Using capture-recapture methods to estimate influenza hospitalization incidence rates

Goundappa K. Balasubramani1 | Mary Patricia Nowalk2 | Lloyd G. Clarke3 | Klancie Dauer1 | Fernanda Silveira4 | Donald B. Middleton5 | Mohamed Yassin6 | Richard K. Zimmerman2

1Department of Epidemiology, University of Pittsburgh, Pittsburgh, Pennsylvania, USA
2Department of Family Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA
3Department of Pharmacy, Division of Infectious Diseases/Pharmacy Department, UPMC Health System, Pittsburgh, Pennsylvania, USA
4Department of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA
5Department of Medical Education, UPMC St. Margaret, Pittsburgh, Pennsylvania, USA
6Infection Control Department, UPMC Mercy, Pittsburgh, Pennsylvania, USA

Correspondence
Mary Patricia Nowalk, Department of Family Medicine, University of Pittsburgh, 4420 Bayard St., Suite 520, Pittsburgh, PA 15260, USA.
Email: tnowalk@pitt.edu

Funding information
National Institutes of Health, Grant/Award Number: UL1TR001857; Centers for Disease Control and Prevention, Grant/Award Number: 5U01IP001035-02

Abstract

Background: Accurate population estimates of disease incidence and burden are needed to set appropriate public health policy. The capture-recapture (C-R) method combines data from multiple sources to provide better estimates than is possible using single sources.

Methods: Data were derived from clinical virology test results and from an influenza vaccine effectiveness study from seasons 2016–2017 to 2018–2019. The Petersen C-R method was used to estimate the population size of influenza cases; these estimates were then used to calculate adult influenza hospitalization burden using a Centers for Disease Control and Prevention (CDC) multiplier method.

Results: Over all seasons, 343 influenza cases were reported in the clinical database, and 313 in the research database. Fifty-nine cases (17%) reported in the clinical database were not captured in the research database, and 29 (9%) cases in the research database were not captured in the clinical database. Influenza hospitalizations were higher among vaccinated (58%) than the unvaccinated (35%) in the current season and were similar among unvaccinated (51%) and vaccinated (49%) in the previous year. Completeness of the influenza hospitalization capture was estimated to be 76%. The incidence rates for influenza hospitalizations varied by age and season and averaged 307–309 cases/100,000 adult population annually.

Conclusion: Using C-R methods with more than one database, along with a multiplier method with adjustments improves the population estimates of influenza disease burden compared with relying on a single-data source.

KEYWORDS
capture-recapture, hospitalization, influenza, influenza burden, influenza incidence

INTRODUCTION

Policy makers and planners need accurate estimates of the incidence or prevalence of diseases and health conditions to anticipate, prevent, and mitigate the effects of those diseases. The decentralized nature of U.S. healthcare makes overall population burden estimates difficult to calculate. Thus, policy makers must rely on population sampling for estimates, leaving true population burden unknown. The accuracy of

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. Influenza and Other Respiratory Viruses published by John Wiley & Sons Ltd.
population estimates depends largely upon the quality of sampling, which in turn is dependent upon many factors including consistency of the population being sampled across all capture occasions,\textsuperscript{3} and affected individuals having contact with the healthcare system to allow for enumeration. Cases of unreported disease are more difficult to account for.

Statistical methods to improve population disease incidence and prevalence detection include the capture–recapture (C-R) method. The Lincoln–Petersen (Petersen) method was the earliest C-R method for estimating population size. It was developed for field studies of animals in which only a sample of the population can be caught and marked (captured). Frequently, animals are re-caught (recaptured), and this method allows for recaptured animals to improve the population estimate.\textsuperscript{2} In health-related research, C-R uses the overlap of subjects from two or more data sources and log-linear methods to more accurately estimate true population disease burden.\textsuperscript{3} C-R has the advantage of using both research and clinical databases to measure the same data, thereby creating a fuller perspective; however, the best method to adjust for complex denominators in urban areas with competing health systems is not straightforward within C-R.

The Centers for Disease Control and Prevention (CDC) has developed methods to estimate population influenza burden (L Kim, personal communication, 2020) that account for some of the complexities of multicenter study design and the incomplete nature of surveillance data of the Hospitalized Adult Influenza Vaccine Effectiveness Network (HAIVEN) study. The current study used available data from a single health system, HAIVEN methods and adjustments, and C-R calculations, to estimate adult influenza hospitalization burden in Allegheny County in Southwestern Pennsylvania.

