



인도네시아 팔렘방에 기록에 없는 한국인 “위안부” 14명 있었을 수 있다

패트릭 볼 (Patrick Ball)*
신희석 (Ethan Hee-Seok Shin)†
양혜린 (Hyerin Yang)‡

2018년 12월 26일

개요

제2차 세계대전 중 인도네시아 팔렘방 지역에는 얼마나 많은 한국인 여성들이 일본군 성노예로 끌려갔을까? 현존하는 기록의 부족으로 제2차 세계대전 중 “위안부”의 전체 규모나 지역별 규모를 정확히 추정하기는 어렵다. 하지만 우리는 역사의 기록으로 남은 “위안부” 명부들을 바탕으로 통계적 분석을 실시하였고, 이를 근거로 명부들에서 적어도 한 번은 식별되는 123명을 포함하여 대략 137명의 한국인 “위안부”가 팔렘방에 있었을 것으로 추정한다.¹ 여기서 주목할 것은 현존하는 3개의 명부 중 어디에도 기록되지 않은 “위안부” 14명이 더 있을 수 있음을 시사한다는 점이다.

우리는 “위안부” 문제를 오랫동안 연구해온 강정숙 박사로부터 “위안부”를 이름으로 식별할 수 있는 3개의 명부(B, C, D 리스트)를 입수하였다. 각 명부에서 확인 가능한 동일 인물을 식별하는

* [Human Rights Data Analysis Group](#) (인권데이터분석그룹, 미국 샌프란시스코)

† [Transitional Justice Working Group](#) (전환기정의워킹그룹, 한국 서울)

‡ [Transitional Justice Working Group](#) (전환기정의워킹그룹, 한국 서울)

¹ 양측검정에 따른 95%의 신뢰구간은 124명에서 198명이다. 자세한 설명은 아래 [‘다크 피겨\(Dark Figure, 암흑 수치\)의 추정’](#) 부분 참조.

기록(이름)을 고리로 이 명부들은 서로 연결되어 있다. ² 명부들에 중복 기재된 교차기록(intersections)을 분석하면 명부들에 기록된 여성들과 기록되지 않은 여성들을 포함한 팔렘방의 “위안부” 전체 규모를 추정하는 것이 가능하다. 즉, 이 통계 기법은 기록에 없는 여성들의 수, 즉 “다크 피겨(Dark Figure, 암흑 수치)”를 추정할 수 있게 해준다.

우리가 사용한 기법은 다중시스템추정(MSE; Multiple Systems Estimation)이다. 일반인을 위한 MSE 기법의 설명은 [이 링크](#)를 참조할 수 있고, 보다 구체적인 수학적 해설과 인권 분야에서의 MSE 기법 적용에 대한 설명은 [만리크-밸리어\(Manrique-Vallier\), 프라이스\(Price\), 고데스\(Gohdes\)의 저술\(2013년\)](#)을 참조할 수 있다. 특히 주목할 만한 MSE 기법의 응용 사례로는 현대판 노예제 규모를 추정한 [발레스\(Bales\), 헤스케스\(Hesketh\), 실버만\(Silverman\)의 연구\(2015년\)](#)가 있다. 이하의 내용은 우리가 사용한 데이터와 이 데이터를 이용하여 팔렘방의 “위안부” 추정치를 도출한 방법을 설명한다.

역사적 배경

1930년대 중국 침략에 나선 일본군은 “위안소”라 미화되어 불린 군 매춘소의 확장과 운영을 관장하였다.³ 위안소는 중국 전선에서 끝없는 소모전의 수렁에 빠진 일본 군인들을 위해 점령지에 세워졌다. 1937년 중국과의 전면전, 1941년 미국, 영국, 네덜란드와의 전면전 개시로 군 매춘소 수요는 급증하였다. 이에 일제는 업자들을 앞세워 주로 1910년부터 1945년까지 일제 식민지였던 조선에서 성인 및 미성년 여성 최대 20만 명의 인신매매를 몰래 장려하였다([강정숙, 2010](#)). 전례가 없는 “위안부” 제도는 “집단강간, 강제 유산, 치욕과 신체 훼손, 사망, 궁극적으로 자살까지 초래한 성폭력에 이르기까지 20세기 최대 규모의 인신매매 사례 중 하나”였다.⁴ 조선인

² 복수의 명부에 중복 기재된 고유 인물을 연결하는 것을 “기록 연결(record linkage)”, “데이터베이스 중복제거(database deduplication)”이라 한다. 기술적이지 않은 입문을 위해서는 [이 링크](#) 참조. 더 기술적 설명을 위해서는 [Christen\(2012\)](#) 참조.

