Look at Your *Old* Men Dying: Long-Run Effect of Civil War Experience on Mortality

Mikko Myrskylä and Torsten Santavirta*

March 4, 2024

Abstract

We study the long-run effect of early adulthood political crime court sentence on mortality using an examiner design and data on 6,961 prisoners of war during the Finnish Civil War of 1918. Our descriptive analysis shows a statistically significant adverse association and a dosage response between length of sentence to prison terms and the loss of citizen rights and old-age mortality. Our instrumental variable design exploits the variation in sentence tendency across quasi-randomly assigned judges. We document an effect of sentence length on old-age mortality and a dosage response that is six times larger than the estimated associations. Our causal estimates have a local average treatment effect interpretation for the compliers.

JEL Classification: D74, J14, N34, N44
Keywords: Civil war, Ageing, Mortality, Examiner design

*Myrskylä: Max Planck Institute for Demographic Research, myrskyla@demogr.mpg.de. Santavirta (corresponding author): University of Helsinki, P.O. Box 17 (Arkadiankatu 7). FI-00014 University of Helsinki. Finland, torsten.santavirta@helsinki.fi. Acknowledgements: We are grateful for the invaluable research assistance provided by Ilkka Jokipii, Virva Liski and Sanni Saarinen. We thank Joe Doyle, Martti Kaila, Giovanni Mastrobuoni, Matti Mitrunen, Jan Stuhler and Jim Sullivan for insightful comments. Both authors acknowledge financial support from Kone Foundation (grant number 80-37727).

1 Introduction

Civil wars and internal conflicts are arguably more harmful to their victims and involved societies than any other social phenomenon (Garfield, 2008; Blattman and Miguel, 2010; de Groot et al., 2022). They are associated with elevated morbidity and mortality among their veterans long after the end of the armed conflict (Pizarro, Silver and Prause, 2006; Costa, 2012; Costa, Yetter and DeSomer, 2018). The recovery phase of civil wars is often associated with retributions, community rejection, and investments in restoration of order and demonstration of state capacity through punishments for political crimes, which may themselves be adversely associated with life outcomes (McKenzie, 2006; De los Rios Hernández, 2023). Understanding the legacy of civil war is important for designing effective post-conflict recovery. Though the effects of civil war on human capital (e.g., through foregone education, malnutrition, and prison camp exposure) are thought to be more persistent than those on economic legacy (physical capital and investment), much of the existing research has focused on the latter (Goldin and Lewis, 1975; Hutchinson and Margo, 2006; Blattman and Miguel, 2010; Miguel and Roland, 2011; Riaño and Caicedo, 2024).

Three major challenges have hindered empirical advances on the long-run health consequences of civil war veterans. First, data generated in conflict zones are rarely systematically collected and properly stored in the archives of a country's statistical agency. Even long after the conflict, when conditions have begun to stabilize, the authorities responsible for data collection tend to revert to peacetime recording rather than retrospectively documenting the tragedy and individual narratives through, for example, representative surveys (Blattman and Miguel, 2010). Second, contrary to warfare more generally, enlistment to either (or one of many) combatant party in a civil war is, by definition, not universal, but rather highly selective. Thus potential confounding to exposure is not mitigated in such a way as universal enlistment of entire cohorts may be in settings such as World War 2 (WW2) (Ichino and Winter-Ebmer, 2004; Akbulut-Yuksel, 2014; Kesternich et al., 2014; Braun and Stuhler, 2023). For this reason, comparing the outcomes of civil war veterans and non-veterans call for particularly rigorous research designs in identifying causal effects of exposure. Third, biological processes imply that effects of events such as stress, malnutrition, and disease exposure (and in the prisoner of war context, extreme starvation), could be delayed and appear with a lag of as much as several decades (Pope and Wimmer, 1998; Abramitzky, Boustan and Eriksson, 2014; Black et al., 2015; Aizer et al., 2016; Meriläinen, Mitrunen and Virkola, 2022). Establishing relationships between such conditions and events in early adulthood and late-age outcomes require longitudinal data sets that span many years. All three challenges complicate the study of the causal effects of civil war.

In this paper, we examine the causal effect of sentences to prisoner of war (POW) camps and loss of citizen rights on old-age mortality using data on veterans of the socialist Red Guard who faced internment in the aftermath of the Finnish Civil War. The human catastrophe that unfolded in the POW camps – where prisoners were, according to the important bourgeois industrialist Mr Hjalmar Linder, "dying like flies," a situation he predicted "[...] will not be defendable in the future" (Linder, 1918, reprinted in Nya Argus, 2011) – was of similar proportions to some of the more systematic internment camp institutions of the 20th century, or even the US Civil War POW camps (average camp mortality was 13.5% during roughly four months)(Upton, 1980; Paavolainen, 1971; Tepora and Roselius, 2014; Costa, 2012). We ascertain survivors from complete POW camp archives, follow up all ex-POWs who survived until age 75, and exploit a policy that generated exogenous variation in prison sentence length; namely, the quasi-randomization of examiners in the special courts set up to deal with offences committed during the war. We document a large adverse causal effect and a statistically significant dosage response of prison sentence length on mortality among the POWs who survived to age 75 (hereafter 75+ mortality). Our estimated causal effect of sentence length on old-age mortality is six times larger than the estimated associations. This adverse causal effect seems to be driven by biological rather than socioeconomic processes. We show that a local average treatment effect interpretation can be placed on our causal estimates.

The literature documenting patterns of health problems among ageing civil war veterans and ex-POWs converges in suggesting long-lasting associations of war with health among the exposed (Dent et al., 1989; Page and Brass, 2001; Pizarro, Silver and Prause, 2006; Costa, 2012). Longitudinal data from the medical records of ex-POWs shows evidence that the adverse effects of the extreme health insult manifest themselves with a lag of up to 50 years, with POWs demonstrating a survival advantage in the first three decades after exposure, possibly because of selective mortality (Page and Brass, 2001). While there is evidence of a biological basis (through, e.g., extreme starvation) to this excess old-age mortality (Page and Ostfeld, 1994), social processes transmitted through, e.g., truncated socioeconomic attainment, community rejection, or stigma, may also be relevant (for elaboration of potential social pathways, see for example Daza, Palloni and Jones (2020)). Indeed, the main shortcoming of the current literature is that no causal effects have been identified, but rather only associations adjusting the comparisons of outcomes in POWs versus non-POWs for age, race, rank (and sometimes socioeconomic characteristics (Costa, 2012)).

Our study offers the first causal evidence on the health consequences of POW exposure exploiting micro-level data and quasi-experimental variation in exposure. This is a significant contribution to the existing body of work, perhaps as meaningful as the introduction, more than three decades ago, of experimental methods to study the consequences of Vietnam veteran status on mortality and labor market outcomes (Hearst, Newman and Hulley, 1986; Angrist, 1990) and the human capital consequences of child soldiering in the developing country context of Uganda (Blattman and Annan, 2010). To add to this, our study contributes to the literature on the long-run health effects of extreme starvation by providing a local average treatment effect of exposure and in particular by documenting a dosage response. We are aware of only one study before ours, namely van den Berg, Pinger and Schoch (2016), that uses an instrumental variable design to estimate the local average treatment effect of famine on later life health, for which the authors document four times larger estimates as compared to their estimated associations. Specifically, van den Berg, Pinger and Schoch (2016) highlight the importance of scaling estimates by compliance even in the case of nation-wide famine episodes.

To our knowledge, ours is also the first study to exploit an examiner design in a political offence court context. Examiner designs have meanwhile been used to estimate local average treatment effects in other settings, for example in work examining the consequences of foster care (Doyle, 2007), incarceration (Aizer and Doyle, 2015), pre-trial detention (Dobbie, Goldin and Yang, 2018), and disability insurance receipt (Dahl, Kostøl and Mogstad, 2014).[1]

[1]See Table 1 of Frandsen, Lefgren and Leslie (2023) for a comprehensive list of studies using an examiner design.

2 Historical Background

The Finnish Civil War was fought in 1918 between January 27 and May 15 (only one year into the independence of the country from Russia), between the revolutionary Red Guards formed by working class socialists and the government lead White Army the backbone of which was the volunteers of the conservative White Civil Guards. The level of confrontation between the socialists and the middle classes was mounting up in the turmoil of the independence process and the February Revolution in Russia in 1917, eventually escalating into violent confrontations late in 1917 and early January 1918. The war began as a socialist revolution in Helsinki and with simultaneous actions taken by the Whites in Ostrobotnia on the western coast and quickly permeated the southern half of the country into a full scale Civil War. After some three and a half months of bloody battles, harvesting some 11,000 lives, the last Red Guard troops surrendered and victory was declared by the White Army.

2.1 Finnish POW Camps and the Political Offence Court

After the war, some 80,000 surviving Red Guards were held in the roughly 60 prison camps that had been set up set up across the country exposing them to aggressive retributions. Some 9,000 captured Reds lost their lives in arbitrary killings in the aftermath of the war mainly carried out by local members of the White Civil Guard organization. As these retributions (known as the White Terror) started weakening the credibility of the government, it responded by setting up a Political Offence Court system to deal with offenses committed by Red Guards, only, during the armed conflict. So called *offences against the state* were tried in a total of 141 local political offence courts (mostly located in connection to the POW camps), that consisted of five members each, two of whom (chairman and first member) had to have formal legal training, one being a military representative and the remaining two being mostly teachers, civil servants or entrepreneurs (Appendix Table C.2 provides the occupational distribution of each of the five members across courts).[2] The courts would "decide according to their own conviction what is the truth of the matter" and verdicts of these courts could not be appealed (Upton, 1980). The courts had to deal with on average 6.9 cases per day in the on average 96 days during which the they remained operational (Leinberg, 1923). The composition of the two members with formal legal training changed on average once during those 96 days (by either the identity of the chairman or the first member changing).