2 | METHODS

The study was approved by the University of Pittsburgh IRB and consists of a three-phase analytic plan. The phases are as follows: (1) C-R to more accurately estimate influenza cases; (2) statistical analyses for population burden based on the CDC HAIVEN network methods (Lindsay Kim, MD, personal communication); and (3) an adjustment to this resultant population burden using C-R incidence estimates. The HAIVEN population burden methods may not account for the richness of the clinical virology data available in our particular locale. Our adjustment to the HAIVEN methods was intended to capitalize on both the robust nature of the HAIVEN methods and the richness of the local virology data.

2.1 | Phase 1: Statistical analyses for C-R

Data used for this analysis were collected from two sources: (1) the local health system’s clinical surveillance software system (Theradoc\textsuperscript{®}), which extracts virology test results from the electronic medical record (EMR); and (2) research data from selected hospitals participating in the HAIVEN study.

An IRB-approved honest broker extracted a data list from Theradoc\textsuperscript{®} of Allegheny County residents who received an inpatient clinical respiratory viral panel (RVP) test at two to five (depending upon the season) general acute care hospitals in the health system during the study period that included the 2015–2016 through 2018–2019 influenza seasons. This list also contained basic demographic data of race, sex, and age and is henceforth called the “clinical” database. The clinical database was prepared for analysis by limiting it to data from hospitals in which research enrollments were taking place for each influenza season. For example, in 2015–2016, two hospitals were enrolling participants, whereas in 2018–2019, there were five participating hospitals. Secondly, data were limited to the periods during which research enrollments were taking place. Thirdly, patients <18 years of age were eliminated. Finally, patients were separated into influenza cases and non-cases. The “research” database was derived from adults who were recruited from the hospitals during the 2015–2016 through 2018–2019 influenza seasons for the HAIVEN study that only included inpatients ≥18 years of age. Detailed study methods for the HAIVEN study have been described elsewhere.\textsuperscript{4} Briefly, patients aged ≥18 years admitted with an acute respiratory infection (ARI) including cough or worsening symptoms of a respiratory illness beginning within 10 days were enrolled. Patients who had been enrolled in the prior 14 days were ineligible. Following informed consent, study staff collected respiratory specimens (nasal and throat swabs from patients) for influenza virus testing (including virus type and subtype) by reverse-transcription polymerase chain reaction (RT-PCR) or used results from a clinical RVP test, if available. Demographic data were obtained from interview. Vaccination status was based on documented receipt of each year’s influenza vaccine from the local electronic health record and/or the Pennsylvania Statewide Immunization Information System (PA-SIIS).

The following variables were used in the calculations: $M =$ number of cases identified in the clinical database; $n =$ number of cases identified in the research database; $m =$ number of cases identified in both databases (matched); $N_1 =$ number of cases reported only in the clinical database; $N_2 =$ number of cases reported only in the research database; $X =$ number of cases missing/not captured in either database.

Summary statistics of the demographic and clinical characteristics were determined for the patients found in the matched database. The number of observed influenza cases in the two databases and the Petersen’s C-R method were used to estimate influenza incidence ($\hat{N}$).\textsuperscript{5}

\begin{equation}
\text{Petersen estimate } \hat{N} = \frac{(\text{Observed clinical cases (M)}) \times (\text{Observed research cases (n)})}{(\text{Observed cases from matched database (m))}}
\end{equation}

The variance and 95% confidence intervals (CIs) were calculated for the C-R estimates using the formulae:

\begin{equation}
\text{Variance } \hat{N} = \frac{M + n + N_1 + N_2}{m^2}
\end{equation}
The C-R calculations were made assuming that (1) the population is closed; that is, there was no outmigration or loss to follow-up because the capture and recapture would have usually occurred during the same hospitalization. Calculation of completeness of reporting by the two sources of the C-R method is determined by calculating the number of missing cases, $X$

$$X = \frac{(N_1 \times N_2)}{m}. \quad (4)$$

An example of a C-R estimate is shown in Table S1.

Secondly, it is assumed that the populations are homogeneous; that is, each hospitalized patient has the same and constant probability of being captured by any combination of the databases. Thirdly, it is assumed that the clinical database and the HAIVEN research database are independent of each other. That is, the population estimate assumes that the probability of being captured by one source does not affect the probability of being captured by the other source. Independence can be tested by calculating the probability of influenza positives being captured by both databases. If that probability is equal to the product of the marginal probabilities of being influenza positive captured by clinical and research databases, then the samples are independent.

