Data Appendix: "The Social Origins of Female Combatants"

The *Women in Armed Rebellions Dataset* (WARD) contains information on women's participation in armed rebel, insurgent, and terrorist movements. This appendix provides information on the construction and coding of the dataset. We limit our discussion to the variables included in the manuscript.

Defining the Sample of Rebel Organizations:
In order to facilitate data collection, the WARD dataset relies on the Uppsala Conflict Data Program's (UCDP) Dyadic Dataset (Harbom, Lotta, Erik Melander, and Peter Wallensteen (2007) as the initial sample of cases for which to collect data. Datasets created by the UCDP are among the most frequently used datasets for analyzing civil conflict. Because it relies on this existing dataset to define the sample, WARD adopts the UCDP definition of armed conflict: *a contested incompatibility that concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25 battle-related deaths*[1] The UCDP dataset further identifies specific armed groups involved in such "contested incompatibilities" with the state. The UCDP defines these groups as *any non-governmental group of people having announced a name for their group and using armed force to influence the outcome of the stated incompatibility*. These organizations represent the relevant actors for which the creators of WARD collected data on female combatants.[2]

The WARD currently includes groups active between 1979 and 2009. Ultimately, sufficient data was collected to make coding decisions for 211 armed opposition organizations from this initial sample of relevant cases.[3] This represents approximately 75% of the relevant groups listed in the UCDP Dyadic Dataset.

Defining Female Combatants:
WARD draws on existing studies (e.g., Viterna 2013) in conceptually distinguishing among three categories of women connected to rebel organizations: *collaborators*, *non-combatant members*, and *combatants*. The first groups, *collaborators,* include those persons who agree with the group's goals and offer direct or indirect assistance to the organization (food, information, shelter, moral and political support) but are not formally associated with the organizations. This group is not relevant to the data collection efforts. Rather, it focuses on the subset of women that were formal members of the organization: *non-combatant members* and *combatants*.[4]

WARD includes information only on women who had formal ties to the organization and actively and routinely participated (voluntarily or involuntarily) in the conduct of the group's military activities (broadly defined). The definitions of these two groups are based on language commonly used in disarmament, demobilization, and reintegration (DDR) programs sponsored

[1] See the ACD codebook for more details on the definitions of specific terms.
[2] Coups and related military factions are excluded from the WARD.
[3] We note, however, that in our analyses missing values in the predictors notably reduces the number of observations.
[4] The data used in the manuscript at hand reflects only the female participants included in the latter category.

by the United Nations and related international organizations. The definitions used by the agency UN Women in their literature on women in the DDR process distinguish between *female combatants* from *females associated with armed groups* (FAAGs) (UN Women 2012, 22-23).[5] Specifically, they define these categories as follows:

> *Female combatants*: Women and girls who participated in armed conflicts as active combatants using arms.
>
> *Female supporters/females associated with armed forces and groups* (FAAGs): Women and girls who participated in armed conflicts in supportive roles, whether coerced or voluntarily. These women and girls are economically and socially dependent on the armed force or group for their income and social support. Examples include: porters, cooks, nurses, spies, administrators, translators, radio operators, medical assistants, public information workers, camp leaders or women/girls used for sexual exploitation.

In the WARD, *Female Combatants* represent the subset of female members of a rebel organization who are armed and who directly participate in combat activities sanctioned by the rebel leadership. This category includes women employed in frontline combat, female suicide bombers or assassins, female members of armed auxiliary and civil defense forces, and other female members of the organization who receive military training, carry combat weapons, and would be expected to participate in combat operations if called upon. It excludes women who served in noncombat support roles such as fundraisers, recruiters, couriers, cooks, and porters but did not directly engage directly in combat (e.g., FAAGs).

Identifying Source Material:
Because there is no organized location for data on female combatants, the creators conducted an extensive search of news reports, academic accounts, and international and non-governmental organization reports. They enlisted student assistants (graduate and undergraduate) to identify and collect relevant sources. In most cases, two researchers were assigned to a given case in order to maximize the amount of data collected and to serve as a check on the reliability of the information gathered in the searches.