The random assignment of judge, or in our case judge-pair as we consider the chairman-first member combination, is generated by the random order of processing cases (and random date of changing judge-pair composition). Given the short time span and the massive congestion of cases, it is unlikely that courts would know the date on which a particular case would be tried, suggesting that changes of judge-pair compositions were random to the individual POW. Official documents show that the foremost reason for the heavy workload and tight schedule of the courts emanated from the tens of thousands of POWs whose internment was lacking legal legitimacy in the absence of legal trial (Upton, 1980). Anecdotal evidence suggests that the courts were working under extreme pressure, the maximum number of cases handled per day in one court being as high as 44 (Leinberg, 1923). All the POWs accused for participating in the

[2]Originally, 145 courts were established but court numbers 54, 103, 104 and 116 never became operational. Further, court number 133 only sentenced to youth detention centers and dismissals from office and court number 145 was ambulating and lacks the same level of documentation as the rest. These two courts are excluded from our data.

Red Guards would be investigated and when seen justified, brought to trial. Of all roughly 76,000 accused some 68,000 were sentenced to prison (μ=4 years) and 554 were sentenced to death of which 265 were executed. The sentenced individuals (bar 61 cases) also lost their rights of citizenship (Leinberg, 1923). Eventually, shorter sentences were suspended and the longer ones reduced by various amnesties between 1918-1921, rendering the prison spells essentially to the on average 166 days of prison camp internment, a punishment that for as many as roughly 12,000 POWs came to cost his or her life.[3] Over the period of four months during which the 20 main camps remained operational, some 11,800 prisoners died of malnutrition or disease spread (e.g., Spanish flu or dysentery) in these overcrowded, unsanitary camps. Mortality is estimated to have been as high as 31% in the most notorious and one of the largest camps called *Tammisaari* (Paavolainen, 1971).[4] Mostly, the POWs' death cause was reported generically as "death on camp" (41%), whereas "death due to disease" accounted for 23% of the cases and "starvation" for 4% (authors' own calculation base on the War Victims 1914-1922 database (National Archives of Finland, 2022). Historical epidemiologic work suggests that the Spanish flu was the (primary or secondary) death cause of roughly 4,500 POWs (Mäkelä, 2007).

An unusual feature of the Finnish Civil War, reflected in our data, is that 7.5% of all those who joined the Red Guard were women, roughly 5,500 in total (Leinberg, 1923). It is of course a matter of definition as to the extent that women (be it

[3]Authors' own calculation based on the War Victims 1914-1922 database made available by The National Archives of Finland (National Archives of Finland, 2022). The average prison camp duration (μ=165.7) is based on the date of internment and the date of release of each individual in the study sample. See Appendix Table A.2 for a timeline of the amnesties.

[4]Authors' own calculation suggest a relative death toll in Tammisaari of 35%, i.e., even higher than the one reported by Paavolainen (1971). This calculation is based on the documented 3,023 POW casualties at the camp and the maximum number of prisoners at the camp, 8,597 (National Archives of Finland, 2022).

unintentionally) participate in wars. In our context however, 222 of the total 1,003 women in our sample are reported to have worn arms. Even though only very few of them (n=14) are reported to have fought in battles, all 1,003 were imprisoned and their cases of participation in the insurrection were tried in court.

3 Data

Our sample of members of the Red Guard was constructed by linking two data sources: a registry of compensation claims by former members of the Red Guard combined to the complete archive of individual-level prosecution acts of the Political Offence Court dating back to 1918.

In 1973 the POWs of the Red Guard were rehabilitated by the Finnish Government. Everyone who was interned at a prison camp in 1918 and prosecuted by the Political Offence Court was entitled to a compensation (ranging from between roughly 1,300 Euros in 2023 currency to roughly 3,200 Euros depending on the duration of POW internment). The base population of the Red Guard data set is a registry stored at the National Archives of Finland containing all 12,000 filed compensation claims in 1973 to the Ministry of Social Affairs. After screening, roughly 11,000 claims were approved and the alleged POW duration tested by verifying the camp exposure from the White Army's prison administration archives and the registry of the Political Offence Court (both stored in the National Archives of Finland). We linked the registry of compensation claims manually based on first names, surnames, birth date and birth place to the registry of Political Offence Court Acts in which all individual acts of the prosecutions in 1918 and 1919 of Red Guards are included and to the complete archive of prison camp records. In total 7,939 successful linkages were made. From these linked acts, all case- and individual-level information available was collected.

All 7,939 linked acts were brought to the Population Register to identify the social security number, again based on first names, surnames, birth date and birth place and, whenever observed in the compensation claim, self-reported social security number. Based on identified social security numbers of 7,907 individuals at the Population Register of Finland and after removing 209 duplicates, we successfully linked these data to administrative records at Statistics Finland, including the death cause registry up until 2018 (recording date of death of the complete Finnish population) at Statistics Finland for a sample of 7,502 individuals. After excluding those who died before age 75 and the individuals who were sentenced to death or lifetime in prison we end up with a study sample of 6,961 of which 5,958 were men and 1,003 were women. We complement our data with available aggregate-level data on the population of all Red Guard POWs (Leinberg, 1923). Appendix Table A.1 presents the summary statistics.

4 Experimental Setting and Research Design

4.1 Empirical Model and Instrument

For ex-POW i, consider a model that relates 75+ mortality measured as age in years at death (based on number of days at exact date of death) to number of years sentenced to prison by the political offence court c and judge pair j:

$$(\textit{Age at death})_{icj} = \beta_0 + \beta_1(\textit{Sentence length})_{icj} + \beta_2 \mathbf{X}_{icj} + \varepsilon_{icj}, \tag{1}$$

where $\mathbf{X}_{icj}$ is a vector of case- and individual-level predetermined control variables (see Table 1 for the complete set), and ε_{icj} is an error term. In principle, equation (1) could be estimated with ordinary least squares (OLS) but the key problem for inference is that parameter β_1^{OLS} is likely to be biased by the correlation between sentence length and unobserved individual-level characteristics that themselves are correlated with mortality. For example, higher-rank positions tend to be filled by individuals with better human capital attributes who are more apt to shoulder leadership duties.[5] The methodological concern then becomes that judges may be more likely to pass long prison sentences to highly ranked officers who may on average live longer than lower-ranked guardsmen and -women in the counterfactual state of no sentences. At its most extreme, OLS estimates may in this scenario be biased towards a finding that a harsher sentence prolonged longevity. To address this issue, we estimate the causal impact of sentence length using judge sentence tendency of a quasi-randomly-assigned judge pair as an instrumental variable (IV) for sentence length.

We estimate judge-pair sentence tendency for each case in our data by taking the average sentence length in all other cases the assigned judge pair has ever handled, adjusted for court effects. As we show in Figure 1, some judge pairs are systematically more stringent than others within the same court which generates exogenous variation in the sentence length. We perform two-stage least squares (2SLS) with equation (1) as the second stage and the following regression-based analogue to Figure 1 as the first stage:

$$(\mathit{Sentence\ length})_{icj} = \gamma_0 + \gamma_1(\mathit{Sentence\ tendency})_{icj} + \gamma_2\mathbf{X}_{icj} + \epsilon_{icj}, \qquad (2)$$

with the goal of consistently estimating the parameter β_1. The instrument $\mathit{Sentence\ tendency}_{icj}$ is the mean sentence length in all other cases (leaving one, i.e.,

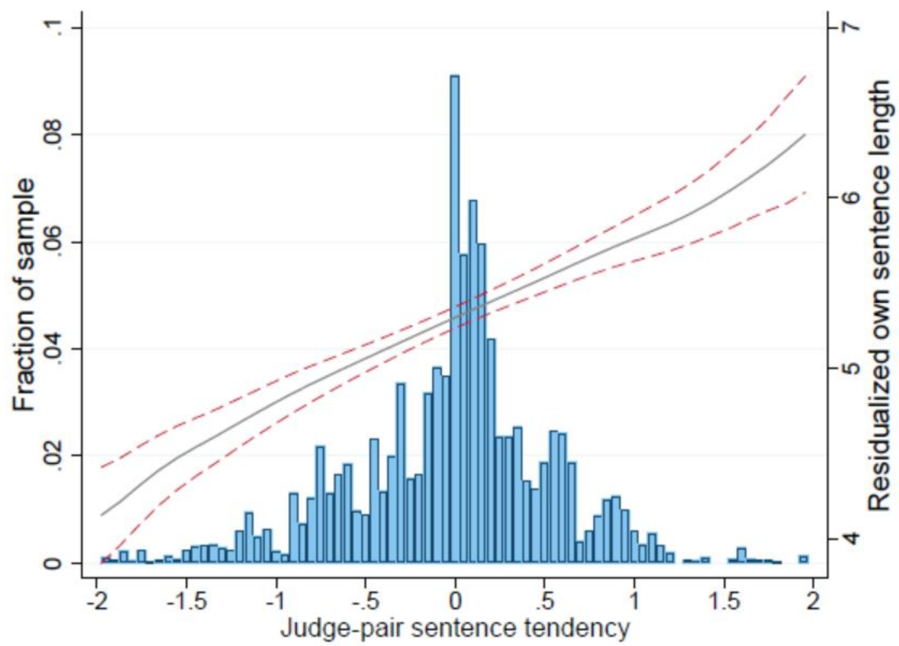
[5]See (Lee, 1999) for a careful empirical exploration of how selective assignment of missions in the Union Army during the US Civil War indirectly affected mortality while in service.

individual i's sentence length, out) handled by the judge pair j that also handled the case of individual i (Appendix B provides the details for the construction of judge-pairs' *Sentence tendency*$_{icj}$). We think of β_1^{2SLS} in equation (1) as the causal effect of prison sentence length on 75+ mortality, the sentence capturing, apart from appalling prison terms, various deleterious forms of repression (prison sentences were paired with loss of rights of citizenship and, following the release, by community rejection, oppression by the state security forces and government agencies in general (Upton, 1980).

