A Statistical Response to "Traumatic Injury Mortality in the Gaza Strip from Oct 7, 2023, to June 30, 2024: A Capture–Recapture Analysis"

Professor Abraham Wyner
Department of Statistics and Data Science
Wharton School of University of Pennsylvania

Abstract

A January 9th Lancet(Jamaluddine, 2025) study on traumatic injury mortality in the Gaza Strip employs a three-list capture-recapture model, widely used for estimating partially observed populations using multiple random samples. But the methodological framework and assumptions underlying this study raise significant concerns about the reliability and accuracy of its conclusions. The authors estimate that the number of recorded decedents substantially undercounts the true population of traumatic deaths by approximately 35,000, an amount larger by far, than the number of recorded deaths. In this response, we show that this surprisingly large estimate has two causes: 1) a methodological problem and 2) the inclusion of a relatively small but highly influential subset of incomplete data. We propose an alternative estimate of the undercount that is substantially lower: only 7.8 thousand. We also project a combined total population of decedents, civilians and combatants, that closely matches the total reported by Gaza's Ministry of Health.

Introduction

The capture–recapture study utilized three data sources to estimate mortality. The first source, hospital morgue data ('hospital list'), includes detailed records of 22,368 identified decedents reported by the Ministry of Health (MoH) from October 2023 to June 2024, with retrospective efforts to reduce the initial proportion of unidentified deaths. The second source, an online MoH survey ('survey list'), was launched in January 2024 to collect mortality data from Palestinians inside and outside Gaza through social media platforms, capturing retrospective information dating back to October 2023. The third source is social media obituaries ("social media list") that were manually collected from widely used social media platforms.

To be clear, throughout this response, we stipulate the assumption that all three lists of casualties and related data are indeed accurate, even though their accuracy cannot be independently verified due to the ongoing conflict in the region. If the underlying data has been falsified or altered in any substantive way, all the work that relies on it would be undermined.

Under standard assumptions, a three-list capture-recapture model requires 8 parameters: seven to fully specify the probability distribution for the list-membership combinations and one for the unknown population size. In the absence of covariates, a sufficient statistic for the data is the count of observable combinations of list membership. Since there are only 7 observable combinations and there are 8 parameters, the population size cannot be estimated without introducing a dimension-reducing assumption that not only permits the introduction of continuous covariates, but importantly, also models the dependence structure among the lists. In the typical ecological or epidemiological uses of multiple systems capture-recapture methods, the modeling is guided by the existence of bonafide random samples and designed independence among multiple samples, which assures methodological validity.

In this study, samples are not only not random, they are also dependent, arguably by design, but certainly in fact, since the names of the decedents on the hospital lists were widely distributed and publicly updated multiple times. This is in stark contrast to the decedent samples obtained in other conflicts Lum (2010) Ball and Asher. (2002) for which the identities of decedents were not widely disseminated and not constructed in part by the public using social media. For example, the survey list began as a Google doc widely understood to have been created in part for the purpose of soliciting names of decedents not collected by the hospital. This causes a strong negative dependency readily observable in the data with a simple calculation. First, let $I(i, j \mid k)$ denote the event that an entry in list k is also on list i and list j. We calculate $P(I(1, 2 \mid 3))$ first using a model that assumes lists 1 and 2 are conditionally independent given membership in list 3:

$$P(I(2, 1 \mid 3)) = P(I(2 \mid 3)) \cdot P(I(1 \mid 3))$$

We then compare this to the full distribution empirical count

$$P(I(1, 2 \mid 3)) = P(I(2 \mid 3)) \cdot P(I(1 \mid 3, 2))$$

which is the proportion of entries on list 3 in the intersection of all three lists. To measure conditional dependence, we must stratify by age and sex. We consider the 10099 males with ages 15-44. Assuming conditional independence of list 1 and list 3 we would expect 160 on all three lists; there are only 96. For this stratum of sex and age, it follows that

$$\frac{P(I(1 \mid 3, 2))}{P(I(1 \mid 3))} = \frac{96}{160},$$

which means that among the entries on both list 3 and list 2, far fewer are on list 1 than expected using a model that assumes independence between lists 1 and 2 given membership on list 3. This negative dependence pattern is very strong and is present not only in this highly populated stratum, but also in every age-sex grouping.

The bottom line: without a valid and reliable way of simplifying the complete joint distribution between and among the lists, all estimates are, to varying degrees, just guesses.

