There may have been 14 undocumented Korean "comfort women" in Palembang, Indonesia

Patrick Ball[*]
Ethan Hee-Seok Shin[†]
Hyerin Yang[‡]

26 December 2018

Overview

How many Korean women were held in sexual slavery by the Japanese military during World War Two in Palembang, Indonesia? The shortage of surviving records has made it difficult to estimate with precision the total number of World War II-era "comfort women," locally or globally. Including the records of women who have been documented plus those who were never documented, we estimate that there were likely a total of approximately 137 Korean "comfort women", including the 123 who were identified on at least one of the lists, in Palembang during World War Two.[1] Crucially, our analysis estimates the existence of approximately 14 women who were not documented on any of the three lists currently available.

We have three lists identifying the women by name, lists B, C, and D, provided by longtime "comfort women" researcher Dr Kang Jeong-sook. These lists have been linked such that we know which records on which lists refer to the same women.[2] The intersections among the lists can be used to estimate the total number of "comfort women" held in Palembang, including both those who were documented on these lists and those who have were not

---

[*]Human Rights Data Analysis Group

[†]Transitional Justice Working Group

[‡]Transitional Justice Working Group

[1]The two-tailed equal probability 95% credible interval is [124; 198]. The estimate is explained in the Estimating the Dark Figure section.

[2]Linking unique people across duplicated records in multiple lists is called "record linkage" or "database deduplication." For a non-technical introduction, see this page; for a more technical explanation, see Christen (2012).

documented on these lists. This technique therefore enables us to estimate the "dark figure," the number of women whose names were not recorded.

The estimates are made using multiple systems estimation (MSE). This link provides a non-technical introduction. For a very slightly more mathematical explanation and a description of the method's applications in human rights, see Manrique-Vallier, Price, and Gohdes (2013). Most relevantly, MSE has been used recently in estimates of the scale of modern slavery (Bales, Hesketh, and Silverman 2015). The rest of this memo explains what the data are, and how we used the data to make this estimate.

Historical Context

As the Imperial Japanese Army (IJA) embarked upon the armed conquest of China in the 1930s, it supervised the expansion and operation of military brothels, euphemistically called "comfort stations."[3] The comfort stations were located in the occupied territories to be used by Japanese soldiers bogged down in a war of attrition. The all-out war—with China from 1937 and with the United States, Britain, and the Netherlands from 1941—expanded the demand for military brothels. In resposne, Imperial Japan secretly sanctioned the trafficking of up to 200,000 women and girls, mostly from Korea, which was under Japanese colonial rule in 1910–1945, by utilizing private contractors (Kang 2010). The unprecedented "comfort women" system "included gang rape, forced abortions, humiliation, and sexual violence resulting in mutilation, death, or eventual suicide in one of the largest cases of human trafficking in the 20th century."[4] Some Korean "comfort women" ended up in Palembang, the capital of South Sumatra province, occupied by Japanese forces during 1942–1945.[5]

Data

We have three lists available for this analysis:

---

[3]For the general historical background, see Hicks (1995) and Yoshiaki (2000).

[4]United States House of Representatives Resolution 121, 110th Congress (2007-2008), July 30, 2007, A resolution expressing the sense of the House of Representatives that the Government of Japan should formally acknowledge, apologize, and accept historical responsibility in a clear and unequivocal manner for its Imperial Armed Forces' coercion of young women into sexual slavery, known to the world as "comfort women", during its colonial and wartime occupation of Asia and the Pacific Islands from the 1930s through the duration of World War II; European Parliament resolution of 13 December 2007 on Justice for the 'Comfort Women' (sex slaves in Asia before and during World War II) (B6-0542/2007) . See also Canada House of Commons Motion 291, 39th Parliament, 2nd Session, November 28, 2007.

