



HHS Public Access

Author manuscript

Eval Health Prof. Author manuscript; available in PMC 2024 July 31.

Published in final edited form as:

Eval Health Prof. 2018 December ; 41(4): 474–492. doi:10.1177/0163278717725372.

Estimation of the Percentage of Newly Diagnosed HIV-Positive Persons Linked to HIV Medical Care in CDC-Funded HIV Testing Programs

Guoshen Wang¹, Yi Pan¹, Puja Seth¹, Ruiguang Song¹, Lisa Belcher¹

¹Division of HIV/AIDS Prevention, Centers for Disease Control and Prevention, Atlanta, GA, USA

Abstract

Missing data create challenges for determining progress made in linking HIV-positive persons to HIV medical care. Statistical methods are not used to address missing program data on linkage. In 2014, 61 health department jurisdictions were funded by Centers for Disease Control and Prevention (CDC) and submitted data on HIV testing, newly diagnosed HIV-positive persons, and linkage to HIV medical care. Missing or unusable data existed in our data set. A new approach using multiple imputation to address missing linkage data was proposed, and results were compared to the current approach that uses data with complete information. There were 12,472 newly diagnosed HIV-positive persons from CDC-funded HIV testing events in 2014. Using multiple imputation, 94.1% (95% confidence interval (CI): [93.7%, 94.6%]) of newly diagnosed persons were referred to HIV medical care, 88.6% (95% CI: [88.0%, 89.1%]) were linked to care within any time frame, and 83.6% (95% CI: [83.0%, 84.3%]) were linked to care within 90 days. Multiple imputation is recommended for addressing missing linkage data in future analyses when the missing percentage is high. The use of multiple imputation for missing values can result in a better understanding of how programs are performing on key HIV testing and HIV service delivery indicators.

Keywords

HIV testing; multiple imputation; linkage to care; estimation; program evaluation

Introduction

More than 1.1 million persons are living with HIV (PLWH) in the United States, and 15.0% are unaware of their HIV infection at the end of 2014 (Singh, Song, Satcher Johnson, McCray, & Hall, 2017). Certain groups are disproportionately affected by HIV, such as gay men, bisexual men, and other men who have sex with men as well as racial/ethnic minority populations, specifically Blacks or African Americans and Hispanics or Latinos (Centers for

Reprints and permission: sagepub.com/journalsPermissions.nav

Corresponding Author: Guoshen Wang, Division of HIV/AIDS Prevention, Centers for Disease Control and Prevention, 1600 Clifton Road NE; MS E-59, Atlanta, GA 30329, USA. ifd1@cdc.gov.

Declaration of Conflicting Interests

The author(s) declared no potential conflict of interest with respect to the research, authorship, and/or publication of this article.

Disease Control and Prevention, 2015b). HIV testing and linking HIV-positive persons to HIV medical care are crucial first steps in the HIV continuum of care. From 2010 to 2014, the rate of HIV diagnoses has remained stable (Centers for Disease Control and Prevention, 2015b). Despite this success, increased efforts are needed to decrease HIV infection through prevention and treatment efforts.

Early diagnoses, linkage to HIV medical care, and retention in care are important for decreasing viral load and reducing morbidity and mortality among persons living with HIV (Cohen et al., 2011; The Insight Start Study Group, 2015). Self-report data have indicated that persons engage in less frequent risky sex when they know they have HIV, but ongoing prevention services are needed for persons who know they are HIV positive (Marks, Crepaz, Senterfitt, & Janssen, 2005). Once patients are linked and retained in HIV medical care, this leads to a decrease in viral load, decrease in HIV transmission to others, and ultimately a decrease in HIV incidence (Cohen et al., 2011; Gardner, McLees, Steiner, Del Rio, & Burman, 2011).

The Centers for Disease Control and Prevention (CDC) funds state and local health departments to collect data on service provision of HIV testing, linkage to HIV medical care, and other HIV-related services as part of the National HIV Prevention Program Monitoring and Evaluation (NHM&E) system. CDC receives, analyzes, and disseminates these data in order to evaluate CDC-funded HIV testing programs. Issues with missing data or invalid responses negatively affect a thorough and reliable evaluation of the HIV testing programs. Data quality is important because higher quality data increase the credibility of the data and the confidence in the use of the data (USAID, 2009). Beginning in 2010, CDC implemented data quality assurance monitoring and grantee feedback for HIV testing data. However, despite improvements in data quality, monitoring and evaluation of HIV testing and prevention programs continue to be challenging because of missing program data.

Missing data create challenges for determining accurate linkage percentages and progress made in linking HIV-positive patients to HIV medical care. Many current CDC reports and publications have provided a minimum and maximum percentage to describe linkage because of missing data (Centers for Disease Control and Prevention, 2014, 2015a, 2016; Seth, Wang, Collins, & Belcher, 2015). For the minimum percentage, the denominator includes missing data and likely underestimates the linkage percentage, whereas for the maximum percentage, the denominator excludes missing data and potentially overestimates the linkage percentage (Seth, Wang, et al., 2015). In using this approach, the range is large and no point estimation and confidence intervals (CIs) are provided. The maximum percentage that is currently calculated for linkage is exactly the same as the methods of a complete case analysis (He, 2010; Liu & De, 2015), where only observations without any missing values in variables are used to calculate the final indicator estimate. However, additional methods and approaches to address missing values, such as multiple imputation, need to be examined to ensure the most robust data analysis and interpretation.