4.2 Instrument Relevance

Figure 1 provides a graphical representation of equation (2) by plotting a flexible local linear regression of residualized individual sentence length against residualized judge-pair sentence tendency (each residualized by removing court fixed effects), overlaid over the distribution of the latter (See Appendix Table C.1 for the raw sentence distribution for the study sample and for the complete population of accused). The individual sentence length is monotonically and approximately linearly increasing in judge leniency (solid gray line). A 10 percentage point increase in the judge pair's average sentence tendency is associated with an approximately 2.2 month longer sentence for the left out individual. Panel C of Table 2 presents formal first stage estimates of γ_1 of equation (2). The histogram in the background reveals a wide spread in sentence tendency. Controlling for court-specific effects, the residualized judge-pair sentence tendency (leaving one out) measures -0.76 at the 10th percentile and 0.94 at the 90th percentile, implying a more than one year longer sentence by moving from the most lenient to the most stringent judges (one year being a 30%

Figure 1: Effect of Judge-Pair Sentence Tendency on Actual Sentence Length



Note: The solid gray line is a flexible local linear regression of residualized (removing the court-specific fixed effects) measure of leave-out judge-pair sentence tendency and the residualized (again, removing the court-specific fixed effects) individual sentence length. To recover the original scale of the y-axis we add the mean sentence length to each residual of individual sentence length. The histogram of residualized judge-pair sentence tendency (x-axis) is shown in the background (height of the bars scaled to fractions so that their sum equals 1). Dashed red lines represent 95% confidence intervals.

increase relative to the sentence length average of 3.4 years).

One may wonder why some judges-pairs are systematically more stringent than others. While we do not know much about the internal logic and reasoning of particular court compositions, we do know that they all were faced with a new task, namely passing sentences en masse for state treason. On-the-job learning may lead judge-pairs that were allowed to stay intact for longer to systematically converge towards more or less stringent tendencies. We do not observe the complete count of cases which they handled but we do observe the conviction date of each POW in our data which allows us to calculate the rank order of each case. We find that judge-pairs become slightly more stringent the longer they have worked together, although experience of the pair accounts for only a tiny fraction of the variation in sentence tendency across judge-pairs (a$R^2 = 0.01$).

4.3 Instrument Exogeneity

Our causal analysis of the effect of sentence length on 75+ mortality for Finnish 1918 Civil War POWs exploits the random assignment of judge-pairs at the 145 Political Offence Courts that handled roughly 75,000 cases in the roughly eight months from June 1918 through January 1919 during which they were operational (see section 2.1 for the institutional background). Quasi-random assignment of the judge pairs to cases within the court where the judge served is verified empirically in Table 1. Column (1) documents that case characteristics, i.e., past activities as ascertained by the court itself through interrogation of the accused (and confirming his or her account thorough investigations) or formally reported into the court by the White Guard Staff in the accused's home community (such as year of joining the Workers Association, year of joining the Red Guard and whether the accused served therein as an officer) and individual-level characteristics such as age and marital status as of 1918 are highly predictive of sentence length. Column (2) examines whether our measure of judge-pair sentence tendency can be predicted by these same characteristics. Even though the set of characteristics are highly predictive of case outcomes, they are not statistically related to the stringency of the judge assigned to a case: none of the ten variables is statistically significant at the 5% significance level, and the variables are not jointly significant either (p= 0.3095). In fact, Table 1 verifies randomization just the same way as RCT studies tend to verify compliance with randomization. Taken together, both the institutional context as described in 2.1 and the empirical evidence reported in this subsection confirm that the stringency of the randomly assigned judge-pair (as measured by the leave-out sentence tendency of all other cases that the judge-pair handled) is an exogenous IV for actual sentence length.

A side note is warranted when discussing the results of Table 1. The estimates might lead one to think that the courts were internally fair by not putting weight on education or occupational status, rather on fundamentals directly related to the events of 1918 and the political activity preceding it. Anecdotal evidence however suggests that accounts on the prisoner's prewar political activity, ascertained by the courts through a detailed questionnaire sent out to the local White Civil Guard, were grossly arbitrary and could be related to any prior grudges in the local community held against the accused Red Guardsmen and -women (Tepora and Roselius, 2014).

5 Results

5.1 Ordinary Least Squares Estimates

We begin our presentation of results by providing regression-based evidence of the statistically significant adverse association between the sentence length and mortality for Finnish 1918 Civil War POWs who survived to age 75 (Figure 2a).

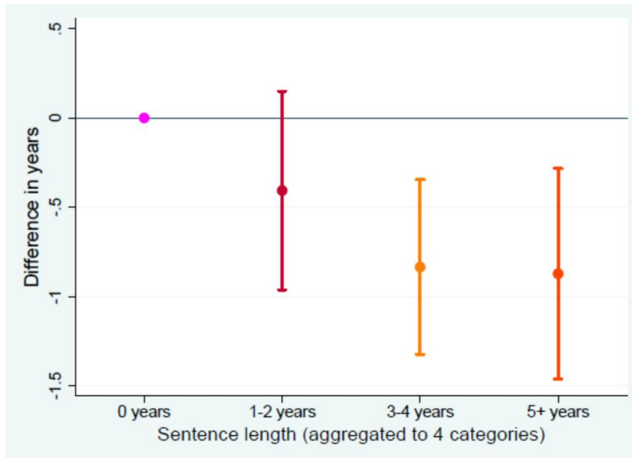
Table 1: Test of Random Assignment of Cases to Judge-Pairs.

	Dependent variable:	
	Sentence length	Judge-pair sentence tendency
Court examiner interrogations		
Female	-0.517	-0.239
	(0.714)	(0.300)
Age	0.066***	0.001
	(0.011)	(0.003)
Years of schooling	-0.001	-0.003
	(0.021)	(0.008)
Occupational status (HISCAM)†	-0.001	0.001
	(0.004)	(0.001)
Married	-0.373**	0.070
	(0.185)	(0.058)
Children	0.109	-0.059
	(0.202)	(0.067)
Joined workers association prior to 1917	0.161*	0.018
	(0.094)	(0.025)
Joined Red Guard prior to 1918	0.324	0.057*
	(0.188)	(0.032)
Officer (or military training)	1.342***	-0.042
	(0.157)	(0.042)
Local Civil White Guard questionnaire		
Agitation (for armed revolution)	1.310***	0.039
	(0.106)	(0.038)
Strike activity prior to 1918	0.581***	0.030
	(0.150)	(0.038)
Joint F-test	[0.0000]	[0.3095]
Observations	4,472	4,472

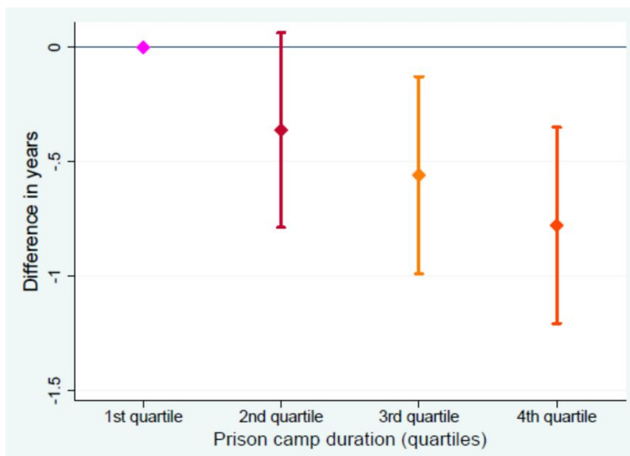
Note: † Occupational status is measured using a continuous HISCAM scale (Lambert et al., 2013) that assigns a score to each historical HISCLASS occupational category based on social interactions. The analyses use the restricted sample including only cases of courts that changed judge-pairs at least once (n=4,472). There are 198 judge-pairs in 76 courts in this restricted sample. All models include court fixed effects and POW camp-by-gender fixed effects‡. The p-value reported at the bottom of the columns is for an F-test of the joint significance of the variables listed in the rows with (in parentheses) standard errors clustered at the judge-pair level. The sample is left censored at age 75. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

‡ Even though we observe detailed information on sentence dates of each case within each court and on the exact dates of each particular court composition, we do not observe the assignment mechanism of a Red Guardsman or -woman to a court. The courts were placed in the proximity to the largest prison camps and often, one town would host several courts. For example, 14 courts in total operated in the town of Tammisaari (courts number 77-82, 111-115 and 117-120). It seems as good as arbitrary to which of the 14 courts a POW placed in the prison camp of Tammisaari was assigned but sorting based on unobserved characteristics cannot be excluded.

Figure 2: Impact of War Exposure Intensity on Old-Age Mortality Among the Survivors Alive at Age 75+.



(a) Sentence length (years), n=6,961



(b) Duration (days) quartiles, n=6,693

Note: The coefficients (and 95 percent confidence intervals) reported in the sub-figures come from OLS regressions (one regression for each sub-figure). Both regression models include the following control variables: the complete set of case- and individual-level control variables reported in Table 1, POW camp-by-gender fixed effects and court fixed effects. The sample is left censored at age 75. The confidence intervals are based on standard errors that were clustered at the judge-pair level.

There is some evidence of a dosage response, although caution must be taken with such an interpretation as not all coefficients are statistically significantly different from each other (Appendix Table A.3 reports the coefficients and F-test of equality of coefficients of Figure 2). The POWs who received at least a five year long sentence died on average roughly 11 months younger than the POWs who were found unguilty and were released upon receiving their verdict, a difference corresponding to a roughly 20% elevated mortality risk. A similar adverse association and dose response is found between the duration quartiles (in terms of the duration distribution of days) of POW camp exposure and mortality (Figure 2b). POWs surviving to age 75 and belonging to top quartile of POW duration length (on average 369 days) died on average 9 months earlier than the POWs belonging to the bottom quartile of POW duration length (on average 55 days). Despite the premature closure of the camps, the correlation between sentence length and length of POW camp exposure is high ($\rho = 0.64$), which we take as suggestive evidence of the adverse sentence length and old age mortality association being conferred through POW camp exposure.