Independence equality: $\frac{m}{N} \times \frac{M}{N} = \frac{M \times n}{N \times N} \times \frac{m}{N} \times \frac{N}{N}$. \quad (5)

Independence was tested for the 3-year total samples and the 15 subpopulations derived by stratifying on demographic factors (age, sex, and race), influenza season, vaccination status, and prior vaccination status (Table S2).

### 2.2 Phase 2: Statistical analyses for population burden based on HAIVEN methods

The Pennsylvania Health Care Cost Containment Council (PHC4) provided data for ARI-specific hospitalizations based on CDC ARI ICD codes in all county hospitals for all 4 years of the study.

The HAIVEN methods to calculate disease burden estimates were used as follows:

Influenza burden estimate = \frac{Adjusted number of cases}{Adult population in Allegheny County}. \quad (6)

Proportion of all ARI cases in all county hospitals represented by the study-specific hospitals was determined using quarterly data from PHC4.

$r$ = cases identified through research enrollment (research cases) among Allegheny County residents.

$V_1$ = number of ARI hospitalizations from county residents who are enrolled in the research database

$V_2$ = number of ARI hospitalizations in study-specific hospitals during influenza months among Allegheny County residents in PHC4 database.

$V_3$ = number of ARI hospitalizations in all county hospitals during the same time period among Allegheny County residents in PHC4 database; rationale is that both $V_2$ and $V_3$ should come from the same database.

$V_4$ = number of influenza cases in study-specific hospitals during research enrollment period from clinical database.

$V_5$ = number of influenza cases in study-specific hospitals over the entire year from clinical database; rationale is that both $V_4$ and $V_5$ should come from the same database.

### Adjusted number of cases

\( Adjusted \ number \ of \ cases = research \ cases \ (r) \times \left( \frac{1}{\text{Proportion of all ARI enrolled in research database}} \right) \times \left( \frac{1}{\text{Proportion of all ARI in study - specific hospitals}} \right) \times \left( \frac{1}{\text{Proportion of all influenza cases detected during research enrollment period}} \right). \quad (7) \)

### 2.3 Phase 3: Combination of C-R and HAIVEN methods for adjusted population burden

To incorporate the C-R method into the HAIVEN methods to account for cases estimated by C-R but not due to the enrollment fraction, the following modification of Equation 7 was used:
For influenza burden calculations by race, the county population used as the denominator was the total adult population of the county multiplied by 0.78 for Whites and 0.13 for Blacks, representing their relative proportions of the population. Data were analyzed using SAS version 9.4 (SAS Institute, Cary, NC, USA).

\[
\text{Adjusted number of cases} = \text{research cases} \times \left( \frac{C - \text{restimated cases}(\hat{N})}{\text{Total Observed cases} (m + N1 + N2)} \right) \times \left( \frac{1}{\text{Proportion of all ARI in study - specific hospitals}} \right) \times \left( \frac{1}{\text{Proportion of all influenza cases detected during research enrollment period}} \right)
\]

(8)

**TABLE 1** Demographic characteristics of patients identified in the matched database (N = 1,310)

| Variable                                     | n (%)  |
|----------------------------------------------|--------|
| Age group                                    |        |
| 18–49 years                                  | 249 (19.0) |
| 50–64 years                                  | 449 (34.3) |
| 65–74 years                                  | 311 (23.7) |
| 75+ years                                    | 301 (23.0) |
| Race                                         |        |
| White                                        | 848 (64.7) |
| Black                                        | 421 (32.1) |
| Other, unknown                               | 41 (3.2) |
| Sex                                          |        |
| Female                                       | 824 (62.9) |
| Male                                         | 486 (37.1) |
| Season                                       |        |
| 2016–2017                                    | 356 (27.2) |
| 2017–2018                                    | 449 (34.3) |
| 2018–2019                                    | 505 (38.5) |
| Vaccination status                           |        |
| Unvaccinated                                 | 447 (34.1) |
| Vaccinated \(\geq 14\) days prior to illness onset | 766 (58.5) |
| Vaccinated \(< 14\) days prior to illness onset | 97 (7.4) |
| Prior year vaccination (total for all seasons) |        |
| No                                           | 645 (49.2) |
| Yes                                          | 665 (50.8) |
| Influenza status                             |        |
| No                                           | 973 (74.3) |
| Yes                                          | 337 (25.7) |

