondents.

This sample selection process leaves 2,286 unique neighborhoods in our data set. Panel (b) of Figure 3 shows the subset of neighborhoods included in our sample. We merge these 2,286 neighborhoods with the 1995, 2000, 2010, and 2015 census statistics and measure contemporary, local-level demographic and socioeconomic variables to facilitate the empirical analysis.

Crime Data: The neighborhood-level crime statistics are derived from the official statistics of the Tokyo Metropolitan Government, publicly available as a part of the Tokyo Government open-data catalogue.[13] A notable feature of the crime data is its disaggregated nature. The data set contains neighborhood-level geocoding that is readily compatible with the census unit for the 2004-2018 period, as well as monthly and annual records of crimes disaggregated into several major categories listed by the National Police Agency.

Micro-geography of Tokyo: Our coding of the micro-level variations of geographic conditions relies on the MLIT National Land Numerical Information.[14] We measure elevation, ruggedness, railway length, and number of stations as of 1945 within individual census units using the MILT's spatial data sets. Note that we extract the neighborhood-level information on physical geography (e.g., elevation) from the contemporary data sets assuming that these features remain almost unchanged during the study period.[15] To our knowledge, disaggregated physical geographic information of prewar Tokyo is not available. In contrast, we

[12] This preprocessing distinguishes our study from prior studies that treat whole regions or geographic units as a scope of analysis regardless of whether all subsets actually had a positive chance to be treated.

[13] Specifically, we scraped crime-rate records provided by the Tokyo Crime Prevention Network. Available at: http://www.bouhan.metro.tokyo.jp, as of March 14, 2019.

[14] Available at: http://nlftp.mlit.go.jp/ksj-e/index.html. Accessed March 21, 2019.

[15] One might wonder if the bombing itself affected physical geography. Yet this assumption was unlikely to occur in the firebombing of Tokyo, in which the perpetrator primarily used "blast-free" incendiary bombs.

use historical spatial data to code artificial geographical features of the study area. These include both buildings (e.g., train stations) and natural features (e.g., rivers) that existed in the WWII period.

4 Variables

To capture the multifaceted nature of the raid damages, we use various socioeconomic indicators as our outcome variables. First, we use the logged number of reported crimes. Crime rates are thought to reflect the urban landscape as well as neighborly ties, both of which should be strongly correlated with raid damages. Using the official statistics of the Tokyo Metropolitan Government, the neighborhood-level crime counts are categorized into five mutually exclusive categories: "violent offenses," "non-burglary thefts," "burglary thefts," "felonious offenses," and "other crimes." We use the logged numbers of these five count variables as our outcome variables. For the regression models of these crime outcomes, we add the logged contemporary population count as an additional independent variable in Equation (2) specified below.[16]

Second, we use the logged (overall, male, and female) unemployment rates, logged ratio of professional and executive workers, logged average residency (in years), and logged years of education. These variables are intended to capture how the raid damages affected the later economic development or accumulation of human capital in the neighborhoods. With a few exceptions, these variables are recorded every 5 years.[17]

Finally, we use the logged male-to-female ratio of children younger than 5 years old (i.e., preschool age in Japan) as the outcome variable for our primary falsification test. It is hard

[16] We add 1 to the raw counts and then log-transform the crime count variables. Note that the crime count variables are coded based on the locations where the crimes *occurred*. The crime count variables do not reflect information on the neighborhood where victims or perpetrators reside or the locations of police stations to which crime claims were submitted. For example, if someone residing in neighborhood A files a claim to a police station located in neighborhood B that her bicycle was stolen in neighborhood C, only the corresponding crime count (non-burglary theft) in neighborhood C increases by 1.

[17] The logged years of education are available only in 2000 and 2010, and the logged average residency is available only in 2000, 2010, and 2015. See Appendix C for the detailed definition of these outcome variables.

to imagine that the raid damages would have any impact on the contemporary gender ratio of preschool children, and this likely non-finding is what this *negative control* type approach intends to validate (Arnold and Ercumen, 2016). A finding of no association between raid damage and preschool child sex ratio provides additional credibility that our primary results are unlikely to be attributed to common unmeasured confounders, measurement errors, or self-selection.

Our key explanatory variable is measured as the fraction of destroyed areas relative to overall residential areas within each neighborhood (henceforth *damage ratio*), which is conceptually defined as follows:

$$\text{Damage Ratio} = \frac{\text{Destroyed residential area}}{\text{Overall residential area}}. \quad (1)$$

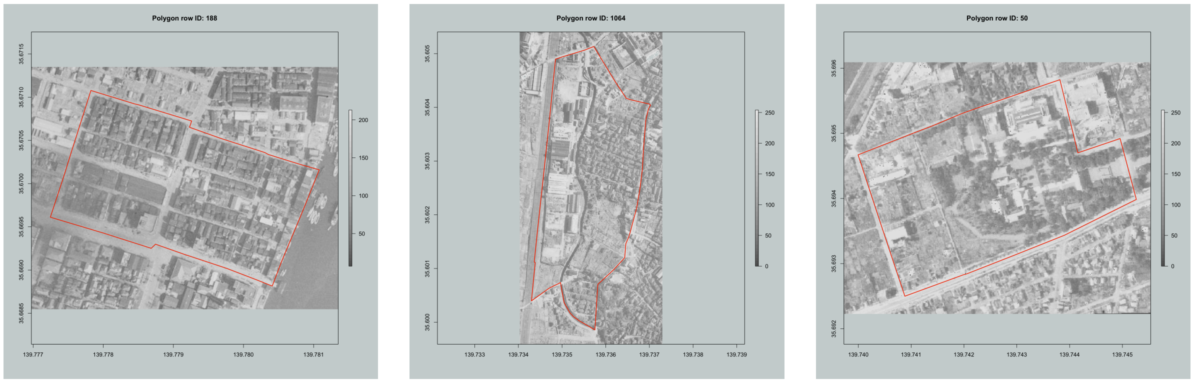
Note that the denominator only includes the residential area. Other areas such as parks, water areas (rivers, ponds, etc.), fields, schools and commercial buildings are not included.

The raid damages are evaluated by human-coding. Specifically, one of the authors (Harada) evaluated the ratio of both the destroyed and undamaged residential areas to the land area in a polygon (henceforth *the residential area ratio*) first, and then evaluated the damage ratio. This process was repeated for each of 2,286 polygons. We chose this strategy because it provides the most accurate measurements compared to other alternatives.[18] These variables were evaluated with the values ranging from 0% to 100% in increments of 10%, which produced eleven-unit scale variables that we treat as continuous variables for most of the analysis.[19]

To show how damage ratio values assigned to each neighborhood differ qualitatively, we present aerial photos of three neighborhoods in Figure 4. Each of these photos was assigned a different value of the residential area ratio and the damage ratio. For example, the middle panel shows that the left part of the neighborhood is not a residential area and is slightly

[18] See Appendix D for details of the coding procedure.

[19] As a robustness check, we also perform the analysis with the dichotomized raid variable and with the model including each value of the raid damages as a separate dummy variable.



(a) Residential Area Ratio = 1, Damage Ratio = 0

(b) Residential Area Ratio = 0.6, Damage Ratio = 0.5

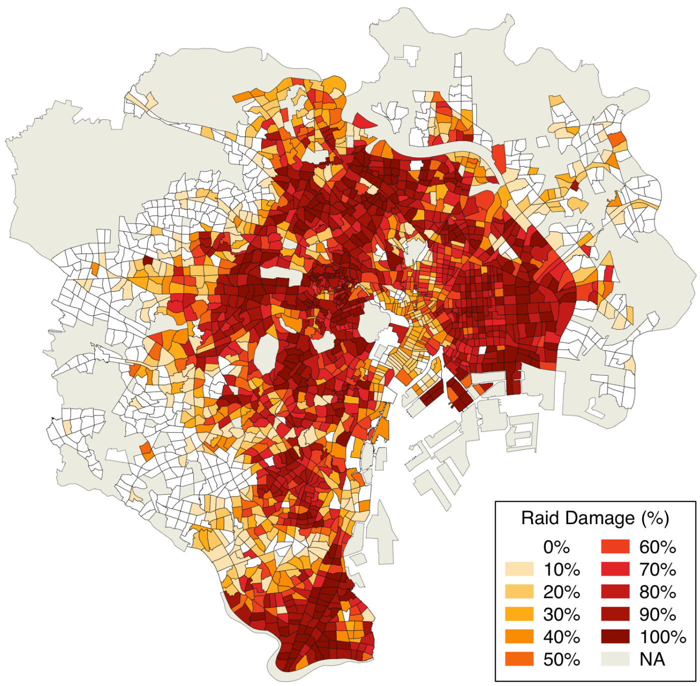
(c) Residential Area Ratio = 0.3, Damage Ratio = 1

Figure 4: Examples of the evaluation of residential area ratio and damage ratio.
Note: The red solid segments in individual panels represent the neighborhood boundaries overlaid onto aerial images.

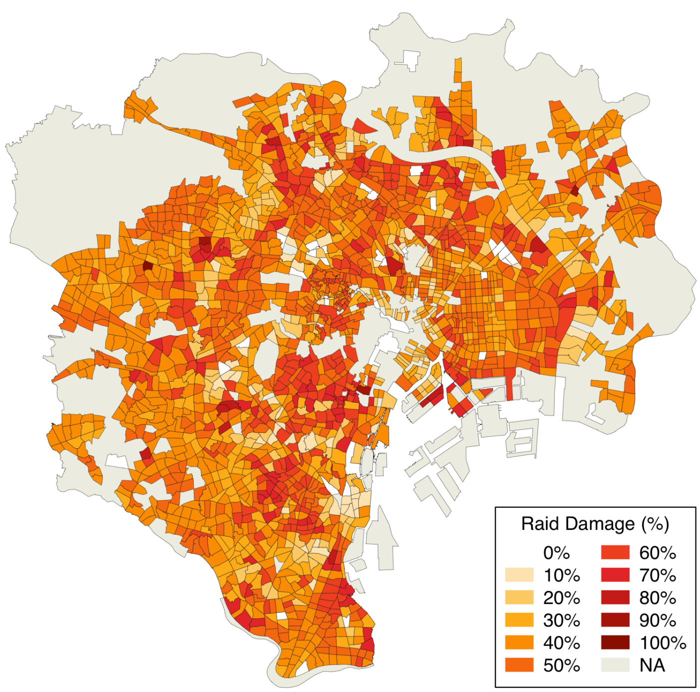
smaller than the residential area to the right, so its residential area ratio is evaluated as 0.6. Within the residential area, about half of the area looks destroyed by fire, so its damage ratio is evaluated as 0.5.

Panel (a) of Figure 5 shows the result of the raw evaluation of the damage ratio. The darker areas indicate a higher value of the damage ratio. It shows that most of the damages are concentrated on the southern area near the border separating Tokyo and Kanagawa prefectures, and the areas surrounding the Imperial Palace (large blank space in the center). A comparison to Figure 1 reveals that damages are also concentrated in the neighborhoods with high residential area ratios. Especially, the damages east of the Imperial Palace—the downtown (*Shitamachi*) area—were severe, which corresponds to the fact that the first low-altitude bombing on Tokyo is also known as the Bombing of Downtown. The map also shows some seemingly miscoded polygons (e.g., moderately bombed areas away from the city center). Robustness checks presented later will show that the substantive findings remain unchanged. Also, the residualized map in Figure 5(b) shows the seemingly exogenous variations of the damages.

The following pretreatment covariates and time-constant geographic characteristics are



(a) Raw Human Coding



(b) Residualized Human Coding

Figure 5: Distribution of the Air Raid Damages across Tokyo's 23 Wards

Note: Shading of the neighborhoods indicates the damage ratio in percentage scale. The neighborhoods excluded from our sample are left blank. See Section 5 for the control variables used in residualization.

selected as control variables: the geo-coordinates (as smoothed terms of centroid longitude and latitude), the human-coded residential ratio, the logged area, the number of neighboring polygons (neighborhoods), the logged mean elevation, the logged mean slope (degrees), the logged distance from the closest rivers, the logged length of the railway, the logged length of the railway of neighboring polygons, the number of train stations, the number of train stations in neighboring polygons, the logged distance from the Imperial Palace, and the logged distance from the nearest aiming point.[20] The last two control variables are particularly important as they are all of the known treatment assignment mechanisms, so we include them as smoothed terms. Finally, fixed effects for Tokyo's 23 wards control for all observed and unmeasured heterogeneities across wards that are constant during the 70 years or so between when the air raids took place and when the outcome variables were recorded. We do not control for all other variables observed during this period because they can be thought of as consequences of the air raids, and thus their inclusion would invite post-treatment bias into our analysis.

5 Estimation

In our estimation models, we regress one of the outcome variables in a given year on the damage ratio, conditional on pretreatment covariates and time-constant geographic characteristics as well as ward fixed effects. We estimate the following regression models with smoothing splines, and these estimates are similar to each other. We fit a semiparametric model with smoothing splines using a Generalized Additive Model (GAM, Wood, 2006) with the contemporary neighborhood population as weights.[21]

[20] We added 1 to the original values before log-transformation.

[21] A GAM with a two-dimensional spline term of longitude and latitude is often called a geo-additive model in the spatial statistics literature (Kammann and Wand, 2003).

Our estimation models are formally written as:

$$\mathbb{E}(y_{js}) = \tau Damage_{js} + \boldsymbol{x}_{js}\boldsymbol{\beta} + f_1(IP_{js}) + f_2(AP_{js}) + f_3(\text{Lon}_{js}, \text{Lat}_{js}) + \alpha_s, \qquad (2)$$

where y_{js} is the outcome variable, the subscripts j and s are neighborhoods (*Cho-chomoku*) and Tokyo's 23 wards, $Damage_{js}$ and τ are the eleven-point scale damage ratio and its coefficient, $\boldsymbol{x}_{js}$ is a $1 \times K$ vector of the geographic covariates, $\boldsymbol{\beta}$ is a $K \times 1$ vector of their coefficients, f_1, f_2, and f_3 are spatial tensor-product spline terms, IP is the logged distance from the Imperial Palace, AP is the logged distance from the closest aiming point, and α_s denotes ward fixed effects. With these extensive control variables including the two key treatment assignment mechanisms, coupled with the sample restriction based on the prewar habitation status, if the conditional independence assumption holds, our quantity of interest, τ, represents the average treatment effects (ATE) as well as the average treatment effects on the treated (ATT).[22]

Figure 6 shows the results of balance checks. Each pretreatment covariate on the vertical axis is plotted against the raid damages.[23] The figure shows that the regression lines of almost all panels are close to flat, indicating that the treatment assignments are well balanced. However, the regression line of the residential area ratio (top-left plot) has a noticeable positive slope, indicating that the damages were concentrated on dense residential areas. To examine whether this imbalance introduces any bias to our main findings, we later perform a subsample analysis as a robustness check.

We also performed an additional balance check using the fire hazard ratio estimates based on the flammability rating of central Tokyo in 1942. The flammability was rated by a fire insurance company based on the types of building materials and building densities. This

[22]This is because the sample consists of the neighborhoods that had a conditionally equal chance of being bombed. If one deems some subsample, say the neighborhoods within 10-km radius, more plausible as the target of bombing, the subsample estimates would be considered as the ATTs.

[23]Specifically, the damage ratio on the horizontal axis is residualized by the control variables, and the control variable on the vertical axis is residualized by all other control variables. In order to residualize the variables by the same set of right-hand-side variables, we used the polynomial specification rather than the smoothed-term specification. The variables on the plots are all standardized for comparison.