Multiple imputation is a statistical procedure for addressing missing data in statistical analyses (Graham & Hofer, 2000; Graham, Hofer, Donaldson, MacKinnon, & Schafer, 1997; Little & Rubin, 2002; Little & Schenker, 1995). Multiple imputation replaces each

missing value with a set of plausible values that preserve the statistical distribution of the imputed variable and the relationship with other variables in the imputation model. The multiply imputed data sets are analyzed using standard procedures for complete data. Results from these analyses are then combined to get the final estimates (Rubin, 1987, 1996). Multiple imputation is a popular approach to address missing data in many fields, such as the Blood Transfusion Service of Namibia (NAMBTS) Nationally Representative Census (Liu & De, 2015), the Risk Factor Redistribution of the National HIV/AIDS Surveillance Data (Harrison, Kajese, Hall, & Song, 2008), and the Cardiovascular Health Study (Arnold & Kronmal, 2003). However, to our knowledge, the multiple imputation method has not been used for linkage to HIV medical care program data. Therefore, the primary objective of this article is to compare the complete case analysis and multiple imputation approaches to address missing values on linkage to HIV medical care data from CDC-funded HIV testing programs. This comparison will provide an understanding of the better approach to estimate percentages for linkage to HIV medical care within any time frame and within 90 days resulting in enhanced program monitoring and improved program performance.

Methods

Data Source

CDC funds 61 state and local health departments to implement HIV testing and prevention services programs and to collect and report related data as part of CDC's NHM&E system. Data for each CDC-funded HIV testing event are collected by local service providers and submitted to CDC biannually without personal identifiers via a secure, online CDC-provided system. CDC uses these data for monitoring and evaluation of HIV testing and HIV-related service delivery. Linkage to HIV medical care is a variable of public health significance that is assessed among all newly diagnosed HIV-positive persons (Centers for Disease Control and Prevention, 2014, 2015a, 2016; Seth, Figueroa, Wang, Reid, & Belcher, 2015; Seth, Walker, Hollis, Figueroa, & Belcher, 2015; Seth, Wang, et al., 2015). In the current analyses, 2014 data were used from the 61 health department jurisdictions who submitted their data to CDC, but data are limited to newly diagnosed HIV-positives persons (Centers for Disease Control and Prevention, 2016). This data collection effort is considered a nonresearch, program evaluation activity by the CDC; therefore, approval from the institutional review board was not required. The Office of Management and Budget approved this activity.

Definition of Missing/Unusable Data

On the HIV test form, three hierarchical questions are related to the calculation of linkage to HIV medical care. The first level question is "Was client referred to HIV medical care?" (i.e., referral to HIV medical care). If the answer is no for referral to HIV medical care, then a client should be asked the reason they were not referred to medical care (already in HIV medical care or declined). However, if the answer is yes to the first-level question, then the following second-level question should be assessed: "Did client attend the first appointment?" (i.e., linkage within any time frame). If the answer is yes for this second-level question, then a client should respond to the third-level question "Was the first appointment within 90 days of the HIV test?" (i.e., linkage within 90 days). Because of the multiple,

hierarchical levels used to capture linkage to HIV medical care, the response for each of the questions could be a nonresponse or an unusable option (e.g., don't know, declined, missing). These potential responses generate missing values for linkage-related variables, and this missing pattern is called a monotone missing pattern (Figure 1). Monotone missing data indicate where the variables can be arranged so that all $Y = Y_{j+1}, \dots, Y_k$ are missing for cases where Y_j is missing, for all $J = 1, \dots, K - 1$

Data Cleaning

With CDC-funded HIV testing program data, some contradictions exist among these hierarchical questions related to linkage. Before applying multiple imputation to the linkage data, the data are cleaned following certain criteria. First, if the question assessing linkage to HIV medical care within 90 days (third-level question) has values, then linkage within any time frame (second-level question) is considered “yes” and not categorized as a missing value. If linkage within any time frame (second-level question) has a value, then referral to medical care (first-level question) is considered “yes.” If reason client not referred to medical care has a value, then referral to medical care (first-level question) is considered “no.” Finally, if both linkage to medical care and “reason client not referred to medical care” have values, then the value of linkage to medical care will overwrite the value of “reason client not referred to medical care.”

Data Analyses Before Using Multiple Imputation

The percentage of referral to care, linkage within any time frame, and linkage within 90 days is calculated using complete case scenario. The formulas for the definition of linkage within any time frame are shown below:

$$\text{Complete case percentage} = \frac{\text{linked persons}}{\text{all newly diagnosed HIV-positive persons} - \text{linkage missed persons}}$$

Multiple Imputation

Multiple imputation requires specification of a statistical model and is considered a modern approach for addressing missing data. The multiple imputation approach draws a random sample of the missing values from its posterior predictive distribution. This process results in valid statistical inferences that properly reflect the uncertainty due to missing values. Specifically, multiple imputation follows three distinct phases (Harrison et al., 2008; He, 2010): (1) impute missing values under an appropriate model incorporating random variation, repeat m times, generating m data sets; (2) analyze each data set separately to obtain desired parameter estimates and standard error; and (3) combine results of the m analyses by computing the mean of the m parameter estimates and a variance estimate that includes both within and between imputation variations. The number of imputations, m , is determined by the amount of data with missing values and the desired relative efficiency.

The multiple imputation method maintains the original variability of the missing data by creating imputed values that are based on variables correlated with the missing data and the reason as to why the data are missing. Generating iterations of the missing data

and observing the variability between the imputed datasets account for uncertainty due to missing values (Harrison et al., 2008; Little & Rubin, 2002).

