Capture/Recapture Methods

Brandon del Pozo, PhD, MPA, MA

THE WARREN ALPERT
Medical School
BROWN UNIVERSITY

Capture recapture is a sampling technique used to **estimate population size.**

Sampling method	Description	Example
Capture recapture	Collecting a sample data from one location at different points in time, marking individuals to estimate a population size	A sample of woodlice were captured, marked and released. Another sample of woodlice was captured 5 days later and the number of marked woodlice was counted.

Advantages	Disadvantages
Estimate population size Track population changes (e.g. seasonal, health, climate change) Tracking over time.	Individuals have to remain local to the area of research with a definite boundary (no radical changes in the population due to births / deaths / migration). Markers are not lost or removed.

AJPH OPEN-THEMED RESEARCH

Estimated Prevalence of Opioid Use Disorder in Massachusetts, 2011–2015: A Capture–Recapture Analysis

Joshua A. Barocas, MD, Laura F. White, PhD, Jianing Wang, MSc, Alexander Y. Walley, MD, MSc, Marc R. LaRochelle, MD, Dana Bernson, MPH, Thomas Land, PhD, Jake R. Morgan, PhD, Jeffrey H. Samet, MD, MA, MPH, and Benjamin P. Linas, MD, MPH

THE WARREN ALPERT Medical School
BROWN UNIVERSITY

Methods: We performed a multisample stratified capture-recapture analysis to estimate OUD prevalence in Massachusetts. Individuals identified from 6 administrative databases for 2011 to 2012 and 7 databases for 2013 to 2015 were linked at the individual level and included in the analysis. Individuals were stratified by age group, sex, and county of residence.

Limitations and Strengths

Some limitations of our study merit attention. The methodology we used is an imperfect tool for estimation and is bound by several underlying assumptions: (1) the capture probabilities for different databases are independent, (2) the probability of capture by a given data set is assumed to be the same for each individual in the population, (3) the population does not change from capture to capture, and (4) matches are accurately identified. We addressed these assumptions in the following ways.

The prevalence of opioid use disorder in Kentucky's counties: A two-year multi-sample capture-recapture analysis

Katherine Thompson [a,1], Joshua A. Barocas [b,*,1], Chris Delcher [c,d], Jungjun Bae [c,d], Lindsey Hammerslag [e], Jianing Wang [f], Redonna Chandler [g], Jennifer Villani [g], Sharon Walsh [h,i], Jeffery Talbert [e,j]

Methods

We performed a multi-sample capture-recapture analysis to estimate OUD prevalence in Kentucky in 2018 and 2019. We utilized four statewide datasets that were linked at the individual level: 1) Registry of Vital Statistics, 2) Emergency Medical Services (EMS), 3) Kentucky's Prescription Drug Monitoring Program (PDMP), and 4) Kentucky Medicaid. We included persons aged 18–64 years who resided in Kentucky between 2018 and 2019. We identified individuals with administrative data consistent with OUD in each of the datasets, including a fatal opioid-involved overdose (Vital Statistics), EMS runs for suspected opioid overdose, receipt of buprenorphine for OUD treatment (PDMP), or Medicaid claims for OUD. Observed and estimated counts of OUD cases and prevalence of OUD among the adult population in Kentucky.

THE WARREN ALPERT Medical School
BROWN UNIVERSITY

Research article | Open Access | Published: 07 July 2014

Estimating the number of homeless deaths in France, 2008–2010

Cécile Vuillermoz, Albertine Aouba, Lise Grout, Stéphanie Vandentorren, Fanny Tassin, Layla Vazifeh, Walid Ghosn, Eric Jougla & Grégoire Rey