³ [Hicks \(1995\)](#) 및 [Yoshiaki \(2000\)](#) 참조.

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](#).

“위안부” 중 일부는 1942년부터 1945년까지 일본군이 점령한 남수마트라주의 주도인 팔렘방에 보내졌다.⁵

데이터

상기한 바와 같이 우리는 통계적 분석을 위하여 3개의 명부를 사용하였다.

제2차 세계대전 중 일본군은 행정처리를 위하여 일본이나 조선에서 부재 중인 군인원에 대하여 유수명부(留守名簿)를 작성, 관리하였다. 조선인은 집 주소, 본적(本籍), 조선식 이름으로 구별할 수 있다. 팔렘방의 제9육군병원(제9육군병원)은 해당 명부를 관리하였다. 일본의 항복 선언 일주일 후인 1945년 8월 22일 제9육군병원 유수명부에 “임간(임시 간호부)”으로 추가된 78명의 조선인 여성은 모든 가능성을 따져봤을 때 “위안부”였을 것으로 보인다. 이 중 일부는 “위안부”로 식별되었고, 전장에 파견 가능한 숙련된 조선인 간호사는 거의 없었기 때문이다. 제9육군병원 유수명부를 B 리스트라 한다.

1945년 항복 후 일본군은 무장해제되었고 해외 파병인원은 일본 본토로 귀환하였다. 일본 당국은 귀환자를 확인하기 위하여 복원명부(復元名簿)를 만들었는데, 제9육군병원도 복원명부를 작성하였다. 유수명부와 마찬가지로 이 복원명부에 “임간부(임시 간호부)”로 기재된 조선 여성들은 “위안부”였던 것으로 간주할 수 있다. 제9육군병원 복원명부를 C 리스트라 한다.

1945년 패망으로 일제의 한반도 식민통치는 종식되었다. 인도네시아의 한인들은 자신들의 권익 주장을 위하여 조선인회를 결성하였다. 팔렘방조선인회는 자체적으로 팔렘방 지역의 한인 명부를 작성하였다(원본은 없고, 당시 강석재가 수첩에 기록한 명단에 기초). 팔렘방조선인회명부를 D 리스트라 한다.

교차기록 (Intersections)

3개 리스트의 조사에서 제일 먼저 살펴볼 것은 데이터들간의 교차 또는 중복되는 기록들이다. 즉, 리스트에 기록된 여성들이 1개의 리스트에서만 기록되었는지, 2개 혹은 3개의 리스트에 중복으로 기록되었는지를 확인하는 것이다. 달리 말하자면, 어느 정도의 여성이 어느 리스트에 기록되었는지를 확인하는 것이다.

그림1은 B, C, D 리스트에서 중복되어 나타나는 기록이 얼마나 되는지를 보여준다.⁶ 가장 좌측에 있는 막대그래프는 D 리스트에만 점이 찍혀 있고 B 리스트와 C 리스트는 공란으로 되어 있다.

⁵ 대한민국 국무총리 소속 일제강점하강제동원피해진상규명위원회, “인도네시아 동원 여성 명부에 관한 진상조사 (조사책임: 강정숙)” 2009년 11월 19일. 본 보고서에서 분석한 3개의 명부에 대한 설명과 기록된 명단은 이 진상조사 보고서의 25-82쪽 참조.

이는 D 리스트에는 이름이 발견되었지만 나머지 B 리스트와 C 리스트에는 이름이 발견되지 않는 여성의 수치 데이터를 그래프로 표시한 것이다.