Modeling Concerns

The authors consider a variety of models, ranging from complete independence among the lists (requiring 3 parameters) to partial dependence using up to 6 parameters that account for pairwise interactions. These models critically rely on untestable assumptions of at least partial independence between lists. Since the joint relationship among the lists cannot be estimated, it is impossible to reliably estimate the undercount without accurate exogenous information about the full joint dependence relationships among the samples.

The authors present a headline estimate of approximately 35,000 decedents that are not on any of the 3 lists. This significantly exceeds the total number of identified decedents (29,271). This estimate is accompanied by a wide 95% confidence interval (26,000–50,000), highlighting extreme variability. Alarmingly, this figure appears to have been selected from among multiple models, with alternative estimates ranging from as low as 20,000 to as high as 48,000. Notably, the individual model most aligned with the reported result (36,906, Row 1 in Table 2 of Jamaluddine (2025)) assumes complete list independence.

The author's estimate of the undercount is formed as a model average using AIC-penalized likelihoods to assign weights to 8 different models. A closer look at the models in Table 2 of Jamaluddine (2025) reveals that 6 of the 8 models are assigned substantive weights, even those that do not account for the known strong negative dependence between the hospital list and the survey list. These models also have high undercount estimates. Different penalty terms would result in different estimates, but there is no way to know which model is truly correct without a three-way interaction term.

The authors use other approaches; a Bayesian model average, which the authors disclose only in the supplementary materials, yields a far lower estimate of 21,000 missing decedents, an estimate that falls entirely outside the confidence interval presented in the main paper. This inconsistency underscores the fragility of the reported estimates and raises serious questions about the transparency and appropriateness of the model selection process.

Data Quality Problem

The problem with the paper is not limited to the modeling alone. A careful examination of the data itself, carefully curated and generously provided by the authors, reveals another glaring problem which inflates the undercount. Although the authors report that the covariate age of death is missing from 957 (30%) entries on the social media list, we found that not a single entry among these 957 is successfully matched to an entry on either of the other two lists. In contrast, of the more than 2000 entries in the social media list with age and sex, more than 1500 (77%) have matches in at least one of the other two lists. These 957 entries, for which match determination could not be established due to missingness, should have been removed. At a minimum, the problem should have been disclosed, especially since exclusion would have an impact on the modeling of the joint distributions.

The impact of inclusion is substantial. When these 957 entries are removed from the social media list, the overlap diagram (Figure 3 in Jamaluddine (2025)) is radically different. Specifically, the social media circle (Figure 1a) overlaps the other lists by 54%. When missing data entries are removed, the circle of social entries (Figure 1b) overlaps by 77%. These 957 are a relatively small number of names, but they have an exceptionally large influence on the estimate.

An Alternative Approach

We demonstrate the impact of these errors by creating another estimate that:

1. Addresses the dependency structure between the hospital list and the survey list.

2. Removes entries with missing gender or age.

We sidestep the problem of modeling the extreme dependence between the hospital list and the survey list by combining lists. This is a well-known technique that goes back to Marks, Seltzer, and Krótki in 1974 (Marks. et al., 1974). In a two-list data problem, the dependence cannot be further identified without introducing untestable assumptions, so we proceed by assuming the independence between the social media list and the combined list. This is just a reasonable and necessary step; but if there were negative dependence, the undercount estimate would be too high, and if there were positive dependence, the undercount estimate would be too low.

The independence assumption allows us to use the Peterson capture-recapture estimator to project the population size, which we apply to four different age strata that broadly reflect groups with large differences in the likelihood that decedents are combatants (Table 3). The total projected number of decedents is 36,078 which is 7,821 more than the 28,257 unique entries (Table 1 and Figure 2). This is very close to the official number of 37,877 decedents reported by the Ministry of Health (an agency that has many reasons to overcount the decedents). We note, for completeness, that if instead of removing the 957 entries on list 3 we could have chosen to include them, but with a match probability model derived from the observed match probability derived from data with age recorded. This would have increased the total projection by about 1240, which is an even closer to the official count. Finally, a further calculation reveals that in all age strata, males are more abundant. The male bias is small (15%) for the youngest age group and largest(112%) for men of fighting age (Table 2).

References

Ball, Patrick, W. B. F. S. J. D. and Asher., J. (2002). Killings and refugee flow in kosovo march-june 1999. *Report to the International Criminal Tribunal for the Former Yugoslavia*.

