Replication Materials for "Racial Inequality in War"

Connor Huff, Eric Min, and Robert Schub

Last updated July 10, 2025

This document contains all information regarding the files necessary to replicate all results in the main manuscript and the appendix.

The document is divided into two main sections. In the first, we list and briefly describe the function of each item in the replication materials. In the second, we describe every covariate in the CSV files that contain relevant data.

Technical Specifications **2**

Overview of Files **2**

CSV Files 2

R Files 2

Variables in CSV Files **3**

wwi_unit_type_data.csv 3

wwi_unit_period_data.csv 4

Technical Specifications

All analyses were performed on a PC with Windows 11 Pro, using R version 4.4.0.

Overview of Files

CSV Files

File	Description
wwi_unit_type_data.csv	Dataset containing information on all distinct unit types deployed by the United States Army to Europe during World War I, created using the *Order of Battle of the United States Land Forces in the World War*.
wwi_unit_period_data.csv	Dataset containing information on United States Army infantry casualties suffered in Europe during World War I, created using casualty lists, disaggregated at the half-month level.

R Files

File	Description
load_packages.R	Installs and loads all necessary packages.
racial_inequality_in_war_analysis.R	Generates all results, tables, and figures in the main manuscript and appendix. Data used: • /data/wwi_unit_type_data.csv • /data/wwi_unit_period_data.csv

Variables in CSV Files

wwi_unit_type_data.csv

Dataset containing information on all distinct unit types deployed by the United States Army to Europe during World War I, created using the *Order of Battle of the United States Land Forces in the World War*.

- *id:* Unique identifier for a unit type. This is a combination of the *unit* and *subunit* columns.
- *unit:* The overall type of unit, as designated by the *Order of Battle*.
- *subunit:* The subtype of the unit, as designated by the *Order of Battle*.
- *combatrole:* Whether the unit served a combat or support role.
- *n_overseas:* The total number of units deployed to Europe.
- *n_white_overseas:* The number of White units deployed to Europe.
- *n_black_overseas:* The number of Black units deployed to Europe.
- *white_personnel_overseas:* The estimated number of White troops deployed to Europe based on average unit sizes. Note that this measure is only produced for unit types that are displayed in Figure 1.
- *black_personnel_overseas:* The estimated number of Black troops deployed to Europe based on average unit sizes. Note that this measure is only produced for unit types that are displayed in Figure 1.
- *white_personnel_overseas_1000s:* The estimated number of White troops deployed to Europe based on average unit sizes, expressed in thousands. Note that this measure is only produced for unit types that are displayed in Figure 1.
- *black_personnel_overseas_1000s:* The estimated number of Black troops deployed to Europe based on average unit sizes, expressed in thousands. Note that this measure is only produced for unit types that are displayed in Figure 1.

wwi_unit_period_data.csv

Dataset containing information on infantry casualties suffered by individual United States Army deployed to Europe during World War I, created using casualty lists, disaggregated at the half-month level.

- *period:* The half-month period, recorded as YYMMHH, where HH represents which half of the month is referenced. (As an example, 180502 refers to the second half of May 1918.)
- *periodct:* A number corresponding to *period*, used for creating plots where half-months are on the x-axis.
- *core:* A dummy variable for whether the half-month was part of the core (pre-armistice) phase of World War I. Half-months up to and including the first half of November 1918 qualify.
- *post:* A dummy variable for whether the half-month was part of the post-armistice phase of World War I. Half-months starting with and including the second half of November 1918 qualify.
- *unit:* The unit.
- *black:* A binary variable indicating whether the unit was Black.
- *pioninf:* A binary variable indicating whether the unit was a pioneer infantry unit.
- *dead_total:* The total number of soldiers who died in a unit during the relevant half-month. Note that this total includes all deaths in the unit, regardless of race. This distinction is most relevant in including White commanders who may have died while leading all-Black units.
- *dead_spec:* The total number of soldiers who died in a unit during the relevant half-month, conditional on the soldier having the unit's designated race. This distinction is most relevant in excluding White commanders who may have died while leading all-Black units.
- *combat_death:* The total number of combat-related deaths suffered by a unit during the relevant half-month.
- *noncombat_death:* The total number of non-combat-related deaths suffered by a unit during the relevant half-month.
- *cause_unknown:* The total number of deaths with unknown causes suffered by a unit during the relevant half-month.