Missing data matters in participatory syndromic surveillance systems: comparative evaluation of missing data methods when estimating disease burden

Kristin Baltrusaitis, Craig Dalton, Sandra Carlson, Laura F. White

Center for Biostatistics in AIDS Research, Harvard T.H. Chan School of Public Health, Boston, United States of America
Hunter New England Population Health, Wallsend, Australia
Hunter Medical Research Institute, Newcastle, Australia
Department of Biostatistics, Boston University School of Public Health, Boston University, Boston, United States of America

Kristin Baltrusaitis
651 Huntington Avenue, Boston, MA 02115
kbaltrus@sdac.harvard.edu

3200

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Participatory Syndromic Surveillance, Digital Epidemiology, Missing Data, Missing Not at Random, Multiple Imputation, Influenza Surveillance

ABSTRACT

Introduction

Traditional surveillance methods have been enhanced by the emergence of online participatory syndromic surveillance systems that collect health-related digital data. These systems have many applications including tracking weekly prevalence of Influenza-Like Illness (ILI), predicting probable infection of Coronavirus 2019 (COVID-19), and determining risk factors of ILI and COVID-19. However, not every volunteer consistently completes surveys. In this study, we assess how different missing data methods affect estimates of ILI burden using data from FluTracking, a participatory surveillance system in Australia.

Methods

We estimate the incidence rate, the incidence proportion, and weekly prevalence using five missing data methods: available case, complete case, assume missing is non-ILI, multiple imputation (MI), and delta (δ) MI, which is a flexible and transparent method to impute missing data under Missing Not at Random (MNAR) assumptions. We evaluate these methods using simulated and FluTracking data.

Results

Our simulations show that the optimal missing data method depends on the measure of ILI burden and the underlying missingness model. Of note, the δ-MI method provides estimates of ILI burden that are similar to the true parameter under MNAR models. When we apply these
methods to FluTracking, we find that the $\delta$-MI method accurately predicted complete, end of season weekly prevalence estimates from real-time data.

**Conclusion**

Missing data is an important problem in participatory surveillance systems. Here, we show that accounting for missingness using statistical approaches leads to different inferences from the data.

**INTRODUCTION**

Over the past two decades, traditional surveillance methods have been enhanced by the emergence of online participatory syndromic surveillance systems that collect health-related digital data in near-real time.[1] Through these systems, participants volunteer to regularly report syndromic, health information via online or mobile communication technologies. These systems complement traditional healthcare-based surveillance systems by reducing the time delay associated with visiting a healthcare provider and capturing individuals who do not seek medical care.[2] The first of these systems, de Grote Grieppmeting, or the Great Influenza Survey, started in 2003 in the Netherlands and Belgium,[3] and multiple systems throughout Europe,[4] Australia (AU),[5] the United States (US),[6] Mexico, and Brazil have followed. More recently, participatory syndromic surveillance systems have been deployed to track symptoms of Coronavirus 2019 (COVID-19) in the community.[7]

These systems actively engage the public in reporting and providing timely information about disease trends within the community, while providing a mechanism for community members to become "citizen-scientists". However, because not every participant reports regularly the population at risk can be temporally inconsistent and include systematic biases. For example, in post-influenza season surveys, over 30% of users from the U.S. system, Flu Near You (FNY), reported they were at least slightly more likely to report if they have had symptoms. Furthermore, during the influenza season, FNY users were significantly less likely to consistently submit reports if they reported symptoms during their first report following registration. Of note, the reporting habits of users vary by system. For example, approximately 22% of FNY users submitted at least half of the symptom reports during the 2013-2014 influenza season, whereas the AU system, FluTracking, reported that during the 2017 influenza season, of the participants who completed a survey during the first four survey weeks, 69% completed all available surveys, and 82% completed over 90% of available surveys.

The most common approach for addressing the inconsistencies in the user reporting is to select a cohort of "active users", where the definition of "active user" varies by system and study, and assume that all missing reports were asymptomatic. However, no study has assessed how this deterministic assumption affects estimates or systematically compared approaches through simulations.

Here, we draw upon statistical methods for missing data and assess how different missing data methods, such as complete case and imputation-based methods, affect estimates of ILI burden using both simulated data and data from FluTracking.

