Supplementary Appendix for: Survey Estimates of Wartime Mortality*

Gary King[†] Mitsuru Mukaigawara[‡]

July 29, 2025

*Please see GaryKing.org/sibs for the current version of the paper and this Supplementary Appendix.

[†]Albert J. Weatherhead III University Professor, Institute for Quantitative Social Science, Harvard University; GaryKing.org, King@Harvard.edu.

[‡]Ph.D. candidate in Government, Harvard University, MitsuruMukaigawara.com, mitsuru_mukaigawara@g.harvard.edu.

A1 Influence Function of the G-K Estimator

In this supplementary appendix, we build on the robust statistics literature to derive the influence function of the G-K estimator. We first formally define the contaminated distribution and the influence function, which captures the effect of contamination on the overall estimates. Although the influence function does not directly capture the specific impact of contaminated responses that we are interested in, it provides intuition for how the concept of maximum influence in the main text is constructed.

Following Casella and Berger (2001), consider a sample of sibling-level mortalities $M_1, M_2, \cdots, M_J$ drawn from a distribution with cumulative distribution function F^S (i.e., the distribution of mortalities at time 2). For notational simplicity and to follow conventions in the robust statistics literature, denote a statistic $T = (M_1, \cdots, M_J)$ as $T(F^S)$. We define a contaminated distribution of mortalities as follows.

Definition 1 (δ-Contaminated Distribution of Mortalities)**.** *M_j follows a δ-contaminated distribution of mortality, $M_j \sim F_\delta$ if*

$$M_j \sim \begin{cases} F^S & \text{with probability } 1-\delta \\ x & \text{with probability } \delta, \end{cases}$$

where $\delta \in [0,1]$ and $x \in [0,1)$.

As Casella and Berger (2001) notes, F_δ is a mixed distribution of F^S and a point (mass) $x \in [0,1)$. With Definition 1, we next define the influence function, which examines the degree to which a contaminated observation affects a test statistic.

Definition 2 (Influence Function (Casella and Berger, 2001))**.** *The influence function of a statistic $T(F^S)$ at a point x is*

$$IF(T,x) = \lim_{\delta \to 0} \frac{T(F_\delta) - T(F^S)}{\delta}.$$

Intuitively, the influence function measures the marginal effect of adding an infinitesimal fraction of contaminated data to the overall estimate.

Recall that the G–K estimator is a statistic based on sibling-level mortalities. Definition 2 therefore prompts us to derive its influence function.

Theorem 1 (Influence Function of the G-K Estimator)**.** *Denote the sample mean of M_j under F^S as $\overline{M}^S$. The influence function of the G-K estimator is:*

$$IF(\tilde{q}, x) = \frac{\frac{x\left(\frac{1}{1-\overline{M}^S}+\frac{\hat{\xi}}{J}\right)-\left(\frac{\overline{M}^S}{1-\overline{M}^S}+\frac{\hat{\xi}}{J}\right)}{1-x}+\frac{\hat{\xi}}{J}}{\left(\frac{1}{1-\overline{M}^S}+\frac{\hat{\xi}}{J}\right)^2}.$$

Proof. To prove Theorem 1, it is useful to introduce the concept of the Gâteaux derivative.

Definition 3 (Gâteaux derivative (Gateaux, 1913))**.** *Let F and G be a cumulative density function. Then, the Gâteaux derivative of a statistic T in the direction G is*

$$L(x) = \lim_{\delta \to 0} \frac{T\left((1-\delta)F + \delta G\right) - T(F)}{\delta}.$$

Definitions 2 and 3 show that the influence function is a Gâteaux derivative. A key implication is that the theorem of calculus valid for Gâteaux derivatives also applies to influence functions, such as chain and quotient rules. With this, we now derive the influence function of the G-K estimator and prove Theorem 1.