Panel A of Table 2 presents the OLS estimates of estimating equation (1) in which sentence length enters as a continuous variable. On average, one extra year of sentence length advances death by roughly one month (column (2)).

5.2 Instrumental Variable Estimates

Panels B and C of Table 2 report the effect of sentence length on 75+ mortality of POWs using our IV approach, estimated by the two-stage least square (2SLS) estimator. Panel B presents both the OLS and the 2SLS estimates for the IV sample that includes only those 76 courts in which the judge-pair changed at least once (for a maximum of six times) in the period during which it remained operational (removing 2,482 individuals sentenced in 63 courts from the full sample). The OLS results in the IV sample are reassuringly similar (only less precise) as those in the full sample (column (2)). The 2SLS results show a substantial adverse effect of sentence length on 75+ mortality, the point estimates being roughly six times larger than the OLS estimates (column (4) vs. column (2)). Moreover, the 2SLS results lend further support of the dosage response documented in subsection 5.1, though this time assigning it a causal interpretation. The regression based first stage analysis (Panel C) confirms the graphical evidence of the instrument relevance (Figure 1), the relationship between the judge-pair sentence tendency and own sentence length being both substantially and statistically significant (F=26.86).[6]

The estimated magnitude of the effects is substantial. Take for example the coefficient -0.462 (0.238) in column (4) of Panel B in Table 2. This is the marginal effect of a one year longer sentence. The standard deviation (SD) in sentence length for men is 2.4 years (Appendix Table A.1). Hence for a one SD increase in sentence length, remaining life expectancy at age 75 is estimated to be shortened by 1.1 years, or to increase mortality rate by approximately 25%. This corresponds to (or is even larger than) population estimates of associations for differences in some key social predictors of mortality, e.g., the difference between growing up in single-parent versus two-parent households, primary school versus higher educated parents, manual job versus white collar job (Elo, Martikainen and Myrskylä, 2014).

We document in Appendix E.1 that all our results remain qualitatively unchanged when considering the quasi-random assignment of only the chairman of the court (i.e.,

[6] The F-statistic of a bivariate first stage regression is 97.79. These F-statistics should however be treated with some caution. As Hull (2017) points out, even though we use the jackknife (sort of grouping) estimator in the first stage, the true dimensionality of the instrument is the total number of judges, not one. Therefore, our first-stage F-statistics may be understating the instrument dimensionality and hence weak instrument bias of the 2SLS estimator towards the OLS may go undetected. It is however not straightforward to run a judge fixed-effects first stage regression that would be analogous to the jackknife estimator. The closest analogy would be a Split-Sample IV (Santavirta and Stuhler, Forthcoming).

excluding the legally qualified first court member of the judge pair). On average, the chairman changed at least once in 57 (41%) courts.

5.3 Interpreting the IV Estimates

We show proof of instrument relevance (Figure 1 and Panel C of Table 2) and instrument exogeneity (Table 1). Assuming constant treatment effects, relevance and exogeneity guarantee instrument validity, which would be sufficient to place a causal interpretation on β_1. Under heterogeneous treatment effects, in order to recover the local average treatment effect (LATE), i.e., the causal impact for the compliers, an additional condition must hold: the impact of judge assignment on the sentence length is monotonic across the accused POWs. This monotonicity assumption implies that accused who receive long sentences by lenient judge-pairs would always receive equally long or longer sentences by more stringent judge-pairs (and vice versa). While this assumption is not directly testable, Frandsen, Lefgren and Leslie (2023) show that this strict (pairwise) monotonicity can be relaxed to an average monotonicity assumption while still recovering a weighted average of individual treatment effects. This average monotonicity assumption requires that the data contain only complier groups where the covariance between judge sentence tendency and own sentence length is positive. We follow Dobbie, Goldin and Yang (2018) and provide an informal test of this assumption by reporting significant positive firststage regressions across subgroups (Table E.4). We further implement an indirect joint test of instrument exogeneity and pairwise monotonicity in Stata 18 via the package `testjfe` (Frandsen, Lefgren and Leslie, 2023). We fail to reject the null hypothesis that the instrument exogeneity

Table 2: The Effect of Sentence Length on 75+ Mortality.

	Dependent variable: Age at death (years)				
	OLS		2SLS		
	(1)	(2)	(3)	(4)	(5)
Panel A: Full sample (n=6,961)					
Sentence length (yrs)	-0.073**	-0.078**			
	(0.028)	(0.031)			
Adjusted R^2	0.140	0.140			
Panel B: IV sample (n=4,472)					
Sentence length (yrs)	-0.071*	-0.067*	-0.453**	-0.462 *	-0.471 **
	(0.037)	(0.040)	(0.226)	(0.238)	(0.240)
Income in 1970					0.058 ***
					(0.011)
Adjusted R^2	0.148	0.148	0.106	0.107	0.113
Panel C: First stage (n=4,472)			Dep. var.: Sentence length		
Judge-pair sentence tendency			0.376***	0.355***	0.355***
			(0.077)	(0.069)	(0.068)
F-test of excluded instrument			24.10	26.86	26.92
Case- and individual controls		Yes		Yes	Yes
Mediation analysis					Yes

Note: Panel A uses the baseline sample (n=6,961). Panels B and C present results based on the restricted sample that only includes cases of courts that changed judges-pair composition at least once (n=4,472). There are 198 judge-pairs in 76 courts in this restricted sample. All models control for age in 1918, POW camp-by-gender fixed effects and court fixed effects. Taxable income in 1970 is rescaled by dividing it by 100. Models reported in columns (2), (4) and (5) additionally include the complete set of case- and individual controls reported in Table 1. The sample is left censored at age 75. Standard errors (in parentheses) are clustered at the judge-pair level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

and monotonicity jointly hold (p=0.202).[7] Thus, we take this as evidence of our 2SLS identifying the LATE of sentence length for the marginal individuals who would not have received a long sentence had it not been for being assigned to a stringent judge, i.e., the so called compliers. Our IV estimates do not inform about the effect of sentence length on mortality for the so-called always takers and never takers, i.e., those who would not have received a different sentence length had their case been assigned to a more (or less) lenient judge-pair. This suggests due caution in extrapolating our causal estimates to the population at large (n=76,000) or to other settings.

Our IV approach would in principle allow the estimate of marginal treatment effects (MTE) as the instrument of judge sentence tendency varies (Heckman and Vytlacil, 2005). In a context such as ours, where the latent variable shifting the treatment is highly unclear (compared to, e.g., the foster care context, in which the latent variable is clearly defined, i.e., abuse and neglect (Doyle, 2007)), the value added of estimating the MTE is limited.

Appendix E.2 presents robustness of the 2SLS results using a split sample IV estimator and a 2SLS estimator that runs a true judge fixed effects regression (equivalent to an inclusive grouping estimator with coefficient one) in the first stage, including the exhaustive set of judge fixed effects. Reassuringly, the point estimates of sentence length are much larger than the OLS (-0.518 and -0.251) and significant at the 10% and 5% level, respectively.

[7] Table E.5 tests for the joint null hypothesis that instrument exogeneity and monotonicity hold across different combinations of knots and Bonferroni weights using the suggested quadratic spline (controlling for court fixed effects).

6 Discussion

6.1 Mechanisms

For our results to guide policy, it is important to understand whether the adverse effect of sentence length operates through biological processes that still play a role more than 50 years after exposure, or whether it is conferred through social processes. To address this, we compare the β_1^{2SLS} of equation (1) with the equivalent estimate in an otherwise similar specification, but that adds taxable income in 1970 (including capital income and pensions) as an explanatory variable. An attenuated and less precisely estimated slope coefficient of sentence length in such a specification would suggest that social processes, captured by income, play a role in the effect of sentence length on mortality. One advantage of our research design is that including a mediator in the specification should not confound the estimates of the key causal effect, whereas this would be a major concern in a descriptive analysis, given how intertwined the health and socioeconomic status of veterans tend to be (Lee, 2005). In this sense, our study is closely related to work that addresses mechanisms through a mediation analysis in a context in which the total causal effect is cleanly identified (see e.g., Kelly (2011)). However, it is important to note that to satisfy the (almost heroic) identifying assumptions required to identify both the direct effect of sentence length (holding income constant) and the indirect effect of sentence length that goes through income would require quasi-randomization of income as well or alternatively a (rather morbid) double RCT (Imai et al., 2011; Heckman and Pinto, 2015).

As we report in column (5) of Table 2, we do not find evidence of social factors mediating the effect of sentence on mortality. The coefficient of sentence length remains largely unchanged by including taxable income as an explanatory variable in the estimating equation. We infer from this disconnect between taxable income and the sentence length effect on mortality that the adverse effect of sentence length more likely operates through biological processes.

6.2 Strengths and Limitations

Beyond the linking of unique historical records to contemporary administrative data, the main strength of our study is that both the anecdotal and empirical evidence presented thus far suggest quasi-random assignment of cases to judge-pairs. We lend further support to this assumption by using a similar balancing test as that presented in column (2) of Table 1 to empirically document that the first case after a change in the judge pair does not differ along case or individual characteristics from cases with higher rank orders (F-test: 0.73). A caveat with this test is that we can only proxy the first case after a change in judge-pair composition, as we do not observe the population of cases.