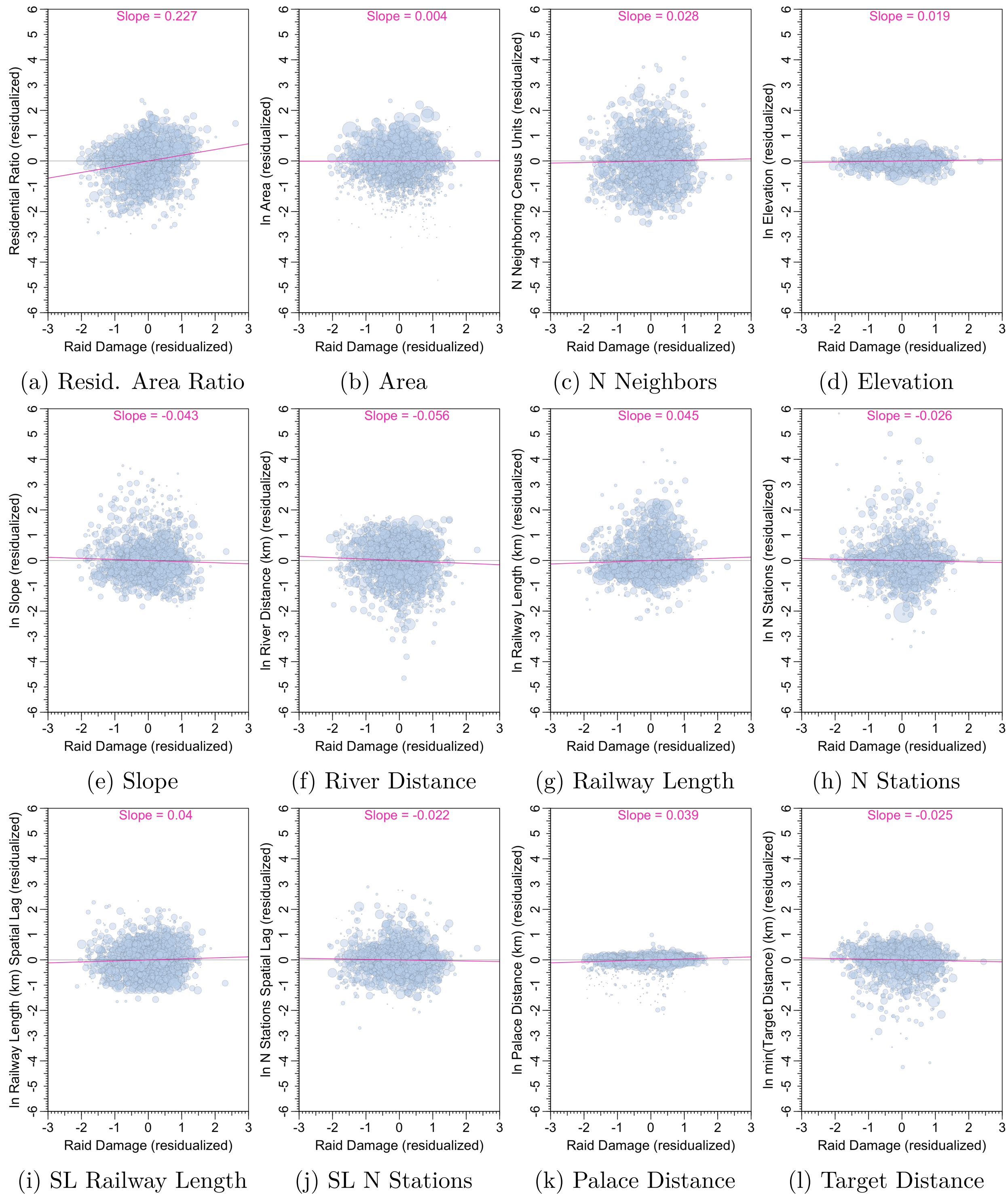
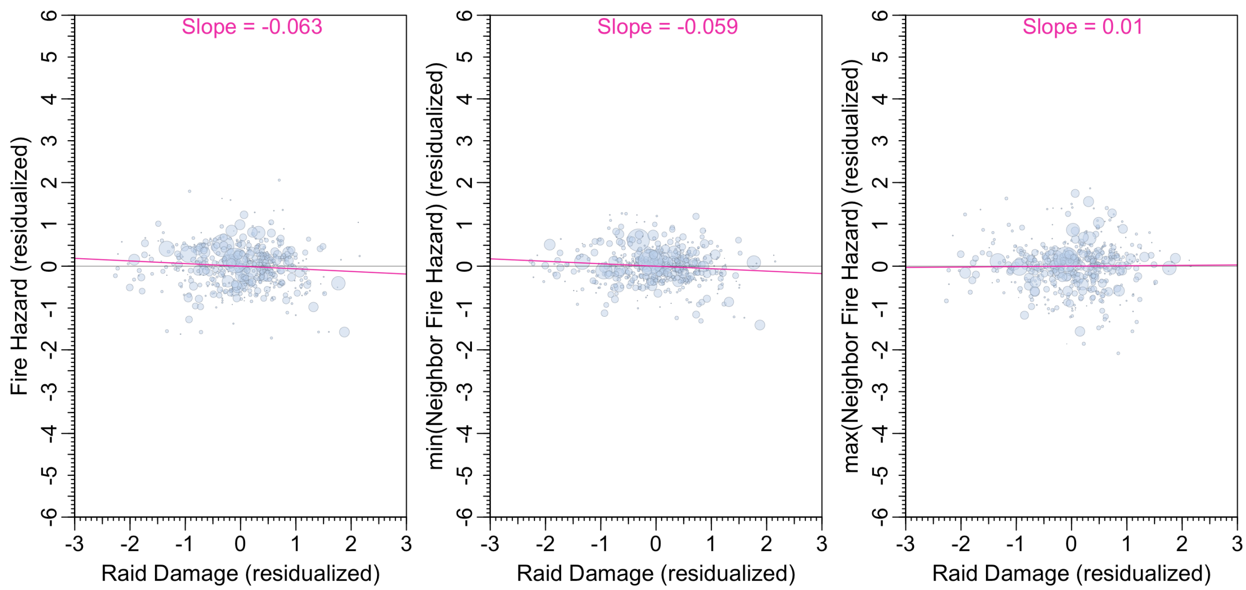


Figure 6: Balance Plots for the Pretreatment Covariates

Note: SL = (first-order) spatial lag. Each panel plots the regression of one of the pretreatment covariates on the damage ratio variable. The damage ratio variable on the horizontal axis is residualized by all of the right-hand-side variables excluding itself, and the covariate on the vertical axis is residualized by all of the right-hand-side variables excluding itself and the damage ratio. The size of each dot is proportional to the current population count in thousands (in the 2010 census) of each neighborhood. Solid lines and texts indicate linear regression fits.



(a) Hazard Ratio (HR) (b) Min. HR, neighboring (c) Max. HR, neighboring

Figure 7: Balance Plots for the Pretreatment Flammability Measurements

Note: See note in Figure 6.

intelligence was one of the many information materials the USAAF collected in planning the firebombing of Tokyo.[24] Because low income households at that time often lived in wooden shacks in densely built-up areas where fires could easily spread, the differences in the raid damages may simply reflect differences in the socioeconomic status of the households. However, the balance check in Figure 7 shows that the damage ratio neither correlates with the fire hazard ratio nor with the minimum and maximum hazard ratios of neighborhoods.[25]

6 Results

6.1 Baseline Results for Crime Statistics

First, we investigate how the raid damages affect contemporary crime patterns. Because the crime statistics are counted based on where the crimes occurred, and perpetrators do

[24]See Appendix E for details of the coding process of the flammability data.

[25]Due to the limited geographical coverage of the original hazard rate estimates and missing values in other pretreatment variables, the sample size includes only a subset of our neighborhood polygons, with the number of observations remaining 527.

Table 1: GAM Estimates of the Raid Effects on Contemporary Crime Rates, 2010

	Outcome variable:				
	(1) Non-burglary	(2) Felonious Offenses	(3) Burglary	(4) Violent Offenses	(5) Others
Damage	0.1212**	0.0586**	−0.1912***	0.0523	−0.0049
	(0.0582)	(0.0291)	(0.0574)	(0.0625)	(0.0513)
Pretreatment Covariates	✓	✓	✓	✓	✓
Spatial Spline	✓	✓	✓	✓	✓
Ward FE	✓	✓	✓	✓	✓
Observations	2,153	2,153	2,153	2,153	2,153
Adjusted R^2	0.4216	0.0699	0.2830	0.2380	0.4091
Log Likelihood	−2,224.2570	−864.1996	−2,185.1470	−2,386.5260	−1,937.1240
UBRE	1,175.7570	332.2310	1,133.8320	1,366.9370	900.5307

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. All outcome variables are logged.
Standard errors in parentheses. Constants not reported for brevity.
UBRE: Unbiased Risk Estimator.

not necessarily commit crimes in the neighborhoods where they reside, a reasonable way to interpret the raid effects is whether the bombing made the neighborhoods more or less likely to be targeted by criminal activities, which closely relates to the community's capacity of monitoring—how much the residents are acquainted with each other and how much the residents know who lives where or which property belongs to whom (Becker, 1968). Or, neighborhoods might introduce security cameras in high-crime areas if they are able to overcome the classical problem of public goods provision, and community members bear accompanying private costs (Olson, 1965).[26] In either case, this analysis illuminates the relationship between the bombing damage and communities' ability to perform collective efforts in the present day.[27]

Table 1 reports the baseline GAM estimates for the 2010 crime statistics using the model specification of Equation (2), and Figure 8 replicates and summarizes the treatment effect

[26]Indeed, Tokyo Metropolitan Government subsidizes neighborhood associations that introduce security cameras for crime prevention (National Police Agency, 2019).

[27]Some existing work in criminology directly connects crime (especially juvenile crime) to deterioration in local institutions that facilitate community-level behavioral norms, including neighborhood newspapers, festivals, clubs, and other organizations (Becker, 1988).

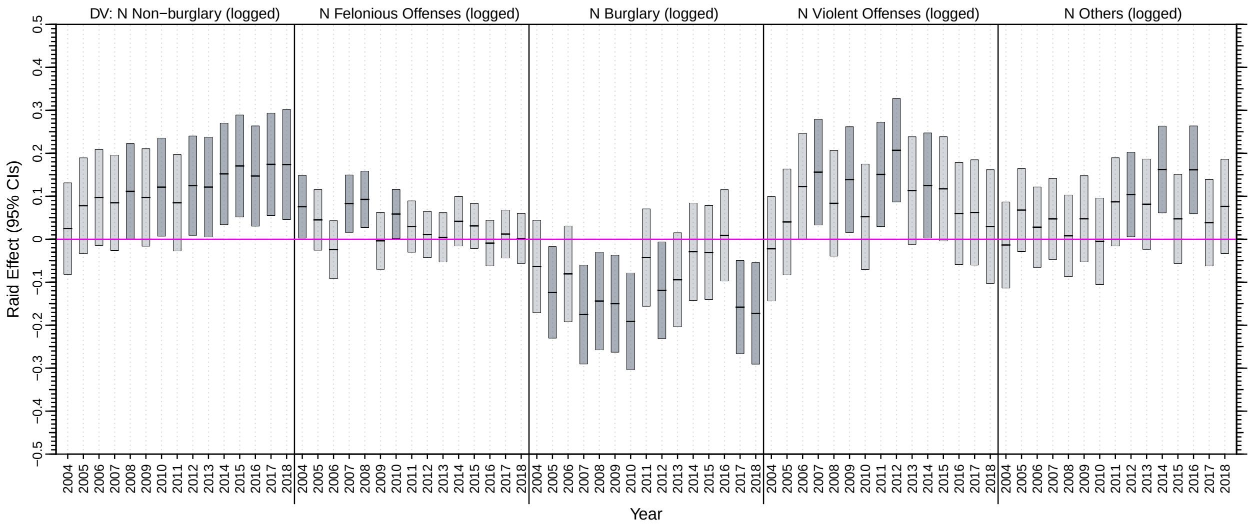


Figure 8: Geo-additive Model Estimates of Damage Ratios on Logged Crime Numbers
Note: The labels on the top indicate the five crime categories (separated by vertical segments). Black stripes indicate the point estimates, and gray vertical bars represent the corresponding 95% CIs.

estimates across the annual subsamples ranging from 2004 to 2018. The labels in the top row of Figure 8 represent the outcome variables.[28] The stripes in the center and the bars in Figure 8 represent the estimated treatment effects of the raid damages and the corresponding 95% confidence intervals (CIs) from 2004 to 2018 (horizontal axis). The figure shows that the damage ratio has a negative association with burglaries, but is positively associated with all other categories, especially non-burglary crimes. The coefficients for the crime categories other than non-burglary thefts fluctuate because the number of crimes is small. The results indicate that, with the exception of burglary, the air raids made the affected neighborhoods more susceptible to crime.

The baseline coefficient estimates for the 2010 crime statistics in Table 1 suggest that the substantive impacts, as well as coefficient signs, vary across crime categories. The coefficient estimates indicate that a 10 percentage-point (p.p.) increase in the raid damage yields 1.2 p.p. (equivalent to 0.39 count) increase in non-burglary thefts and 1.9 p.p. (equivalent to 0.045 count) decrease in burglary thefts, while the effects on the other two crime categories

[28]See Appendix F for the information on the breakdown of each crime category.

remain statistically insignificant.[29]

One may reasonably wonder why the coefficients for the logged number of burglaries show negative and statistically significant effects in most years. On one hand, it is unlikely that burglaries would be the only type of crime that is not influenced by the raid damages; on the other hand, burglaries are the only type of crime for which the location of the victims' residence nearly always corresponds with the location of the crime. We suggest at least two probable explanations for why burglaries are more likely in the neighborhoods that were less damaged by the air raids. First, the land in these undamaged neighborhoods is often divided into many irregularly shaped small lots owned by different landlords, which discourages larger redevelopment projects due to high transaction costs. As a result, such areas are typically overcrowded, with old wooden housing and insufficient home security systems. Second, community members develop closer relationships in undamaged neighborhoods, which may promote sociable and trusting living styles that are not necessarily desirable from a crime prevention perspective. Indeed, popular accounts suggest that it is a common custom for residents to leave their doors unlocked so that neighbors can easily drop by to visit.[30]

6.2 Baseline Results for Socioeconomic Outcomes

Next, we analyze how the bombings affect contemporary socioeconomic outcomes at the neighborhood level. The positive relationships between damages and crime rates imply that socially undesirable behavior might be suppressed in undamaged neighborhoods, where intimate social ties persist. If the WWII air raids undermined neighborhood-level social and economic performance, the damages would be positively associated with unemployment rates, while negatively associated with educational attainment and the ratio of executive and professional workers. In contrast, we expect damages to have no discernible impact on the

[29] The calculation of crime counts is based on the sample means of each crime category: 31.7 for non-burglary thefts and 2.35 for burglary thefts. For example, the count of non-burglary thefts is calculated as follows: $0.121 \times 31.7 \times 0.1 \sim 0.39$. The analysis of felonious offences fails to retain substantial significance. Given the comparatively small mean instances of 0.226 (in the 2010 data set) and the coefficient of 0.0586, a 10 p.p. increase in the raid damage is associated with only a 0.0013 increase in felonious offences.

[30] A mediation analysis we report later suggests support for the latter story.

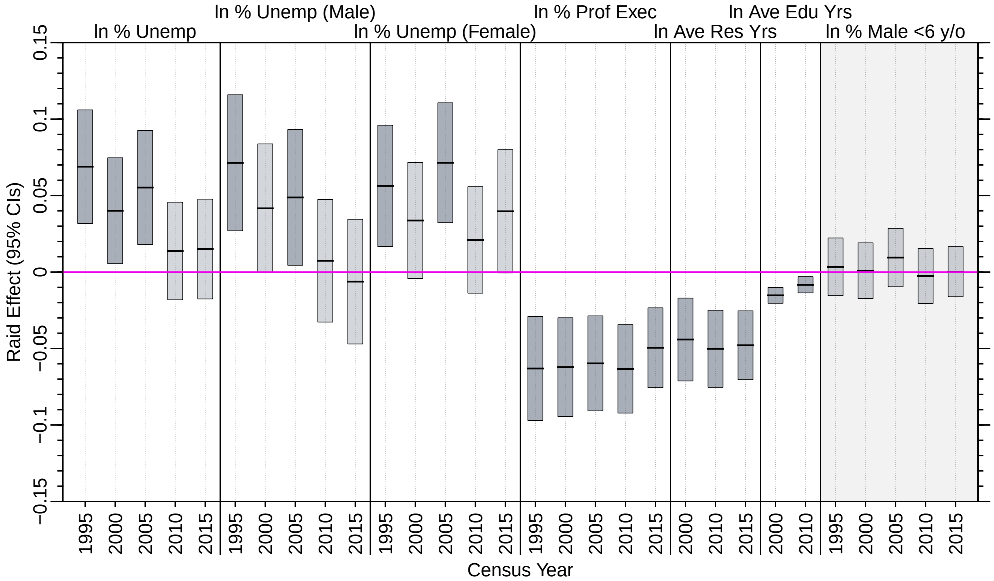


Figure 9: Geo-additive Model Estimates of the Regressions of Damage Ratios on Logged Socioeconomic Outcomes

Note: Socioeconomic variables are logged unemployment rates, logged ratio of executives and high-skilled workers, logged average years spent living in their current residence, logged educational attainment and logged male sex ratio for children under 6 years old. Black stripes indicate the point estimates, and vertical bars represent the corresponding 95% CIs.

male-female ratio for preschoolers as it hardly reflects the social or economic performance of the residents and neighborhoods.

Figure 9 reports the estimated treatment effects of the damage ratio on the various socioeconomic variables. The results mostly correspond with expectations. The figure shows that the raid damages positively affected the logged unemployment rates in several years, but had negative impacts on the logged ratio of executive and professional workers, the logged average years spent living in their current residence and the logged average years of education. The logged male sex ratio for children under 6 years old has a coefficient near zero as expected. Overall, the raid damages generally reduce the share of residents with a high socioeconomic status.

Based on the baseline results with the 2000 census reported in Table 2, a 10 p.p. increase in damages due to the air raids raises the overall unemployment rates by 0.4% which is

Table 2: GAM Estimates of the Raid Effects on Contemporary Socioeconomic Outcomes, 2000

	Outcome variable:						
	% Unemp (All)	% Unemp (Male)	% Unemp (Female)	% Prof & Exec	Ave Res (Yrs)	Ave Edu (Yrs)	% Male < 6 y/o
Damage	0.0400**	0.0416*	0.0332*	−0.0624***	−0.0442***	−0.0152***	0.0009
	(0.0176)	(0.0215)	(0.0194)	(0.0165)	(0.0138)	(0.0026)	(0.0093)
Pretreatment Covariates	✓	✓	✓	✓	✓	✓	✓
Spatial Spline	✓	✓	✓	✓	✓	✓	✓
Ward FE	✓	✓	✓	✓	✓	✓	✓
Observations	2,158	2,158	2,158	2,158	2,158	2,158	2,158
Adjusted R^2	0.3974	0.3958	0.2312	0.7057	0.3333	0.7873	0.0026
Log Likelihood	350.3820	-73.9729	147.3735	485.9518	868.8769	4,447.2630	1,503.5600
UBRE	95.4246	141.4024	115.1605	84.1430	59.0033	2.1411	32.7446

Note: *p<0.1; ***p* <0.05; ****p* <0.01. All outcome variables are logged.
Standard errors in parentheses. Constants not reported for brevity. UBRE: Unbiased Risk Estimator.

roughly equivalent to 0.018 p.p.[31] Similarly, a 10 p.p. increase in raid damages reduces the ratio of executive and professional workers by 0.62% (equivalent to 0.133 p.p), the average residency in years by 0.44% (equivalent to 25 days) and the average years of education by 0.15% (equivalent to 7.5 days).[32] These effects are not substantively large, yet surprising given their persistence.

6.3 Robustness Checks

We performed several robustness checks. First, we estimated our main model replacing the smoothed terms with the fifth-ordered polynomials (Figures A.2(a) and (b)).[33] Second, we dichotomized the damage ratio variable at its mean (≈ 0.482) and used it as the treatment variable (Figures A.3(a) and (b)). Third, because the residential area ratio is somewhat unbalanced in the balance check, we estimated our regression limiting the sample to neighborhoods for which the prewar residential area ratio is 70% or higher—since these were the primary target of the indiscriminate bombings (Figures A.4(a) and (b)). Fourth, we estimated our regression model excluding the neighborhoods located beyond a 10 km ($\approx$ 6.2

[31] The figure is based on the sample mean of the unemployment rate in 2000, which is 4.34 p.p.

[32] The sample means of the ratio of the executive and professional workers, the average years of residency and the average years of education in 2000 are 20.1 p.p., 13.4 years and 14.9 years respectively.

[33] The order of the polynomials was determined such that the degrees of freedom of the polynomials are approximately the same as the estimated degrees of freedom in GAM.

miles) radius from the centroid of the Imperial Palace (Figures A.5(a) and (b)). As indicated by a red circle in Figure 1, a 10-km radius roughly covers the most densely populated areas. For all of these specifications, the treatment effect estimates show substantively the same patterns.