Student researchers were instructed use the following search protocols in searches conducted in Google Scholar, Google, and Lexis-Nexis (in that sequence).

First search (using rebel group name and aliases):

(LTTE OR "Tamil Tigers" OR "Liberation Tigers of Tamil Eelam") AND (girl OR women OR woman OR female) AND (Fighter OR Guerrilla OR Combatant OR Rebels)

Second Search (using country or conflict name):

[5] While the UN differentiates among different groups, it has begun to advocate for the expansion of the definition of "combatant" to include FAAGs in DDR programs because women that participated in armed groups in the latter category are often denied the benefits associated with DDR simply because they did not carry guns or participate directly in the production of violence.

(Sri Lanka) AND (girl OR women OR woman OR female) AND (Fighter OR Guerrilla OR Combatant OR Rebels)

Following web searchers, students were instructed to search potentially relevant books available in the university library.

Following these searches, students were then instructed to include pertinent information in the appropriate location on the coding sheets provided along with citations and the full bibliographical information for the source.

Constructing and Coding the Data:
The variable *Female Combatants* is a binary indicator coded 1 if coders located evidence that the group included female combatants and zero if there was no such evidence. Coding the presence of *Female Combatants* within an organization required evidence that women actively participated in the rebel group in the manner described above. For instance, coders identify groups as having female combatants if they found evidence of any of the following:

- Women using arms during offensive or defensive actions.
- Women operating artillery or anti-aircraft weapons against enemy targets
- Armed women's units (including auxiliary forces) that ever participated in conflict
- Women detonating mines or other explosives against enemy or civilian targets
- Women conducting assassinations
- Women conducting suicide bombings

In order to increase the validity of the coding, at least three independent sources are required to make a determination on a case.

The WARD also includes a variable accounting for the prevalence of women that met the above criteria included in a given rebel group. *Female Combatant Prevalence* is a categorical indicator accounting for the estimated proportion of women that comprised the group's combat force. The four-category indicator accounting for *Female Combatant Prevalence* is constructed as follows:

0) No evidence of Female Fighters
1) Low (< 5%)
2) Moderate (5 - 20%)
3) High (> 20%)

WARD includes a categorical indicator rather than a numeric estimate because the source data are sometimes not fined grained enough to create accurate counts or percentages of female combatants. While specific numeric or percentage estimates exist, this is not always the case. Sources occasionally provide only qualitative descriptions of the extent of women's participation (e.g., "rare", "small numbers", "very few"). This is particularly true of groups that appear to have few combatants. Furthermore, different source materials sometimes provide different estimates of the numbers of women serving as combatants. We return this point below.

Conceptually, this variable reflects a categorical version of the previously described binary *Female Combatant* variable. It is important to note that available sources do not always cleanly differentiate between *Female Combatants* and FAAGs when providing estimates of the number of women within the armed opposition movement. Coders attempt to ensure that they only include women who actively participated in conflict activities as combatants within this indicator. In some cases, however, it is likely that we over- or under-estimate the prevalence of female combatants in some cases. For example, if coders located numerous references that state that more than 20% of the FARC guerrillas are female.[6]

The source data also include numerous references that large numbers of women participate in direct combat activities in large numbers. Based on the available evidence, coders were to able confidently conclude that at least 20% of FARC's combat forces were *female combatants*, and they therefore code the group as "3" on this indicator. By contrast, while several sources that note that perhaps a third of POLISARIO's members were women, the available evidence suggests that women rarely participated directly as "active combatants using arms." Rather, the evidence suggests that the large majority of female POLISARIO members occupied support roles (porters, spies, recruiters, cooks, etc.), which is consistent with the definition for FAAGs but not for *female combatants*. Many other women were formally associated with the military and political apparatuses of the movement, but there is no compelling evidence that they participated in combat. POLISARIO is therefore coded as "0", which reflects *no evidence* of female combatants.[7]