In general, missing at random (MAR) is one of the assumptions to perform the traditional multiple imputation (Rubin, 1987, 1996). When the data are MAR, multiple imputation can lead to consistent, asymptotically efficient, and asymptotically normal estimates. Multiple imputation can be extended to nonignorable missingness as well. The model used to generate the imputed values must be “correct” in some sense (i.e., must include all anticipated predictor variables), and the model used in the analysis must be consistent with the model used in the imputation (Harrison et al., 2008; Little & Rubin, 2002; Rubin, 1996).

Identifying important covariates in the imputation model is critical. We generally cannot be sure whether data really are MAR or whether the missingness depends on unobserved predictors or the missing data itself. The fundamental difficulty is that these potential “lurking variables” are unobserved by definition and therefore can never be ruled out. In practice, as many predictors as possible are included in a model so that the MAR assumption is reasonable. The correlations are calculated by treating the missing values as a separate category in those predictor variables while using the data set with available outcome variables.

Among newly diagnosed HIV-positive persons, the correlation of covariates (i.e., region, HIV test setting, number of tests, age, race/ethnicity, and gender) with reported linkage information (i.e., referral to HIV medical care, linkage within any time frame, and linkage within 90 days of the HIV test) was tested. These covariates are similar to those used with HIV surveillance data risk factor redistribution. Cramer’s V statistic is used to measure the correlation between linkage variables and covariates (Kendall & Stuart, 1979). Covariates having a Cramer’s V statistic $>.04$ with any linkage variable were included in the analysis (Harrison et al., 2008). Based on the proportion of data with missing values (4% for referral to HIV medical care, 13% for linkage within any time frame, and 29% for linkage within 90 days) and a desired relative efficiency of 95%, each missing value was imputed 10 times. Imputation was implemented through three steps. First, 10 values for each missing value of referral to HIV medical care were imputed. Second, attended first medical appointment was imputed among those who were known to be referred to HIV medical care or unknown but classified as referred to HIV medical care through imputation in the first step, one for each of the 10 data sets generated from the first step. Finally, first appointment within 90 days was imputed among those who were known to attend the first medical appointment or unknown but classified as attended first medical appointment through imputation in the second step, one for each of the 10 data sets generated from the second step. Each imputed data set included all three variables with either known or imputed values.

Comparisons were conducted among the percentage of referral to HIV medical care, linkage to HIV medical care within any time frame, and linkage to HIV medical care within 90 days before and after using the multiple imputation method in the overall and subgroup population. Relative differences in percentage were also calculated to evaluate the size of the difference between estimates from the complete case and multiple imputation analyses,

which is the difference between the two estimates (the complete case estimate minus the multiple imputation estimate) divided by the complete case estimate.

Results

There were 12,472 newly diagnosed HIV-positive persons from CDC-funded HIV testing events in 2014. After data cleaning, the missing percentage was 4% (552/12,472) for referral to HIV medical care, 13% (1,655/12,472) for linkage within any time frame, and 29% (3,666/12,472) for linkage within 90 days. The 2014 data are a monotone missing pattern in nature (Table 1). In the subgroup population, the missing percentage range was 0.0–22.9% for referral to HIV medical care, 1.3–43.2% for linkage within any time frame, and 6.5–52.0% for linkage within 90 days (Appendix Table A1).

The variables retained and used in the multiple imputation models to impute values for referral to HIV medical care, linkage within any time frame, and linkage within 90 days were region of tests reported, test setting of where an HIV test was conducted, number of HIV tests, age at diagnosis, and race/ethnicity. Gender was not selected in the final imputation model. Covariates were selected based on Cramer's V statistic; for example, for region, the Cramer's V statistic was 0.150, 0.140, and 0.285 for referral to HIV medical care, linkage within any time frame, and linkage within 90 days, respectively (Table 2).

The results indicated that among newly diagnosed HIV-positive persons, before imputation, 94.5% (11,262/11,920) were referred to care, using the completed complete case analysis method. After imputation, the percentage for referral to HIV medical care was 94.1% (95% CI: [93.7%, 94.6%]) (Table 3). For linkage within any time frame, before imputation, 88.4% (9,559/10,817) of newly diagnosed HIV-positive persons were linked within any time frame, using the complete case analysis method. After imputation, the percentage for linkage within any time frame was 88.6% (95% CI: [88.0%, 89.1%]; Table 3). For linkage within 90 days, before imputation, 81.8% (7,200/8,806) of newly diagnosed HIV-positive persons were linked within 90 days, using the complete case analysis method. After imputation, the percentage for linkage within 90 days was 83.6% (95% CI: [83.0%, 84.3%]; Table 3).