METHODS
FluTracking

FluTracking is an online health surveillance system of influenza in AU and, as of 2018, New Zealand and Hong Kong in 2021. Launched in 2006, the FluTracking system has grown to include over 150,000 participants during the peak weeks of COVID-19 surveillance across Australia and New Zealand in 2020.[25] At registration, FluTracking users provide basic demographic information, including month and year of birth, sex, postcode of residence, Indigenous status, highest level of education, and whether or not they work directly with patients in a healthcare setting. After registration, users complete weekly surveys for themselves and other household members about ILI symptoms including fever, cough, and/or sore throat. Users who report any symptoms are asked follow-up questions about absenteeism from work or normal duties, visits to health care providers, and results of laboratory tests. All users are also asked about influenza vaccination. Symptom surveys are sent every Monday, however, unlike other participatory surveillance systems, FluTracking participants have the option to complete missed surveys up to five weeks previous, called "retrospective reports".

For this study, ILI is defined as report of both fever and cough, with or without sore throat. Descriptive statistics for participant characteristics, including age, sex, household status, and influenza vaccination status, are displayed as median (25th percentile, 75th percentile) for continuous variables and n (%) for categorical variables for all participants who submitted at least one symptom report during the 2016, 2017, and 2018 influenza seasons. Although FluTracking collects data from the beginning of May through mid-October, we use reports submitted during the southern hemisphere’s influenza surveillance season, defined as
Morbidity and Mortality weeks 25 through 41, or approximately late-June through mid-
October.

Measures of Influenza-Like Illness Burden

As recommended by the World Health Organization (WHO), we use the incidence rate (IR) as
one measure of ILI burden.[26] The IR is equal to the number of incident ILI reports, defined as
a report of ILI in which ILI was not reported the previous week, divided by the total person-time
reported by participants:

\[
IR = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} Y_{it}}{\sum_{i=1}^{N} \sum_{t=1}^{T}(1-R_{it})} \times 10,000
\]  

(1)

where

\[
Y_{it} = \begin{cases} 
1, & \text{if ILI reported at time } t \\
0, & \text{otherwise}
\end{cases}
\]

\[
R_{it} = \begin{cases} 
0, & \text{if } Y_{it} \text{ observed} \\
1, & \text{otherwise}
\end{cases}
\]

and \( t \) indexes the week numbers and ranges from 25 to 41.

The rate is expressed per 10,000 person weeks. Because person-time at risk is unavailable for
most routine influenza surveillance data, we also present the incidence proportion (IP) for
comparability across systems. The IP is equal to the number of participants who reported ILI at
least once during the influenza season divided by the total number of participants:

\[
IP = \frac{\sum_{i=1}^{N} Q_{i}}{N}
\]  

(2)

where

\[
Q_{i} = \begin{cases} 
1, & \text{if } \sum_{t=1}^{T} Y_{it} > 0 \\
0, & \text{otherwise}
\end{cases}
\]

The 95% Confidence Intervals (CI) for these estimates are given by
where \( d \) is the number of cases.\[^{[27,28]}\] We calculate these measures of ILI burden for the overall population and by age group (<5, 5-17, 18-49, 50+ years). Finally, we present the weekly prevalence (WP) of ILI at each week, which is calculated by dividing the number of ILI reports by the total number of reports observed,

\[
WP_t = \frac{\sum_{i=1}^{N} Y_{it}}{\sum_{i=1}^{N} (1-R_{it})}, \quad (4)
\]

Because WP is estimated in near-real time, all retrospective reports were assumed missing when calculating estimates. We assess and compare these measures of ILI burden across three influenza seasons (2016, 2017, and 2018).

**Missing Data Methods**

We assess five methods that account for missing data:

1. Available case
2. Complete case
3. Assume all missing reports are non-ILI
4. Multiple imputation (MI)
5. MI with delta (\( \delta \)) adjustment

The first method uses all available cases (i.e., all reports submitted or select all \( Y_{it} \) for \( R_{it} = 0 \)), whereas the second method includes only complete cases (i.e., reports from individuals who submitted all reports or select all individuals, \( i \), where \( R_{it} = 0 \) for all \( t \)). These methods assume that data is Missing Completely at Random (MCAR). In other words, the causes of the missing data are unrelated to the data. The third method assumes that all missing reports are non-ILI
reports (i.e., $P(Y_{it} = 1 | R_{it} = 1) = 0$), similar to past studies. The next two methods use MI methods to produce 10 point estimates, which are aggregated using Rubin's rules with a log transformation to account for non-normality.\cite{29} For the first MI method (method 4), we fit a model assuming Missing at Random (MAR),

$$logit[P(Y_t = 1 | X_t)] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_{3t} + \beta_4 Y_1 + \ldots + \beta_{2+t} Y_{t-1} \quad (5)$$

where

$$X_1 = \text{age}, X_2 = \text{sex}, X_{3t} = \text{vaccination status at week } t.$$