In coding the *Female Combatant Prevalence* variable, WARD coders rely on a relatively restrictive assessment of the available information. As the POLISARIO case illustrates, it is possible that more than a third of its combatants were women (category "3"), but most of the evidence casts doubt on that figure. The creators recognize the inherent imprecision and possibility of error in data collection efforts of this sort. As a result, in addition to the primary *Female Combatant Prevalence* variable, the dataset includes the additional variable *Female Combatant Prevalence (high estimate)*. This variable represents a less restrictive coding of the source material that accounts for the ambiguity and uncertainty inherent in the source material. In this variable, POLISARIO therefore receives a score of "3". This variable also account for cases in which sources include different values for the number of female combatants. For instance, estimates of the proportion of female fighters in the LTTE range from 15% to 30%. WARD uses the more conservative value for the variable *Female Combatant Prevalence* and the higher value for the variable *Female Combatant Prevalence (high estimate)*. This coding procedure permits users to perform robustness checks and reduces concerns that any results might be driven by erroneous coding of challenging or disputed cases.

In practice, roughly 25% of the cases within the sample receive different scores for these two variables. The means and standard deviations are very similar are well: *Female Combatant Prevalence* Mean= 0.74; SD=0.99) and *Female Combatant Prevalence (high estimate)* (Mean=

[6] Estimates range from 25% to as high as 40%. In either case, the FARC exceeded the threshold of 20% for a score of "3" on the *Female Combatant Prevalence* variable.

[7] One source specifically indicated that women did not serve in combat, but other sources are more ambiguous, leaving open the possibility that women might have served in combat.

0.91; SD=1.10). For comparison, we replicated Figures 1 and 2 from the manuscript below but use the "high" estimates of *female combatant prevalence.*

Figure 1a. Distribution of female combatant prevalence in sample (High Estimate)