Finally, we estimated the main model replacing the original continuous *damage ratio* variable with 10 dummy variables each representing a distinct level of damage ratings ranging from 10% to 100% (using 0% as the baseline). The results in Figures A.6(a) and (b) indicate that when compared with the neighborhoods that incurred no damage, the neighborhoods that experienced 90% and 100% destruction of their residential areas generally show the largest impacts on most outcomes. Importantly, these results indicate that any measurement error due to the construction of new houses before the 1946–1948 period when aerial photos were taken has attenuation effects at worst.

7 Discussion

Our main analyses in the previous section show that the firebombing of Tokyo undermined the living environments of the affected neighborhoods. Yet, whether the air raids actually weakened neighborhoods' informal institutions remains a mystery. This section explores the possible mechanisms through a causal mediation analysis and a survey analysis.

7.1 Mediation Analysis

If the bombing produced large vacant lots and killed some of the landowners, this would facilitate subsequent large scale real-estate developments and would significantly change the building compositions of the damaged neighborhoods. Specifically, we presume the stock of high-rise buildings in a neighborhood could be a key mediating factor that links the damage ratio and crime rates. To explore this possibility, we perform sequential *g*-estimation (Acharya, Blackwell, and Sen, 2016a), a class of mediation analysis, using the ratio of six-storied or higher residences in a neighborhood as the mediator.[34] Through this mediation analysis, we shed light on how much of the observed causal effects of the air raids can be attributed to the subsequent differentiation of the building compositions between the more damaged and less damaged neighborhoods.

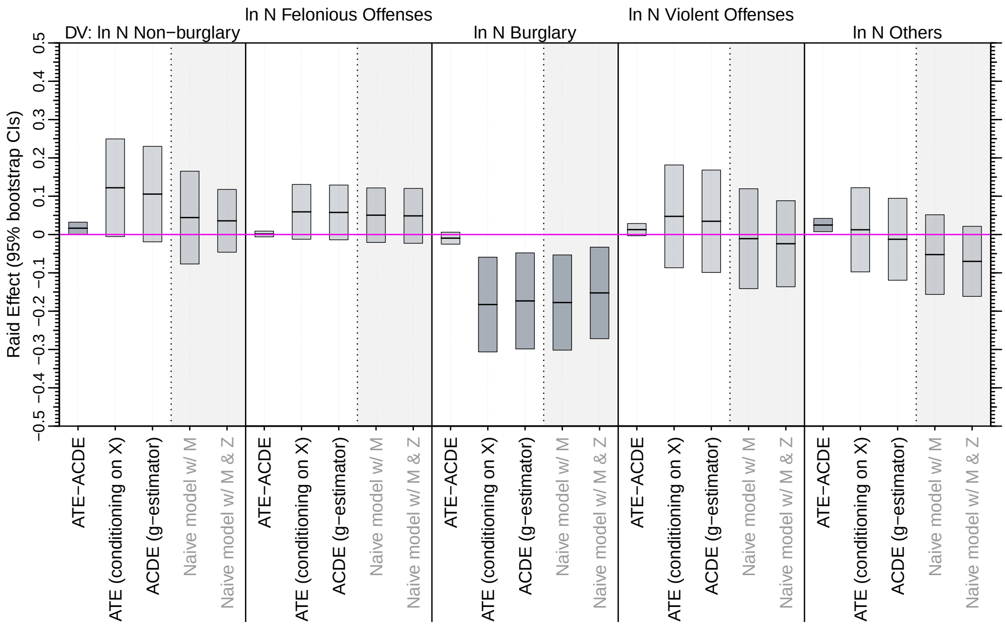
Figure 10 shows the results of the mediation analysis of the crime-related outcome variables in the upper panel and the socioeconomic outcomes in the lower panel. Because some of the socioeconomic outcomes are only available in the full census, the outcome variables are taken from the 2010 census. The five bars of each outcome variable represent different quantities of interest—starting from the left: the net effect of the mediator (ATE − ACDE), the original treatment effect (ATE), the sequential g-estimator (i.e., the direct effect of the air raids removing the effects of the mediator), the naïve model estimate adding the mediator as a right-hand-side variable, and the naïve model estimate adding the mediator and the intermediate confounders as right-hand-side variables.[35]

Looking at the leftmost bar (labeled as ATE − ACDE) of all outcome variables, the mediator generally has limited effects, with one exception. The ratio of high-rise residences has significant negative effects on the average years of residency because almost all high-rise residences were built after WWII, and the residents of these buildings naturally came from elsewhere after the war. In Japan, those with high socioeconomic status also tend to live in high-rise residences, so the mediation effects on the ratio of executive and professional workers and the average education years are positive, while the counterparts for unemployment rates are negative. However, the net effects on these outcomes are negligible.

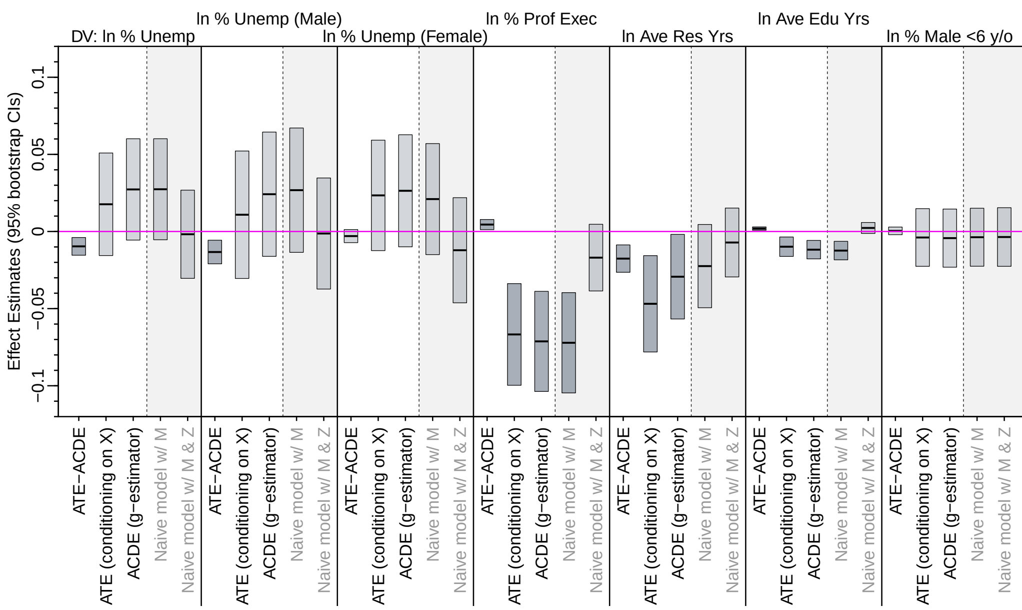
If the contemporary differences in the socioeconomic outcomes between the bombed areas and the undamaged areas are attributable to the differences between the personal attributes typically seen among new movers and the counterpart local inhabitants, the legacies of the air raids can be reduced to individuals rather than communities. Thus, we performed the

[34]See Appendix G for details of the estimation procedure.

[35]The intermediate confounders are the covariates that are observed after the treatment variable and affect the mediator as well as the outcome variable.



(a) Results for crime data



(b) Results for census data

Figure 10: Sequential g-estimation Results for the Effects of Damage Ratios

Note: Sequential g-estimation results for the crime data (Panel (a)) and logged unemployment rates (male and female), logged ratio of executives and high-skilled workers, logged average years spent living in their current residence, logged average educational attainment, and logged ratio of male children younger than 6 years old (Panel (b)), with the ratio of six-storied or higher residences in a neighborhood as a mediator. Five bars for each outcome variable (separated by vertical solid segments) indicate, starting from the left, the net effect of the mediator (ATE − ACDE), the total treatment effect (ATE) reported above, the ACDE estimate derived from the sequential g-estimator, the naïve model (WLS) estimate including the mediator entering as a right-hand-side variable, and the naïve model (WLS) estimate including the mediator and the intermediate confounders as right-hand-side variables, respectively.

same sequential g-estimation, now using the ratio of those who moved to the neighborhood from elsewhere as a mediator. The two panels in Figure A.7 generally give us the same story. That is, the mediator has a large effect on the average years of residency but overall has limited or opposite effects on the other outcome variables.

These mediation analyses offer some important insights for our study. First, each mediator certainly affects people's residency patterns, while this rarely translates into socioeconomic performance. Second, the mediating role of the renewed urban landscape in the relationship between the air raids and the contemporary outcomes is minor. For example, the difference in security systems between high-rise and low-rise residences are now unlikely to be the mediating factor that explains the deterrent effect of burglary thefts of the bombing, which highlights human-related factors such as overconfidence in neighborly ties. Third, population influxes also play a minor role in generating the long-run treatment effects. That is, the differences in the contemporary outcomes most likely reflect differences in the raid damages to neighborhood communities.

7.2 Survey Analysis

So, what did the air raids change about neighborhoods? To answer this question, we conducted a geocoded internet survey from September 25 to October 1, 2018 (first wave) and December 6 to 10, 2018 (second wave).[36] The survey was performed through Rakuten Insight, a Japanese internet survey company, and sent to a registered panel of respondents living in Tokyo's 23 wards. Respondents participated in two common behavioral games, the ultimatum game and the dictator game,[37] and were also asked direct questions about their levels of social trust, political participation, hostilities toward the former adversaries in WWII, and whether their families were directly affected by the Tokyo air raids, as well as questions about their socioeconomic status.

An important feature of this survey is that it also asked each respondent's neighborhood

[36]See Appendix H for details of the internet survey.

[37]All players received the rewards promised to be paid from the game (i.e., there was no deception).

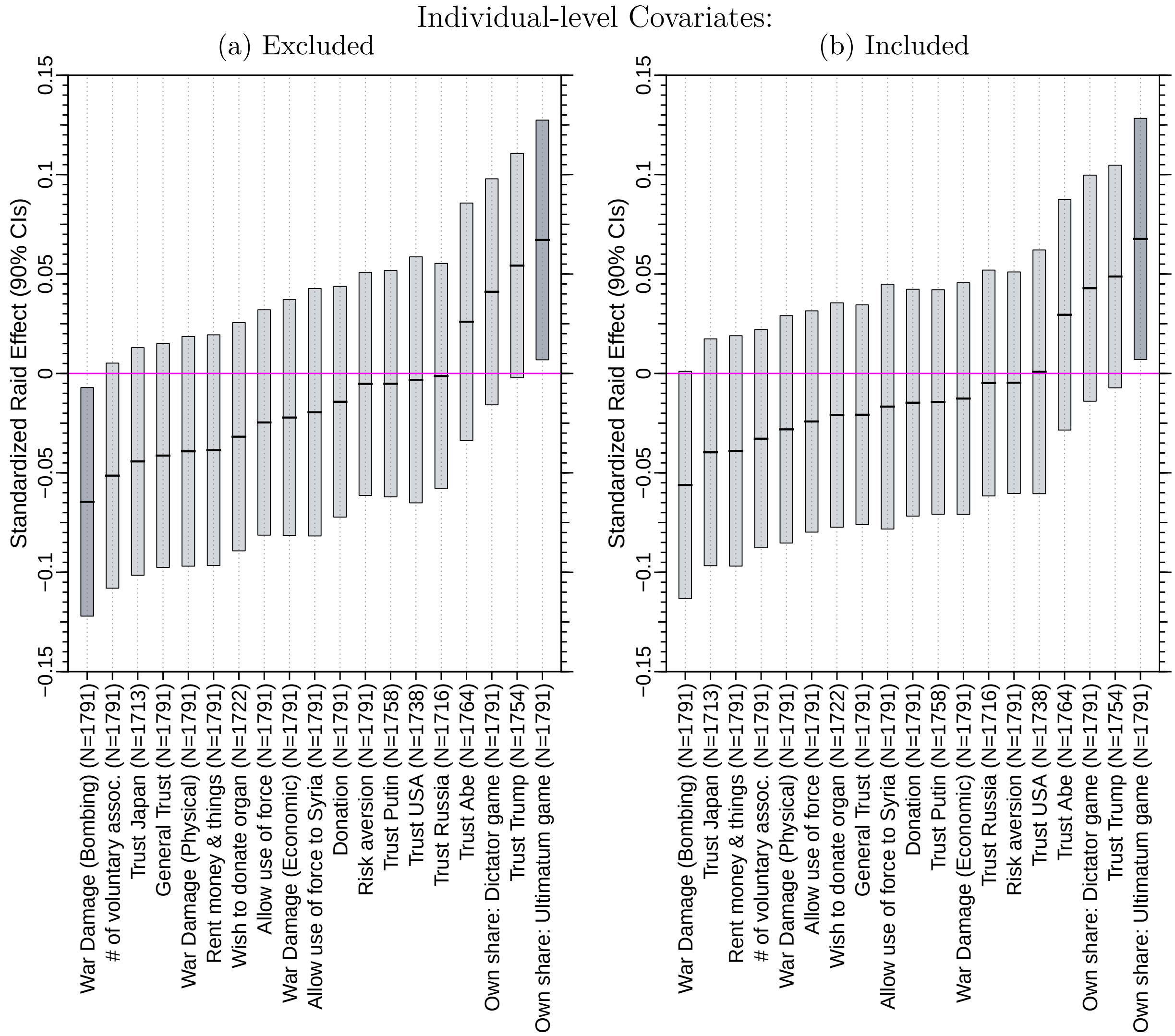


Figure 11: Geo-additive Model Estimates of Regressions of Damage Ratios on Survey Items, with 90% CIs.

Note: Panel (a) represents the regression estimates with individual-level post-treatment covariates, while Panel (b) plots the estimates without post-treatment covariates. Appendix C describes the definitions of the outcome variables.

of residence so that we can merge their responses into the air raid data set using neighborhood as a common key. Across the two waves, 3,423 respondents entered the survey, and 2,626 completed responses, for a response rate of 76.7%. Respondents who did not provide their residential location or finished the survey in less than three minutes are excluded from the sample. These screening processes leave 1,791 respondents, representing 1,267 neighborhoods.

Figure 11 displays the effects of neighborhood-level damage on various outcome variables measured in the survey. The estimation model is the same as in Equation 2, except that the outcome variables are now the survey responses.[38] In addition, we estimate the models with (right panel) and without (left panel) individual-level covariates.[39] Individual characteristics are post-treatment variables, and generally should not be used as control variables; however, they are also strong predictors of preferences, so we estimate and present both specifications. As denoted in Figure 11, sample size varies across outcome variables because some of the 1,791 respondents in the sample did not complete several questions.

In general, the effects are not strong, and we cannot be overconfident in interpreting them. Nevertheless, a few points are worth mentioning in light of existing research and possible theoretical mechanisms. First, the leftmost bars in the plots indicate that the neighborhood-level damage due to the air raids is statistically significant (Panel (a)) or marginally insignificant (Panel (b)) and negatively associated with direct experience of the bombing by respondents' families. This suggests that many families victimized by the air raids do not continue to live in the area, in contrast to situations studied elsewhere, such as Ukraine, in which victims and their descendants continue to live in the afflicted areas, sharing collective memories of the war through generations (Rozenas, Schutte, and Zhukov, 2017; Lupu and Peisakhin, 2017).

Second, the rightmost bars indicate statistically significant and positive effects on the proposed "own shares" in the ultimatum game, and the coefficients for the dictator game have statistically insignificant yet fairly large positive coefficients. In other words, residents in the more heavily bombed areas are more likely to offer a self-serving proposal whether their individual characteristics are taken into account or not. This suggests that the air raids, or the displacement of local communities that they caused, may have destroyed informal institutions underpinning the norms that lead people to socially desirable behaviors, and

[38] Survey questions that are measured in ordinal scale are treated as continuous variables for the purposes of this analysis.

[39] Specifically, the following variables are included: the logged age, sex, the logged years of education, and the logged income, marital status, the logged number of households, and the logged years of residency.

that this legacy has continued until the present day.

Third, although not statistically significant, the barometers of social trust, such as membership in voluntary associations, willingness to rent one's money and things to others, general level of social trust, and willingness to donate one's organs after death, consistently show a negative relationship with the raid damages. Respondents in the most heavily bombed areas are also more likely to trust in the individual leaders of the US and Japan, President Donald Trump and Prime Minister Shinzo Abe, than either country in general.[40]

In all, the survey analysis reveals that heavily bombed neighborhoods tend to have fewer residents whose families were directly affected by the air raids, and lower levels of social trust and altruistic behavior, which may obstruct community-level cooperation and economic development.[41] While we cannot measure the communities' socioeconomic performance per se (Fukuyama, 2001), these pieces of evidence suggest a link between the air raids and the destruction of informal institutions linking community members together. In other words, our survey results are in line with a long-term negative effect of violence that operates through the weakening of local social cohesion (e.g., Wood, 2003, 2008; Nunn and Wantchekon, 2011; Grosjean, 2014). However, our results suggest that this weakening may be due to the displacement of existing local communities (cf., Gilligan, Pasquale, and Samii, 2014).

8 Conclusion

How does exposure to mass political violence shape subsequent social, economic, and political outcomes? The main inferential threat to isolating the long-term causal effects lies in the non-random treatment assignment. In this study, we used historical aerial imagery to develop an original data set capturing the neighborhood-level damages caused by the

[40] A possible interpretation of this pattern, but one which we do not have the data to prove, is that it reflects community-level differences in a reliance on traditional media, especially national newspapers. Households living in traditional communities may be more likely to subscribe to newspapers, and the tone (especially in the *Asahi Shimbun*, is generally critical of the Trump and Abe administrations, in contrast to TV media.

[41] Our results are mostly unchanged by whether the individual covariates are included, and the findings from the mediation analysis suggest that architectural developments and migration patterns of neighborhoods played a limited role.

firebombing of Tokyo in WWII. This data set allowed us to exploit the naturally-occurring exogenous variation in the damages to estimate the long-term effects of the air raids on several socioeconomic outcomes in modern-day Tokyo. We thus provide one of the most disaggregated assessments of war damages, and demonstrate that the indiscriminate bombing across the city indeed left long-lasting negative legacies, even after adjusting for the primary treatment assignment mechanisms and an extensive set of covariates.