Despite the empirical advances of this study, we acknowledge at least two limitations. First, when studying long-run consequences of extreme health insults, one concern is mortality selection leading to excess immediate mortality responses of the most frail tail of the population's health distribution and leaving a population with a more robust health stock and longer survival (Hobcraft, Menken and Preston, 1982). At its most extreme, this might manifest in results that make the acute adverse health shock look health preserving. In the starvation context, the evidence supporting this, or so-called cohort inversion (Preston et al., 1996), is out-weighted by the evidence in favor of debilitating scars among survivors (Almond et al., 2010; Gørgens, Meng and

Vaithianathan, 2012).[8] Given our results of a causal adverse effect, the implication of mortality selection would be that our estimated effect is a lower bound of the actual adverse effect of POW exposure on excess old-age mortality, detectable only if we were to observe the censored part of the health distribution. Note, however, that one of the aforementioned effects may dominate over the other among individuals exposed during a certain stage of life, whereas for another age group, the balance between the two might be the reverse. Costa (2012) finds evidence of dominating mortality selection among American Civil War POWs exposed at ages older than 30. As we start follow-up at age 75 in 1973 (or later), our analysis is almost exclusively restricted to below 30-year olds at exposure due to natural mortality of the oldest cohorts born in the 1880s. In the population of sentenced POWs (n= 67,788), 62.2% were below age 30 whereas 95.8% of ex-POWs in our data were under 30 in 1918 (see Appendix Table D.1 for a comparison of the age distribution in our study sample and the POW population). We see very little sample selection along any other dimensions than age. For example, Appendix Table D.2 shows that the occupational distribution is as good as identical in the study sample and the full POW population.

A second limitation is that our judge sentence tendency measure is calculated based on all other cases examined by the judge-pair within the court appearing in our study sample. We would obviously have liked to be able to construct the judge-pair sentence tendency measure in the full POW population. Appendix D documents, however, a remarkable similitude with respect to sentence length for both genders between our study sample and the POW population, occupational distribution of court members, and the duration of judge-pair spells. We take this as evidence that our judge sentence

[8] See Almond (2006) and Bozzoli, Deaton and Quintana-Domeque (2009) for analytical formalizations of the tension between cohort inversion and scarring in determining the health consequences of insults using observational data.

tendency measure is representative of the actual judge-pairs' sentence tendency.

6.3 Concluding Remarks

In an effort to identify the long-run causal effects of having been sentenced to prison camp during the Finnish Civil War of 1918, this study uses detailed primary data on sentences, prison camp exposure of each Red Guard, and the political offence court institution. We document quasi-random assignment of accused prisoners of war to judges and use an examiner design to identify the causal effect of sentence length on long-run mortality for those who survived through age 75. We find a substantial and statistically significant adverse causal effect of sentence length on old age mortality among the complier population, whose sentences were lengthened as a result of being assigned a judge with a tendency to pass stricter sentences. The dosage response implied by the local average treatment effect is roughly six times larger than the one implied by the association estimated by ordinary least squares. This causal effect seems to be conferred mainly through biological rather than socioeconomic processes. Taken together, our results suggest that more attention in designing rehabilitative policy should be spent on identifying differential risk based on dosage of exposure to extreme insults such as prison camp internment.

The longitudinal dimension of the data, spanning 100 years and following the completion of survival of the entire 1918 civil war cohort of young men and women, allows us to shed light on the long-run effect of exposure to extreme conditions during late adolescence and early adulthood. By providing the first quasi-experimental evidence on the effects of POW camp exposure, our results corroborate earlier descriptive evidence of long-lasting debilitating health consequences for young civil war cohorts of exposure to prison camp terms (Costa et al., 2017; Costa, Yetter and DeSomer, 2018, 2020).

Lastly, our findings can also be interpreted as a side effect of the demonstration of state capacity and the restoration of order through enforcement of a strict penal code for political crimes such as state treason. Indeed, though the political offence courts and the sentences passed by them may have served the purpose of reestablishing peace and deterring subsequent uprisings, the humanitarian legacy of the punishments linger on among survivors 50 years after the military victory of the White Army and the end of the war.[9]

[9]See Fergusson et al. (2023) for a discussion on the relationship between the penal code for quelling rebellion and maintaining order.

References

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. 2014. "A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration." *Journal of Political Economy*, 122(3): 467–506. 1

Aizer, Anna, and Joseph J. Doyle. 2015. "Juvenile Incarceration, Human Capital, and Future Crime: Evidence from Randomly Assigned Judges." *The Quarterly Journal of Economics*, 130(2): 759–804. 1

Aizer, Anna, Shari Eli, Joseph Ferrie, and Adriana Lleras-Muney. 2016. "The Long-Run Impact of Cash Transfers to Poor Families." *American Economic Review*, 106(4): 935–71. 1

Akbulut-Yuksel, Mevlude. 2014. "Children of War: The Long-Run Effects of Large-Scale Physical Destruction and Warfare on Children." *The Journal of Human Resources*, 49(3): 634–662. 1

Almond, Douglas. 2006. "Is the 1918 Influenza Pandemic Over? Long-Term Effects of In Utero Influenza Exposure in the Post-1940 U.S. Population." *Journal of Political Economy*, 114(4): 672–712. 8

Almond, Douglas, Lena Edlund, Hongbin Li, and Junsen Zhang. 2010. "Long-Term Effects of Early-Life Development: Evidence from the 1959 to 1961 China Famine." *The Economic Consequences of Demographic Change in East Asia*, 321–345. University of Chicago Press. 6.2

Angrist, Joshua D. 1990. "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records." *The American Economic Review*, 80(3): 313–336. 1

Black, Dan A., Seth G. Sanders, Evan J. Taylor, and Lowell J. Taylor. 2015. "The Impact of the Great Migration on Mortality of African Americans: Evidence from the Deep South." *American Economic Review*, 105(2): 477–503. 1

Blattman, Christopher, and Edward Miguel. 2010. "Civil War." *Journal of Economic Literature*, 48(1): 3–57. 1

Blattman, Christopher, and Jeannie Annan. 2010. "THE CONSEQUENCES OF CHILD SOLDIERING." *The Review of Economics and Statistics*, 92(4): 882–898. 1

Bozzoli, Carlos, Angus Deaton, and Climent Quintana-Domeque. 2009. "Adult Height and Childhood Disease." *Demography*, 46(4): 647–669. 8

Braun, Sebastian, and Jan Stuhler. 2023. "Exposure to War and its Labor Market Consequences over the Life Cycle." Unpublished. 1

Costa, Dora L. 2012. "Scarring and Mortality Selection Among Civil War POWs: A Long-Term Mortality, Morbidity, and Socioeconomic Follow-Up." *Demography*, 49(4): 1185–1206. 1, 6.2

Costa, Dora L., Heather DeSomer, Eric Hanss, Christopher Roudiez, Sven E. Wilson, and Noelle Yetter. 2017. "Union Army Veterans, All Grown Up." *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 50(2): 79–95. 6.3

Costa, Dora L., Noelle Yetter, and Heather DeSomer. 2020. "Wartime health shocks and the postwar socioeconomic status and mortality of union army veterans and their children." *Journal of Health Economics*, 70: 102281. 6.3

Costa, Dora, Noelle Yetter, and Heather DeSomer. 2018. "Intergenerational transmission of paternal trauma among US Civil War ex-POWs." *Proceedings of the National Academy of Sciences of the United States of America*, 115: 11215–11220. 1, 6.3

Dahl, Gordon B., Andreas Ravndal Kostøl, and Magne Mogstad. 2014. " Family Welfare Cultures *." *The Quarterly Journal of Economics*, 129(4): 1711–1752. 1, B

Daza, Sebastian, Alberto Palloni, and Jerrett Jones. 2020. "The Consequences of Incarceration for Mortality in the United States." *Demography*, 57(2): 577–598. 1

de Groot, Olaf J, Carlos Bozzoli, Anousheh Alamir, and Tilman Brück. 2022. "The global economic burden of violent conflict." *Journal of Peace Research*, 59(2): 259–276. 1

De los Rios Hernández, Isabela. 2023. ""Enough Already!": The Challenges of Restorative Justice in War Crimes." *Harvard International Review.* 1

Dent, Owen F., Bruce Richardson, Sue Wilson, Kerry J. Goulston, and Catherine W. Murdoch. 1989. "Postwar mortality among Australian World War II prisoners of the Japanese." *Medical Journal of Australia*, 150(7): 378–382. 1

Dobbie, Will, Jacob Goldin, and Crystal S. Yang. 2018. "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges." *American Economic Review*, 108(2): 201–40. 1, 5.3, B

Doyle, Joseph J., Jr. 2007. "Child Protection and Child Outcomes: Measuring the Effects of Foster Care." *American Economic Review*, 97(5): 1583–1610. 1, 5.3

Elo, Irma T., Pekka Martikainen, and Mikko Myrskylä. 2014. "Socioeconomic status across the life course and all-cause and cause-specific mortality in Finland." *Social Science & Medicine*, 119: 198–206. 5.2

Fergusson, Leopoldo, Javier Mejia, James A Robinson, and Santiago Torres. 2023. "Constitutions and Order: A Theory and Evidence from Colombia and the United States." National Bureau of Economic Research Working Paper 31501. 9

Frandsen, Brigham, Lars Lefgren, and Emily Leslie. 2023. "Judging Judge Fixed Effects." *American Economic Review*, 113(1): 253–77. 1, 5.3, E.5

Garfield, Richard. 2008. "The Epidemiology of War." In *War and Public Health.* Oxford University Press. 1

Goldin, Claudia D., and Frank D. Lewis. 1975. "The Economic Cost of the American Civil War: Estimates and Implications." *The Journal of Economic History*, 35(2): 299–326. 1

Gørgens, Tue, Xin Meng, and Rhema Vaithianathan. 2012. "Stunting and selection effects of famine: A case study of the Great Chinese Famine." *Journal of Development Economics*, 97(1): 99–111. 6.2

Hearst, Norman, Thomas B. Newman, and Stephen B. Hulley. 1986. "Delayed Effects of the Military Draft on Mortality." *New England Journal of Medicine*, 314(10): 620–624. PMID: 3945247. 1