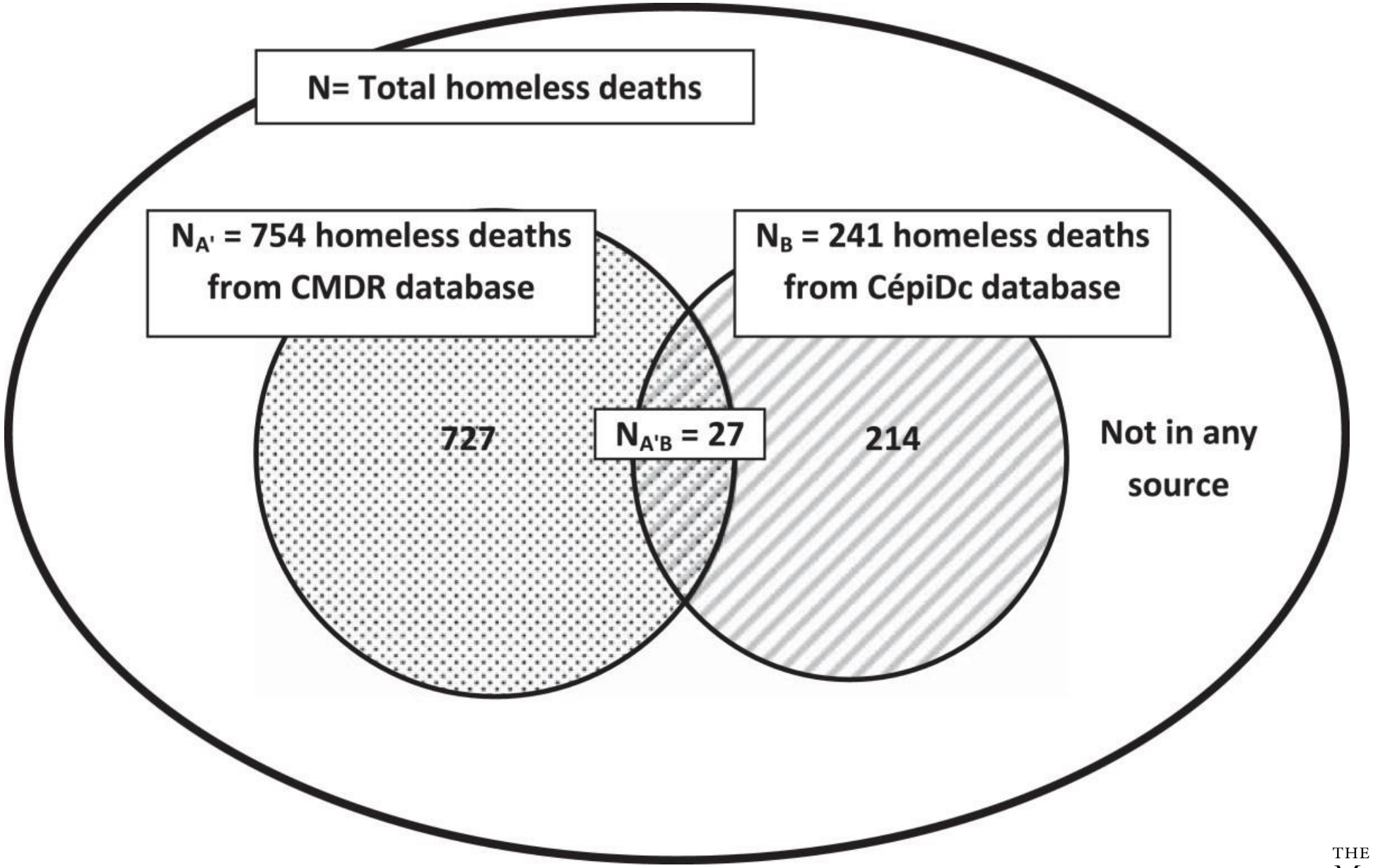
BMC Public Health **14**, Article number: 690 (2014) | Cite this article

Methods

We used the capture-recapture method to estimate the number of homeless deaths in France using two independent sources. An associative register of homeless deaths was matched with the national exhaustive database of the medical causes of death, using several matching approaches based on various combinations of the following variables: gender, age, place of death, date of death.

Results

The estimated number of homeless deaths between 2008 and 2010 was 6730 (95% CI: [4381–9079]), a number greatly underestimated by the two sources considered separately (less than 20%).



THE WARREN ALPERT Medical School
BROWN UNIVERSITY

Estimating the number of homeless deaths in France, 2008–2010

Cécile Vuillermoz, Albertine Aouba, Lise Grout, Stéphanie Vandentorren, Fanny Tassin, Layla Vazifeh, Walid Ghosn, Eric Jougla & Grégoire Rey

THE WARREN ALPERT Medical School BROWN UNIVERSITY

Applying the capture-recapture method requires some general validity conditions:

– *Independence of sources* (the probability that an observation is in one of the two sources does not depend on the probability of it being in the other source): Since different actors produce the sources, the independence is plausible by construction. In the absence of more than two sources, the qualitative assessment of the dependency between the sources could not be implemented.

– *Adequate matching* (deaths designated by a source can be matched to those reported by another source without mismatched data): Given the strict rules of anonymity of the CépiDc database, this assumption can hardly be assessed.

– *Capture homogeneity* (all persons in the population have the same chance of being observed in any source): The homogeneity of the capture was studied by comparing the distributions of homeless deaths from source A and source B, by gender, age, season, place and region of death. When the distribution of the deaths for those variables was significantly different, stratified estimation was undertaken [24].

– *Closed population (*no movement of subjects within the population): since the study population consisted of dead people, this assumption was fulfilled.