Under MAR, the probability that a value is missing depends only on the observed data and not the value itself. Because this assumption may not be valid for this data, we also perform MI using an $\delta$ adjustment, which is a flexible and transparent method to impute missing data under Missing Not at Random (MNAR) assumptions.\cite{30} For MNAR, the probability that a value is missing depends on the unobserved data. The $\delta$-MI method (method 5) uses (5), however, prior to imputing the missing data, a fixed quantity, $\delta$, is added to the linear predictor of the regression model,

$$logit[P(Y_t = 1 | X_t)] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_{3t} + \beta_4 Y_1 + \ldots + \beta_{2+t} Y_{t-1} + \delta R_t. \quad (6)$$

Here, $\delta$ represents the difference in log-odds of ILI between participants who did not report compared with participants who did report.

Estimation of delta

Because the log-odds of ILI for participants who did not report are unknown, we use the retrospective reports to estimate this value and create annual $\bar{\delta}$ estimates. In other words,

$$\bar{\delta} = log \left[ \frac{odds(P(Y=1|R=1))}{odds(P(Y=1|R=0))} \right] \quad (7)$$

where,
Negative values of $\delta$ indicate that participants were less likely to report ILI for retrospective reports compared with reports submitted during the same week. Data are analyzed using R, version 3.3.2, and both MI methods are fit using the Multivariate Imputation by Chained Equations (MICE) package.[31,32]

Simulations

We evaluate the missing data methods assuming three missingness scenarios: MCAR, MAR, and MNAR, using simulated data. The data is simulated using a three-step process. First, 1000 FluTracking populations (n=30,000 each) are simulated using the characteristics (age group, sex, and vaccination status) of the 2016 influenza season participant population. Simulated participants are assigned an age group, sex, vaccination status, and 17 weeks of symptom reports, $Y_t$. These weekly symptom reports are simulated using a multinomial distribution, where $n$, which is Poisson distributed with an age-group specific mean, represents the total number of ILI reports for the participant and $q = \{q_1, \ldots, q_{17}\}$ is the vector of weekly percentage of the 2016 sentinel general practitioner consultations that were ILI as reported by AU’s Department of Health.[33] We simulate 17 missingness indicators, $R_{it}$, to reflect distribution of FluTracking participant reports (Supplemental Figure 1), using three missingness scenarios:

1. MCAR

$$R_{it} \sim \text{Binomial}(n = 17, p_i) \quad (7)$$

$$p_i = \begin{cases} 
0.05 \text{ with probability } 0.8 \\
\text{Uniform}(0.1, 0.95) \text{ with probability } 0.2 
\end{cases}$$
2. MAR

\[ R_{it} \sim Bernoulli(p_{it}) \quad (8) \]

\[ p_{it} = \frac{e^{Z_i + \gamma R_{it-1}}}{1 + e^{Z_i + \gamma R_{it-1}}} \]

\[ Z_i = \begin{cases} 
\log\left(\frac{0.05}{1 - 0.05}\right) & \text{with probability } 1 - W_i \\
\text{Uniform}\left(\log\left(\frac{0.1}{1 - 0.1}\right), \log\left(\frac{0.95}{1 - 0.95}\right)\right) & \text{with probability } W_i 
\end{cases} \]

\[ W_{it} = \frac{e^{\beta_0 + \beta_1 x_{1t} + \ldots + \beta_7 x_{7t}}}{1 + e^{\beta_0 + \beta_1 x_{1t} + \ldots + \beta_7 x_{7t}}} \]

\[ X = \{\text{age group (X_1 - X_9), sex, vaccination status}\} \]

\[ \gamma = -0.2 \text{ (i.e., } R_{it} \text{ is less likely to be observed if } R_{it-1} = 1) \]

3. MNAR

\[ R_{it} \sim Bernoulli(p_{it}) \quad (9) \]

\[ p_{it} = \frac{e^{Z_i + \gamma R_{it-1} + \delta Y_{it}}}{1 + e^{Z_i + \gamma R_{it-1} + \delta Y_{it}}} \]

\[ \delta = 0.3 \text{ (i.e., } R_{it} \text{ is more likely to be observed if } Y_{it} = 1) \]