Heckman, James J., and Edward Vytlacil. 2005. "Structural Equations, Treatment Effects, and Econometric Policy Evaluation." *Econometrica*, 73(3): 669–738. 5.3

Heckman, James J., and Rodrigo Pinto. 2015. "Econometric Mediation Analyses: Identifying the Sources of Treatment Effects from Experimentally Estimated Production Technologies with Unmeasured and Mismeasured Inputs." *Econometric Reviews*, 34(1-2): 6–31. PMID: 25400327. 6.1

Hobcraft, John, Jane Menken, and Samuel Preston. 1982. "Age, Period, and Cohort Effects in Demography: A Review." *Population Index*, 48(1): 4–43. 6.2

Hull, Peter. 2017. "Examiner Designs and First-Stage F Statistics: A Caution." Unpublished note. 6

Hutchinson, William K., and Robert A. Margo. 2006. "The impact of the Civil War on capital intensity and labor productivity in southern manufacturing." *Explorations in Economic History*, 43(4): 689–704. 1

Ichino, Andrea, and Rudolf Winter-Ebmer. 2004. "The Long-Run Educational Cost of World War II." *Journal of Labor Economics*, 22(1): 57–86. 1

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. 6.1

Kelly, Elaine. 2011. "The Scourge of Asian Flu: In utero Exposure to Pandemic Influenza and the Development of a Cohort of British Children." *The Journal of Human Resources*, 46(4): 669–694. 6.1

Kesternich, Iris, Bettina Siflinger, James P. Smith, and Joachim K. Winter. 2014. "The Effects of World War II on Economic and Health Outcomes across Europe." *The Review of Economics and Statistics*, 96(1): 103–118. 1

Lambert, Paul S., Richard L. Zijdeman, Marco H. D. Van Leeuwen, Ineke Maas, and Kenneth Prandy. 2013. "The Construction of HISCAM: A Stratification Scale Based on Social Interactions for Historical Comparative Research." *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 46(2): 77–89. 1, A.1

Lee, Chulhee. 1999. "Selective Assignment of Military Positions in the Union Army: Implications for the Impact of the Civil War." *Social Science History*, 23(1): 67–97. 5

Lee, Chulhee. 2005. "Wealth Accumulation and the Health of Union Army Veterans, 1860–1870." *The Journal of Economic History*, 65(2): 352–385. 6.1

Leinberg, Georg. 1923. "Tilastollinen Tutkimus 1918 Vuoden Valtiorikollisista [Engl. Statistical Research on the 1918 Political Crimes]." *Suomen Virallinen Tilasto XXIII [Engl. Official Statistics of Finland]*, Oikeustilasto 32 [Engl. Yearbook of Justice Statistics]. 2.1, 3, C.1, D.1, D.2

Linder, Hjalmar. 1918. "Ur Arkivet: Till Red. af Hbl. Nog med blodbad! [Engl. From the Archive: Letter to the Editor of Hufvudstadsbladet. Enough bloodshed!]." In *Nya Argus*. Vol. 104(5-6)2011, 126-127, , ed. Trygve Söderling. Publisher: Garantiföreningen för Nya Argus r.f. 1

McKenzie, Robert Tracy. 2006. "Retribution and Reconciliation." In *Lincolnites and Rebels: A Divided Town in the American Civil War*. Oxford University Press. 1

Meriläinen, Jaakko, Matti Mitrunen, and Tuomo Virkola. 2022. "Famine, Inequality, and Conflict." *Journal of the European Economic Association*, 21(4): 1478–1509. 1

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

Mäkelä, Pentti. 2007. *Vuosien 1917–19 kulkutaudit, espanjantauti ja vankileirikatastrofi. Historiallisepidemiologinen näkökulma Suomen väestön korkeaan tautikuolleisuuteen [Engl. Epidemics in 1917-1919, Spanish Flu and the Prison Camp Catastrophe. A Historical Epidemiological View on Finland's High Disease Mortality].* Suomi:Valtioneuvoston kanslia. 2.1

National Archives of Finland. 2022. "War Victims 1914-1922." [Online; accessed 15 August 2023]. 2.1, 3, 4

Paavolainen, Jaakko. 1971. *Vankileirit Suomessa 1918 [Engl. Prison Camps in Finland 1918].* Tammi. 1, 2.1, 4

Page, William F., and Adrian M. Ostfeld. 1994. "Malnutrition and subsequent ischemic heart disease in former prisoners of war of world war II and the Korean conflict." *Journal of Clinical Epidemiology*, 47(12): 1437–1441. 1

Page, William F., and Lawrence M. Brass. 2001. "Long-Term Heart Disease and Stroke Mortality among Former American Prisoners of War of World War II and the Korean Conflict: Results of a 50-Year Follow-Up." *Military Medicine*, 166(9): 803–808. 1

Pizarro, Judith, Roxane Cohen Silver, and JoAnn Prause. 2006. "Physical and Mental Health Costs of Traumatic War Experiences Among Civil War Veterans." *Archives of General Psychiatry*, 63(2): 193–200. 1

Pope, Clayne L., and Larry T. Wimmer. 1998. "Aging in the Early 20th Century." *The American Economic Review*, 88(2): 217–221. 1

Preston, Samuel, Irma Elo, Ira Rosenwaike, and Mark Hill. 1996. "African-american mortality at older ages: Results of a matching study." *Demography*, 33(2): 193–209. 6.2

Riaño, Juan Felipe, and Felipe Valencia Caicedo. 2024. "Collateral Damage The Legacy of the Secret War in Laos*." *The Economic Journal*, ueae004. 1

Santavirta, Torsten, and Jan Stuhler. Forthcoming. "Name-Based Estimators of Intergenerational Mobility." *Economic Journal.* 6

Tepora, Tuomas, and Aapo Roselius. 2014. *The Finnish Civil War 1918: History, Memory, Legacy.* Leiden, The Netherlands:Brill. 1, 4.3

Upton, Anthony F. 1980. *The Finnish Revolution 1917-1918.* Minneapolis:University of Minnesota Press. 1, 2.1, 4.1

van den Berg, Gerard J., Pia R. Pinger, and Johannes Schoch. 2016. "Instrumental Variable Estimation of the Causal Effect of Hunger Early in Life on Health Later in Life." *The Economic Journal*, 126(591): 465–506. 1

van Leeuwen, M.H.D., and I. Maas. 2011. *Hisclass: A Historical International Social Class Scheme. G - Reference,Information and Interdisciplinary Subjects Series*, Leuven University Press. D.2

Appendix For Online Publication

A Descriptive Statistics

Table A.1: Summary Statistics

	Male		Female	
	Mean	SD	Mean	SD
Exposure				
Sentence length (years)	3.77	2.40	1.96	1.47
Prisoner of War camp duration (days)*	176.22	197.67	102.38	87.39
Court examiner interrogations				
Age in 1918	21.81	3.86	20.77	3.52
Years of schooling†	3.22	1.57	3.36	1.52
Occupational status (HISCAM)	51.48	7.45	47.79	7.24
Married	0.12	0.32	0.05	0.21
Children	0.10	0.29	0.06	0.23
Joined workers association prior to 1917	0.16	0.37	0.11	0.31
Joined Red Guard prior to 1918	0.06	0.23	0.01	0.12
Officer (or military training)	0.08	0.27	0.03	0.16
Local Civil White Guard Questionnaire				
Agitation (for armed revolution)	0.21	0.40	0.19	0.39
Strike activity prior to 1918	0.10	0.30	0.08	0.27
Contemporary administrative records				
Age at death	83.83	5.36	85.48	5.84
Taxable income in 1970‡	519.72	768.07	156.19	363.45
Observations	5,958		1,003	

Note: *We observe prisoner of war camp duration for 5,744 men and 949 women. † We observe years of schooling for 5,245 men 890 women. In our data, unlike in contemporary administrative data, years of schooling is based on the reported precise number of completed years. ‡ We observe annual taxable income in 1970 for 5,824 men and 977 women. In order to economize with our observations, we replace missing years of schooling with 1 year of schooling and missing taxable income in 1970 with zero income and dummy out these missing observations by including an indicator variable each for missing value for years of schooling and income.
Occupational status is measured using a continuous HISCAM scale (Lambert et al., 2013) that assigns a score to each historical HISCLASS occupational category based on social interactions. SD stands for standard deviation.

Table A.2: Amnesties of Sentences Passed by the Political Offence Court

Date	Amnesties
October 30, 1918	All sentences up to 4 years
December 7, 1918	All sentences up to 6 years
June 19, 1919	All prison sentences (excluding murder, manslaughter and arson)
June 19, 1919	Release on parole of 12 Members of Parliament
January 30, 1920	Restoration of citizen rights to all released
January 20, 1921	Full amnesty of remaining prisoners*

Note: * Full amnesty on January 20, 1921 implied in practice converting the prison sentences to conditional imprisonment and shortening the sentence to half of the remaining time. Citizen rights would be restored only after the conditional imprisonment term.

Table A.3: Impact of War Exposure Intensity on Old-Age Mortality Among the Survivors Alive at Age 75+.