Our main findings are threefold. First, the most heavily damaged neighborhoods still suffer from higher crime, higher unemployment rates, lower worker skills, and lower education among residents. Second, the mediation analysis suggests that contemporary or post-treatment factors, such as postwar urban development and population influxes, played a limited role in generating the lasting effects of the WWII air raids. Finally, the original online survey demonstrates that residents in the heavily-damaged neighborhoods tend to show lower levels of social trust and altruistic behaviors. The minor roles of postwar factors and the insights from the survey shed light on the underlying mechanisms. Although the evidence remains indirect, our findings generally suggest that the destruction of local communities as informal institutions, rather than urban redevelopment or population change, operates as the key factor in the causal pathway linking historical catastrophes and contemporary outcomes.

Our findings speak to the broader literature of institutional political economics in several ways. First, our study contributes to the emerging research agenda on the deep historical roots of contemporary social, economic, and political outcomes. While it is often difficult to identify the long-run causal effects of historical events and decisions, our original, micro-level data set alleviates several important methodological challenges. The micro-level nature of the data allows us to exploit plausibly exogenous variation in damages, thereby minimizing the major inferential threat of non-random treatment assignment.

Second, a related implication lies in the role of social or informal institutions in facilitating community performance and crime prevention. Our empirical findings underscore the role of local communities as a likely causal mechanism in determining observed crime rates. The air raids weakened locally oriented communities and informal institutions that deter crimes and contribute to socially desirable outcomes. A general implication follows that the displacement of community members and the destruction of community ties, whether intended or not, may be followed by deteriorated community performance. Crime prevention and other policy efforts would benefit from paying careful attention to the role of local communities.

Finally, our doubly remote-sensed approach that combines historical archives with remote-sensing and GIS techniques offers a new empirical strategy for future studies. Despite the increasing availability of historical records and imagery, the application of such non-traditional data remains limited in related fields. Our methodological approach opens the way for this strategy to uncover the causal effects of institutions, culture, and history.

References

Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2001. The colonial origins of comparative development: An empirical investigation. *American Economic Review* 91 (5):1369–1401.

Acharya, Avidit, Matthew Blackwell, and Maya Sen. 2016a. Explaining causal findings without bias: Detecting and assessing direct effects. *American Political Science Review* 110 (3):512–529.

———. 2016b. The Political Legacy of American Slavery. *Journal of Politics* 78 (3):621–641.

Arnold, Benjamin F, and Ayse Ercumen. 2016. Negative control outcomes: A tool to detect bias in randomized trials. *JAMA* 316 (24):2597–2598.

Bauer, Michal, Christopher Blattman, Julie Chytilová, Joseph Henrich, Edward Miguel, and Tamar Mitts. 2016. Can War Foster Cooperation? *Journal of Economic Perspectives* 30 (3):249–274.

Becker, Carl B. 1988. Report from Japan: Causes and Controls of Crime in Japan. *Journal of Criminal Justice* 16:425–435.

Becker, Gary S. 1968. Crime and Punishment: An Economic Approach. *Journal of Political Economy* 76 (2):169–217.

Bellows, John, and Edward Miguel. 2006. War and institutions: New evidence from Sierra Leone. *American Economic Review* 96 (2):394–399.

———. 2009. War and Local Collective Action in Sierra Leone. *Journal of Public Economics* 93 (11-12):1144–1157.

Blattman, Christopher. 2009. From Violence to Voting: War and Political Participation in Uganda. *American Political Science Review* 103 (2):231–247.

Brakman, Steven, Harry Garretsen, and Marc Schramm. 2004. The Strategic Bombing of German Cities during World War II and its Impact on City Growth. *Journal of Economic Geography* 4:201–218.

Collier, Paul, V.L. Elliott, Håvard Hegre, Anke Hoeffler, Marta Reynal-Querol, and Nicholas

Sambanis. 2003. *Breaking the Conflict Trap: Civil War and Development Policy.* A World Bank Policy Research Report.

Davis, Donald R, and David E Weinstein. 2002. Bones, bombs, and break points: The geography of economic activity. *American Economic Review* 1269–1289.

Dell, Melissa, and Pablo Querubin. 2018. Nation Building Through Foreign Intervention: Evidence from Discontinuities in Military Strategies. *Quarterly Journal of Economics* 133 (2):701–764.

Donaldson, Dave, and Adam Storeygard. 2016. The View from Above: Applications of Satellite Data in Economics. *Journal of Economic Perspectives* 30 (4):171–198.

Fedman, David, and Cary Karacas. 2012. A cartographic fade to black: Mapping the destruction of urban Japan during World War II. *Journal of Historical Geography* 38 (3):306–328.

Fukuyama, Francis. 2001. Social capital, civil society and development. *Third World Quarterly* 22 (1):7 – 20.

Gilligan, Michael J., Benjamin J. Pasquale, and Cyrus Samii. 2014. Civil War and Social Cohesion: Lab-in-the-Field Evidence from Nepal. *American Journal of Political Science* 58 (3):604–619.

Grosjean, Pauline. 2014. Conflict and social and political preferences: Evidence from World War II and civil conflict in 35 European countries. *Comparative Economic Studies* 56 (3):424–451.

Kammann, E. E., and Matthew P. Wand. 2003. Geoadditive models. *Journal of the Royal Statistical Society. Series C: Applied Statistics* 52 (1):1–18.

Kocher, Matthew Adam, Thomas B. Pepinsky, and Stathis N. Kalyvas. 2011. Aerial Bombing and Counterinsurgency in the Vietnam War. *American Journal of Political Science* 55 (2):201–218.

Lowes, Sara, Nathan Nunn, James A. Robinson, and Jonathan L. Weigel. 2017. The Evolution of Culture and Institutions: Evidence From the Kuba Kingdom. *Econometrica* 85 (4):1065–1091.

Lupu, Noam, and Leonid Peisakhin. 2017. The Legacy of Political Violence across Generations. *American Journal of Political Science* 61 (4):836–851.

Michalopoulos, Stelios, and Elias Papaioannou. 2016. The Long-Run Effects of the Scramble for Africa. *American Economic Review* 106 (7):1802–1848.

Miguel, Edward, and Gérard Roland. 2011. The long-run impact of bombing Vietnam. *Journal of Development Economics* 96 (1):1–15.

Miguel, Edward, Sebastián M Saiegh, and Shanker Satyanath. 2011. Civil war exposure and violence. *Economics & Politics* 23 (1):59–73.

National Police Agency. 2019. White Paper on Police 2018 [Keisatsu Hakusho Heisei 30 Nendo Ban]. *Tech. rep.*

Nunn, Nathan, and Leonard Wantchekon. 2011. The slave trade and the origins of mistrust in Africa. *American Economic Review* 101 (7):3221–3252.

Okuzumi, Katsumoto, Yoshishige; Saotome. 2007. *Shinban Tokyo wo Bakugeki Seyo – Beigun Sakusen Ninmu Houkokusho ha Kataru (Bomb Tokyo New Edition – The Analysis of the Tactical Missions Report).* Tokyo: Sanseido.

Olson, Mancur. 1965. *The logic of collective action: Public goods and the theory of groups.* Cambridge, MA: Harvard University Press.

Pape, Robert A. 1996. *Bombing to Win: Air Power and Coercion in War.* Ithaca, NY: Cornell University Press.

Polmar, Norman. 2004. *The Enola Gay: The B-29 that dropped the atomic bomb on Hiroshima.* Potomac Books, Inc.

Rozenas, Arturas, Sebastian Schutte, and Yuri Zhukov. 2017. The political legacy of violence: The long-term impact of Stalin's repression in Ukraine. *The Journal of Politics* 79 (4):1147–1161.

Saotome, Katsumoto. 1981a. *Nihon No Kūshū 10 Hokan Shiryō-hen (Air Raid in Japan, v. 10: Data).* Tokyo: Sanseido.

———. 1981b. *Nihon No Kūshū 3 Tōkyō (Air Raid in Japan, v. 3: Tokyo).* Tokyo: Sanseido.

Searle, Thomas R. 2002. "It made a lot of sense to kill skilled workers": The firebombing of Tokyo in March 1945. *Journal of Military History* 66 (1):103.

USAAF. 10 March 1945. Tactical Mission Report, Mission No. 40.

Watanabe, Yōkichi. 2013. *Senjika no nihonjin to tonarigumi kaihō [Japanese and the Neighborhood Associations during the War]*. Tokyo: Gentosha Renaissance.

Williams, Josette H. 2007. The Information War in the Pacific, 1945: Paths to Peace. *US Central Intelligence Agency*, https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/csi-studies/studies/vol46no3/article07.html .

Wood, Elisabeth Jean. 2003. *Insurgent Collective Action and Civil War in El Salvador*. Cambridge: Cambridge University Press.

———. 2008. The Social Processes of Civil War: The Wartime Transformation of Social Networks. *Annual Review of Political Science* 11:539–61.

Wood, Simon N. 2006. *Generalized additive models: An introduction with R*. Boca Raton, FL: Chapman & Hall.

Zhukov, Yuri M., and Roya Talibova. 2018. Stalin's terror and the long-term political effects of mass repression. *Journal of Peace Research* 55 (2):267–283.

Online Appendix

A Appendix Figures and Tables

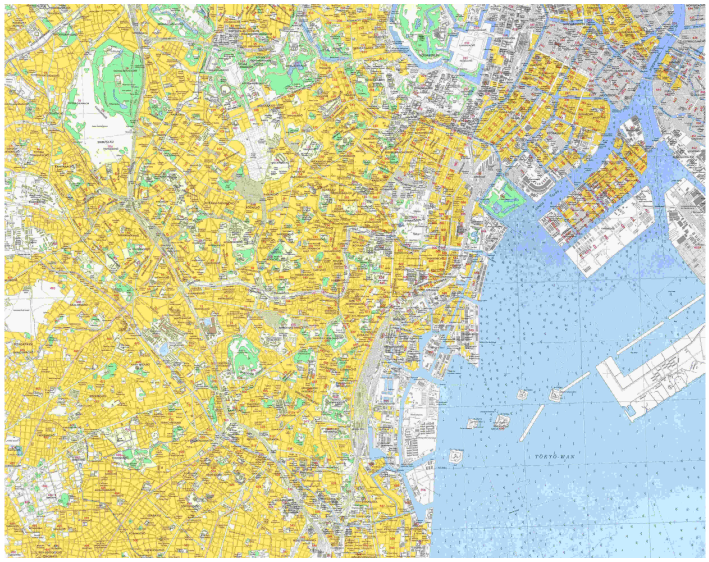
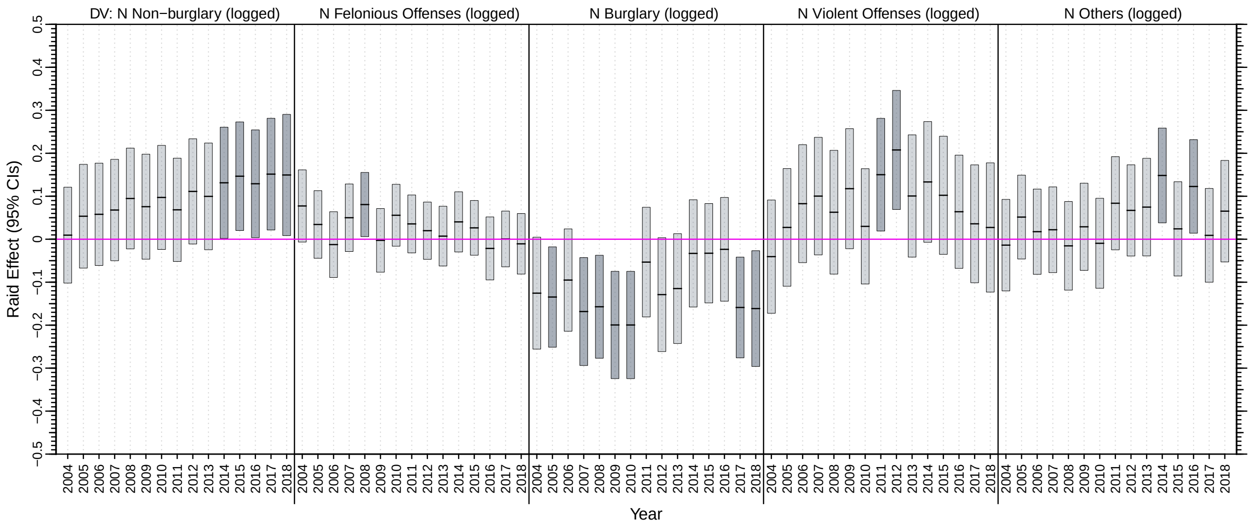


Figure A.1: Map by the US Army Map Service showing the residential areas of prewar Tokyo (Nihombashi area). Yellow-shaded areas indicate residential areas.

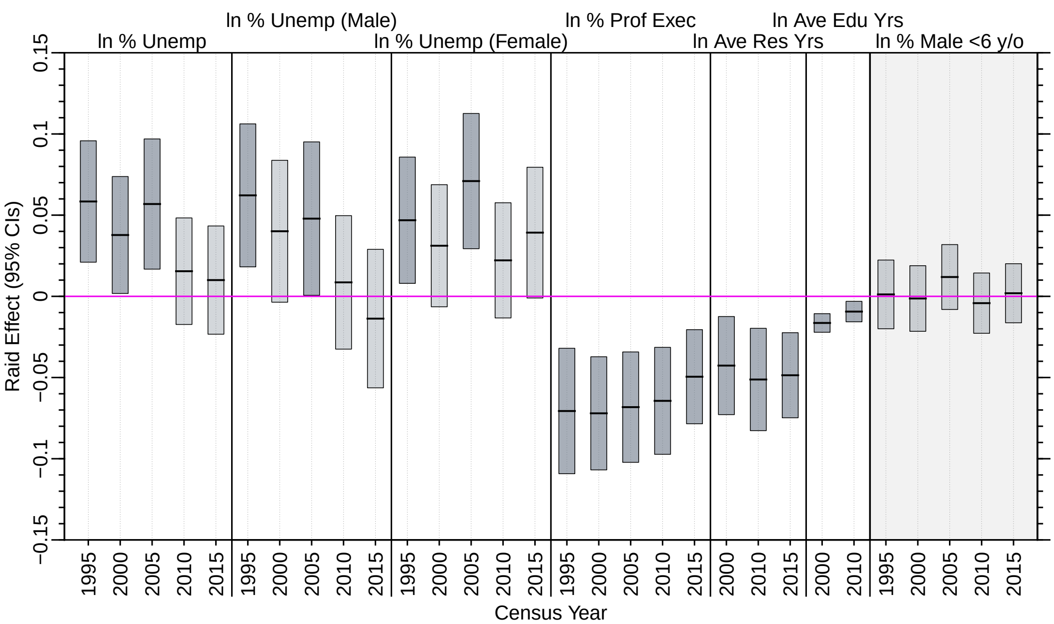
Table A.1: List of the major low-altitude bombings in Tokyo city area.

Date	Target	# of Bombers	Qty. Bombs	# of Deaths
March 10, 1945	East	325	1,665 t	83,793
April 13-14, 1945	North	352	2,120 t	2,459
April 15, 1945	Southern Edge	340	1,879 t	841
May 24, 1945	South	562	3,646 t	762
May 25-26, 1945	West	502	3,302 t	3,651

Notes: The figures are based on Saotome (1981). *Target* represents the area within Tokyo. *Qty. Bombs* includes both ordinary bombs and incendiary bombs. *# of Deaths* does not include missing persons.



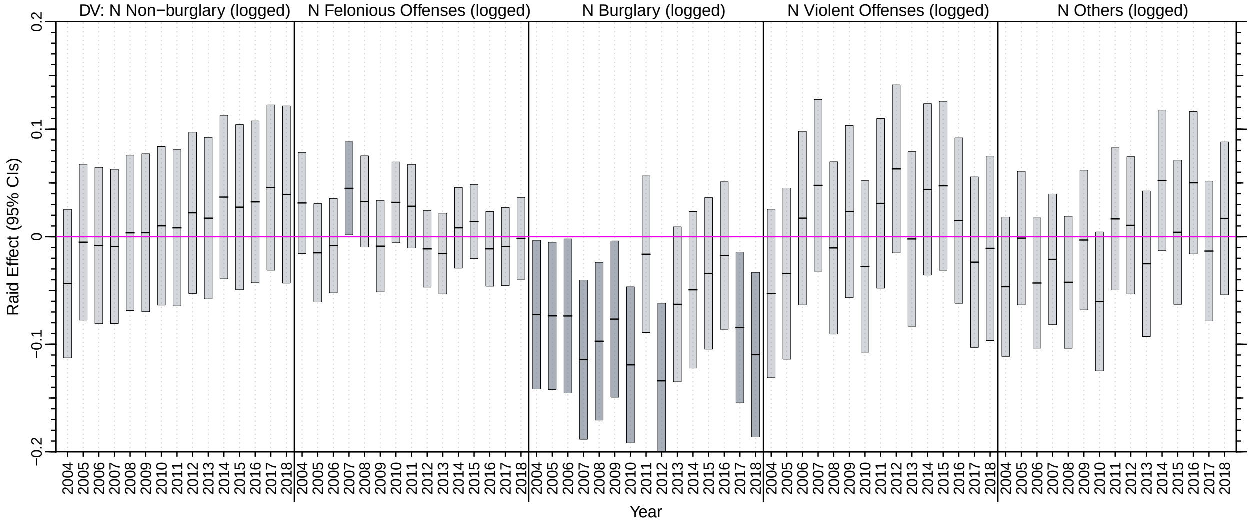
(a) Y: logged crime counts of different crime types



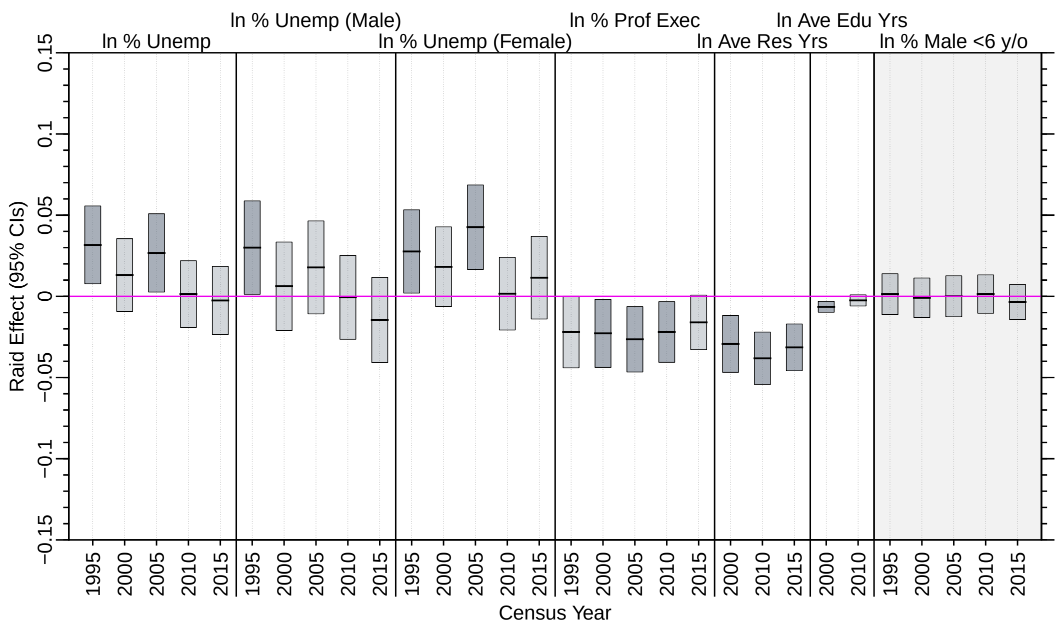
(b) Y: logged unemployment rates (male and female), logged ratio of executives and high-skilled workers, logged average years spent living in their current residence, logged average educational attainment, and logged ratio of male children younger than 6 years old

Figure A.2: Fifth-ordered polynomial model estimates of regressions on damage ratios.