The values of \( \beta_0 \) through \( \beta_7 \) are estimated for FluTracking data using multivariate logistic regression, similar to [23]. We define \( \delta \) equal to 0.3. However, we also present a sensitivity analysis that assesses how varying the MNAR assumption affects IR estimates (\( \delta \) equal to 0.3, 0.8, 1.3, 3.0, and 5.0). Finally, each of the missing data methods described in the previous section is applied to produce overall and age-specific IR and IP estimates, and the overall WP estimates for each simulated dataset. The estimates from each missing data method are compared to the original simulation parameters qualitatively using violin plots and...
quantitatively Normalized Root Mean Square Errors (NRMSE) normalized by the original parameter,

$$NRMSE = \sqrt{\frac{\sum_{i=1}^{T} (\theta_i - \theta)^2}{T}}$$  \hspace{1cm} (10)

for $T = 1000$. In this equation, B represents the measure of ILI burden (i.e., IR, IP, or WP).

RESULTS

Simulations

Incidence Rate

As shown in Figure 1, under each missingness scenario, assuming that all missing reports are non-ILI underestimates the IR and results in the largest NRMSE (Table 1). Under MCAR and MAR scenarios, available case and MI methods have smaller NRMSEs compared with the complete case and δ-MI methods. However, under MNAR, IR estimates using the δ-MI method have smaller NRMSEs compared with all other methods.

Table 1. Normalized Root Mean Square Errors, expressed as percentage, by age group for Incidence Rates (IR) andIncidence Proportions (IP), and overall for Weekly Prevalence (WP) under Missing Completely at Random (MCAR), Missing At Random (MAR), and Missing Not at Random (MNAR) scenarios.

<table>
<thead>
<tr>
<th>Measure of ILI Burden</th>
<th>Missingness Scenario</th>
<th>Age Group</th>
<th>Missing Data Available</th>
<th>Missing Are Non-ILI</th>
<th>Complete Case</th>
<th>Multiple Imputation</th>
<th>δ Multiple Imputation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incidence Rate MCAR</td>
<td>Overall</td>
<td>0.44</td>
<td>14.02</td>
<td>1.52</td>
<td>0.73</td>
<td>2.51</td>
<td></td>
</tr>
<tr>
<td>&lt;5</td>
<td>1.94</td>
<td>14.05</td>
<td>6.82</td>
<td>2.80</td>
<td>3.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-17</td>
<td>1.17</td>
<td>14.07</td>
<td>4.00</td>
<td>1.69</td>
<td>2.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-49</td>
<td>0.65</td>
<td>14.01</td>
<td>2.45</td>
<td>1.02</td>
<td>2.57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50+</td>
<td>0.67</td>
<td>14.06</td>
<td>2.43</td>
<td>1.05</td>
<td>2.68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAR Overall</td>
<td>0.56</td>
<td>14.59</td>
<td>1.73</td>
<td>0.83</td>
<td>2.57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;5</td>
<td>2.23</td>
<td>19.47</td>
<td>7.41</td>
<td>3.45</td>
<td>4.36</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Under all missingness scenarios, the complete case and imputation methods outperform the available case method (Figure 2, Table 1). For MCAR and MAR scenarios, the MI method has the smallest NRMSEs, the δ-MI method underestimates the IP, and the complete case method has the largest variation in estimates. Similar to IR, the δ-MI method is the best approach when data are MNAR. Estimates for the available case method are not shown because this method is the same as assuming that all missing reports are non-ILI when estimating the IP.
When estimating the WP, the available case method outperforms the other methods under each missingness scenario (Figure 3, Table 1). The method that assumes all missing reports are non-ILI underestimates the WP under each missingness scenario, and the $\delta$-MI method outperforms MI when data are MNAR. Because the complete case population is unknown until the end of the season, we did not assess this missing data method.

Sensitivity Analysis

Sensitivity Analysis results for calculation of overall and age group specific IRs under five MNAR scenarios are shown in Supplemental Figure 2. As the value of $\delta$ increases (i.e., ILI reports become less likely to be missing), the method that assumes missing reports are non-ILI becomes a better missing data method.