Overall, there were significant differences between the complete case analysis and multiple imputation results for linkage within 90 days. There were no significant differences between the complete case analysis results and multiple imputation results for referral to HIV medical care and linkage within any time frame (Table 3). The relative difference had a range from –1.3 to 2.27% in referral to HIV medical care, –4.48 to 12.15% in linkage to HIV medical care within any time frame, and –30.64 to 16.12% in linkage to HIV medical care within 90 days. A difference is considered as significant if the complete case analysis result is not covered by the 95% CI from multiple imputation result. However, with the subgroup analyses, there were significant differences between the complete case analysis results and multiple imputation results for linkage within any time frame and within 90 days. For example, for tests conducted in an emergency room, 76.8% of newly diagnosed HIV-positive persons were linked within any time frame, using the complete case analysis method. After imputation, the percentage for linkage within any time frame was 79.6% (95% CI: [77.1%, 82.2%]). The relative difference was –3.65%. For tests conducted among

persons aged 20–29, 81.3% of newly diagnosed HIV-positive persons were linked within 90 days, using the complete case analysis method. After imputation, the percentage for linkage within 90 days was 83.0% (95% CI: [81.9%, 84.1%]). The relative difference was –2.09%. For tests conducted among Blacks, 78.4% of newly diagnosed HIV-positive persons were linked within 90 days, using the complete case analysis method. After imputation, the percentage for linkage within 90 days was 80.6% (95% CI: [79.5%, 81.7%]). The relative difference was –2.81%. For tests conducted among Hispanics, 85.6% of newly diagnosed HIV-positive persons were linked within 90 days, using the complete case analysis method. After imputation, the percentage for linkage within 90 days was 87.5% (95% CI: [86.1%, 88.9%]). The relative difference was –2.22%. For tests conducted in an emergency room, 59.9% of newly diagnosed HIV-positive persons were linked within 90 days, using the complete case analysis method. After imputation, the percentage for linkage within 90 days was 76.2% (95% CI: [73.4%, 79.0%]; Table 3). The relative difference was –27.21%.

Discussion

To our knowledge, this is the first article to evaluate programmatic linkage to HIV medical care data utilizing the multiple imputation method. Overall, the percentage referred to HIV medical care and linked within any time frame after imputation was not significantly different than the percentage found using the complete case analysis method. The missing percentage was 4% for referral to HIV medical care and 13% for linkage within any time frame. However, the percentage linked within 90 days was 83.6% (95% CI: 83.0%, 84.3%) after imputation, which is significantly different than the percentage found using the complete case analysis method (81.8%). The missing percentage was high at 29% (3,666/12,472) for linkage within 90 days. These findings suggest that the multiple imputation method is potentially a valid approach to address missing linkage data received from CDC-funded HIV testing programs, particularly when the missing percentage is high (> 15%). But in the current study, referral to HIV medical care, linkage within any time frame, and linkage within 90 days have a hierarchical relationship, and these variables need to be considered together when using the multiple imputation method to address missing program data on linkage.

The multiple imputation and complete case analysis results were not significantly different for referral, including by subgroup, and linkage to medical care in any time frame indicating that either the missing scheme might be missing completely at random (MCAR) or, as suggested previously, the missing data percentage is low. Subgroup differences for linkage to medical care in any time frame suggest that MCAR might not be applicable for subgroups. The relative differences were small in referral to HIV medical care and linkage to HIV medical care within any time frame because their missing proportion were small. However, complete case analysis generally has major deficiencies (He, 2010; Little & Rubin, 2002). The results can be biased when data are not MCAR. In addition, the reduction in statistical power by discarding cases is a major drawback (He, 2010; Liu & De, 2015). Other statistical methods for addressing missing values have been actively pursued in recent years, including maximum-likelihood (ML) estimation (Enders, 2001), Bayesian estimation (Oba et al., 2003), and full-information maximum-likelihood (FIML) estimation (Allison, 2012), all of which are based on the assumption that data are MAR or MCAR. The relevant discussion

can be found in references (Enders, 2001; Little & Rubin, 2002; Liu & De, 2015; Schafer & Graham, 2002).

Among CDC-funded HIV testing program data, the percentage of missing data and the percentage of newly diagnosed HIV-positive persons linked within any time frame and linked within 90 days have improved each year since 2011 (Centers for Disease Control and Prevention, 2013, 2014, 2015a, 2016). However, national HIV prevention goals were updated in 2015 to include linkage to HIV medical care within 30 (rather than 90) days. As programs report on linkage for this shorter time period, missing data may increase necessitating the use of imputation methods for program evaluation. Moreover, it is possible that these same methods could be applied to other bars of the HIV continuum of care (e.g., retention in HIV care).

The findings from the current analyses were similar to the maximum percentages reported in the literature (calculated using the complete case definition as used here) for linkage from the national-level HIV testing program data (Centers for Disease Control and Prevention, 2016; Seth, Wang, et al., 2015). This findings suggest that the maximum percentages, rather than the minimum percentages, reported in previous testing reports are a more accurate linkage percentage among CDC-funded HIV testing programs.

This study is not without limitations. For starters, because data are cleaned based on a hierarchical relationship among three variables, this process could generate bias, as some assumptions are made about the data depending on values received in one of the three variables. Additionally, data entry errors may contribute to missing data or errors in one of the three linkage-related variables. Unmeasured factors that are not included in the imputation model, such as grantees' capacity, may contribute to missing data. Furthermore, analyses were conducted on linkage data received from CDC-funded grantees. Therefore, these results and methods may not be generalizable to linkage data from other sources. Finally, differences that are statistically significant may not be clinically or practically significant because small difference can be statistically significant when sample size is large enough. On the other hand, a large difference could be statistically insignificant due to lack of testing power when sample size is small. Although the overall sample size is large, the sample sizes for some subgroups could be small.