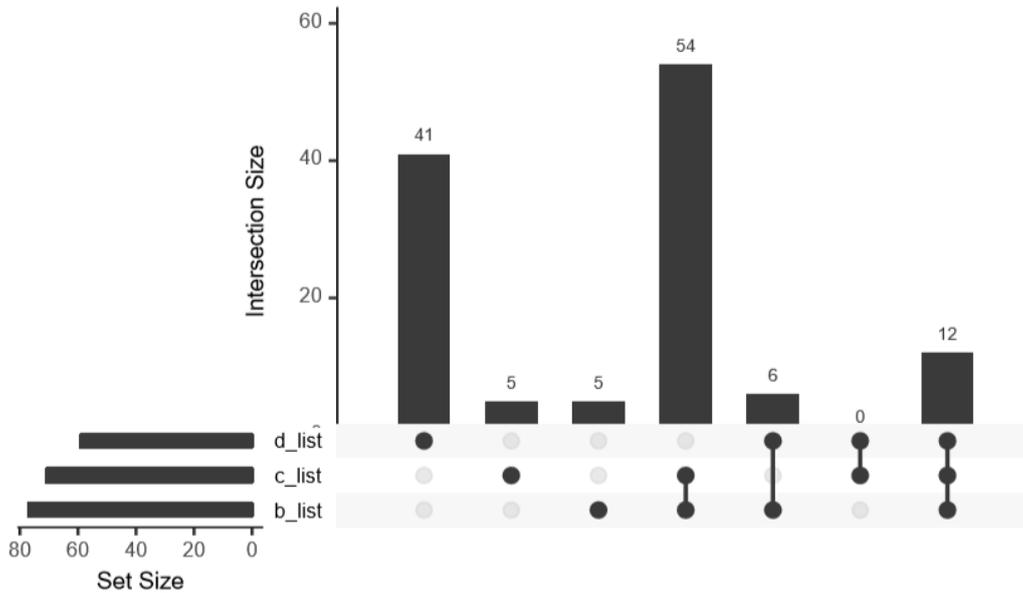


그림1: 명부별 고유기록과 명부간 교차기록으로 구분한 여성 수

가장 우측 막대그래프에 표기된 12명은 B, C, D 리스트 모두에서 이름이 중복으로 발견되었다는 것을 의미한다. 이 막대그래프 아래의 3개의 점들을 굵은 선으로 연결되어 있는데, 이는 12명이 3개의 리스트 모두에서 기록이 발견되는 교차기록(intersection)임을 표시한 것이다. 가장 많은 수가 확인된 부류는 B 리스트와 C 리스트에서 공통으로 이름이 발견되지만 D 리스트에서는 이름이 발견되지 않는 부류로, 이 곳에 해당하는 여성은 54명에 이른다.

다크 피겨(Dark Figure, 암흑 수치)의 추정

그림1은 B, C, D 리스트에 얼마나 많은 여성들이 기록되었는지를 보여준다.⁷ 하지만 3개 리스트 어디에도 기록되지 않은 “위안부”는 몇 명일까? 이 수치는 기록으로는 파악할 수 없으므로, 추정할 수밖에 없다.

이 수치를 추정하기 위하여, 우리는 다음과 같은 직관을 이용한다. 두 개의 암실(two dark rooms)이 있다고 가정하자. 깜깜한 암실 안을 들여다볼 수는 없고, 암실의 크기를 가늠하기 위하여 주어진 유일한 도구는 몇 개의 작은 고무공들이다. 고무공들은 벽이나 천장과 부딪힐 때는 아무런 소리를 내지 않지만 고무공들이 부딪힐 때에는 “클릭(click)”이라는 작은 소리를

⁶ 이 그림은 [UpSetR\(업셋 알\)](#) 패키지로 추출하였다. [렉스\(Rex\)와 게هل렌보르\(Gehlenborg\)의 2014년 논문](#)을 참조할 수 있다.

⁷ 다크 피겨 추정에는 인디애나대학교 통계학 교수 [대니얼 만리크-밸리어\(Daniel Manrique-Vallier\)](#)의 [Bayesian Non-Parametric Latent-Class Capture-Recapture](#) 소프트웨어를 사용하였다.