FluTracking

During the 2016, 2017, and 2018 influenza seasons, 29,671, 32,778, and 43,389 unique participants submitted at least one symptom report between week 25 and week 41, respectively. Across all influenza seasons, approximately 60% identified as female, and the median age of participants ranged from 47 to 49 years. The largest age group was 50+ years, followed by 18 to 49, 5 to 17, and finally <5. More than half were primary users who submitted reports on their own behalf, and 59%, 61%, and 67% of participants reported that they received the influenza vaccination during the 2016, 2017, and 2018 influenza seasons, respectively (Table 2).

**Table 2.** Descriptive statistics of the FluTracking cohort and participant reporting habits during the 2016, 2017, and 2018 influenza seasons. Continuous variables are displayed as median (25th percentile, 75th percentile) and categorical variables are displayed as n (%).
<table>
<thead>
<tr>
<th>Variable</th>
<th>2016</th>
<th>2017</th>
<th>2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants</td>
<td>n</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>Male</td>
<td>11,153 (37.59)</td>
<td>12,665 (38.64)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>17,267 (58.19)</td>
<td>19,277 (58.81)</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>&lt;5 (0)</td>
<td>5 (0.02)</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>&gt;1,245 (4.22)</td>
<td>831 (2.54)</td>
</tr>
<tr>
<td>Age (years)</td>
<td>Median (Q1, Q3)</td>
<td>47 (31, 58)</td>
<td>48 (31, 59)</td>
</tr>
<tr>
<td>Age Group (years)</td>
<td>&lt;5</td>
<td>963 (3.25)</td>
<td>1,062 (3.24)</td>
</tr>
<tr>
<td></td>
<td>5 to 17</td>
<td>3,387 (11.42)</td>
<td>3,813 (11.63)</td>
</tr>
<tr>
<td></td>
<td>18 to 49</td>
<td>11,812 (39.81)</td>
<td>12,306 (37.54)</td>
</tr>
<tr>
<td></td>
<td>50+</td>
<td>13,509 (45.53)</td>
<td>15,597 (47.58)</td>
</tr>
<tr>
<td>Household</td>
<td>Primary user</td>
<td>17,525 (59.06)</td>
<td>19,170 (58.48)</td>
</tr>
<tr>
<td></td>
<td>Household member</td>
<td>12,146 (40.94)</td>
<td>13,608 (41.52)</td>
</tr>
<tr>
<td>Vaccinated</td>
<td>Yes</td>
<td>17,526 (59.07)</td>
<td>19,966 (60.91)</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>12,145 (40.93)</td>
<td>12,812 (39.09)</td>
</tr>
<tr>
<td>Reports</td>
<td>Total</td>
<td>452,627 (100)</td>
<td>498,465 (100)</td>
</tr>
<tr>
<td></td>
<td>Non-ILI</td>
<td>442,817 (97.83)</td>
<td>486,380 (97.58)</td>
</tr>
<tr>
<td></td>
<td>ILI</td>
<td>9810 (2.17)</td>
<td>12,085 (2.42)</td>
</tr>
<tr>
<td>Reports Submitted</td>
<td>Total</td>
<td>398,496 (88.04)</td>
<td>432,733 (86.81)</td>
</tr>
<tr>
<td></td>
<td>Non-ILI</td>
<td>389,706 (97.79)</td>
<td>422,139 (97.55)</td>
</tr>
<tr>
<td></td>
<td>ILI</td>
<td>8790 (2.21)</td>
<td>10,594 (2.45)</td>
</tr>
<tr>
<td>Retrospective</td>
<td>Total</td>
<td>54,131 (11.96)</td>
<td>66,096 (13.26)</td>
</tr>
<tr>
<td></td>
<td>Non-ILI</td>
<td>53,111 (98.12)</td>
<td>64,605 (97.74)</td>
</tr>
<tr>
<td></td>
<td>ILI</td>
<td>1020 (1.88)</td>
<td>1491 (2.26)</td>
</tr>
<tr>
<td>Reports per Week</td>
<td>Median (Q1, Q3)</td>
<td>26,570</td>
<td>29,370</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(26,470, 26,860)</td>
<td>(29,200,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>29,520)</td>
</tr>
<tr>
<td>Reports per Participant</td>
<td>Median (Q1, Q3)</td>
<td>17 (16, 17)</td>
<td>17 (16, 17)</td>
</tr>
<tr>
<td>delta (δ)</td>
<td></td>
<td>-0.158</td>
<td>-0.082</td>
</tr>
<tr>
<td>OR (e^δ)</td>
<td></td>
<td>0.85</td>
<td>0.92</td>
</tr>
</tbody>
</table>
The descriptive statistics of the reporting habits of participants during the 2016, 2017, and 2018 influenza seasons are also shown in Table 2. The total number of symptom reports submitted during the influenza season increased from approximately 450,000 in 2016, to almost 500,000 in 2017, and finally to over 650,000 in 2018. The median number of weekly reports also increased from approximately 26,000 in 2016 to 39,000 in 2018. During each influenza season, the median number of reports per participant was 17 (16, 17), indicating that more than half of participants submitted a symptom report each week (Supplemental Figure 1). While most reports were submitted within one week of the symptom report date, a larger proportion of these reports were ILI compared to the retrospective reports, resulting in a negative $\delta$.