Practice Implications

The use of multiple imputation to adjust for missing data and estimate the linkage percentage in CDC-funded HIV testing program data is desirable (Seth, Wang, et al., 2015). Multiple imputation produces unbiased parameter estimates if the model assumption is satisfied. In addition, multiple imputation methods are available in easy-to-use software (He, 2010), such as the SAS procedure multiple imputation (SAS Institute Inc., 2013) with a discriminant function analysis. The current analyses revealed that the multiple imputation method worked in addressing missing data for linkage to HIV medical care in 90 days. The usefulness of this method for assessing program performance on key HIV testing and HIV service delivery indicators should be further investigated by imputing missing values for other HIV testing data variables, such as referral to HIV partner services, referral to HIV prevention services, and risk factors.

Acknowledgments

The authors gratefully acknowledge Janet Heitgerd, PhD, and the Prevention Program Branch at the Centers for Disease Control and Prevention for monitoring of the HIV testing programs.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Appendix

Table A1.

Missing Percentage in Subgroup Population After Data Cleaning, 2014.

Characteristics	Missing of Referral to HIV Medical Care (%)	Missing of Linkage to HIV Medical Care Within any Time frame (%)	Missing of Linkage to HIV Medical Within 90 Days (%)
Age at test (years)			
13–19	6.0	13.1	33.1
20–29	4.6	13.3	29.9
30–39	4.2	12.7	28.3
40–49	4.1	13.3	28.6
50+	4.3	14.1	29.4
Other	10.0	23.3	50.0
Race/ethnicity			
White	4.3	13.3	28.6
Black	5.1	14.3	30.2
Hispanic or Latino	3.4	11.4	28.6
Other	3.1	11.7	28.3
Setting			
Inpatient facilities	0.0	1.3	6.5
Outpatient facilities	4.1	12.4	24.0
Emergency room	4.6	21.3	52.0
HIV counseling and testing site	3.9	9.9	27.2
Community setting	7.7	24.6	38.1
Correctional facilities	3.3	6.8	19.0
Other	4.7	9.7	40.2
Gender			
Male	4.4	13.3	29.4
Female	5.0	13.1	30.0
Other	2.7	11.9	24.3
Region ^a			
Region 1	9.2	19.2	20.4
Region 2	5.7	9.9	11.6
Region 3	22.9	43.2	44.6
Region 4	0.4	7.0	7.2
Region 5	1.3	6.6	27.7

Characteristics	Missing of Referral to HIV Medical Care (%)	Missing of Linkage to HIV Medical Care Within any Time frame (%)	Missing of Linkage to HIV Medical Within 90 Days (%)
Region 6	5.1	23.1	23.1
Region 7	1.2	3.3	51.4
Region 8	5.5	13.6	23.6
Region 9	1.2	30.2	36.0
Region 10	4.5	15.0	15.3
Overall	4.4	13.3	29.4

^aRegion 1: Connecticut, Massachusetts, Maine, New Hampshire, Rhode Island, Vermont; Region 2: New Jersey, New York State, New York City, Pennsylvania, Philadelphia; Region 3: Illinois, Chicago, Indiana, Michigan, Ohio, Wisconsin; Region 4: Iowa, Kansas, Minnesota, Missouri, North Dakota, Nebraska, South Dakota; Region 5: District of Columbia, Delaware, Florida, Georgia Atlanta, Maryland, Baltimore, North Carolina, South Carolina, Virginia, West Virginia; Region 6: Alabama, Kentucky, Mississippi, Tennessee; Region 7: Arkansas, Houston, Louisiana, Oklahoma, Texas; Region 8: Arizona, Colorado, Idaho, Montana, New Mexico, Nevada, Utah, Wyoming; Region 9: Alaska, California, Los Angeles, San Francisco, Hawaii, Oregon, Washington; Region 10: Puerto Rico, US Virgin Islands.

References

- Allison PD (2012). Handling missing data by maximum likelihood. Paper presented at the SAS Global Forum 2012, Orlando, FL.
- Arnold AM, & Kronmal RA (2003). Multiple imputation of baseline data in the cardiovascular health study. *American Journal of Epidemiology*, 157, 74–84. Centers for Disease Control and Prevention. (2013). HIV Testing at CDC Funded [PubMed: 12505893]
- Sites, United States, Puerto Rico and the U.S. Virgin Islands, 2011. Retrieved August 24, 2015, from www.cdc.gov/hiv/pdf/HIV_testing_report_2011_12.13.13_Version3.pdf
- Centers for Disease Control and Prevention. (2014). HIV Testing at CDC-Funded Sites, United States, Puerto Rico, and the U.S. Virgin Islands, 2012. Retrieved August 24, 2015, from www.cdc.gov/hiv/pdf/2012_HIV_Testing_Report_01-29-15.pdf
- Centers for Disease Control and Prevention. (2015a). CDC-Funded HIV Testing: United, States, Puerto Rico and the U.S. Virgin Islands, 2013. Retrieved August 24, 2015, from www.cdc.gov/hiv/pdf/library/reports/cdc-hiv-CDCFunded_HIV_Testing_UnitedStates_Puerto_Rico_USVI_2013.pdf
- Centers for Disease Control and Prevention. (2015b). Diagnosis of HIV infection in the United States and dependent areas, 2014. Retrieved December 2, 2015, from www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-report-us.pdf
- Centers for Disease Control and Prevention. (2016). CDC-Funded HIV Testing: United, States, Puerto Rico and the U.S. Virgin Islands, 2014. Retrieved May 19, 2016, from www.cdc.gov/hiv/pdf/library/reports/cdc-hiv-funded-testing-us-puerto-rico-2014.pdf
- Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, ... Team HS (2011). Prevention of HIV-1 infection with early antiretroviral therapy. *New England Journal of Medicine*, 365, 493–505. doi:10.1056/Nejmoa1105243 [PubMed: 21767103]
- Enders CK (2001). A primer on maximum likelihood algorithms available for use with missing data. *Structural Equation Modeling: A Multidisciplinary Journal*, 8, 128–141. doi: 10.1207/S15328007SEM0801_7
- Gardner EM, McLees MP, Steiner JF, Del Rio C, & Burman WJ (2011). The spectrum of engagement in HIV care and its relevance to test-and-treat strategies for prevention of HIV infection. *Clinical Infectious Diseases*, 52, 793–800. doi:10.1093/cid/ciq243 [PubMed: 21367734]
- Graham J, & Hofer S (2000). Multiple imputation in multivariate research. In Little SK & Baumert J (Eds.), *Modeling longitudinal and multiple-group data: Practical issues, applied approaches, and specific examples* (pp. 201–218). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