[5]See (Republic of Korea) Commission on Verification and Support for the Victims of Forced Mobilization under Japanese Colonialism [일제강점하강제동원피해진상규명위원회], "The Fact-Finding Investigation of the Lists of Women Mobilized to Indonesia [인도네시아동원여성명부에관한진상조사] (Investigator in Charge: Kang Jeong-sook. Approved by the Commission on November 19, 2009" [in Korean]. For the description of the three lists analyzed in this report and the "comfort women" named in the three lists, see pp. 25–82.
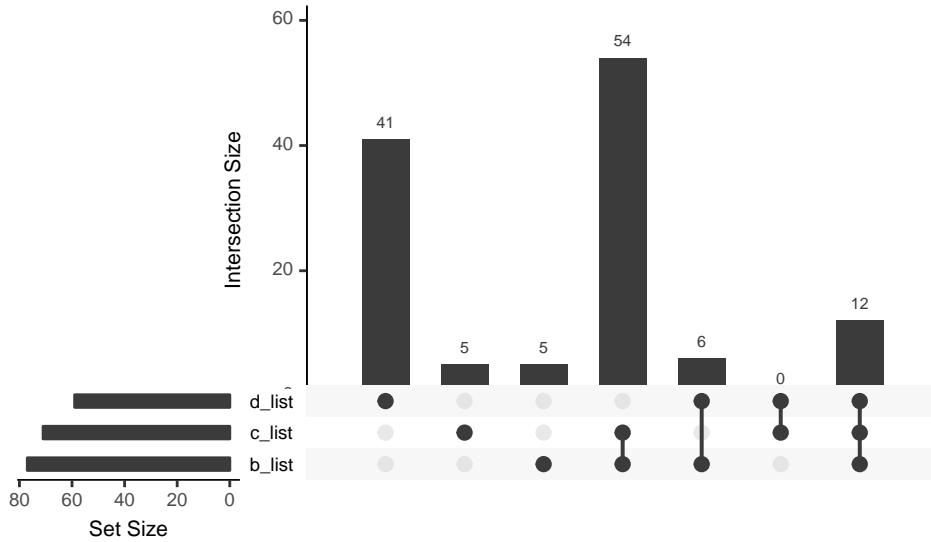
Figure 1: Number of women on lists, by intersection type

During the war, the IJA kept "home-absence lists (rusu-meibo 留守名簿)" for its military personnel, so called as they were absent from their homes in Japan or Korea, for administrative purposes. Koreans can be distinguished from their home address, location of family register (honseki 本籍), and distinctive Korean name style. The 9th Army Hospital in Palembang kept such a list. The 78 Korean women who were added to that list as "temporary nurses" on 22 August 1945, one week after Japan's announcement of surrender, were in all likelihood "comfort women" because a number of them have been identified as "comfort women" and there were few trained Korean nurses to be dispatched to war zones. This is B List.

After Japan's surrender in 1945, the IJA was demobilized and its personnel overseas returned to mainland Japan. The Japanese authorities created the "repatriation list (fukugen-meibo 復元名簿)" to account for the returnees. The 9th Army Hospital in Palembang kept such a list as well. Like the home-absence list, it can be assumed that the Korean women listed in the repatriation list as "temporary nurses" were "comfort women." This is C List.

With its defeat in 1945, Imperial Japan lost its colonial possession of Korea. The Koreans in Indonesia formed their own association to voice their interests. They compiled their own list of Korean in the Palembang area. This is D List.

Intersections

The first exploration of the datasets is to look directly at their intersections or overlaps, that is, how many women were documented by only one list, by two lists but not the third, or by all three lists. Asked differently, which women were documented on which lists?

Figure 1 considers the B, C, and D collections, showing the number of records found in each
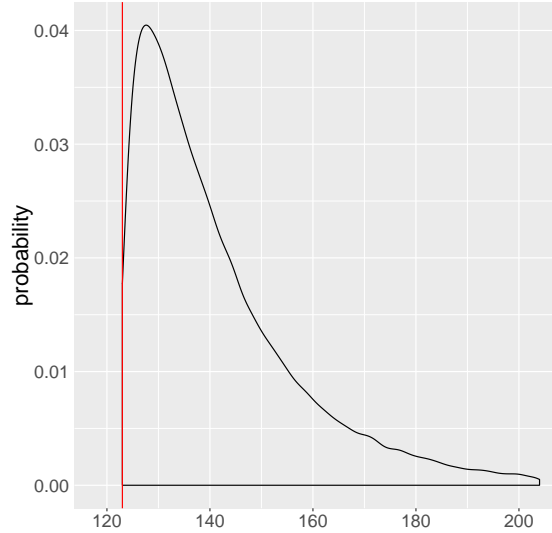
Figure 2: Estimated number of "comfort women," including those never documented, with the documented total shown in red

of the combination of list overlaps.[6] The leftmost bar has a dot in the `d_list` row at the bottom of the graph. This indicates that these data are found in list D, but the empty circles in the other two rows indicate that the people found in D are not found in B or C.

The rightmost bar indicates that for 12 people, their data is found in lists B, C, and D. The dots are joined by a heavy line, indicating the intersection status of these records: these are the records found in all three datasets. The largest category are the 54 people found in the B and C lists but not the D list.

Estimating the Dark Figure

Figure 1 shows how many women were documented on lists B, C, and D.[7] But how many were never documented? By definition, this number is not known, so we have to estimate it.

To estimate, we use the following intuition: imagine two dark rooms. We cannot see inside them, and the only tool we have to explore their size is a handful of small rubber balls. The balls do not make any sound when they hit the walls or ceilings, but they make a small sound – *click* – when they hit each other. We throw the balls into the first room, and listen: click, click, click. We gather the balls and throw them into the second room with equal force: click. Our intuition is that the second room is larger because the balls are able to spread out and therefore strike each other less frequently.