Although the exact value of $\delta$ varies by season, the corresponding ORs of reporting ILI for a retrospective report compared with reporting ILI for a report submitted the same week are all less than 1. The OR was 0.85 in 2016, 0.92 in 2017, and 0.80 in 2018.

As shown in Supplemental Figure 1, most FluTracking participants register before week 25, however, the percentage of participants lost to follow-up increases as the season progresses, as shown by the bright green bars. The percentage of retrospective reports (dark green) is fairly consistent through the season, but the proportion of missing reports (light green) appears to increase until mid-season, at which point it slowly decreases as more participants are lost to follow-up (bright green). These patterns are consistent across all influenza seasons.

Incidence Rate

Supplemental Table 1 and Supplemental Figure 3 display overall and age-group specific IRs and 95% CIs, expressed as number of ILI reports per 10,000 person weeks, by influenza season. Although the 2017 influenza season had higher IRs compared with the 2016 and 2018 seasons,
the general patterns in estimates are consistent across all seasons and age groups. The method that assumes that all missing reports are non-ILI has the lowest IR estimates, whereas available case and MI methods have the highest IR estimates. As expected, IR estimates from the $\delta$-MI method are slightly less than estimates from the MI method without the $\delta$ adjustment, reflecting that missing reports are less likely to be ILI. In most age groups, IR estimates from the complete case method are similar or slightly greater than estimates from the method that assumes all missing reports are non-ILI.

Incidence Proportion

The overall and age-group specific IPs and 95% CIs for each influenza season are shown in Supplemental Table 2 and Supplemental Figure 4. Similar to IR estimates, IPs estimates from the method that assumes all missing reports are non-ILI and complete case method are less than IP estimates from the MI and $\delta$-MI methods. However, the differences in IP estimates appear to be less pronounced compared with the differences in IR estimates.

Weekly Prevalence

Near-real time WP estimates are shown in Supplemental Figure 5. We also present the complete data with the retrospective reports for comparison. The method that assumes all missing reports are non-ILI results in WP estimates lower than the other methods. WP estimates from the MI method are slightly larger than WP estimates from the available case method and the $\delta$-MI method. The estimates from these two methods are similar to the complete data.

DISCUSSION

As we have shown, participatory surveillance systems, due to their design, suffer from
substantial missing data that is likely informative. We have described five ways to handle missing data, drawing upon traditional statistical practice, and shown how these perform under different missingness scenarios. Our simulations show that the optimal missing data method depends on the measure of ILI burden and the underlying missingness model. Of note, the δ-MI method provides estimates of ILI burden that are similar to the true parameter under MNAR scenarios. When we apply these methods to a participatory surveillance system in Australia, we find that the δ-MI method accurately predicted end of season WP estimates from real-time data.

In 2020, Liu et al. estimated WP of ILI in AU using a Bayesian approach that adjusts FluTracking estimates by the probability of an individual reporting in each week given their past reporting.[34] Similar to our approach, they show that estimates of ILI burden are affected by models that correct for user behavior and that users are more likely to report when ill. However, their weekly behavior-adjusted estimates of prevalence were less than the WP estimates that assume all missing reports were non-ILI during the 2017 influenza season. This difference may indicate that their model assumes that the weekly percentage of ILI in the FluTracking population is greater than the weekly percentage of ILI in the general AU population. Their approach also differs because they do not leverage the information within retrospective reports.

While AU estimates of IRs and IPs for ILI are not currently available, laboratory confirmed influenza is a nationally notifiable disease in AU.[35] In 2016 the age-specific rate estimates of laboratory confirmed influenza ranged from 24.67 to 123.72 per 10,000 population.[35] As expected, these estimates are lower than the ILI estimates from FluTracking because the
reported number of cases represent only a proportion of the total cases in the community, that is, only those cases for which health care was sought, a test conducted and a diagnosis made, followed by a notification to health authorities and does not include the many other viral and bacterial causes of ILI. But the patterns in age-group specific estimates are similar, as are the weekly trends.