- Graham J, Hofer S, Donaldson S, MacKinnon D, & Schafer J (1997). Analysis with missing data in prevention research. In Bryant WM & West SG (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 325–366). Washington, DC: American Psychological Association.
- Harrison KM, Kajese T, Hall HI, & Song R (2008). Risk factor redistribution of the national HIV/AIDS surveillance data: An alternative approach. *Public Health Reports*, 123, 618–627. [PubMed: 18828417]
- He Y (2010). Missing data analysis using multiple imputation: Getting to the heart of the matter. *Circulation: Cardiovascular Quality and Outcomes*, 3, 98–105. doi:10.1161/CIRCOUTCOMES.109.875658 [PubMed: 20123676]
- Kendall M, & Stuart A (1979). *The advanced theory of statistics: Vol. 2—Inference and relationship* (Vol. 2). New York, NY: Macmillan.
- Little R, & Rubin D (2002). *Statistical analysis of missing data* (2nd ed.). Hoboken, NJ: John Wiley & Sons.
- Little R, & Schenker N (1995). Missing data. In Arminger CC & Sobel ME (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 39–76). New York, NY: Plenum Press.
- Liu Y, & De A (2015). Multiple imputation by fully conditional specification for dealing with missing data in a large epidemiologic study. *International Journal of Statistics in Medical Research*, 4, 287–295. doi: 10.6000/1929-6029.2015.04.03.7 [PubMed: 27429686]
- Marks G, Crepaz N, Senterfitt JW, & Janssen RS (2005). Meta-analysis of high-risk sexual behavior in persons aware and unaware they are infected with HIV in the United States: Implications for HIV prevention programs. *Journal of Acquired Immune Deficiency Syndromes*, 39, 446–453. [PubMed: 16010168]
- Oba S, Sato M. a., Takemasa I, Monden M, Matsubara K. i., & Ishii S (2003). A Bayesian missing value estimation method for gene expression profile data. *Bioinformatics*, 19, 2088–2096. doi: 10.1093/bioinformatics/btg287 [PubMed: 14594714]
- Rubin D (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: John Wiley & Sons.
- Rubin D (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, 91, 473–489.
- SAS Institute Inc. (2013). *SAS/STAT® 13.1 User's Guide The MI Procedure*. Cary, NC: SAS Institute Inc.
- Schafer JL, & Graham JW (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7, 147–177. [PubMed: 12090408]
- Seth P, Figueroa A, Wang G, Reid L, & Belcher L (2015). HIV testing, HIV positivity, and linkage and referral services in correctional facilities in the United States, 2009–2013. *Sexually Transmitted Diseases*, 42, 643–649. doi:10.1097/OLQ.0000000000000353 [PubMed: 26462190]
- Seth P, Walker T, Hollis N, Figueroa A, & Belcher L (2015). HIV testing and service delivery among Blacks or African Americans—61 health department jurisdictions, United States, 2013. *MMWR Morbidity and Mortality Weekly Report*, 64, 87–90. [PubMed: 25654608]
- Seth P, Wang G, Collins NT, & Belcher L (2015). Identifying new positives and linkage to HIV medical care—23 testing site types, United States, 2013. *MMWR Morbidity and Mortality Weekly Report*, 64, 663–667. [PubMed: 26110836]
- Singh S, Song R, Satcher Johnson A, McCray E, & Hall I (2017). HIV incidence, prevalence, and undiagnosed infections in men who have sex with men. Paper presented at the Conference on Retroviruses and Opportunistic Infections 2017, Seattle, WA. Retrieved from www.hivdent.org/_CROI2017/HIV%20Incidence.pdf
- The Insight Start Study Group. (2015). Initiation of antiretroviral therapy in early asymptomatic HIV infection. *New England Journal of Medicine*, 373, 795–807. doi:10.1056/NEJMoa1506816 [PubMed: 26192873]
- USAID. (2009). Performance monitoring & evaluation TIPS: Data quality standards. Retrieved Decemebr 8, 2015, from <http://usaidprojectstarter.org/sites/default/files/resources/pdfs/TIPS-DataQualityStandards.pdf>