3 | RESULTS

The viral test result analytic databases are shown in Figure 1. The clinical database consisted of 8,994 patients of whom 7,684 patients were unmatched; the research database consisted of 2,154 patients of whom 844 were unmatched patients; and 1,310 patients were found in both databases (matched). Demographic characteristics of the patients found in the matched database are shown in Table 1. The
highest proportion of the group was 50–64 years old (34.3%) with less than one quarter each among patients who were 65–74 and +75 years old and one fifth who were 18–49 years old. The patients were predominantly White (64.7%), female (62.9%) and vaccinated ≥14 days prior to illness onset (58.5%). Half of them had been vaccinated in the previous season and 25.7% were influenza cases.

Table 2 shows the observed cases among persons hospitalized with a cough illness in the clinical, research and matched databases, and C-R estimated influenza hospitalizations over all seasons, by season and by other factors.

The HAIVEN population influenza burden estimates (using Equation 7) and the HAIVEN + C-R (using Equation 8) population influenza burden estimates were calculated using the values shown in Table 2. Over all three influenza seasons, the average incidence rates for hospitalized influenza in the research hospitals were 307–309/100,000 (HAIVEN and HAIVEN + C-R, respectively) (Table 3). The lowest rates over all seasons were 17/100,000 (for both HAIVEN and HAIVEN + C-R) among 18–49 year olds. The highest seasonal rates for the entire adult population were 494–497/100,000 (HAIVEN and HAIVEN + C-R, respectively), in 2017–2018, an especially severe influenza season. Although there were two to three times as many cases among Whites as Blacks, the influenza burden per 100,000 population were 10%–30% higher among Whites than Blacks.

### DISCUSSION

This study used C-R methods to calculate influenza incidence among hospitalized patients and then used two methods for population burden: CDC’s HAIVEN burden estimate methods and a combination of HAIVEN and C-R estimates. C-R has been adapted from ecological studies and used in a wide array of health-related studies including Alzheimer’s disease, heart attack, HIV infection, gun injury, pediatric disease surveillance, gastric cancer, and norovirus infections.7–12 The Petersen C-R method using two sources of data in this study is a special case of the generalized C-R method for estimating burden using multiple data sources. In this study, estimates were nearly identical between observed and estimated incidence. By comparison, in a C-R study of norovirus cases, the combined databases yielded incidence at a rate 2.5 times the level of the highest individual database.12 The assumptions of the Petersen estimator allow for any amount of overlap in the databases. We have shown in Table S2 that

| TABLE 2 | Estimated population influenza hospitalizations using the capture-recapture method |
|---------|-----------------------------------------------------------------------------|
|         | Observed influenza cases | C-R estimated influenza cases N (95% CI) |
|         | Clinical (M) | Research (n) | Matched (m) | Total (m + N1 + N2) |
| 3-year total | 308 | 313 | 284 | 337 | 339 (336, 342) |
| Age group |         |         |         |         |         |
| 18–49 years | 55 | 54 | 50 | 59 | 59 (58, 60) |
| 50–64 years | 94 | 95 | 85 | 104 | 105 (103, 107) |
| 65–74 years | 82 | 82 | 75 | 89 | 90 (88, 92) |
| 75+ years | 77 | 82 | 74 | 85 | 85 (84, 86) |
| Race |         |         |         |         |         |
| White | 191 | 199 | 179 | 211 | 212 (210, 214) |
| Black | 103 | 100 | 92 | 111 | 112 (110, 114) |
| Sex |         |         |         |         |         |
| Female | 191 | 195 | 175 | 211 | 213 (210, 216) |
| Male | 117 | 118 | 109 | 126 | 127 (125, 129) |
| Season |         |         |         |         |         |
| 2016–2017 | 84 | 82 | 77 | 89 | 89 (88, 90) |
| 2017–2018 | 139 | 138 | 127 | 150 | 151 (149, 153) |
| 2018–2019 | 85 | 93 | 80 | 98 | 99 (97, 101) |
| Vaccination status |         |         |         |         |         |
| Unvaccinated | 124 | 124 | 111 | 137 | 139 (136, 142) |
| Vaccinated | 165 | 169 | 155 | 179 | 180 (178, 182) |
| Prior vaccination |         |         |         |         |         |
| No | 167 | 167 | 153 | 181 | 182 (180, 184) |
| Yes | 141 | 146 | 131 | 156 | 157 (155, 159) |