낸다. 첫 번째 암실에 고무공들을 던지면 “클릭”, “클릭”, “클릭” 세 번 소리가 들린다. 고무공들을 수거하여 두 번째 암실에서 동일한 힘으로 고무공들을 던졌더니, 이번에는 “클릭” 한 번만 소리가 들린다. 직관에 따르면 두 번째 암실이 첫 번째 암실보다 넓기 때문에, 고무공들이 더 멀리 흩어질 수 있고, 서로 부딪히는 빈도가 줄어들 수 있음을 알 수 있다.

우리는 데이터베이스들을 모든 “위안부” 여성들로 구성된 “암실”에 “던져놓는다”. 별개 데이터베이스들에 동일한 인물이 기록되어 있다면, 이는 “암실” 안에서 고무공들이 부딪혀서 “클릭” 소리를 내는 것과 같다. 우리는 1개 이상의 리스트에 기재된 사람들의 수치와 데이터베이스들의 전체 크기를 사용하여 전체 “위안부” 여성의 수치를 추정할 수 있다. 여기에는 어느 리스트에도 기록되지 않은 여성들까지 포함된다.

이 접근법에 대한 기술적, 역사적 소개는 [럼\(Lum\), 프라이스\(Price\), 뱅스\(Banks\)의 2016년 논문](#)을 참조할 수 있다. 우리가 이용한 특정한 기법은 [맨리크-벨리어\(Manrique-Vallier\)가 개발하였다\(2016년 논문 참고\)](#). 여기서 주목할 것은 아래의 추정치는 우리가 추정치에 대해 얼마나 확실할 수 있는지에 대한 분석까지 포함하고 있다는 점이다.

[그림2](#)는 추정치와 불확실성을 함께 보여준다. 곡선은 각 “위안부” 전체 추정치가 맞을 확률(probability)을 보여준다. 왼쪽 끝의 붉은색 세로축은 3개의 리스트에서 최소 한 번 이상 확인된 여성의 수(123명)를 표시한 것이다. “위안부” 전체 수치가 이미 기록에서 확인된 123명보다 적을 확률은 0%일 수밖에 없으므로, 전체 수치는 123명보다 적을 수 없다.

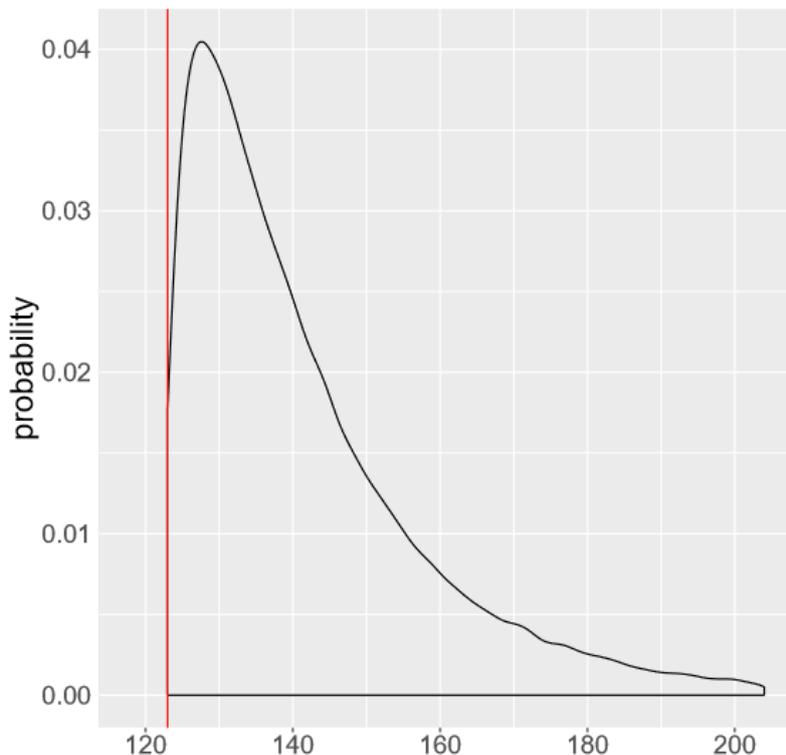


그림2: 기록에 없는 “위안부”를 포함한 “위안부” 전체 추정치. 기록된 수치는 붉은 선으로 표시됨.