Was the client referred to HIV medical care?									
<input type="checkbox"/> No →	Reason the client not referred to HIV Medical Care? <input type="checkbox"/> Client Already in Care <input type="checkbox"/> Client Declined Care								
<input type="checkbox"/> Yes →	Did the client attend the first appointment?								
<input type="checkbox"/> Don't Know	<table border="1"> <tbody> <tr> <td><input type="checkbox"/> Pending</td> <td rowspan="3">First medical appointment within 90 days of the HIV test?</td> </tr> <tr> <td><input type="checkbox"/> Confirmed: Accessed Service →</td> </tr> <tr> <td><input type="checkbox"/> Confirmed: Did Not Access Service</td> </tr> <tr> <td><input type="checkbox"/> Lost to Follow-Up</td> <td rowspan="3"><input type="checkbox"/> No <input type="checkbox"/> Yes <input type="checkbox"/> Don't Know</td> </tr> <tr> <td><input type="checkbox"/> No Follow-Up</td> </tr> <tr> <td><input type="checkbox"/> Don't Know</td> </tr> </tbody> </table>	<input type="checkbox"/> Pending	First medical appointment within 90 days of the HIV test?	<input type="checkbox"/> Confirmed: Accessed Service →	<input type="checkbox"/> Confirmed: Did Not Access Service	<input type="checkbox"/> Lost to Follow-Up	<input type="checkbox"/> No <input type="checkbox"/> Yes <input type="checkbox"/> Don't Know	<input type="checkbox"/> No Follow-Up	<input type="checkbox"/> Don't Know
<input type="checkbox"/> Pending	First medical appointment within 90 days of the HIV test?								
<input type="checkbox"/> Confirmed: Accessed Service →									
<input type="checkbox"/> Confirmed: Did Not Access Service									
<input type="checkbox"/> Lost to Follow-Up	<input type="checkbox"/> No <input type="checkbox"/> Yes <input type="checkbox"/> Don't Know								
<input type="checkbox"/> No Follow-Up									
<input type="checkbox"/> Don't Know									

Figure 1. Assessment items for linkage to care (Part of 2014 HIV test form template, Form Approved: OMB. No. 0920-0696, Exp. Date 3/31/2016).

Table 1.

Frequency Analysis of Missing Variables Among Newly Diagnosed HIV-Positive Persons in the United States, Before Imputation, 2014.

Referral to HIV Medical Care	Attended First Medical Appointment	Attended First Medical Appointment Within 90 Days	Newly Diagnosed HIV-Positive Persons
<i>a</i>	<i>a</i>	<i>a</i>	552
Yes	<i>a</i>	<i>a</i>	1,103
Yes	Yes	<i>a</i>	2,011
Yes	Yes	Yes	7,200
Yes	Yes	No	348
Yes	No	No	600
No	No	No	658
Grand Total			12,472

^aMissing.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2.

Cramer's V , and p Value Statistics for HIV Testing Variables Among Newly Diagnosed HIV-Positive Persons in the United States, 2014.

Variable	Correlation of Variables Associated With a Complete Referral to HIV Medical Care ($n = 11,920$)		Correlation of Variables Associated With a Complete Clients Who Attended First Medical Appointment ($n = 10,159$)		Correlation of Variables Associated With a Complete Clients Who Attended First Medical Appointment Within 90 days ($n = 7,548$)	
	Cramer's V Statistic	p Value	Cramer's V Statistic	p Value	Cramer's V Statistic	p Value
Region	.1506	<.0001	.1397	<.0001	.2845	<.0001
Testing site type	.1536	<.0001	.0370	<.0311	.0434	<.0275
Number of HIV tests	.1231	<.0001	.1082	<.0001	.0337	<.0137
Age group at test year	.0420	<.0008	.0211	<.4798	.0439	<.0124
Race/ethnicity	.0264	<.0404	.0768	<.0001	.0563	<.0001
Gender group	.0165	<.1984	.0097	<.6225	.0076	<.8041

Table 3.

Comparison Among Subgroup Between Complete Case Analysis Result and Imputation Method Result for Referral to HIV Medical Care, Linkage to HIV Medical Care Within Any Time frame and Linkage Within 90 Days in 2014.