This study has several limitations. FluTracking has a highly engaged user population with a smaller proportion of missing reports compared with other systems. Because imputation methods may not be feasible with large amounts of missing data, other systems may need to select a cohort of highly engaged users before implementing MI methods. FluTracking is also unique because it provides users with the opportunity to complete missing surveys. This system accommodation not only adds approximately 10% more weekly reports, but also provides a way to estimate $\delta$ for imputation. Additionally, in our MNAR simulation model, because the value of $\delta$ is known, the resulting $\delta$-MI model can be properly parameterized. In reality, this value is unknown, and the estimated value is based only on retrospective reports, which may be affected by recall bias. However, our sensitivity analysis shows that under modest changes in $\delta$, for example, increasing $\delta$ (OR) from 0.3 (1.35) to 0.8 (2.22) or 1.3 (3.67), the $\delta$-MI method still outperforms the method that assumes missing reports are non-ILI. Finally, only one value of $\delta$ was used for imputation for all age groups and sex. In the future, this parameter can be easily updated and adapted during an influenza season to provide real time, demographic-specific estimates. A Bayesian approach to estimation can also be used.

National estimates of influenza burden in the population are essential to understand the overall global burden of influenza disease.[37] Alternative data sources, such as FluTracking, have the
potential to complement traditional sentinel systems by capturing a population not routinely included among the other healthcare-based systems and identifying healthcare seeking behaviors and testing rates among ill persons. [2] While these surveillance systems now play an important role in surveillance, data analysis needs to be handled carefully to draw appropriate and useful conclusions. The $\delta$-MI method provides a straightforward, interpretable approach to appropriately handling missing data in participatory surveillance systems.
ABBREVIATIONS

AU: Australia
CI: Confidence Interval
COVID-19: Coronavirus 2019
FNY: Flu Near You
IL: Influenza-Like Illness
IP: Incidence Proportion
IR: Incidence Rate
MAR: Missing at Random
MCAR: Missing Completely at Random
MI: Multiple Imputation
MICE: Multivariate Imputation by Chained Equations
MMWR: Morbidity and Mortality Weekly Report
MNAR: Missing Not at Random
NRMSE: Normalized Root Mean Square Error
US: United States of America
WHO: World Health Organization
WP: Weekly Prevalence

ACKNOWLEDGEMENTS

The authors acknowledge all the participants who contributed their time and information to the FluTracking system.

COMPETING INTERESTS
The authors have no competing interests to declare.

FUNDING

The FluTracking surveillance system is funded by the Australian Government Department of Health.

REFERENCES


2. Smolinski MS, Crawley AW, Olsen JM, Jayaraman T, Crawley AW. Participatory Disease Surveillance: Engaging Communities Directly in Reporting, Monitoring, and Responding to Health Threats Corresponding Author: 2017; 3.


5. Moberley S, Carlson SJ, Durrheim DN, Dalton CB, DN. FluTracking: Weekly online

Communicable Diseases Intelligence 2019


http://dx.doi.org/10.1016/j.amepre.2016.10.038.

http://publichealth.jmir.org/2017/2/e18/


27. Kirkwood BR, Sterne J. Essentials of Medical Statistics. 2nd ed. Blackwell Science Ltd,


36. National Notifiable Diseases Surveillance System. 2016. Available at:


FIGURE LEGENDS

Figure 1. Distributions of overall and age-group specific Incidence Rates simulated under three missingness scenarios: Missing Completely at Random (MCAR), Missing at Random (MAR), and Missing Not at Random (MNAR). Dotted line represents the original simulated parameter.

Figure 2. Distributions of overall and age-group specific Incidence Proportions simulated under three missingness scenarios: Missing Completely at Random (MCAR), Missing at Random (MAR), and Missing Not at Random (MNAR). Dotted line represents the original simulated parameter.

Figure 3. Time series of Weekly Prevalence simulated under three missingness scenarios: Missing Completely at Random (MCAR), Missing at Random (MAR), and Missing Not at Random (MNAR).
Weekly Prevalence

Week

Adjustment Method
- Original
- Available Case
- Missing non-ILI
- MI
- Delta MI