Note: The polynomial specification adds the 20 interaction terms of the longitude and latitude, and five polynomials for each of the logged distances from the Imperial Palace, and the logged minimum distance from the aiming points, to the model.



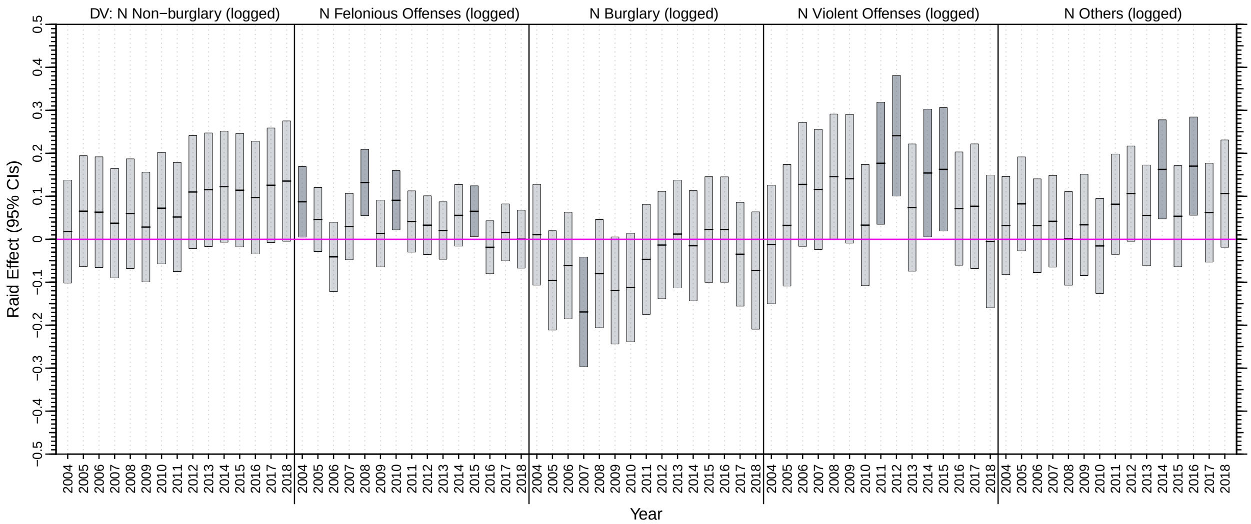
(a) Y: logged crime counts of different crime types



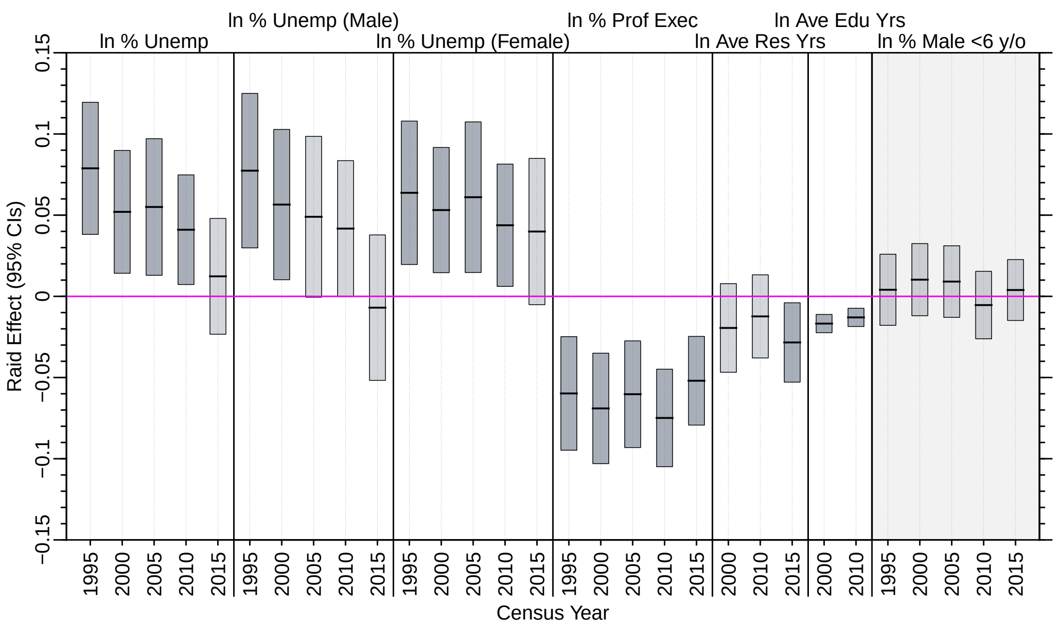
(b) Y: logged unemployment rates (male and female), logged ratio of executives and high-skilled workers, logged average years spent living in their current residence, logged average educational attainment, and logged ratio of male children younger than 6 years old

Figure A.3: Geo-additive model estimates of regressions on binary damage indicator.

Note: The damage ratio variable is split at its mean (≈ 0.482) to dichotomize the variable.

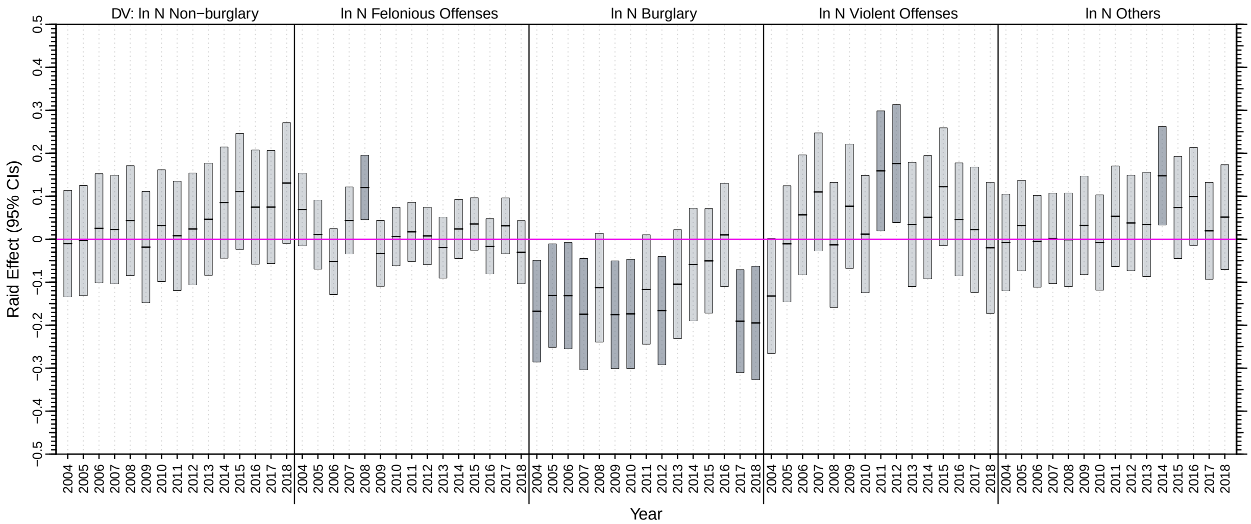


(a) Y: logged crime counts of different crime types

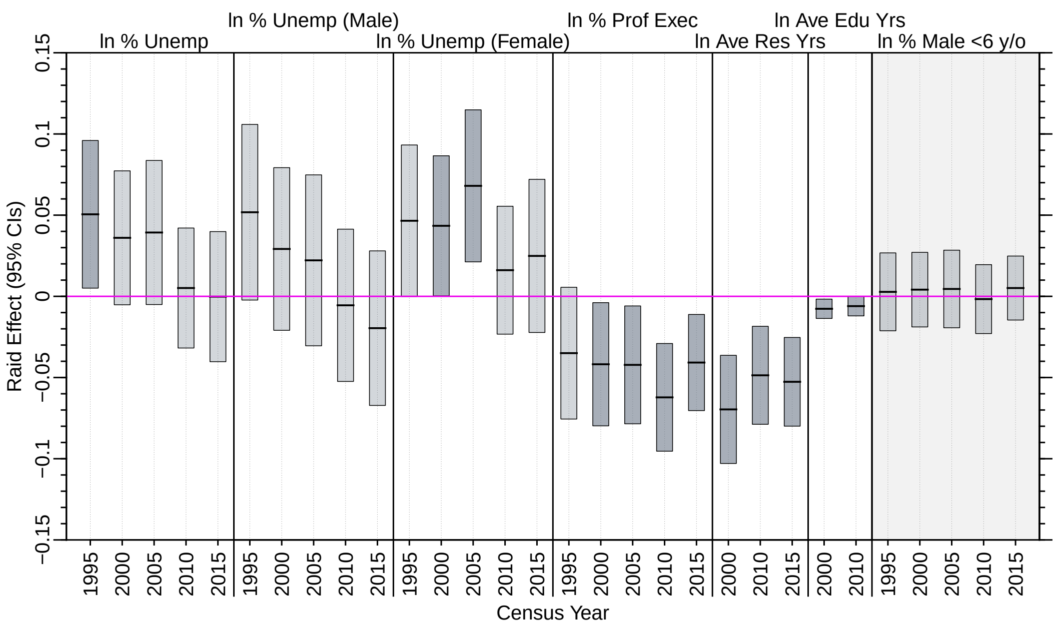


(b) Y: logged unemployment rates (male and female), logged ratio of executives and high-skilled workers, logged average years spent living in their current residence, logged average educational attainment, and logged ratio of male children younger than 6 years old

Figure A.4: Geo-additive model estimates of regressions on damage ratios using the subsample of neighborhoods where the prewar residential area ratio was 70% or greater.



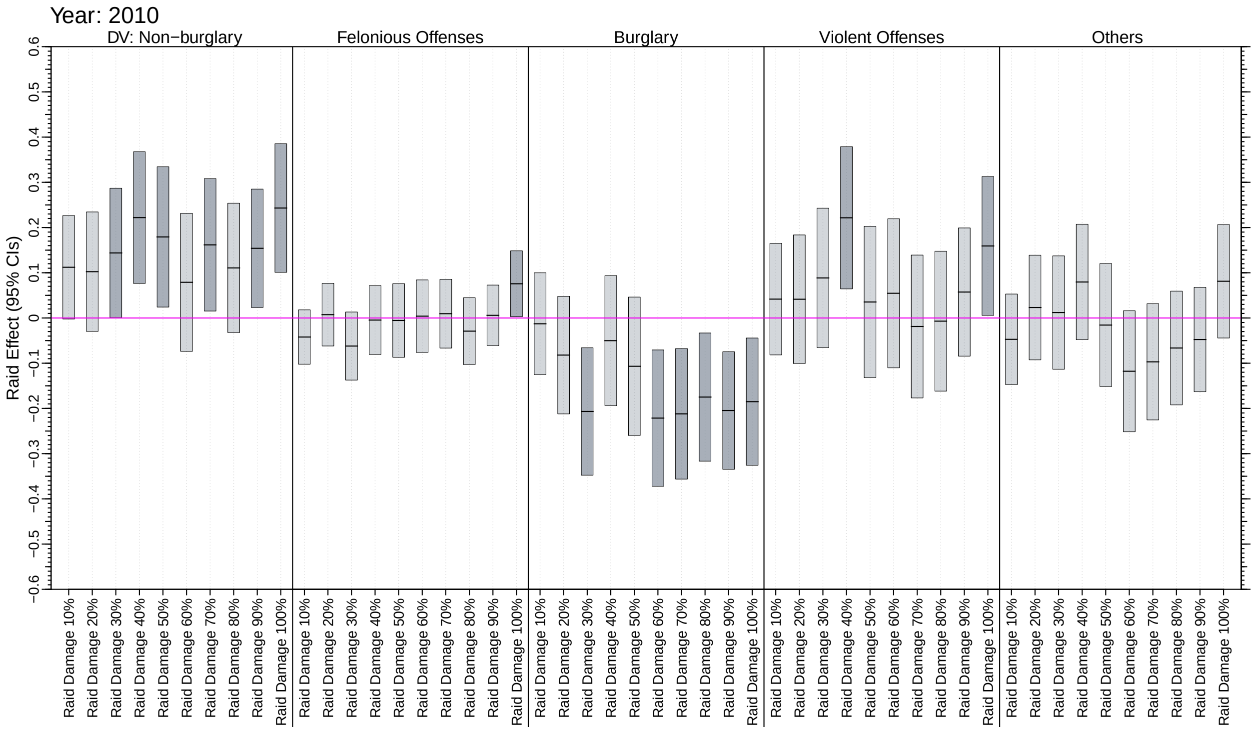
(a) Y: logged crime counts of different crime types



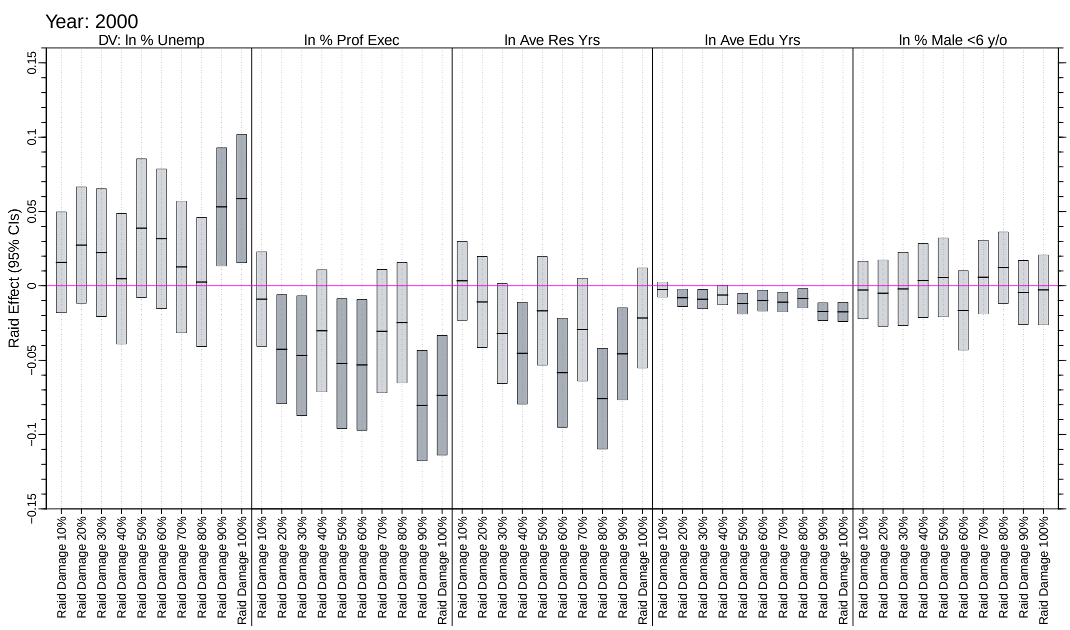
(b) Y: logged unemployment rates (male and female), logged ratio of executives and high-skilled workers, logged average years spent living in their current residence, logged average educational attainment, and logged ratio of male children younger than 6 years old

Figure A.5: Geo-additive model estimates of regressions on damage ratios using the subsample defined as the neighborhoods within 10 km distance from the centroid of the Imperial Palace.

Note: The results are estimated from the data set of the year 2010.