Characteristics	Referral to HIV Medical Care		Linkage to HIV Medical Care Within any Time frame		Linkage to HIV Medical Within 90 days	
	CCA	MI	CCA	MI	CCA	MI
Age at test (years)						
13-19	96.5%	95.8% [93.7%, 97.9%]	89.3%	89.1% [85.9%, 92.3%]	80.4%	83.4% [79.1%, 87.8%]
20-29	94.9%	94.5% [93.0%, 95.2%]	88.5%	88.6% [87.7%, 89.4%]	81.3%*	83.0% ^a [81.9%, 84.1%]
30-39	94.6%	94.4% [93.6%, 95.3%]	88.8%	89.2% [88.1%, 90.3%]	82.7%*	84.5% ^a [83.2%, 85.9%]
40-49	94.4%	94.0% [92.9%, 95.1%]	88.6%	88.5% [87.0%, 90.0%]	83.1%	84.3% [82.5%, 86.1%]
50+	92.5%	92.2% [90.8%, 93.6%]	86.8%	87.2% [85.5%, 88.9%]	80.7%*	83.5% ^a [81.6%, 85.5%]
Other	85.2%	86.3% [73.6%, 99.0%]	82.6%	86.3% [73.6%, 99.0%]	60.0%	74.3% [57.5%, 91.2%]
Race/ethnicity						
White	95.5%	95.3% [94.5%, 96.1%]	90.5%	90.7% [89.6%, 91.9%]	84.5%	85.6% [84.1%, 87.15]
Black	94.0%	93.5% [92.9%, 94.2%]	86.2%	86.3% [85.4%, 87.3%]	78.4%*	80.6% ^a [79.5%, 81.7%]
Hispanic or Latino	94.3%	94.2% [93.3%, 95.1%]	90.6%	91.0% [89.9%, 92.1%]	85.6%*	87.5% ^a [86.1%, 88.9%]
Other	95.3%	95.2% [93.5%, 96.9%]	89.7%	90.1% [87.7%, 92.5%]	85.3%	87.2% [84.4%, 89.9%]
Setting						
Inpatient facilities	98.7%	98.7% [98.7%, 98.7%]	97.3%	96.6% [92.1%, 100.0%]	93.1%	91.4% [84.9%, 98.0%]
Outpatient facilities	96.2%	95.9% [95.3%, 96.5%]	89.4%	89.4% [88.6%, 90.3%]	83.5%	83.9% [82.9%, 84.9%]
Emergency room	84.1%	83.7% [81.4%, 86.1%]	76.8%*	79.6%* [77.1%, 82.2%]	59.9%*	76.2% ^a [73.4%, 79.0%]
HIV counseling and testing site	96.0%	95.8% [95.0%, 96.6%]	90.9%	90.9% [89.7%, 92.0%]	85.2%	86.4% [85.1%, 87.7%]
Community setting	91.5%	90.6% [88.9%, 92.4%]	84.3%	85.6% [83.5%, 87.7%]	76.0%*	79.9% ^a [77.5%, 82.3%]
Correctional facilities	97.0%	96.8% [95.2%, 98.4%]	90.1%	90.1% [87.4%, 92.8%]	83.1%	84.5% [81.2%, 87.8%]
Other	92.8%	92.8% [91.0%, 94.5%]	87.8%	88.0% [85.9%, 90.2%]	79.6%*	85.1% ^a [82.5%, 87.6%]
Gender						
Male	94.7%	94.3% [93.9%, 94.8%]	88.5%	88.7% [88.0%, 89.3%]	81.9%*	83.6% ^a [82.9%, 84.4%]
Female	93.8%	93.4% [92.3%, 94.6%]	88.0%	88.2% [86.7%, 89.7%]	81.5%*	83.6% ^a [81.9%, 85.3%]
Other	93.2%	93.0% [89.5%, 96.4%]	85.9%	86.5% [81.7%, 91.3%]	80.1%	82.5% [77.2%, 87.8%]

Characteristics	Referral to HIV Medical Care		Linkage to HIV Medical Care Within any Time frame		Linkage to HIV Medical Within 90 days	
	CCA	MI	CCA	MI	CCA	MI
Region ^b						
Region 1	99.1%	98.9% [97.3%, 100%]	96.9%	96.4% [93.8%, 99.0%]	96.0%	96.4% [93.8%, 98.9%]
Region 2	96.3%	96.2% [95.3%, 97.2%]	90.5%	90.1% [89.4%, 92.1%]	92.2%*	87.5% ^a [86.0%, 89.1%]
Region 3	88.0%	86.0% [83.3%, 88.7%]	93.8%*	82.4%* [79.5%, 85.3%]	71.1%	71.8% [68.7%, 74.9%]
Region 4	94.8%	94.9% [92.8%, 96.9%]	83.1%	83.3% [79.7%, 86.9%]	81.6%	81.8% [78.1%, 85.5%]
Region 5	97.1%	97.0% [96.5%, 97.5%]	93.2%	93.3% [92.6%, 94.1%]	89.8%*	92.1% ^a [91.3%, 92.8%]
Region 6	95.3%	95.2% [93.2%, 97.1%]	79.7%	80.0% [76.2%, 83.7%]	58.3%*	48.9% ^a [44.3%, 53.4%]
Region 7	88.5%	88.5% [87.0%, 89.9%]	81.6%	81.9% [80.1%, 83.7%]	54.5%*	71.2% ^a [69.1%, 73.4%]
Region 8	93.8%	93.4% [91.4%, 95.4%]	81.1%	80.4% [77.3%, 83.6%]	77.2%	79.4% [76.2%, 82.6%]
Region 9	94.2%	94.2% [92.7%, 95.7%]	86.0%*	89.2%* [87.3%, 91.2%]	81.8%*	87.2% ^a [85.1%, 89.3%]
Region 10	96.0%	95.9% [93.6%, 98.3%]	95.1%	95.3% [92.7%, 97.9%]	95.1%	95.3% [92.7%, 97.9%]
Overall	94.5%	94.1% [93.7%, 94.6%]	88.4%	88.6% [88.0%, 89.1%]	81.8%*	83.6% ^a [83.0%, 84.3%]

^aSignificantly different between CCA result and MI result. CCA: complete case analysis; MI: multiple imputation;

^b**Region 1:** Connecticut, Massachusetts, Maine, New Hampshire, Rhode Island, Vermont; **Region 2:** New Jersey, New York State, New York City, Pennsylvania, Philadelphia; **Region 3:** Illinois, Chicago, Indiana, Michigan, Ohio, Wisconsin; **Region 4:** Iowa, Kansas, Minnesota, Missouri, North Dakota, Nebraska, South Dakota; **Region 5:** District of Columbia, Delaware, Florida, Georgia, Atlanta, Maryland, Baltimore, North Carolina, South Carolina, Virginia, West Virginia; **Region 6:** Alabama, Kentucky, Tennessee; **Region 7:** Arkansas, Houston, Louisiana, Oklahoma, Texas; **Region 8:** Arizona, Colorado, Idaho, Montana, New Mexico, Nevada, Utah, Wyoming; **Region 9:** Alaska, California, Los Angeles, San Francisco, Hawaii, Oregon, Washington; **Region 10:** Puerto Rico, US Virgin Islands.