	Explanatory variables			
Dep.var.	0 years (yrs)	1-2 yrs	3-4 yrs	5+ yrs
Age at death (n=6,961)	-	-0.406	-0.835***	-0.872***
		(0.282)	(0.248)	(0.300)
F-test: 0 vs. 1-2 years		p=0.151		
F-test: 1-2 yrs vs. 3-4 yrs			p=0.028	
F-test: 1-2 yrs vs. 5+ yrs				p=0.045
	1st quartile (qrt)	2nd qrt	3rd qrt	4th qrt
Age at death (n=6,693)	-	-0.361*	-0.558**	-0.778 ***
		(0.216)	(0.219)	(0.218)
F-test: 1st vs. 2nd quartile		p=0.095		
F-test: 2nd vs. 3rd quartile			p=0.370	
F-test: 2nd vs. 4th quartile				p=0.056

Note: Regression results of Figure 2. The coeffients reported in the top panel and the ones reported in the bottom panel come from different regressions. Both regression models (top and bottom panel) include the following control variables: the complete set of case- and individual-level control variables reported in Table 1, POW camp-by-gender fixed effects and court fixed effects. The sample is left censored at age 75. Standandard errors (in parentheses) are clustered at the judge-pair level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B Instrumental Variable Calculation

We follow Dahl, Kostøl and Mogstad (2014) and Dobbie, Goldin and Yang (2018) and construct our instrument using a residualized (adjusted for court fixed effects) leave-out judge-pair sentence tendency measure. Let the residual sentence length of individual i sentenced in political offence court c, after removing the court fixed effects, be denoted by:

$$(\textit{Sentence length})^*_{icj} = (\textit{Sentence length})_{icj} - \lambda \mathbf{P}_c = Z_{cj} + \upsilon_{ic} \tag{3}$$

where $\mathbf{P}_c$ includes the respective 138 political offence court fixed effects. The residual sentence length, $(\textit{Sentence length})_{ic}$, includes our measure of judge-pair sentence tendency Z_{cj} of judge-pair j in court c, as well as idiosyncratic POW level variation υ_{icj}. For each accused red guardsman or -woman, we then use this residual sentence length to construct the leave-out mean length of sentence of the assigned judge-pair within a court:

$$Z_{icj} = \frac{1}{(n_{kcj} - 1)} \left(\sum_{k=1}^{n_{kcj}} (\textit{Sentence length})^*_{kcj} - (\textit{Sentence length})^*_{icj} \right) \tag{4}$$

where n_{kj} is the number of all cases k handled by judge-pair j in court c. Effectively, we remove the residualized sentence length of ex-POW i, $(\textit{Sentence length})^*_{ic}$, from the sum of residualized sentence lengths of all cases handled by judge-pair j in court c and divide it by $n_{kcj} - 1$ to construct the judge-pair leave-out mean of sentence tendency Z_{icj}. For the main text to be self contained, we rename Z_{icj} by the more illustrative $\textit{Sentence tendency}_{icj}$. We use this residualized version of sentence tendency throughout the analysis though all the analyses would remain the same if we were to use non-residualized sentence tendency (while adjusting for court fixed effects).

C Representativeness of the Court Composition in the Study Sample

Table C.1: Sentence Distribution in the Study Sample vs. the POW Population

	Study sample				Population of accused			
	Male		Female		Male		Female	
Sentence	Count	Share (%)	Count	Share	Count	Share	Count	Share
Unguilty	356	5.5	248	23.7	6,257	8.9	1,533	27.7
1 year	152	2.4	51	4.9	975	1.4	153	2.7
2	1,048	16.2	385	36.8	9,922	14.2	1,792	32.4
3	2,789	43.2	311	29.7	29,871	42.6	1,655	29.9
4	575	8.9	20	1.9	6,092	8.7	152	2.7
5	369	5.7	10	1.0	3,601	5.1	61	1.1
6	227	3.5	6	0.6	2,798	4.0	59	1.1
7	55	0.9	2	0.2	531	0.8	11	0.2
8	477	7.4	12	1.1	4,316	6.2	37	0.6
9-10	218	3.4	2	0.2	2.527	3.6	18	0.3
11-12	92	1.4	-	-	1,204	1.7	9	0.2
14	2	<0.1	-	-	78	0.1	-	-
Youth detention center*	-	-	- -	-	228	0.4	46	1.2
Dismissal from office or fine†	-	-	-	-	206	0.3	5	0.1
Life sentence	61	1.0	-	-	882	1.3	4	0.1
Death sentence	34	0.5	-	-	554	0.8	1	<0.1
Total	6,455	100	1,047	100	70,042	100	5,533	100

Note: We have aggregated the sentence distribution of exact sentence length in years to two-year bins for sentence lengths 9-10 and 11-12 years to match the aggregate-level data available for the full population. We further present the sentence distribution for our complete data (n=7,502) for comparison, not for our study sample of individuals who survived until age 75 and dropping life and death sentences (n=6,961). *Of those sentenced to youth detention centers for an unknown period, 172 (76.3%) and 32 (69.6%) women were underaged (Leinberg, 1923). These are excluded from our data. †Those sentenced to dismissal of office were civil servants of the customs office or correctional institutions Leinberg (1923). Also these are excluded from our data.

Table C.2: Occupational Distribution of Political Offence Court Members

	Chairman	First member	Second member	Third member	Military representative
	Count (%)	Count (%)	Count (%)	Count (%)	Count (%)
			Complete count of court compositions		
Legal training of chairman and 1. member					
Judge	181 (82.6)	107 (53.8)			
Law clerk	33 (15.1)	108 (40.5)			
Lower law degree	5 (2.3)	11 (5.5)			
Top-5 occupations of 2. and 3. member					
Master degree		1 (0.5)	15 (7.6)	13 (7.0)	
Entrepreneur or firm CEO			35 (17.7)	36 (19.3)	
Compulsory school teacher			31 (15.7)	21 (11.2)	
High school teacher					
Foreman			19 (9.6)	18 (9.6)	
Shopkeeper			18 (9.1)	22 (11.8)	
Top-5 military rank of military representative					
Lieutenant colonel					6 (3.1)
Colonel					3 (1.5)
Second lieutenant					102 (52.3)
Lieutenant					55 (29.2)
Master sergeant					24 (12.3)
			Study sample		
Legal training of chairman and 1. member					
Judge	164 (81.6)	95 (51.9)			
Law clerk	32 (15.9)	75 (41.0)			
Lower law degree	5 (2.5)	11 (6.0)			
Top-5 occupations of 2. and 3. member					
Master degree		1 (0.6)	15 (8.2)	12 (7.0)	
Entrepreneur or firm CEO			31 (16.9)	34 (19.8)	
Compulsory school teacher			29 (15.9)	18 (10.5)	
High school teacher			10 (5.5)		
Foreman			17 (9.3)	15 (8.7)	
Shopkeeper				21 (12.2)	
Top-5 military rank of military representative					
Lieutenant colonel					4 (2.3)
Colonel					2 (1.1)
Second lieutenant					95 (54..3)
Lieutenant					50 (28.4)
Master sergeant					20 (11.4)

Table C.3: Judge-Pair Changes in the Complete Count of Court Composition and in the Study Sample

	Complete count		Study sample		IV sample	
Number of changes of judge-pair composition	Number of courts	Avg. days per judge-pair	Number of courts	Avg. days per judge-pair	Number of courts	Avg. days per judge-pair
0	52	183.6	63	174.9	1*	151
1	42	90.7	42	85.7	43	86.1
2	24	60.4	21	55.9	22	53.7
3	14	47.6	8	45.3	7	45.79
4	4	41.1	3	38.5	1	40.6
5	2	40.8	2	32.5	2	32.5
7	1	25.9	-	-	-	-
Total	139		139		76	

Note: Court number 53, 103, 104, 116, 113, and 145 are removed from the complete count. See section 2.1 in the main text for an explanation. The marginal distribution of composition changes in the *complete count* comes directly from the raw spreadsheets of the registry of Political Offence Court, which for each court tracks exact court composition at every exact date (and not from individual-level data). The marginal distributions of composition changes in the *study sample* and the *IV sample* come from our own data matched to the court composition spreadsheet based on date of sentence date from the individual case file of the Public Offence Court. For comparison, the marginal distributions of composition changes of the *study sample* are shown without our sample restrictions, i.e., in the complete sample of 7,502 POWs. In this complete sample of ours, we observe 269 judge-pair compositions in the 139 courts. We show the marginal distribution of composition changes for our *IV sample* applying all sample restrictions made (see Section 3) leaving us with 198 judge-pair compositions in 76 courts. *Note that after applying our baseline sample restriction of having survived through age 75, only one of two court compositions in one court (No 102) that we observe in our complete sample remains in our IV sample of POWs. For the record, court No 102 only ever had two court compositions in the complete count.

D The Study Sample vs. the POW Population

Table D.1: Age Distribution of the Sentenced in the Study Sample vs. the POW Population

Age interval	Sample count	Sample share	Population count	Population share
15-17 years	750	10.9	6,830	10.1
18-20	2,568	37.2	12,504	18.5
21-24	2,303	33.4	10,300	15.2
25-29	990	14.4	12,535	18.4
30-34	249	3.6	8,912	13.2
35-39	31	0.5	6,193	9.2
40-49	7	0.1	7,681	11.3
Total	6,898	100	64,955	96.3

Note: For the full population of accused (n=75,575) Red guardsmen, the age distribution of the sentenced only (n=67,788) is obtained from Leinberg (1923). For comparison, the age statistics for our own sample are shown for the 6,898 sentenced to at least one year in prison (i.e., we drop 604 observations of our complete sample (n=7,502) who were found unguilty).