Note. N1 = number of cases reported only in the clinical database; N2 = number of cases reported only in the research database.

aN = (M + n)/m.
bSum ≠ 337 due to missing data.
| Variable | Study-specific hospitals among Allegheny County residents (source = research database) | ARI cases in study-specific hospitals during influenza months among Allegheny County residents (source = PHC4) | ARI cases in all county hospitals during influenza months among Allegheny County residents (source = PHC4) | Influenza cases in study-specific hospitals during research enrollment period (source = clinical database) | Influenza cases in study-specific hospitals over the entire year (source = clinical database) | Influenza burden |
|----------|---------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------|------------------|
|          | Influenza cases in study-specific hospitals among Allegheny County residents (source = research database)                          | Influenza cases in study-specific hospitals during influenza months among Allegheny County residents (source = PHC4) | ARI cases in all county hospitals during influenza months among Allegheny County residents (source = PHC4) | Influenza cases in study-specific hospitals during research enrollment period (source = clinical database) | Influenza cases in study-specific hospitals over the entire year (source = clinical database) | Cases/100 K Allegheny County adult population (based on equation 7 and using C-R estimates) |
| 3 years total | 385 | 2.154 | 13.863 | 46.450 | 1.345 | 1.487 | 921 | 927 |
| Annual average |  |  |  |  | | | 307 | 309 |
|  | | | | | | | | |
| 2016–2017 | | | | | | | | |
| Overall | 107 | 665 | 4.439 | 16.171 | 289 | 316 | 286 | 286 |
| 18–49 years | 18 | 125 | 6.56 | 1.950 | 63 | 65 | 65 | 65 |
| 50–64 years | 27 | 235 | 1.154 | 3.851 | 58 | 66 | 66 | 66 |
| 65–74 years | 31 | 153 | 1.016 | 3.458 | 59 | 68 | 68 | 68 |
| +75 years | 31 | 152 | 1.613 | 6.912 | 109 | 117 | 117 | 117 |
| White | 77 | 481 | 3.323 | 14.071 | 215 | 236 | 236 | 236 |
| Black | 26 | 166 | 1.082 | 2.066 | 68 | 75 | 75 | 75 |
| 2017–2018 | | | | | | | | |
| Overall | 168 | 656 | 4.135 | 16.353 | 539 | 633 | 494 | 497 |
| 18–49 years | 35 | 142 | 6.12 | 1.955 | 126 | 143 | 143 | 143 |
| 50–64 years | 48 | 213 | 1.039 | 3.838 | 121 | 153 | 153 | 153 |
| 65–74 years | 42 | 157 | 997 | 3.644 | 108 | 124 | 124 | 124 |
| +75 years | 43 | 144 | 1.487 | 6.916 | 184 | 213 | 213 | 213 |
| White | 103 | 398 | 3.051 | 13.899 | 361 | 426 | 426 | 426 |
| Black | 57 | 241 | 1.084 | 2.324 | 159 | 189 | 189 | 189 |
| 2018–2019 | | | | | | | | |
| Overall | 110 | 833 | 5.289 | 13.926 | 517 | 538 | 192 | 194 |
| 18–49 years | 17 | 175 | 828 | 1.725 | 159 | 164 | 164 | 164 |
| 50–64 years | 47 | 335 | 1.481 | 3.360 | 154 | 161 | 161 | 161 |
| 65–74 years | 26 | 188 | 1.233 | 3.299 | 104 | 109 | 109 | 109 |
| +75 years | 20 | 135 | 1.747 | 5.542 | 100 | 104 | 104 | 104 |
| White | 72 | 498 | 3.985 | 11.679 | 316 | 331 | 227 | 227 |
| Black | 34 | 299 | 1.294 | 1.921 | 181 | 188 | 175 | 175 |

Abbreviation: ARI, acute respiratory illness.
the independence assumption is met for the 3-year total and all sub-populations. With few exceptions, each database’s probability of being captured by the other is high, resulting in a considerable overlap and high probability of being captured by both databases.