Jamaluddine, Zeina, e. a. (2025). Traumatic injury mortality in the gaza strip from oct 7, 2023, to june 30, 2024: a capture–recapture analysis. *The Lancet*.

Lum, Kristian, P. M. G. T. e. a. (2010). Measuring elusive populations with bayesian model averaging for multiple systems estimation: a case study on lethal violations in casanare, 1998–2007. *Statistics, Politics, and Policy*.

Marks., E. S., Seltzer, W., and Krótki, K. J. (1974). Population growth estimation: A handbook of vital statistics measurement. *New York: The Population Council*.

Tables

Table 1: Population and Projected Population by Age Group and Gender

Age Group	Gender	Population	Projected Population
0-14	Female	3,610	4,907.0
0-14	Male	4,181	5,140.0
15-44	Female	4,403	6,136.0
15-44	Male	10,220	13,230.0
45-64	Female	2,112	2,830.0
45-64	Male	2,315	2,762.0
65-120	Female	666	900.0
65-120	Male	952	1,263.0

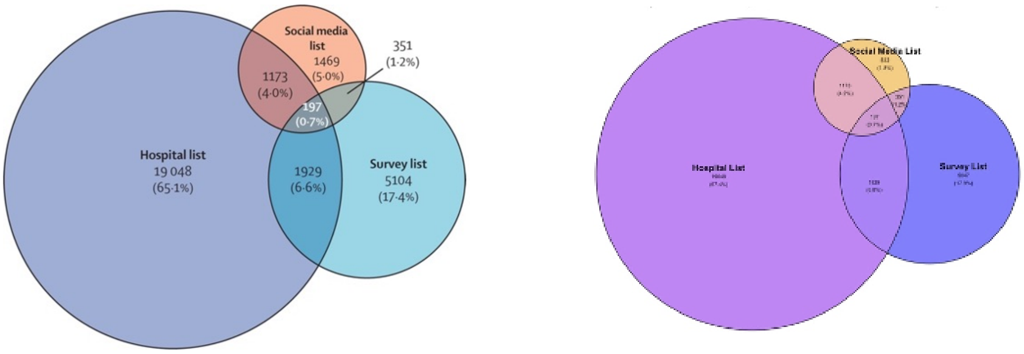
Table 2: Gender Distribution Across Age Groups

Age Category	**Projected Female**	**Projected Male**	**Male Increase (%)**
0-14	3,610	4,181	15.8
15-44	4,843	10,280	112.3
45-64	1,400	2,315	65.4
65-120	646	982	52.0

Table 3: Revised Data Set grouped by Age Category and Gender

Age Group	Gender	Unique Entries	Combined List	Social Media List	Overlap	Overlap Percentage
0-14	Female	3,518	3,153	275	195	0.723
0-14	Male	4,161	4,523	245	211	0.860
15-44	Female	4,403	4,103	264	213	0.776
15-44	Male	16,220	15,283	292	77	0.719
45-64	Female	2,122	1,958	207	152	0.735
45-64	Male	6,186	5,626	239	183	0.763
65-120	Female	546	513	62	46	0.742
65-120	Male	963	944	62	46	0.767

Figures



(a) Original Figure 3 (b) Revised Figure 3

Figure 1: Comparison of data list overlap structure. a) is with all the data including the 957 entries that are missing age, b) revised diagram without missing data

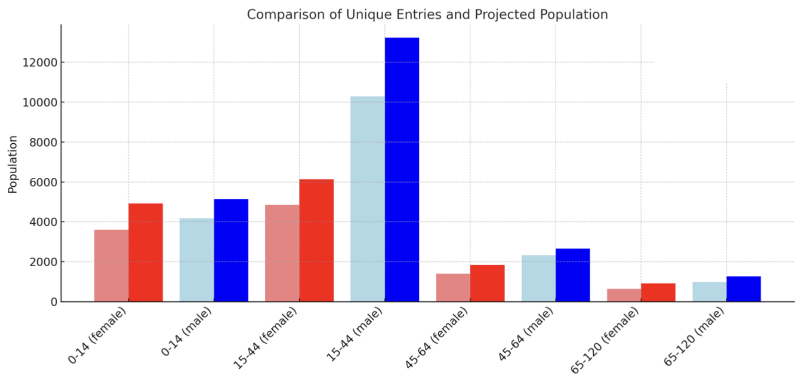


Figure 2: Comparison of number of unique entries to the projected count for every age and sex statum. Blue is male and red is female. The projected count is the darker shade.