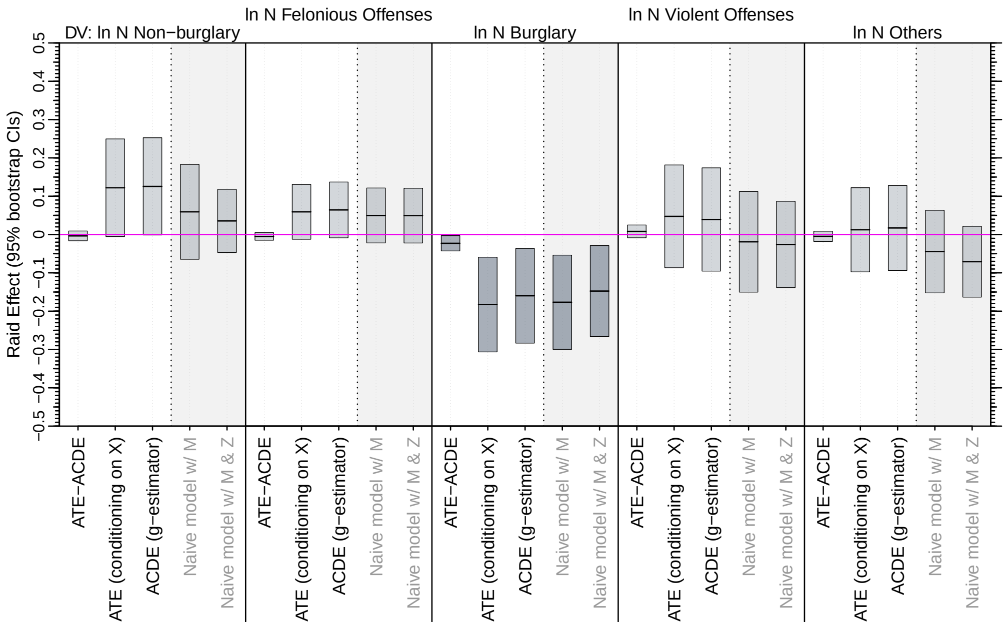
(a) Y: logged crime counts of different crime types



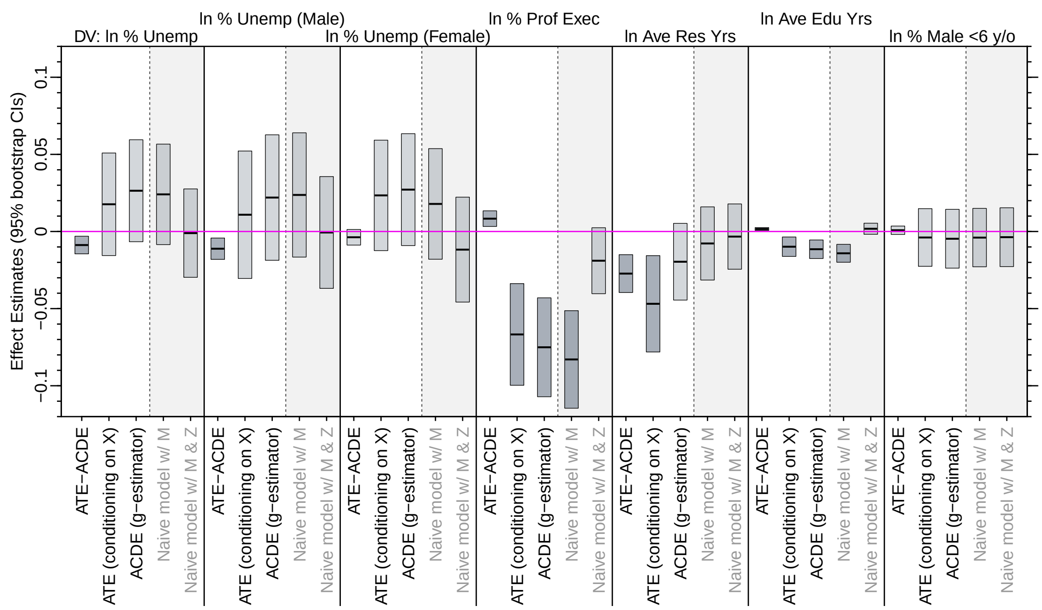
(b) Y: logged unemployment rates (male and female), logged ratio of executives and high-skilled workers, logged average years spent living in their current residence, logged average educational attainment, and logged ratio of male children younger than 6 years old

Figure A.6: Geo-additive model estimates of regressions on the 10 dummy damage variables.

Note: The results are estimated from the data set of the year 2010.



(a) Results for crime data



(b) Results for census data

Figure A.7: Sequential g-estimation results for the effects of the damage ratios

Note: Sequential g-estimation results for the crime data (Panel (a)) and logged unemployment rates (male and female), logged ratio of executives and high-skilled workers, logged average years spent living in their current residence, logged average educational attainment, and logged ratio of male children younger than 6 years old (Panel (b)), with the ratio of those who moved to the neighborhood from elsewhere as a mediator. Five bars for each outcome variable (separated by vertical solid segments) indicate, starting from the left, the net effect of the mediator (ATE − ACDE), the total treatment effect (ATE) reported above, the ACDE estimate derived from the sequential g-estimator, the naïve model (WLS) estimate including the mediator entering as a right-hand-side variable, and the naïve model (WLS) estimate including the mediator and the intermediate confounders as right-hand-side variables, respectively.

B The selection of the aerial photography

The selection of aerial imagery requires striking the balance between the photo quality and the accuracy of the information. In general, the former improves as time proceeds because the USAAF could choose experienced pilots, skilled camera crews, and sunny days for photo-shooting. In contrast, the latter deteriorates as time goes by because of the potential construction of newer buildings in the cleared land, which makes it harder to identify the damaged areas.

We selected most of the aerial photos from those taken in the year 1947 and some from both 1946 and 1948. See Figure B.1 for the histogram of the dates on which the aerial photos for this study were taken. We georeferenced 666 aerial photos so that they cover the entire area of Tokyo's 23 wards, as represented in Figure 3(a).[1] These photos were, however, taken about two years after the firebombing of Tokyo took place. Thus, we assume that the Allies did not rehabilitate specific areas during this two-year period in a way that is correlated with contemporary socioeconomic characteristics.

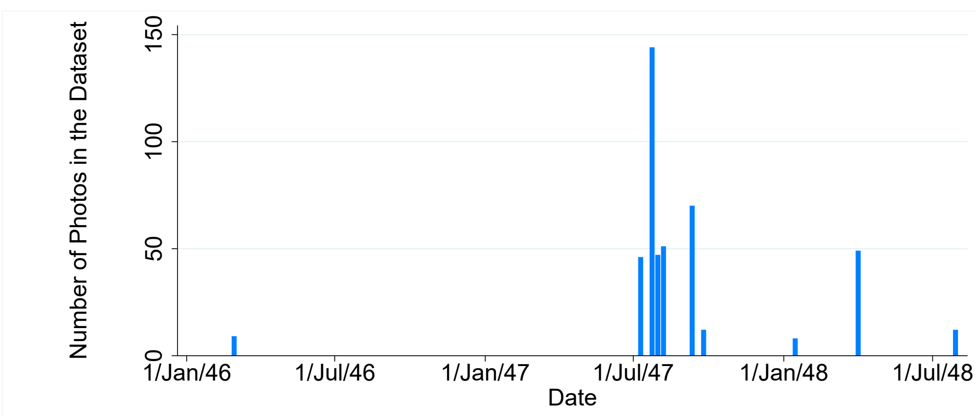


Figure B.1: Histogram of the dates on which aerial photos for this study were taken.

Importantly, this concern makes it inappropriate to use one of the most extensive data sources. Specifically, the GSI developed a series of geo-referenced maps of Tokyo for every five year period since 1936. The geo-referenced aerial imagery right after the war was collected

[1]The complete list of the aerial photography used in our data set is available upon request.

from aerial photography from 1945 to 1950. However, since these data sets were developed to trace the history of cityscapes, not to evaluate the air raid damages, measurement of the damages is a challenge due to newly built houses, especially for the photographs taken in 1948 and after.[2]

Several other potential data sources exist to measure the damages, including *Tokyo-to sensai shi* (Tokyo prefecture(ed.), 2005), *Konsaisu Tōkyō-to 35-ku kubun chizuchō* (Nihon Chizu Kabushiki-gaisha: Reprinted by Tōkyō Kūshū wo Kirokusuru Kai, 1985) and *Nihon toshi sensai chizu* (Daiichi Fukuinshō(ed.), 1983). Furthermore, *Zenkoku shuyō toshi sensai gaikyōzu* (National Archives of Japan Digital Archive, N.d.), the original source of *Nihon toshi sensai chizu*, has been digitized and georeferenced by Takashi Kirimura (URL: https://www.arcgis.com/home/item.html?id=f9ad17ff0a104300ac71ea29dba92457. Accessed March 25, 2019). Although these data sources are valuable, we opt not to use them due to some discrepancies and errors across data sets (owing to the urgency with which they were created in the immediate aftermath of the war), as well as the dichotomous nature of the coding scheme.

C Definitions of the outcome variables

This appendix gives detailed definitions of the outcome variables used in the main analysis and the internet survey.[3] The unemployment rates are calculated with the following formula:

$$\text{Unemployment rates} = 1 - \frac{\text{Number of employed persons}}{\text{Number of employed persons} + \text{Number of unemployed persons}}. \quad \text{(C.1)}$$

The male and female unemployment rates are calculated in the same way. Note that the way the unemployment rates in Japan is calculated differs from the counterpart in the US.

[2] This alternative source is available from the following link: http://maps.gsi.go.jp/.

[3] The English translation of the census is available from the following URL: https://www.stat.go.jp/english/data/kokusei/2010/final_en/final_en.html

One is classified into the unemployed only when one looks for a job *and* cannot get it.

The ratio of the executive and professional workers is calculated by the following formula:

$$\begin{aligned}&\text{Ratio of the executive and professional workers}\\ &\qquad = \frac{\text{Num. professional workers} + \text{Num. executive workers}}{\text{Number of employed persons}}. \quad \text{(C.2)}\end{aligned}$$

The average residency in years is inferred from one of the questions in the census "6. Period of living at the present domicile," which is asked every 10 years until 2010 and becomes every 5 years since 2015. This question has the following 6 choices: (i) since birth, (ii) less than 1 year, (iii) 1 to less than 5 years, (iv) 5 to less than 10 years, (v) 10 to less than 20 years, (vi) 20 years or more. The neighborhood-level census provides the number of those who selected each choice, so the number of those selecting "since birth" is multiplied by the average age of the residents in the neighborhood, and we assigned the middle value to each of the other options, and multiply it by the number of those who selected the corresponding option. Thus, the average years spent living in their current residence are calculated by the following formula:

$$\begin{aligned}&\text{Ave. living years (residency)}\\ &\qquad = \frac{\text{Ave. age} \times (i) + 0.5 \times (ii) + 3 \times (iii) + 7.5 \times (iv) + 15 \times (v) + 25 \times (vi)}{(i) + (ii) + (iii) + (iv) + (v) + (iv)}. \quad \text{(C.3)}\end{aligned}$$

The average years of education are inferred from the "Schooling" question in the census, which is asked only every 10 years and has the following 4 choices: (i) Primary/junior high school, (ii) High school, (iii) Junior/technical college, (iv) University, graduate school. In order for the proportion of the school-age population to affect the statistics, we exclude those who are currently enrolled in school from this calculation. In the same way as the calculation of the average residency in years, we assigned the middle year value to each option and multiply it by the number of those who answered the corresponding option.

Table C.1: The definitions of the outcome variables used in the survey analysis.

Variable Names	Definition
# of voluntary assoc.	The number of options chosen in Q.18 (a) ~(h) and Q.19 (a).
Allow use of force	Q.21 (a) $\to$ 5, (b) $\to$ 4, $\cdots$, (e) $\to$ 1.
Allow use of force to Syria	Q.22 (a) $\to$ 5, (b) $\to$ 4, $\cdots$, (e) $\to$ 1.
Donation	The number of options chosen in Q.16 (a) ~(c).
General Trust	Q.12 (a) $\to$ 1, (b) $\to$ 2, $\cdots$, (g) $\to$ 7.
Own share: Dictator game	The amount selected in Q.29
Own share: Ultimatum game	The amount selected in Q.30
Rent money & things	Q.14 (a) $\to$ 5, (b) $\to$ 4, $\cdots$, (e) $\to$ 1.
Risk aversion	Q.23 (a) & Q.24 (a) $\to$ 4, Q.23 (a) & Q.24 (b) $\to$ 3, Q.23 (b) & Q.25 (a) $\to$ 2 and Q.23 (b) & Q.25 (b) $\to$ 1.
Trust Abe	Q.17 (vi) (a) $\to$ 5, (b) $\to$ 4, $\cdots$, (e) $\to$ 1.
Trust Japan	Q.17 (v) (a) $\to$ 5, (b) $\to$ 4, $\cdots$, (e) $\to$ 1.
Trust Putin	Q.17 (iv) (a) $\to$ 5, (b) $\to$ 4, $\cdots$, (e) $\to$ 1.
Trust Russia	Q.17 (iii) (a) $\to$ 5, (b) $\to$ 4, $\cdots$, (e) $\to$ 1.
Trust Trump	Q.17 (ii) (a) $\to$ 5, (b) $\to$ 4, $\cdots$, (e) $\to$ 1.
Trust USA	Q.17 (i) (a) $\to$ 5, (b) $\to$ 4, $\cdots$, (e) $\to$ 1.
War Damage (Bombing)	The number of options chosen in Q.27 (b), Q.28 (b) and Q.28 (c).
War Damage (Economic)	The number of options chosen in Q.28 (b), Q.27 (c), Q.27 (d) and Q.27 (e).
War Damage (Physical)	The number of options chosen in Q.27 (b), Q.27 (c) and Q.27 (d).
Wish to donate organ	Q.15 (a) $\mid$ Q.15 (b) $\to$ 3, Q.15 (c) $\to$ 2 and Q.15 (d) $\to$ 1

Note: The question numbers correspond with those used in Appendix H.

Those who finished only preschool are excluded from the calculation. The average education years are calculated by the following formula:

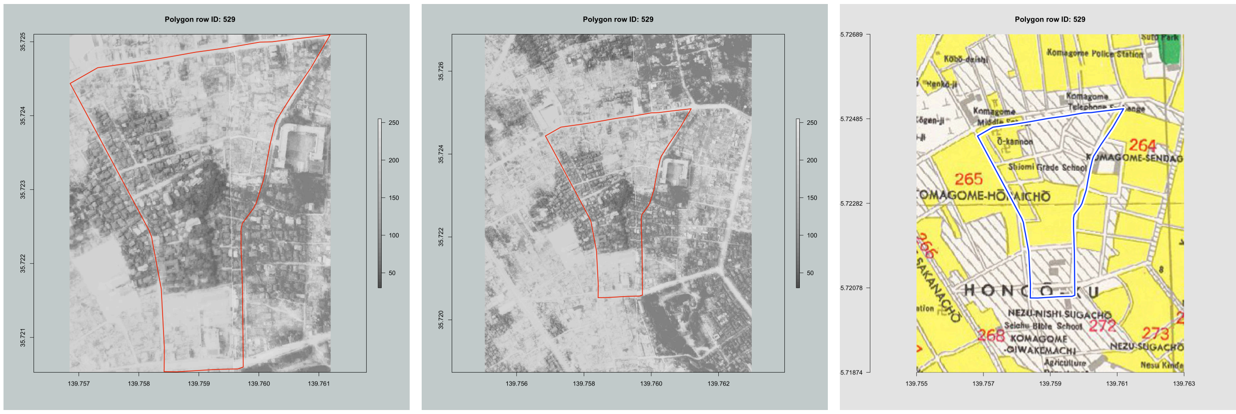
$$\text{Ave. years of education} = \frac{9 \times (i) + 12 \times (ii) + 14 \times (iii) + 17 \times (iv)}{(i) + (ii) + (iii) + (iv)}. \tag{C.4}$$

Finally, Table C.1 shows the definition of the outcome variables used in the internet survey analysis. See also Appendix H for the English translation of the original survey questionnaire.

D Details of human coding

Prior to the human coding, three types of imagery were prepared. The first type of imagery was the aerial photography of the closeup neighborhood polygon (in the left panel of Figure D.1). The boundary of the neighborhood was indicated in the red solid line, and the surrounding area was depicted only in the narrow margin. This was used to determine the damage ratio and the residential area ratio. The second type of imagery was the aerial photography of the neighborhood polygon with 200-meter ($\approx$ 656 foot) margin (in the center panel of Figure D.1). This type of imagery was used to gain additional information of the surrounding environment when classification was difficult. Typical instances include the classification between burnt areas and crop fields or the classification between undamaged houses and trees. The third type of imagery was the AMS map image of the neighborhood polygon with 200-meter margin (in the right panel of Figure D.1). Similar to the second type of imagery, this type of imagery was used to gain additional information for classification. Especially, this map information was primarily used to avoid grave classification errors since burnt areas, residential areas and other non-residential areas were clearly distinguished in the map. However, we used the map information only as a secondary source because many neighborhoods experienced additional damages after the production of this map. For example, the area around the upper left corner of the polygon in this AMS map is in yellow indicating the area was undamaged. However, the corresponding area in the aerial photo shows that the area was actually burnt. In this way, the coding procedure was implemented taking advantage of each type of imagery.

The two variables, the damage ratio and the residential area ratio were evaluated for each neighborhood. First, the residential area ratio was evaluated. The denominator of this variable is the area of the entire polygon excluding water areas, such as seas, rivers, and canals. The numerator is the total area of both the damages and undamaged residential areas. One may think it is hard to determine the original land usage from the burnt area. However, the burnt residential areas typically feature visible small squared traces of the



(a) Aerial Photograph with 0 m margin

(b) Aerial Photograph with 200 m margin

(c) AMS map with 200 m margin

Figure D.1: Three types of images used in human-coding.

burnt houses (though this is also our speculation based on the experience of georeferencing and human-coding). Thus, the coder relied on these signs and the AMS map in deciding whether a given burnt area was originally a residential area.

Second, the damage ratio was estimated. Since the residential area ratio, the denominator, has already been estimated first, the coder only needed to divide the residential area into the burnt and the undamaged areas and estimate the fraction of the former. One issue in this task was the classification of barracks (or buildings that looked like barracks). A few dozen of the aerial photos seem to contain sparsely built barracks in the bombed areas. They might have been built after the air raids cleared the land, or these houses might have escaped from the fire. We cannot determine which is the case from aerial photography. In the study, we classified the barracks into undamaged areas. This is because even if the barracks were built after the air raids, these areas probably incurred lesser damages than the burnt areas with no barracks as these buildings indicate the survival of some people who restarted their daily lives.

It took about one minute for the coder to finish evaluating the damage ratio and the residential area ratio per neighborhood. Moreover, the coder flagged the neighborhoods for which evaluations were difficult. These neighborhoods were later re-evaluated with additional information. Specifically, an airplane usually takes a series of aerial photos, allowing some amount of overlapping regions across photos, and the flight routes of different planes usually overlap each other, too, so that all the planned area is covered by the aerial photography. Thus, several aerial photos are usually available for each neighborhood, and the coder could rely on the clearest photo among them. Most of these neighborhoods were re-evaluated in the same way. After all, most difficulties arise from the impossibility to tell the ground truth. For example, it is difficult to determine whether the small houses near the ocean were fishermen's homes or warehouses to store boats and fishnets.

A reasonable concern would be the possibility of manipulation or the lack of transparency of the coding procedure since an author served as a coder. We chose this strategy primarily because this coding strategy results in the most reliable data compared to machine-coding or human-coding by another coder. Since the coder (Harada) processed most of the georeferencing, he knows the photographic features of each serial of the aerial photography that was taken by a different plane on a different date. In contrast, machine-coding of monotone imagery has many problems, and the resulting classification tends to contain many measurement errors primarily due to the lack of color information. We did not outsource the human-coder primarily because we cannot guarantee the quality of coding. After all, this task is too laborious for anyone other than the authors to maintain a serious commitment while we lack an assessment tool to evaluate one's effort without "ground truth." Thus, the typical principal-agent problem will always overshadow coding by a third party.

To ensure transparency, we disclose our evaluation of which neighborhoods incurred how much of the damages, so that researchers can examine the validity of the coding by referring to the original aerial photography from the GSI website (The Geospatial Information Authority of Japan, N.d.). Furthermore, we will release the georeferenced photography of Tokyo's 23 wards upon the completion of our project, so that researchers can freely replicate or create their own coding.