This study met or partially met four of the assumptions needed for confidence in the reliability of the outcomes. (1) The populations could be considered closed; that is, there was little chance of loss of cases due to outmigration or death. In a disease of short incubation and duration such as influenza, loss to outmigration is minimal. (2) High overlap of cases improves reliability of estimates because of low missed cases. In this study, there was a large overlap of cases between the databases allowing for matching of pairs of cases (same person, both databases). (3) Databases should be homogeneous. The two databases in this study are believed to be homogeneous, in that cases in both have a nearly equal probability of being identified. (4) The data sources should be independent to prevent overestimation or underestimation of missing cases. We have demonstrated their independence. Although they were identified using different methods, the cases in the research database were all hospitalized in the same hospitals from which the clinical database was drawn. This fact does not affect the independence assumptions.

The question about cases not captured by each of the databases deserves comment. The research databases were limited by funding for intensity and duration of recruitment, by volunteer participation, by obtaining specimens later in the course of illness, and by specimen collection and testing techniques. The research specimens, obtained by trained research assistants with a mid-turbinate collection, used the CDC’s quantitative PCR. This test is slightly more sensitive than the clinical qualitative PCR, which used a nasopharyngeal specimen obtained by clinical staff. Thus, a small amount of discrepancy between the two databases was expected.

Because of our confidence in the hospitalization estimates, county-wide incidence rates were calculated. Influenza hospitalization incidence was highest in 2017–2018, an influenza A/H3N2-dominated year and lowest in 2018–2019 a season in which influenza A/H1N1 and A/H3N2 cases were approximately equal. This finding is not unexpected, given influenza A/H3N2’s higher severity than influenza A/H1N1. Using C-R methods, influenza hospitalization incidence estimates among children have ranged from 240/100,000 in 2003 to 286/100,000 for Allegheny County; for 2017–2018, influenza burden was ≈324/100,000 compared with 494–497/100,000 for Allegheny County and for 2018–2019 influenza burden was ≈196/100,000 compared with 172–174/100,000 for Allegheny County. These similarities support the strength of our methods for calculating burden.

4.1 Strengths and limitations

Some inaccuracy of burden estimates based on surveillance data would be expected given its inherent weaknesses, compared with true population-based data. Hence, adjustments were made to account for some of those weaknesses. The methods were further enhanced by using C-R to estimate hospitalized cases. This study is one of only a few C-R papers for influenza hospitalization rates among adults and the only one of which we are aware for these seasons. It was conducted in a complex urban/suburban county with detailed PCR-based clinical virology and in a health system with the majority of the market share of hospitalizations. It has the advantages of comparing and combining the robust HAIVEN and the C-R methods.

Limitations include conducting the study in only one county, although county-specific population burden is a key outcome. The sources of the positive or negative dependence may lead to underestimation or overestimation of the population size. Finally, funding and time constraints limited the duration of active surveillance during the influenza season for the research database.

5 Conclusions

Influenza illness is associated with significant costs that include lost productivity due to absenteeism and presenteeism, lost wages, and costs of medical care. Understanding the burden of influenza hospitalization is important for policy makers to allocate resources for the prevention and treatment of influenza. Petersen’s C-R method, in combination with the CDC HAIVEN burden method, improved population estimates.

Acknowledgment

The Pennsylvania Health Care Cost Containment Council (PHC4) is an independent state agency responsible for addressing the problem of escalating health costs, ensuring the quality of healthcare, and increasing access to healthcare for all citizens regardless of ability to pay. PHC4 has provided data to this entity in an effort to further PHC4’s mission of educating the public and containing healthcare costs in Pennsylvania.

PHC4, its agents, and staff have made no representation, guarantee, or warranty, express or implied that the data—financial, patient, payor, and physician specific information—provided to this entity, are error-free, or that the use of the data will avoid differences of opinion or interpretation.

This analysis was not prepared by PHC4. This analysis was done by the University of Pittsburgh. PHC4, its agents and staff, bear no responsibility or liability for the results of the analysis, which are solely the opinion of this entity.

Author Contributions

GK Balasubramani: Conceptualization; formal analysis. Mary Patricia Nowalk: drafting and editing the manuscript. Lloyd G. Clarke: Data curation. Klancie Dauer: Data curation. Fernanda Silveira:
Project administration. Donald B. Middleton: Funding acquisition; investigation. Mohamed Yassin: Project administration. Richard K. Zimmerman: Conceptualization; funding acquisition.