이제 팔렘방 지역에 성노예로 끌려간 전체 한국인 "위안부"의 추정치를 제시한다. 우리가 제시하는 추정치인 137명은 가능성이 있는 수치들의 중간, 즉 [그림2](#)의 분포도에서 중간값(median)에 해당한다. 실제로 얼마나 많은 여성들이 기록되지 않았는지 정확히 알 수는 없기 때문에 흔히 말하는 "신뢰구간(credible interval)" 내에서 추정치를 추정할 수 있다. 통계적 분석에 따른 신뢰구간 내 추정치는 124명에서 198명까지다. 즉, 입수된 데이터와 MSE 기법에 따르면 실제 "위안부" 추정치가 124명에서 198명까지의 구간에 있을 확률이 95%라는 뜻이다.

신뢰구간이란 제시한 추정치가 불확실한 정도를 표현하는 것이기도 하다. 관찰되지 않은 "위안부" 수가 0보다 적을 수는 없기 때문에 우리는 구간 좌측 끝이 124명인 것은 확실하다. 그러나 우측으로는 불확실성이 따를 수 있으며, 실제 수치는 198명보다 상당히 클 수도 있다.

기록되지 않은 "위안부" 수에 관하여 다음과 같은 추가적인 추정을 제시할 수 있다. 분석에 따르면, 팔렘방 지역에 기록되지 않은 "위안부"가 10명 이상일 확률은 64%이다. 이는 최소 10명 이상의 기록에 없는 "위안부"가 있을 가능성이 그렇지 않을 가능성보다 크다는 것을 뜻한다.

이처럼 "다크 피겨", 즉 기록되지 않은 "위안부"의 수치는 상당할 수 있다. 기록되지 않은 "위안부"의 수가 50명 이상일 확률은 8%이다. 이는 낮은 확률일 수도 있지만 제2차 세계대전 중 팔렘방 지역에 있던 "위안부" 관련 정보의 추가 수집을 정당화할 만큼 큰 확률이라 할 수도 있다.

결론

이상의 분석은 입수된 명부들에 "위안부" 123명이 기재되어 있다는 점과 함께 기존 명부들 어디에도 기재되지 않은 "위안부"가 14명이 더 있을 가능성을 시사한다. 기록되지 않은 "위안부"가 10명 이상일 확률은 65%이다. 총괄하면 우리는 팔렘방 지역에 137명의 "위안부"가 있었을 것으로 추정한다.

다른 지역들에서도 "위안부" 명부가 입수되었지만, 팔렘방 지역처럼 전체 추정치를 도출하기에는 정보가 충분하지 않았다.

우리는 다른 지역들에 대한 추정치 도출을 가능하게 할 "위안부" 명부가 추가로 입수되기를 바란다. 학계 연구자, 정부 공문서 관리자, 또는 개인 중에 "위안부" 명부를 소장한 경우가 있을 수 있다. 이러한 명부가 있다면, cwdarkfigure@gmail.com 으로 보내주실 것을 부탁드립니다.

감사의 글

[메건 프라이스\(Megan Price\) 박사](#)의 설명과 코멘트에 감사드리며, 이 보고서를 한국어로 번역한 이영환(Hubert Y. Lee)에게 감사의 뜻을 전한다. 대한민국 국무총리 소속 일제강점하강제동원피해진상규명위원회 조사관으로 활동하였고, 이 보고서에 쓰인 명부들의 존재를 우리에게

알려주었고, 이 보고서의 작성을 가능하게 한 역사적 배경을 알려주고 귀한 조언을 해준 강정숙 박사에게 특별히 감사드린다.

참고문헌

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](https://doi.org/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' [일본군위안소업자의지위와역할 에관한연구]." 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 Fischhoff, 165–84. Oxford: Oxford UP.

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