E Details of the flammability data

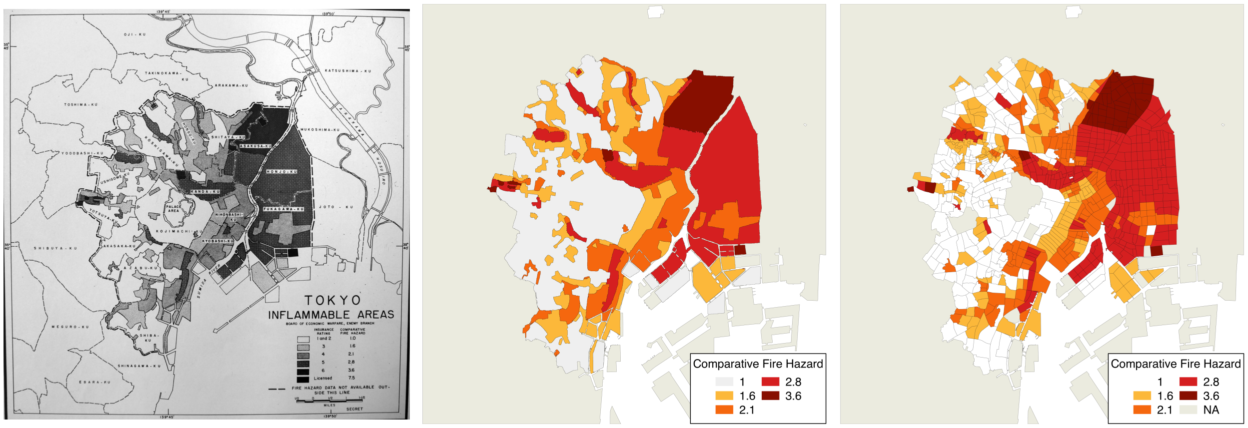
This section explains how we processed the historical image developed by the USAAF to construct the local-level flammability measure of prewar Tokyo for the balance check in Figure 7. The raw historical image, "OSS Map no. 877, Tokyo — Inflammable Areas," was originally developed by the USAAF's Intelligence Section during the 1942–1943 period and obtained from Figure 5 in Fedman and Karacas (2012) for our data construction, which is shown in Figure E.1(a). The image is thus a secondary source material but retains sufficient clarity for further image processing. In this figure, the central area of Tokyo is painted in six different patterns based on the insurance rating from "1 and 2" to "Licensed," and the hazard ratio is calculated for each category setting "1 and 2" as a baseline.[4] The geographic coverage of the flammability data is 15 out of 35 wards of prewar Tokyo, which includes 746 neighborhoods in our census-based data set of contemporary Tokyo.

The image and geoprocessing operations involved several steps.[5] First, as preprocessing, we deleted the letters representing the ward names with the background shades because these letters are a potential source of miscoding while carrying little meaningful information for the coding procedure. Second, we assigned several ground control points (GCPs) to the retouched image and georeferenced the image using the GCPs. Third, we utilized the image segmentation technique to convert the georeferenced image (raster) into a shapefile containing multiple polygons, each representing the regions with distinct insurance ratings.[6] Figure E.1(b) shows the resultant segmented polygons. Finally, we overlaid the contemporary census polygons onto the segmented polygons to measure the flammability scores at the neighborhood level including the two spatially-lagged variables, *min. Neighbor Fire Hazard*

[4] The areas rated as "6" and those rated as "Licensed" are painted too similarly to distinguish each other, so we processed the image regarding the highest rating as "6."

[5] The following tools were used in each operation: Adobe Photoshop CS5.1 for photo retouching, Map Warper (http://mapwarper.net/) for the coding of ground control points, the Geospatial Data Abstraction Library (GDAL) for georeferencing, and the Orfeo Toolbox (Grizonnet, Michel, Poughon et al., 2017) and a series of original R implementations for segmentation and other geoprocessing operations.

[6] We relied on the mean-shift segmentation algorithm for image segmentation.



(a) Original flammability image (b) Segmented polygons (c) Neighborhood-level hazard distribution

Figure E.1: Geoprocessing operations of the flammability data

and *max. Neighbor Fire Hazard*.[7] Figure E.1(c) maps the distribution of the flammability rating overlaid and aggregated at the neighborhood level.

F Classification of crime categories

We created the crime outcome variables in the main analysis based on the classification of crime categories defined by the National Police Agency. The following table describes the detailed crime sub-categories that constitute the five major categories employed as the outcome variables in the empirical analysis.

G Mediation Analysis

An important issue remaining in the main analysis is the role of post-treatment factors in generating the long-run effect of the WWII raid damages. Yet investigations into the role of posttreatment factors and causal pathways generally face a methodological challenge. On the one hand, including post-treatment factors, or the variables affected by the historical treatment and affecting the contemporary outcomes, would result in post-treatment bias

[7]The spatial overlay procedure was executed based on the centroid-coordinates of the neighborhood polygons.

Large Classification	Small Classification	Number of crimes in 2010
Felonious offenses	Robbery	433
	Other felonious offenses	234
Violent offenses	Unlawful assembly with dangerous weapons	1
	Assault	3,397
	Battery	2,479
	Intimidation	258
	Extortion	540
Burglary thefts	Safe-cracking	210
	School break-in	70
	Business office break-in	743
	Store break-in	1,021
	Residence break-in (while absent)	3,221
	Residence break-in (while sleeping)	424
	Residence break-in (while unaware)	202
	Other burglary thefts	698
Non-burglary thefts	Motor vehicle thefts	346
	Motorcycle thefts	4,175
	Bike thefts	40,644
	Vehicle load thefts	6,312
	Vending machine thefts	899
	Construction-site thefts	563
	Pick-pocketing	2,690
	Purse-snatching	1,476
	Bag-lifting	3,308
	Shoplifting	15,035
	Other non-burglary thefts	21,435
Other crimes	Frauds	5,533
	Stealing of lost property	7,474
	Intellectual offenses	997
	Unlawful gambling	28
	Other criminal offenses	16,492

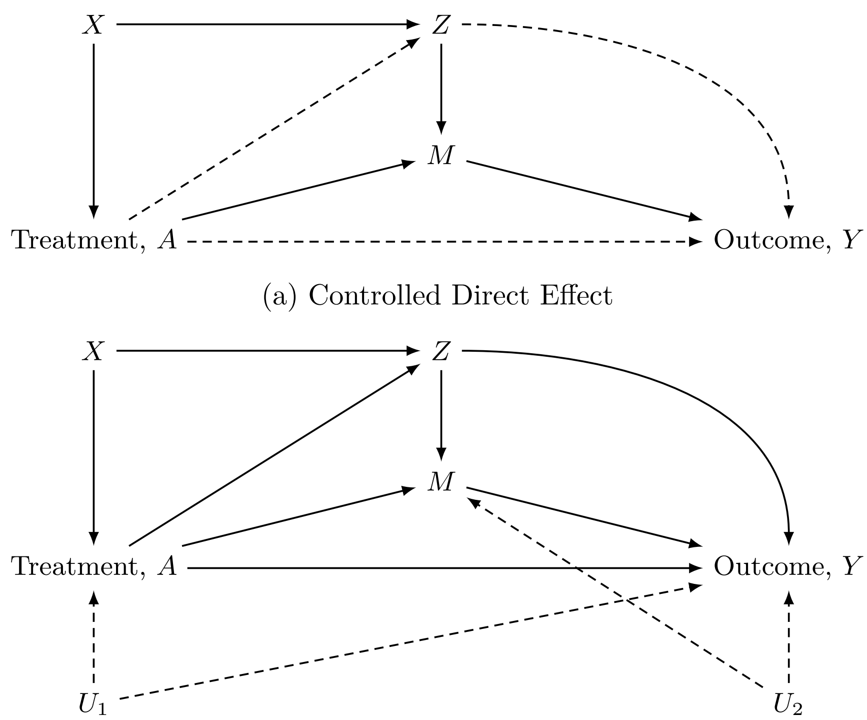
Notes: Translations are based on Police Policy Research Center (N.d.). The numbers of crimes are the total of Tokyo's 23 wards in 2010.

(Rosenbaum, 1984). On the other hand, omitting these factors would result in an indeterminate conclusion on the underlying mechanisms linking the treatment and the contemporary outcomes (Acharya, Blackwell, and Sen, 2016; Imai, Keele, and Yamamoto, 2010; Imai, Keele, Tingley et al., 2011).

G.1 Sequential *g*-estimation

To address the trade-off and decompose the total causal effect, we utilize a two-stage estimator for mediation analysis called the sequential *g*-estimator (Acharya, Blackwell, and Sen, 2016; Joffe and Greene, 2009; Vansteelandt, 2009). A noticeable feature of this two-stage estimator is that it allows for conditioning on intermediate confounders as well as mediators. Intermediate confounders generally refer to the post-treatment variables that are affected by the treatment and confound the mediator-outcome association. Importantly, the presence of intermediate confounders violates the sequential ignorability assumption of the conventional causal mediation analysis (Imai, Keele, and Yamamoto, 2010; Imai, Keele, Tingley et al., 2011). The sequential ignorability assumption involves two separate assumptions: (1) the treatment is ignorable given pretreatment covariates, and (2) the mediator is ignorable given treatment and pretreatment covariates (Imai, Keele, and Yamamoto, 2010, 55). The second condition is violated in the presence of intermediate confounders. However, given the time gap between historical treatment and contemporary outcomes, the no intermediate confounders assumption would be implausible in the current context (Acharya, Blackwell, and Sen, 2016, 526). The sequential *g*-estimator is flexible enough to condition on intermediate confounders and allow us to explore the causal pathways linking the WWII firebombing and the contemporary socioeconomic outcomes in Tokyo.

The quantity of interest in sequential *g*-estimation is the average controlled direct effect (ACDE) of the treatment. The dashed segments in Figure G.1(a) represent the ACDE proportion of the total effect. The ACDE measures the treatment effect with a mediator set to the same fixed value for all units, or the effect *not* mediated or conditioned by the



(a) Controlled Direct Effect

(b) Violation of the Sequential Unconfoundedness Assumption

Figure G.1: Controlled Direct Effect and the Sequential Unconfoundedness Assumption

Note: Acharya, Blackwell, and Sen (2016) Figures 3 and 4. (a) Dashed segments indicate the controlled direct effect. (b) The presence of unobserved confounders U_1 (Treatment $\leftarrow U_1 \rightarrow$ Outcome) and U_2 (Mediator $\leftarrow U_2 \rightarrow$ Outcome) violates the sequential unconfoundedness assumption. In both panels, X, M, and Z denote pretreatment covariates, mediator, and intermediate confounders, respectively.

mediator.[8] The ACDE of treatment A captures the direct effect of changing the treatment values from a to a', with mediator M fixed at some value m,

$$\text{ACDE}(a, a', m) = \mathbb{E}[Y_i(a, m) - Y_i(a', m)], \tag{G.5}$$

where Y_i is an outcome variable. Here, the ACDE measures the *ceteris paribus* average direct effect of *Damage* with the mediator M (specified as population influx or six-storied or higher residences ratio) fixed at a value m for all observations.

Intuitively speaking, the sequential g-estimator transforms ("demediates") the outcome variable by removing the statistical variations that are directly accounted for by the mediator

[8]The following section follows the notation in Acharya, Blackwell, and Sen (2016).

M. It then regresses the transformed outcome variable on the treatment and pretreatment covariates to derive the ACDE. In practice, the estimation involves two stages. We first estimate the effect of mediator M on outcome Y conditioning on treatment A, pretreatment covariates X, and intermediate confounders Z.

$$\mathbb{E}[Y_i|A_i, M_i, X_i, Z_i] = \alpha_0 + \alpha_1 A_i + \alpha_2 M_i + \alpha_3 X_i + \alpha_4 Z_i. \tag{G.6}$$

Here, we derive the consistent estimate of α_2, or the effect of the mediator on the outcome, when M is ignorable given A, X, and Z.

We then subtract the mediator's effect from the observed outcomes to construct counterfactual outcomes as if all observations had the same mediator value,

$$\widetilde{Y}_i = Y_i - \widehat{\alpha}_2 M_i. \tag{G.7}$$

Finally, we derive the ACDE by regressing the transformed outcome variable on the treatment and pretreatment covariates such that

$$\mathbb{E}[\widetilde{Y}_i|A_i, X_i] = \gamma_0 + \gamma_1 A_i + \gamma_2 X_i, \tag{G.8}$$

where γ_1 captures the ACDE. Because the mediation effect has been removed, the remaining covariation between the treatment and demediated outcome is attributable to the direct effect of the treatment *not* through or interacted with the mediator.

Note that the resultant estimate of γ_1 measures the ACDE with the mediator fixed at 0, or ACDE$(a, a', 0)$. Although the ACDE with $m = 0$ (e.g., population influx = 0) is not of interest in the current application, we can derive the estimate with the mediator at which we want to compute the ACDE by simply recentering the observed mediator values. Specifically, in Section 7 of the main text, we recenter the mediator variable, or the ratio of six-storied or higher residences in a neighborhood, to the mean value. The resultant estimate thus measures the ACDE in the counterfactual that the mediator is set to the mean for all neighborhood observations.

G.2 Identification Assumption and Intermediate Confounders

The key identification assumption in sequential g-estimation is the sequential unconfoundedness assumption (Acharya, Blackwell, and Sen, 2016, 519). As represented by the dashed segments in Figure G.1(b), the identification assumption involves two no omitted variables assumptions: (1) there are no omitted variables for the treatment-outcome relationship given pretreatment covariates (absence of U_1), and (2) there exist no omitted variables for the mediator-outcome association given pretreatment covariates, treatment, and intermediate confounders (absence of U_2).

We have already addressed the validity of condition (1) in Sections 2 and 5 of the main text. The largely indiscriminate nature of the US strategic bombing of Tokyo, the careful control for the key components of the treatment assignment mechanisms (i.e., primary aiming points and the Imperial Palace), and the covariate balance broadly confirmed the validity of the first no omitted variables assumption. Condition (2) requires that the mediator and the outcome are conditionally independent given other right-hand side variables in equation (G.6). Although conditioning on intermediate confounders makes the ignorability of the mediator more credible, it is particularly difficult to find a sufficient set of intermediate confounders. To account for this concern, we include the lagged outcome variable as well as other post-treatment variables as intermediate confounders. Other post-treatment variables are also lagged by five or ten years, depending on the availability of the outcome variables.

Combined, the sequential g-estimation reported in Section 7 of the main text includes the following variables as intermediate confounders. To construct the temporally lagged variables, we employ the 2000 full-scale census for the mediation analysis of the socioeconomic outcomes and the 2005 records for the crime count analysis. Note that these variables are log-transformed after adding 0.01 to the original values.

- Lagged outcome variable
- Lagged ratio of foreign population
- Lagged ratio of population aged 15 and younger
- Lagged ratio of population aged 65 and older
- Lagged population density
- Lagged population growth
  - $Population_{2000}/Population_{1995}$ for the socioeconomic outcome analysis
  - $Population_{2005}/Population_{2000}$ for the crime count analysis
- Lagged ratio of single member households
- Lagged ratio of home ownership
- Lagged ratio of apartment houses
- Lagged ratio of workers in tertiary sector
- Lagged growth rate of secondary sector workers
  - $Workers_{2000}/Workers_{1995}$ for the socioeconomic outcome analysis
  - $Workers_{2005}/Workers_{2000}$ for the crime count analysis

As reported in the main text, the socioeconomic outcome variables are taken from the 2010 full-scale census, and crime count outcomes are similarly constructed using the 2010 records for a comparative purpose.[9] The pretreatment covariates, ward fixed effects, and the smoothing terms of neighborhood geocoordinates described in Section 5 are also included in the first- and second-stage estimates. Also note that we rely on the fifth-ordered polynomial (weighted least square) specification of the smoothing terms as in Section 6.3 for computational efficiency. To account for the increased error due to the two-stage estimation procedure, we obtain the standard errors via 5,000 nonparametric bootstraps.

[9] As explained in the main text, the Statistics Bureau of the Ministry of Internal Affairs and Communications provides precisely georeferenced census statistics since the 1995 population census. The Tokyo Crime Prevention Network annually releases the neighborhood-level crime count records in Tokyo since 2004.

H Details of the internet survey

In this appendix, we explain the details of the internet survey. In the following, *cho-chomoku* is referred to as "neighborhood."

H.1 How the survey was performed

Our surveys were conducted from September 25 to October 1, 2018 (first wave) and from December 6 to 10, 2018 (second wave). We performed the survey twice due to financial constraints: since our survey contained economic experiments with compensation, it turned out that the second wave could be conducted after finalizing the cost for the first wave. We recruited the survey respondents living in Tokyo's 23 wards from the registered panel at Rakuten Insight, one of the largest internet survey companies in Japan. We used Survey Monkey as a tool to conduct the internet survey. Since Rakuten Insight only allows the third-party survey tools with the endorsement of their affiliated companies, we used the panel of Rakuten Insight via Survey Research Center Co., LTD, which is another survey company in Japan affiliated with Rakuten Insight.

In order to collect responses from as many neighborhoods as possible, we used the following sampling strategy. Rakuten Insight allows the use of a quota based on 7-digit zip codes. Each zip code contains one or more neighborhoods, and no single neighborhood extends over two or more zip codes. Thus, we first obtained the census polygons in the 2010 census and extracted the information of neighborhood addresses. We then utilized the Google API and neighborhood addresses to assign the corresponding zip codes to individual census polygons.

Then, we determined sampling weights. Rakuten Insight decides the quota of the panels to be invited for the survey for each zip code based on these weights and the target number of the final sample. We decided the weights in proportion to the number of neighborhoods contained in each zip code. For example, the weight of the zip code that contains two neighborhoods is set to twice as much as the zip code that contains only one neighborhood.

The limitation is that since we can only set the weight at the zip code level, we cannot necessarily equalize the final number of respondents at the neighborhood level. Especially, when one zip code contains several neighborhoods and the registered panels are concentrated in a few of these neighborhoods, the respondents in these neighborhoods tend to be over-represented. This unbalance is corrected for using population weight in the estimation.

Exactly how many invitation emails were sent for each zip code is not disclosed because how many respondents are registered to the panel is a business secret. However, Rakuten Insight controls the number of invitations based on past response rates so that the number of complete responses is slightly below the target number of the sample. That is, if one in ten people is expected to participate in and complete the survey, they email the invitation to several times the target number of the sample.