FUNDING INFORMATION
This work was supported by the Centers for Disease Control and Prevention (CDC) (SU01IP001035-02) and by National Institutes of Health (NIH) (UL1TR001857). This work represents the views of the authors and not the CDC or NIH. It is subject to CDC’s public access policy.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are available on request from the lead author balagk@edc.pitt.edu with a proper data use agreement in place. The data are not publicly available due to privacy or ethical restrictions.

ORCID
Goundappa K. Balasubramani https://orcid.org/0000-0001-7221-1825
Mary Patricia Nowalk https://orcid.org/0000-0002-1702-2690

REFERENCES
1. Hwang W-H, Huggins R. An examination of the effect of heterogeneity on the estimation of population size using capture-recapture data. Biometrika. 2005;92(1):229-233.
2. Nichols JD. Capture-recapture models: using marked animals to study population dynamics. Bioscience. 1992;42(2):94-102.
3. Hook EB, Regal RR. Recommendations for presentation and evaluation of capture-recapture estimates in epidemiology. J Clin Epidemiol. 1999;52(917–926):discussion 929-933.
4. Ferdinands JM, Gaglani M, Martin ET, et al. Prevention of influenza hospitalization among adults in the United States, 2015–2016: results from the US Hospitalized Adult Influenza Vaccine Effectiveness Network (HAIVEN). J Infect Dis. 2018.
5. Ackman DM, Birkhead G, Flynn M. Assessment of surveillance for meningococcal disease in New York State, 1991. Am J Epidemiol. 1996;144(1):78-82.
6. Bishop YM, Fienberg SE, Holland PW. Estimating the size of a closed population Chapter 6. In Discrete Multivariate Analysis: Theory and Practice. Springer New York, NY; 2007.
7. Waller M, Mishra GD, Dobson AJ. Estimating the prevalence of dementia using multiple linked administrative health records and capture-recapture methodology. Emerg Themes Epidemiol. 2017; 14(1):3.
8. Razzaq JA, Mawani M, Azam I, Robinson C, Talib U, Kadir MM. Burden of out-of-hospital cardiac arrest in Karachi, Pakistan: Estimation through the capture-recapture method. J Pak Med Assoc. 2018; 68(7):990-993.
9. Poorolajal J, Mohammadi Y, Farzinara F. Using the capture-recapture method to estimate the human immunodeficiency virus-positive population. Epidemiol Health. 2017;39.
10. Post LA, Balsen Z, Spaso N, Vaca FE. Bolstering gun injury surveillance accuracy using capture-recapture methods. J Behav Med. 2019;42(4):674-680.
11. Knowles RL, Smith A, Lynn R, Rahi JS. Using multiple sources to improve and measure case ascertainment in surveillance studies: 20 years of the British Paediatric Surveillance Unit. J Public Health. 2006;28(2):157-165.
12. Hardstaff J, Clough H, Harris J, Lowther J, Lees D, O’Brien S. The use of capture-recapture methods to provide better estimates of the burden of norovirus outbreaks from seafood in England, 2004–2011. Epidemiol Infect. 2019;147.
13. National and Regional Level Outpatient Illness and Viral Surveillance [http://gis.cdc.gov/grasp/fluview/fluportaldashboard.html]
14. Grijalva CG, Weinberg GA, Bennett NM, et al. Estimating the undetected burden of influenza hospitalizations in children. Epidemiol Infect. 2007;135(6):951-958.
15. Grijalva CG, Craig AS, Dupont WD, et al. Estimating influenza Hospitalizations among children. Emerg Infect Dis. 2006;12(1):103-109.
16. Jules A, Grijalva CG, Zhu Y, et al. Estimating age-specific influenza-related hospitalization rates during the pandemic (H1N1) 2009 in Davidson Co TN. Influenza Other Respi Viruses. 2012;6(3):e63-e71.
17. Centers for Disease Control and Prevention Disease burden of influenza [https://www.cdc.gov/flu/about/burden/index.html]

SUPPORTING INFORMATION
Additional supporting information may be found in the online version of the article at the publisher’s website.

How to cite this article: Balasubramani GK, Nowalk MP, Clarke LG, et al. Using capture-recapture methods to estimate influenza hospitalization incidence rates. Influenza Other Respi Viruses. 2022;16(2):308-315. doi:10.1111/irv.12924