Table D.2: Common Occupation Groups for Study Sample vs. the POW Population

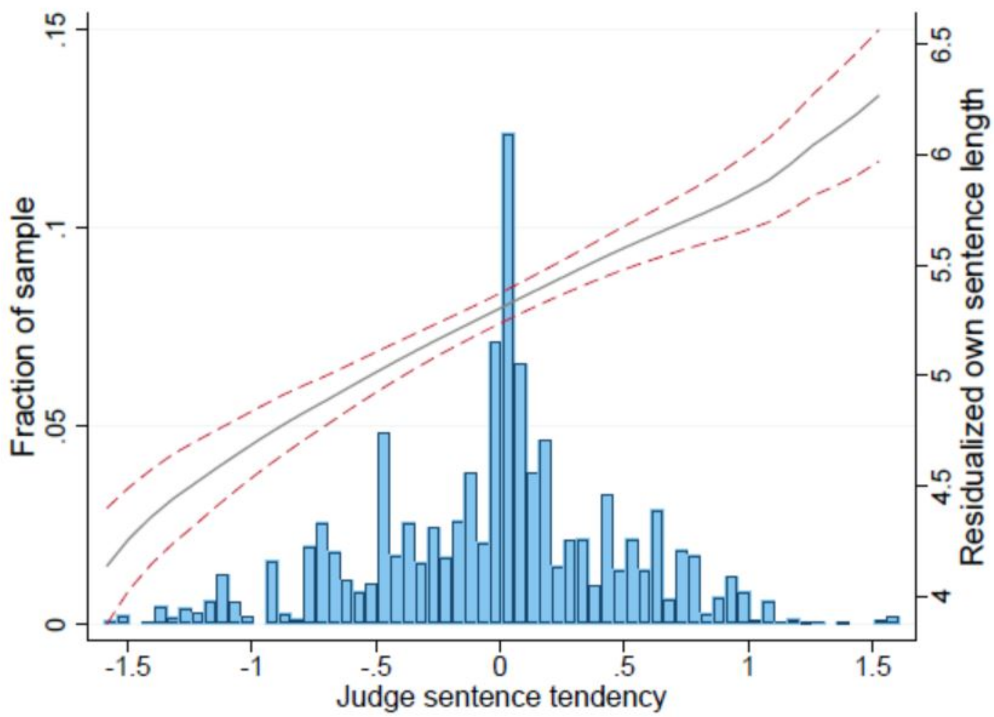
	Study sample			Population of sentenced		
Occupation	Count	Percent	HISCLASS	Count	Percent	HISCLASS
Unskilled laborer	3,423	49.7	11	30,956	45.7	11
Skilled industry worker	1,681	24.4	7 & 9	17,946	26.5	7 & 9
Agricultural laborer	831	12.1	12	6,222	9.2	12
Tenant farmer	613	8.9	10	6647	9.8	10
Sales personnel	58	0.8	5	538	0.8	5
Total (top 5)	6,606	95.8		62,309	91.9	
Outside top 5	292	4.2		5,479	8.1	

Note: HISCLASS is a 12-step classification system indicating the social class of each occupation, 1 being the highest class (van Leeuwen and Maas, 2011). For the full population of accused (n=75,575) Red guardsmen, the occupational distribution of the sentenced only (n=67,788) is obtained from Leinberg (1923). For comparison, the occupational distribution for our own sample are shown for the 6,898 sentenced to at least one year in prison (i.e., we drop 604 observations of our complete sample (n=7,502) who were found unguilty).

E Robustness checks

E.1 Considering the Judge Sentence Tendency of the Court Chairman Only

Figure E.1: Effect of Judge Sentence Tendency on Actual Sentence Length.



Note: The solid gray line is a local linear regression of residualized (removing the court-specific fixed effects) measure of leave-out judge sentence tendency and the residualized (again, removing the court-specific fixed effects) individual sentence length. To recover the original scale of the y-axis we add the mean sentence length to each residual of individual sentence length. The histogram of residualized judge sentence tendency for sentence length (x-axis) is shown in the background (height of the bars scaled to fractions so that their sum equals 1). Dashed red lines represent 95% confidence intervals.

Table E.1: Test of Random Assignment of Cases to Judges.

	Dependent variable:	
	Sentence length	Judge sentence tendency
Court examiner interrogations		
Female	-0.563	-0.113
	(0.743)	(0.292)
Age	0.067***	0.002
	(0.0133)	(0.003)
Years of schooling	-0.011	0.005
	(0.021)	(0.008)
Occupational status (HISCAM)	-0.001	-0.001
	(0.005)	(0.001)
Married	-0.452***	0.076
	(0.169)	(0.056)
Children	0.110	-0.107
	(0.220)	(0.067)
Joined workers association prior to 1917	0.188*	0.018
	(0.107)	(0.024)
Joined Red Guard prior to 1917	0.259	0.064*
	(0.208)	(0.035)
Officer (or prior military training)	1.397***	-0.046
	(0.161)	(0.042)
Local Civil White Guard questionnaire		
Agitation (for armed revolution)	1.303***	0.060
	(0.096)	(0.040)
Strike activity	0.609***	-0.011
	(0.145)	(0.035)
Joint F-test (p-value)	[0.0000]	[0.194]
Observations	3,583	3,583

Note: The analyses use the restricted sample including only cases of courts that changed chairman (judge) at least once (n=3,583). There are 130 judges in 57 courts in this restricted sample. All models include court fixed effects and POW camp-by-gender fixed effects. The p-values reported in square brackets at the bottom of the columns is for an F-test of the joint significance of the coefficients listed in the rows with standard errors (in parentheses) clustered at the judge level. The sample is left censored at age 75. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table E.2: The Effect of Sentence Length on 75+ Mortality.

	Dependent variable: Age at death				
	OLS		2SLS		
	(1)	(2)	(3)	(4)	(5)
Panel A: Sample of cases in courts >0 judge changes (n=3,583)					
Sentence length (yrs)	-0.047	-0.042	-0.489**	-0.495**	-0.469**
	(0.042)	(0.046)	(0.228)	(0.240)	(0.241)
Income in 1970					0.063***
					(0.014)
Adjusted R^2	0.017	0.146	0.100	0.102	0.113
Panel B: First stage (n=3,583)			Dep. var.: Sentence length		
Judge sentence tendency			0.454***	0.428***	0.427 ***
			(0.079)	(0.071)	(0.071)
F-test of excluded instrument			32.85	36.50	36.42
Case- and individual controls		Yes		Yes	Yes
Mediation analysis					Yes

Note: The table presents results based on the restricted sample that only includes cases of courts in which the chairman (judge) at least once (n=3,583). There are 130 judges in 57 courts in this restricted sample. All models control for age in 1918, POW camp-by-gender fixed effects and court fixed effects. Models reported in columns (2), (4) and (5) additionally include the complete set of case- and individual controls reported in Table 1. Taxable income in 1970 is rescaled by dividing it by 100. The sample is left censored at age 75. Standard errors (in parentheses) are clustered at the judge level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

E.2 Robustness of Two Stage Least Squares Results

Table E.3: Additional Variants of the Two Stage Least Squares Estimator

	Split sample	Judge FE
	Second stage	
Explanatory variable:	*Dependent variable: Age at death*	
Sentence length	-0.518*	-0.251**
	(0.269)	(0.100)
	First stage	
	Dependent variable: Sentence length	
Inclusive mean judge sentence tendency	0.373***	
	(0.077)	
F-statistic	23.25	17,254.96
Observations	2,254	4,472

Note: The Split sample column presents results of a split sample IV (SSIV) that estimates the first stage in an auxiliary sample consisting of a randomly drawn 50% subsample and the second stage in the primary sample consisting of the remaining 50% of the original IV-sample (n=4,472). The instrument used in the SSIV is the inclusive mean sentence tendency (not leaving anyone out in the auxiliary data), as leaving auxiliary POWs out makes no sense as long as there is no overlap between the auxiliary and primary data. The individuals in the primary data will indeed be left out of the mean sentence tendency in this way, which is the way we want it. The judge FE column presents results of a 2SLS regression that uses an exhaustive set of judge-pair fixed effects (n=197) as instruments (first stage F-statistic = 17,254.96). The first stage is thus to be compared to an inclusive (and not jackknife) grouping estimator. All models control for the complete set of case- and individual controls reported in Table 1, POW camp-by-gender fixed effects and court fixed effects. The sample is left censored at age 75. Standard errors (in parentheses) are clustered at the judge level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table E.4: First-Stage Regression Results by Sub-Samples

	Baseline	Occupational Status		Age		Years of Schooling	
		≥p(50)	<p(50)	≥p(50)	<p(50)	≥p(50)	<p(50)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sentence tendency	0.355***	0.342***	0.441***	0.320***	0.395***	0.331***	0.473***
	(0.069)	(0.097)	(0.098)	(0.096)	(0.091)	(0.092)	(0.087)
F-statistic	26.86	12.51	20.12	11.20	18.76	12.95	29.54
Observations	4,472	2,270	2,202	2,418	2,054	2,494	1,978

Note: The table reports regression results for samples stratified by occupational status, age, and years of schooling. All strata variables refer to the IV sample (n=4,472). F-statistic refers to the Kleinbergen-Paap F-statistic. All models control for age in 1918, POW camp-by-gender fixed effects and court fixed effects and the complete set of case- and individual controls reported in Table 1. The sample is left censored at age 75. Standard errors (in parentheses) are clustered at the judge-pair level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table E.5: Test of Joint Null of Monotonicity and Exclusion

	5 knots			10 knots		
	$\omega = 1$	$\omega = 0.8$	$\omega = 0.5$	$\omega = 1$	$\omega = 0.8$	$\omega = 0.5$
	(1)	(2)	(3)	(4)	(5)	(6)
Test statistic	135	135	135	134	134	134
d.f.	122	122	122	117	117	117
P-value	0.202	0.253	0.404	0.133	0.166	0.265

Note: Notes: The table presents results from the test proposed in Frandsen, Lefgren and Leslie (2023) for the joint null hypothesis that the monotonicity and exclusion restrictions hold. This null is tested using court fixed effects. Columns (1) to (3) provide the results imposing 5 knots in the quadratic spline function. Columns (4) to (6) provide the results imposing 10 knots in the quadratic spline function. Each column is associated with different weighting schemes between the fit and slope components of the test. Failure to reject the null implies that we cannot reject the hypothesis that the monotonicity and exclusion restrictions jointly hold. The test was implemented in Stata via the package `testjfe`.

Table E.6: Similarity of Cause of Death between Those Found Unguilty and Those Sentenced to 5+ years

	Share (%) within 0 years	Share (%) within 5+ years	t-test
Cancer (C00-C97)	12.0	18.6	<0.001
Cardiovascular (I00-I425, I427-I99)	62.6	52.4	<0.001
Respiratory (J00-J64, J66-J99)	12.2	14.7	0.142
Total	86.8	85.7	