Translating from wildlife to admin data

THE WARREN ALPERT Medical School
BROWN UNIVERSITY

- Capture/Recapture in the case of LEO suicide consists of detecting appropriate cases in various administrative datasets
- The unit of analysis are the bounded areas (e.g., the lake or forest); in this case it would be the United States, or individual states
- The different parts of the “lake” or “forest” are the different bodies of administrative data present within the bounded area
- In wildlife research, the recaptures are prospective; in the examples used here are retrospective recaptures from administrative datasets

Design considerations

THE WARREN ALPERT Medical School BROWN UNIVERSITY

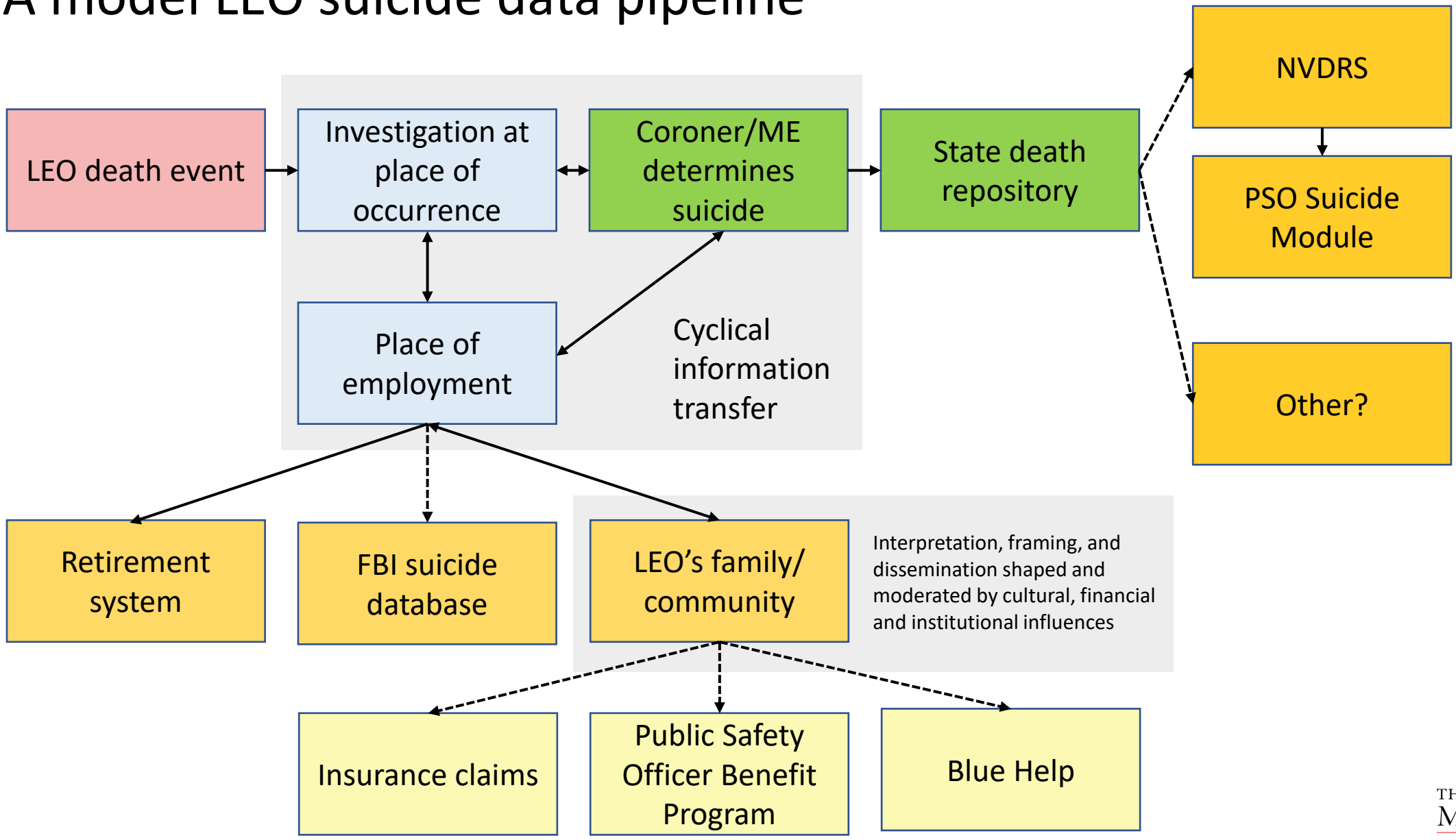
- Law enforcement officers who have completed a suicide during time period ***P*** are a closed population.
- Their distribution across sampling points may not be homogeneous, but a sufficient number of sampling points can help correct for this.
- The most accurate results would likely come from the largest possible preliminary capture, accompanied by several fully cooperative recapture points.
- Success depends on accurately linking data at the individual level in a national setting.

Design limitations

THE WARREN ALPERT Medical School
BROWN UNIVERSITY

- The results are subject to error within a range determined by the quality and size of the initial samples.
- Only estimates aggregate prevalence; does not provide demographic or other data at the individual level.
- Only provides overall prevalence within the closed boundary under study (e.g., state or nation); does not detect skews in the distribution among sub-regions (e.g., city or state).

A model LEO suicide data pipeline



THE WARREN ALPERT Medical School
BROWN UNIVERSITY