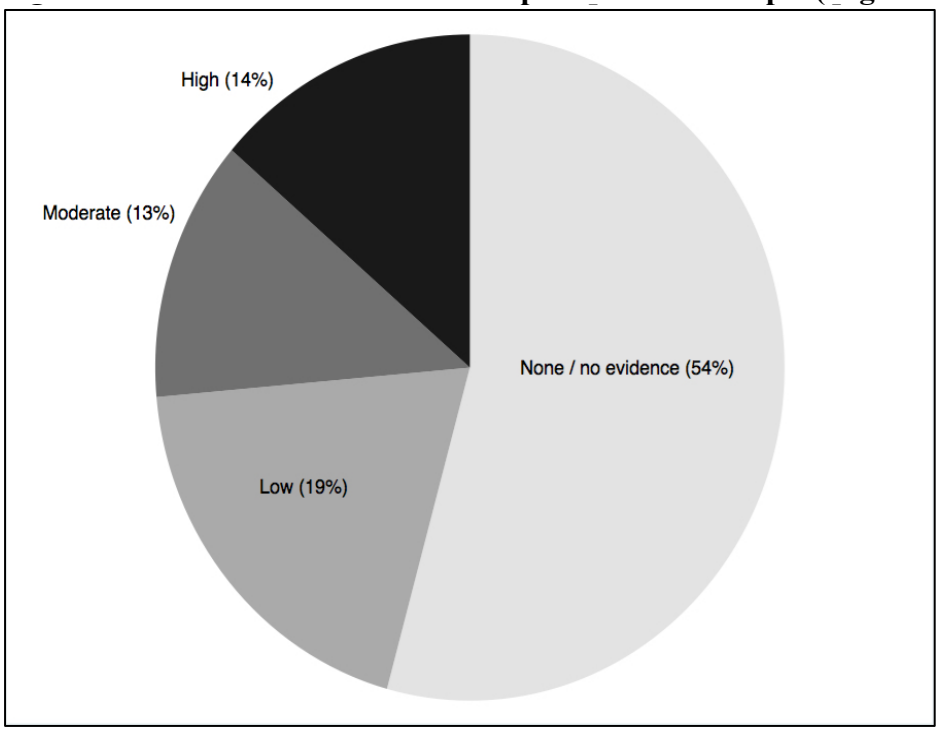
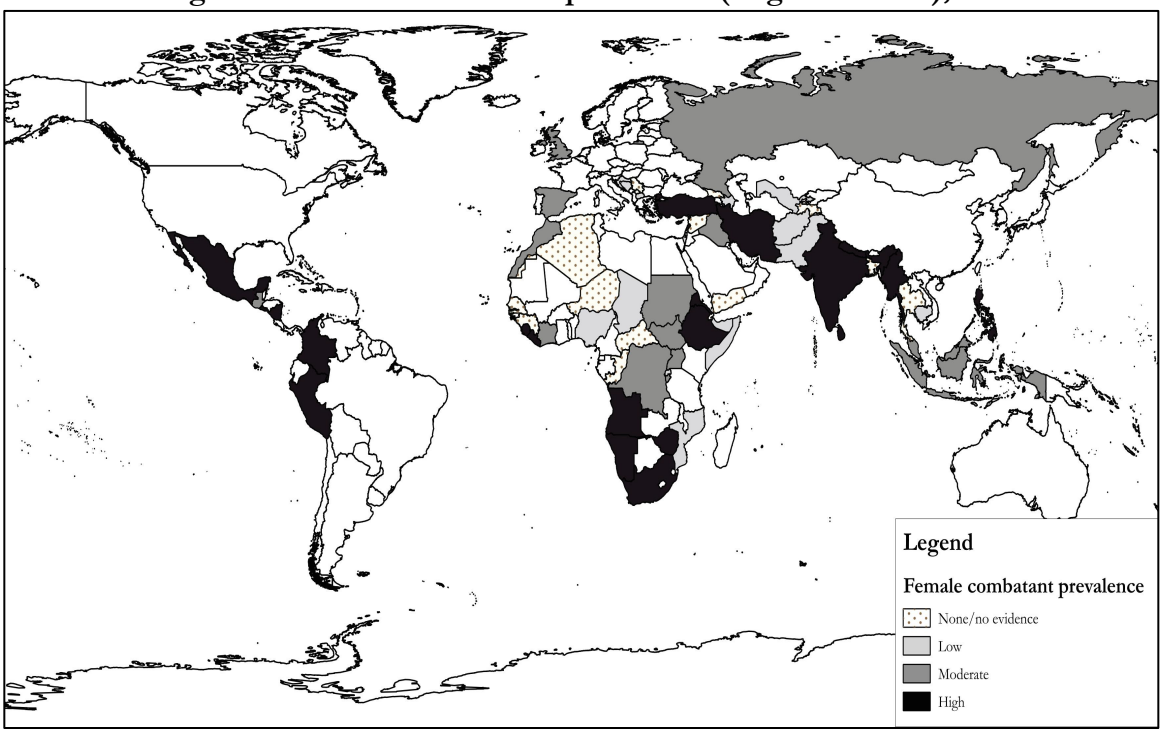


Figure 2a. Female combatant prevalence (High Estimate), 1979-2009



Countries included in Sample:

Afghanistan	Iraq	South Africa
Algeria	Israel	South Africa (Namibia)
Angola	Lebanon	Spain
Azerbaijan	Liberia	Sri Lanka
Bangladesh	Nepal	Sudan
Bosnia and Herzegovina	Nicaragua	Syria
Bosnia-Herzegovina	Macedonia	Tajikistan
Burundi	Malaysia	Thailand
Cambodia	Mali	Turkey
Central African Republic	Mexico	Uganda
Chad	Morocco	United Kingdom
Colombia	Mozambique	Uzbekistan
Cote D'Ivoire	Myanmar (Burma)	Yemen
Democratic Republic of	Niger	Yugoslavia (Kosovo)
Congo (Zaire)	Nigeria	Zimbabwe (Rhodesia)
Djibouti	Pakistan	
Egypt	Papua New Guinea	
El Salvador	Peru	
Ethiopia	Philippines	
Georgia	Republic of Congo	
Guatemala	Russia (USSR)	
Guinea	Rwanda	
India	Senegal	
Indonesia	Sierra Leone	
Iran	Somalia	

*Note: many countries host multiple rebellions.