Using databases, in some sense we "throw" the databases into the "room" of all abducted

---

[6]The graph is from the UpSetR package. See Lex and Gehlenborg (2014).

[7]The software used for the estimates is Bayesian Non-Parametric Latent-Class Capture-Recapture by Indiana University Statistics Professor Daniel Manrique-Vallier.

4

women. When the databases document the same person, it is as if the databases collided, making a click. We can use the number of people documented on more than one list and the total size of the databases to estimate the total number of women, including those not on any of the lists.

A technical and historical introduction to this approach can be found in Lum, Price, and Banks (2013). The specific method we will use here was developed by Manrique-Vallier (2016). Crucially, the estimate below includes an analysis of how certain we can be about the estimate.

The estimate and the uncertainty are shown together in Figure 2. The curve shows the probability at each estimated number of women. Note that the red line on the left indicates the number of women who were observed on one or more of the three lists (123). The probability of any number of women below the observed number must be zero – after all, we observed this number, so there cannot be fewer.

We are presenting an estimate of the total number of women held in sexual slavery. The estimate we present, 137, is center of the likely values, that is, the median of the distribution shown in Figure 2. We don't know how many women might have been undocumented, so the estimate is in a "credible interval," which is a little bit like what is sometimes called a "margin of error." The interval in this case is [124; 198]. This means that with these data and this approach, there is a 95% probability that the true total number of women falls in the range [124; 198].

The credible interval expresses the uncertainty around the estimate. In this case, we are certain about the left side of the interval because we know the number of unobserved women must be no less than zero. However, the right side has more uncertainty: the true number could be considerably larger than our estimate.

This estimate enables us to make a specific probability statement about the number of undocumented "comfort women." The probability that there were 10 or more *undocumented* "comfort women" in Palembang is 0.64. This means that there is a more-than-likely chance that there were at least ten undocumented "comfort women."

The dark figure may be quite substantial: there is a 0.08 probability that the number of undocumented women is *50* or more. While this probability is small, it may be large enough to warrant additional searching for new information about women held in Palembang during World War Two.

Conclusion

This brief analysis shows that there were 123 "comfort women" documented on the lists used here, and there were likely 14 additional "comfort women" who are not documented on these three lists. There is a 0.64 probability that there were 10 or more undocumented "comfort women." In total, we estimate that there were 137 "comfort women" held in Palembang.

There are a few lists of "comfort women" in other regions, but they do not provide sufficient information for estimates like the one we have made for Palembang.

We would welcome more lists of "comfort women" that may enable estimates in other locations. Lists might be in the possession of academic researchers, government archivists, or individuals. These lists can be sent to us at `cwdarkfigure@gmail.com`.

Acknowledgements

References

Bales, Kevin, Olivia Hesketh, and Bernard Silverman. 2015. "Modern Slavery in the UK: How Many Victims?" *Significance* 12 (3): 16–21. doi:10.1111/j.1740-9713.2015.00824.x.

Christen, Peter. 2012. *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection.* Springer Science & Business Media.

Hicks, George. 1995. *The Comfort Women: Japan's Brutal Regime of Enforced Prostitution in the Second World War.* New York: W.W. Norton.

Kang, Jeong-sook. 2010. "A Study for Colonialism Characteristics of Japanese Military 'Comfort Women' System: Focusing on Korean 'Comfort Women' [일본군위안소업자의지위와역할에관한연구]." PhD thesis, Sungkyunkwan University. www.riss.krdatabase.

Lex, Alexander, and Nils Gehlenborg. 2014. "Sets and Intersections." *Nature Methods* 11. Nature Publishing Group: 779. https://doi.org/10.1038/nmeth.3033.

Lum, Kristian, Megan Emily Price, and David Banks. 2013. "Applications of Multiple Systems Estimation in Human Rights Research." *The American Statistician* 67 (4). Taylor & Francis: 191–200.

Manrique-Vallier, Daniel. 2016. "Bayesian Population Size Estimation Using Dirichlet Process Mixtures." *Biometrics* 72 (4). Wiley Online Library: 1246–54.

Manrique-Vallier, Daniel, Megan E. Price, and Anita Gohdes. 2013. "Multiple Systems Estimation: Techniques for Estimating Casualties in Armed Conflicts." In *Counting Civilian*

Casualties, edited by Taylor B. Seyboldt, Jay D. Aronson, and Baruch Fischoff, 165–84. Oxford: Oxford UP.

Yoshiaki, Yoshimi. 2000. *Sexual Slavery in the Japanese Military.* Translated by Suzanne O'Brien. New York: Columbia University Press.