When the number of the responses in a zip code reaches the quota, the other respondents in the zip code can no longer participate in the survey. The respondents include both complete and incomplete responses. Thus, the quota is set with some margin that allows for sample attrition due to incomplete responses. In the case that the number of complete responses does not reach the quota, the invitation emails are sent to the originally selected panels again and to newly selected panels. Once the survey was distributed, we reported the numbers of the total and complete responses every day, and Rakuten Insight controlled how many new invitations would be sent the next day. When the complete responses exceeded the target number of the sample, we reported this to Rakuten Insight and closed the survey by ourselves.

This survey had two questions that practically screened out a quarter of the respondents. First, since this survey asks respondents' addresses, there is a very small possibility for malicious users to identify the personal information of the respondents by combining other data sources. Of course, we will process data anonymization according to the national Census, but when we stated this possibility at the beginning of the survey and asked to agree to proceed, about 7% of the respondents quit the survey. After this question, we explicitly asked their zip code and neighborhood, and then about 16% of the respondents either spontaneously exited by themselves or were eliminated by entering fake information.[10] There were several respondents who dropped out in the middle of the survey. With the two waves of the survey, 3,423 respondents entered the survey, and 2,626 completed responses (response rate = 76.7%). Specifically, the first wave contains 2,667 respondents and 2,062 complete responses, and the second wave contains 756 respondents, and 564 complete responses. Three respondents were dropped from the sample due to failure to merge their reported neighborhoods with the census. Those who finished the survey in less than three minutes were also excluded from the final sample to ensure that the sample only includes respondents with a solid understanding of the instructions of the two economic experiments. These screening processes leave 1,791 respondents in the sample, representing 1,267 neighborhoods.

A response rate of higher than 75% is relatively high. One of the reasons for this high response rates is the maximum amount of the rewards they could get from this survey. The maximum amount of the rewards was 200 Rakuten points plus the baseline rewards for completing the survey, where 1 Rakuten point is roughly equivalent to one Japanese Yen. The amount of the base rewards are not disclosed, but one point per question is probably close to the truth guessing from other surveys conducted in Rakuten Insight. The respondents could receive varying rewards from 0 to 200 points depending on their responses in the two economic experiments, the dictator game and the ultimatum game. Since this survey contained about 30 questions, the maximum rewards would be about 230 points, and "max 230" (or a similar value) was displayed in the respondents' personal webpage when they logged in. Some readers may think this amount is too low given the prices of Japan. This would be the case if we recruited the respondents from outside of the panel. However, for those who are registering at Rakuten Insight, the 230-point offer is relatively very attractive because most other surveys pay only the baseline rewards.

[10] We regard an entry as fake information if a respondent entered a zip code that does not exist or a number of *chome* that does not exist for a zip code entered by the respondent.

H.2 The economic experiments and the preliminary online survey

Our survey contains various questions, including direct questions about levels of social trust, political participation, and hostilities toward the allies in WWII, as well as the respondents' socioeconomic status. However, the most important questions among them are the two kinds of economic experiments, namely the dictator game and the ultimatum game. Dictator game and ultimatum game are among the most popular economic experiments to measure altruism. In the dictator game, the proposer divides a fixed amount of goods with the receiver who has no option but to accept, while in the ultimatum game, the receiver has an option to decline the offer. In such a case, both players receive no reward.

In the dictator game, a proposer's optimal strategy is to take all rewards for him or herself, and this is the Nash equilibrium for this game. In the ultimatum game, the receiver's reward from declining the offer never exceeds the counterpart from accepting the offer regardless of the amount of the proposal. Therefore, accepting any offer weakly dominates declining the offer. Thus, the Nash equilibrium for the ultimatum game is the same as the dictator game. That is, a proposer proposes to take all rewards, and a receiver accepts the offer.

These two games are, however, different in that some proposers consider the possibility of retaliation by the receiver in the ultimatum game. Indeed, some players are known to punish the proposers if they offer an unfair proposal. This characterizes the dictator game as an experiment measuring unconditional altruism and the ultimatum game as an experiment measuring conditional altruism, or reciprocity (Axelrod, 1984).

In the two economic experiments, the survey respondents played the role of a proposer, so who was the receiver? In the internet survey setting it was difficult for the respondents to interact with another person. We simulated the interaction of the economic experiments by preparing the set of responses by the receivers from another preliminary survey. For the two economic experiments, a proposer could propose one's own reward from 0 to 100 in the increment of 5. That is, there were 21 patterns of the proposal. In the dictator game, receivers do not play any role, so we simply asked Question 29 to the proposers and told them someone in another survey would receive the amount they chose. The proposers received the proposed amounts of their own share as promised.

For the ultimatum game, a proposer can propose the same 21 patterns of their own share, and the receiver's response must be prepared for each of 21 cases. In the preliminary survey, we asked respondents who played the role of receiver the following question: "In a separately conducted survey, you are selected as a participant of the game to divide 100 Rakuten points with an opponent. If you are offered the proposal that you receive ( ) points, and your opponent receives 100 – ( ) points, will you agree with the proposal? Note that if you agree, you will get ( ) and your opponent will get 100 – ( ) points, while if you decline, both will get 0 points." The different amount of offer from 0 to 100 in the increment of 5 is entered in ( ). Thus, 21 versions of this question were created. We collected about 45 responses so that about two respondents answered each type of the question in the case that the first respondent did not complete the survey. For each version of the survey, we adopt the first valid response as the receiver's response corresponding to the specific amount of offer. However, when the amounts of the offer were 25, 30, 45, 70, 80 and 85 points, we could not obtain the valid responses from the sample due to a system glitch.[11] Thus, we set the response of these offers to "accept." This procedure was explained in Question 30 before the proposer's making a decision.

H.3 The questionnaire for the online survey

1. Please read the following instruction of the survey and proceed if you agree with it.

[**Instruction**] This survey is conducted to the monitors of the Rakuten Insight (formerly known as the Rakuten Research) who live in Tokyo's 23 wards to study the impacts of the neighborhood characteristics on the social and political attitudes of ordinary citizens. The participation in this survey is voluntary, and you

[11] One respondent seems to notice the amount of rewards offered in the ultimatum game increases on each response and repeated the survey multiple times without completing it. We set the survey tool so that each respondent could take the survey only once, but this respondent somehow escaped from this restriction and repeated the survey until he or she got the desired proposal. We asked Rakuten Insight, but they could not identify the problem either. As a result, several surveys with the different amount of the receiver's offer were skipped.

can terminate the survey whenever you want. This survey is implemented by the principal investigator (PI), [name deleted for anonymous review] with the approval of the IRB board of [name deleted for anonymous review] and is made possible by [name deleted for anonymous review]. Those who take this survey will receive the Rakuten Points based on the number of questions answered. In addition, the final points you earn will vary by a random factor and your responses. The results of this survey and the study based on this survey will be made public on the PI's personal website ([name deleted for anonymous review]) after the statistical processing in such a way that your personal information is not exposed. This survey does not contain any question that makes you reveal your personal information. However, this survey collects some information that helps identify individuals by merging a separately conducted survey that is irrelevant to this survey. Therefore, the use of individual data is limited only when it is necessary to make the data public to ensure the reproducibility of this study after either the data anonymization according to the national census or the deletion of the relevant variables so that it is impossible to identify personal information.

(a) Agree and proceed; (b) Do not agree.

2. This survey investigates how your living place affects your sociopolitical attitudes. First, please enter your 7-digit zip code. Your personal information will not be disclosed by answering this question. (Specifically, the researcher of this survey will never use the data to identify the individuals. If necessary, we will process the data to secure anonymity following the procedure of the national census.) [∗ When the zip code contains more than one Chō-chōmoku, the respondent jumps to one of the questions from 3 to 10 according the number of Chō-chōmokus contained in the zip code.]

3. Which is your *Chōme*? Please choose from the following.

(a) 1 *Chōme*; (b) 2 *Chōme*; (c) Do not answer

4. ~9. ∗ Omitted.

10. Which is your *Chōme*? Please choose from the following.

(a) 1 *Chōme*; (b) 2 *Chōme*; (c) 3 *Chōme*; (d) 4 *Chōme*; (e) 5 *Chōme*; (f) 6 *Chōme*; (g) 7 *Chōme*; (h) 8 *Chōme*; (i) 9 *Chōme*; (j) Do not answer

11. If you forget your zip code, please write the name of the chō-chōmoku you live.

12. Generally speaking, do you think most people can be trusted?
(a) Strongly disagree; (b) Disagree; (c) Somewhat disagree; (d) Neutral; (e) Somewhat agree; (f) Agree; (g) Strongly agree

13. In the past year or so, how many friends asked you for a consultation outside work?

14. How often do you lend your money or belongings to your friends when asked?
(a) Almost always lend; (b) Usually lend; (c) Neutral; (d) Usually not lend; (e) Almost always not lend

15. What is your opinion about donating your organs after your death?
(a) Already expressed my intention to donate in the driver's license.
(b) Already expressed my intention to donate in a form other than the driver's license.
(c) Planning to express my intention to donate.
(d) Not planning to express my intention to donate.
(e) Other.

16. Which of the following forms of donation have you ever done for the disaster relief or social problem? Please choose all that apply.
(a) Financial aid (including Furusato Nōzei tax exemption);
(b) Material aid; (c) Volunteer; (d) None of the above

17. How much do you trust the following actor or country? [∗ This question is repeated for each of (i) the U.S., (ii) Donald Trump (President of the U.S.), (iii) Russia, (iv) Vladimir Putin (President of Russia), (v) Japan, (vi) Shinzo Abe (Prime minister of Japan).]
(a) Trust; (b) Somewhat trust; (c) Neutral; (d) Somewhat not trust; (e) Not trust

18. Which of the following organizations, groups and circles are you affiliated with? Please choose all that apply.
(a) Political organization; (b) Industry organization; (c) Volunteer organization; (d) Civic movement organization; (e) Religious organization; (f) Sport organization; (g) Hobby organization; (h) Consumers' cooperative society (Co-op); (i) None of the above

19. Is your household a member of a neighborhood association?
(a) Yes; (b) No; (c) Do not know

20. It is not unusual for not being able to go to vote for various reasons, but did you go to vote in the general election held on October 22, 2017? (Cf. The voter turnout of this election was 53.64%)
(a) Voted on the election day; (b) Casted an early vote; (c) Did not vote (Abstained); (d) DK/NA

21. How much do you support the use of force or its attempt as a mean of diplomatic solution?
(a) Support; (b) Somewhat support; (c) Neutral; (d) Somewhat not support; (e) Not support;
22. How much do you support the US's attack to the Syrian chemical weapons plant in April, 2018 with UK and France?
(a) Support; (b) Somewhat support; (c) Neutral; (d) Somewhat not support; (e) Not support

23. If given two alternatives, which would you choose? A: Sure to take 4000 yen, B: Get 10000 yen only if you get an odd number on the cast of a dice

(a) Choose A → Q.24; (b) Choose B → Q.25

24. If given two alternatives, which would you choose? A: Sure to take 3000 yen, B: Get 10000 yen only if you get an odd number on the cast of a dice

(a) Choose A; (b) Choose B

25. If given two alternatives, which would you choose? A: Sure to take 5000 yen, B: Get 10000 yen only if you get an odd number on the cast of a dice

(a) Choose A; (b) Choose B

26. When the Pacific War occurred, did you or your family live in Tokyo's 23 wards?

(a) Yes; (b) No; (c) DK/NA

27. What kind of physical or mental damages did you or your family suffer during the Pacific War? Please choose all that apply.

(a) No damage

(b) Family members were killed or disabled due to the air raids

(c) Family members were killed or disabled due to the combat

(d) Family members were killed or disabled due to another reason

(e) DK/NA

28. What kind of economic damages did you or your family suffer during the Pacific War? Please choose all that apply.

(a) No damage

(b) Lost own home due to the air raids

(c) Family business suffered the damage due to the air raids

(d) Assets were condemned by the Allies after the war

(e) Lost assets for other reasonss

(f) DK/NA

∗ The order of the questions 29 and 30 are randomized.

29. Please allocate 100 Rakuten Point between you and an opponent, and decide your share by adjusting the slide bar below. For example, currently the bar is set to 0, so you will receive 0 point and the opponent will receive 100 points. Please read the following if you are interested in the detail. [**Detail**] The opponent's share is calculated as the remaining balance of 100 Rakuten Point after deduction of your share. The opponent's share will be allocated to a randomly chosen respondent of this survey. You will receive the opponent's share from someone, too. Your personal information will not be exposed to other people by answering this question. You will earn the Rakuten Point you get from this question. The total point will be notified at a later date.

30. Please allocate 100 Rakuten Point between you and an opponent, and propose your share by adjusting the slide bar below. In the separately conducted survey, if an opponent agrees with the same proposal as you make, you will receive your proposed share. However, if an opponent declines your proposal, both will receive 0 point. For example, currently the bar is set to 0, so you will propose that you will receive 0 point and the stranger will receive 100 points. If the opponent agrees with your proposal, you will receive 0 point and the stranger will receive 100 points, but if s/he declines both will receive 0 point. Please read the following if you are interested in the detail. [**Detail**] Whether your opponent agrees with your proposal is determined from the result of the separately conducted preliminary survey. Specifically, we ask the following question to the 21 respondents randomly drawn from Tokyo's 23 wards in this preliminary survey: "Question: You are selected as an opponent who receive a proposal of allocating 100 Rakuten Point. If you are offered to receive [NUMBER]points, will you accept this offer? If you accept, you will receive [NUMBER]points, and your opponent will receive [100 - NUMBER]points. If you decline, both you and your opponent will receive no point." [NUMBER] in the above question varies from 0 to 100 in the increment of 5, and 21 patterns of the above question are generated, and each version is allocated to a respondent. For example, if you adjust the slide bar to 0 point, you will make an offer that your opponent will receive 100 points. Therefore, the points you and your opponent receive will be determined based on whether one of 21

respondents whose [point] in the above question is set to 100 accepts this offer. Your personal information will not be exposed to other people by answering this question. You will earn the Rakuten Point you get from this question. The total point will be notified at a later date.

31. What is your gender?

(a) male; (b) female

32. What is your age? Please select by adjusting the slide bar below.

33. What is the last school you enrolled (or are currently enrolling)?

(a) Junior high school; (b) High school; (c) Two year schools after high school; (d) College and university; (e) Graduate school

34. What is your occupation? Please select the closest one.

(a) Office-worker; (b) Self-employed; (c) Family business; (d) Student; (e) Housewives or househusbands; (f) Unemployed; (g) Others

35. What is your household income? Please select the closest one.

(a) Less than 2 million yen; (b) 2 ~4 million yen; (c) 4 ~6 million yen; (d) 6 ~8 million yen; (e) 8 ~10 million yen; (f) 10 ~15 million yen; (g) 15 ~20 million yen; (h) Over 20 million yen

36. What is your current marital status? Please select the closest one.

(a) Never married; (b) Divorced or widowed; (c) Married

37. How many family members are you living together including you? Please select using the adjustment bar below [Max. is set to 10].

38. How long have you lived in the current municipality?

(a) Since my birth; (b) Less than 1 year; (c) 1 ~3 years; (d) 3 ~5 years; (e) 5 ~10 years; (f) 10 ~20 years; (g) 20 ~30 years; (h) Over 30 years

References

Acharya, Avidit, Matthew Blackwell, and Maya Sen. 2016. Explaining causal findings without bias: Detecting and assessing direct effects. *American Political Science Review* 110 (3):512–529.

Axelrod, Robert. 1984. *The evolution of cooperation.* New York: Basic Book, Inc.

Daiichi Fukuinshō(ed.). 1983. *Nihon toshi sensai chizu [The maps of war-damaged cities in Japan].* Tokyo: Hara Shobō.

Fedman, David, and Cary Karacas. 2012. A cartographic fade to black: Mapping the destruction of urban Japan during World War II. *Journal of Historical Geography* 38 (3):306–328.

Grizonnet, Manuel, Julien Michel, Victor Poughon, Jordi Inglada, Mickaël Savinaud, and Rémi Cresson. 2017. Orfeo ToolBox: Open source processing of remote sensing images. *Open Geospatial Data, Software and Standards* 2 (1):15.

Imai, Kosuke, Luke Keele, Dustin Tingley, and Teppei Yamamoto. 2011. Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies. *American Political Science Review* 105 (4):765–789.

Imai, Kosuke, Luke Keele, and Teppei Yamamoto. 2010. Identification, Inference and Sensitivity Analysis for Causal Mediation Effects. *Statistical Science* 25 (1):51–71.

Joffe, Marshall M., and Tom Greene. 2009. Related causal frameworks for surrogate outcomes. *Biometrics* 65 (2):530–538.

National Archives of Japan Digital Archive. N.d. Zenkoku shuyō toshi sensai gaikyōzu [The brief maps for the war-damages of the major cities in Japan]. https://www.digital.archives.go.jp/DAS/pickup/view/category/categoryArchives/0200000000/0203000000/00. [Online; accessed 1-March-2019].

Nihon Chizu Kabushiki-gaisha: Reprinted by Tōkyō Kūshū wo Kirokusuru Kai. 1985. *Konsaisu Tōkyō-to 35-ku kubun chizuchō: sensai shōshitsu kuiki hyōji [Concise map-book for Tokyo's 35 wards showing the burnt areas due to the war-damage].* Tokyo: Nitchi Shuppan.

Police Policy Research Center. N.d. The list of English translations of police terminology

[Keisatsu Yougo Eiyaku Ichiran]. *Tech. rep.*

Rosenbaum, Paul R. 1984. The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society. Series A* 147 (5):656–666.

Saotome, Katsumoto. 1981. *Nihon No Kūshū 10 Hokan Shiryō-hen (Air Raid in Japan, v. 10: Data)*. Tokyo: Sanseido.

The Geospatial Information Authority of Japan. N.d. Chizu Kūchū Shashin Etsuran Sābisu (The Browsing Service for the Maps and Aerial Photos). http://mapps.gsi.go.jp/maplibSearch.do#1. [Online; accessed 27-February-2019].

Tokyo prefecture(ed.). 2005. *Tokyo-to sensai shi [Report of the war-damage of Tokyo prefecture]*. Tokyo: Meigensha.

Vansteelandt, Stijn. 2009. Estimating direct effects in cohort and case-control studies. *Epidemiology* 20 (6):851–860.