Now, rewrite the G-K estimator as follows:

$$T(F^S) = \tilde{q} = \frac{\sum_{j=1}^{J} \frac{\overline{M}^S}{1-\overline{M}^S} + \hat{\xi}}{\sum_{j=1}^{J} \frac{1}{1-\overline{M}^S} + \hat{\xi}} = \frac{J\frac{\overline{M}^S}{1-\overline{M}^S} + \hat{\xi}}{J\frac{1}{1-\overline{M}^S} + \hat{\xi}} = \frac{N(F^S)}{D(F^S)},$$

where $\overline{M}^S$ is the mean of $M_j \sim F^S$. Then, by the quotient rule,

$$IF(\tilde{q}, x) = \frac{IF(N(F^S), x)D(F^S) - N(F^S)IF(D(F^S), x)}{D(F^S)^2}.$$

Notice that $\hat{\xi}$ is estimated externally and thus $IF\left(\hat{\xi}, x\right) = 0$. Therefore, by linearity,

$$\begin{aligned} IF(N(F^S), x) &= IF\left(J\frac{\overline{M}^S}{1-\overline{M}^S}, x\right) + IF\left(\hat{\xi}, x\right) \\ &= \lim_{\delta \to 0} \frac{\left\{(1-\delta)J\frac{\overline{M}^S}{1-\overline{M}^S} + \delta J\frac{x}{1-x}\right\} - J\frac{\overline{M}^S}{1-\overline{M}^S}}{\delta} + 0 \\ &= J\left(\frac{x}{1-x} - \frac{\overline{M}^S}{1-\overline{M}^S}\right). \end{aligned}$$

Similarly,

$$IF(D(F^S), x) = J\left(\frac{1}{1-x} - \frac{1}{1-\overline{M}^S}\right).$$

Therefore,

$$\begin{aligned}
IF(\tilde{q},x) &= \frac{IF(N(F^S),x)D(F^S) - N(F^S)IF(D(F^S),x)}{D(F^S)^2} \\
&= \frac{J\left(\frac{x}{1-x} - \frac{\overline{M}^S}{1-\overline{M}^S}\right)\left(J\frac{1}{1-\overline{M}^S} + \hat{\xi}\right) - \left(J\frac{\overline{M}^S}{1-\overline{M}^S} + \hat{\xi}\right)J\left(\frac{1}{1-x} - \frac{1}{1-\overline{M}^S}\right)}{\left(J\frac{1}{1-\overline{M}^S} + \hat{\xi}\right)^2} \\
&= \frac{\left(\frac{x}{1-x} - \frac{\overline{M}^S}{1-\overline{M}^S}\right)\left(\frac{1}{1-\overline{M}^S} + \frac{\hat{\xi}}{J}\right) - \left(\frac{\overline{M}^S}{1-\overline{M}^S} + \frac{\hat{\xi}}{J}\right)\left(\frac{1}{1-x} - \frac{1}{1-\overline{M}^S}\right)}{\left(\frac{1}{1-\overline{M}^S} + \frac{\hat{\xi}}{J}\right)^2} \\
&= \frac{\frac{x\left(\frac{1}{1-\overline{M}^S} + \frac{\hat{\xi}}{J}\right) - \left(\frac{\overline{M}^S}{1-\overline{M}^S} + \frac{\hat{\xi}}{J}\right)}{1-x} + \frac{\hat{\xi}}{J}}{\left(\frac{1}{1-\overline{M}^S} + \frac{\hat{\xi}}{J}\right)^2}.
\end{aligned}$$

■

A key implication of Theorem 1 is that the influence function of the G–K estimator is bounded from below but *unbounded from above*. Intuitively, by examining the limits of both influence functions as $x \to 0$ and $x \to 1$, we see

$$\lim_{x\to 1} IF(\tilde{q},x) = \infty, \qquad \lim_{x\to 0} IF(\tilde{q},x) = \frac{-\frac{\overline{M}^S}{1-\overline{M}^S}}{\left(\frac{1}{1-\overline{M}^S} + \frac{\hat{\xi}}{J}\right)^2}.$$

This implies that an over-reported (contaminated) observation can distort the G-K estimate substantively, consistent with our statistical intuition.

References

Casella, George and Roger Berger (2001). *Statistical Inference*. 2nd. Cengage Learning. ISBN: 978-0534243128.

Gateaux, René (1913). "Sur les fonctionnelles continues et les fonctionnelles analytiques". In: *Comptes rendus hebdomadaires des séances de l'Académie des sciences* 157, pp. 325–327.