Variability of the Date of HIV Diagnosis: A Comparison of Self-Report, Medical Record, and HIV/AIDS Surveillance Data

Sandra I. McCoy, PhD[1], **Bill Jones, MPH**[2], **Peter A. Leone, MD**[2,3,4], **Sonia Napravnik, PhD**[3,4], **E. Byrd Quinlivan, MD**[4], **Joseph J. Eron, MD**[3,4], and **William C. Miller, MD, PhD**[3,4]

[1] Institute of Business and Economic Research, University of California, Berkeley

[2] North Carolina Department of Health and Human Services

[3] Department of Epidemiology, University of North Carolina at Chapel Hill

[4] Department of Medicine, University of North Carolina at Chapel Hill

Abstract

PURPOSE—Describe and quantify differences among the year of first positive HIV test from patient report, the medical record, and HIV/AIDS surveillance data.

METHODS—We merged two clinic-based studies with overlapping HIV-infected participant populations in North Carolina with the HIV/AIDS Reporting System (HARS) and examined the first positive HIV test year from patient report, the medical record, and HARS. Matches were considered the same year of diagnosis.

RESULTS—The self-reported year of diagnosis had high agreement with the medical record (67% matched exactly and 19% differed by one year, weighted kappa=0.85), although there were wide 95% limits of agreement (−4.0 earlier to 3.9 years later). On average, the dates of diagnosis from patient report and the medical record were earlier than HARS with wide 95% limits of agreement (7.5 years earlier to 6.0 years later for patient report vs. HARS, 7.7 years earlier to 6.0 years later for medical record vs. HARS).

CONCLUSIONS—These measures could not reliably be used interchangeably as there was wide variability in both directions. Although collection of data from patient report or existing sources is convenient, cost-effective, and efficient, there is significant variability between sources.

Keywords

HIV infections; reproducibility of results; HIV serodiagnosis; surveillance; comparative study

CORRESPONDING AUTHOR: Sandra I. McCoy, MPH, PhD, University of California, Berkeley, Institute of Business and Economic Research, F502 Haas School of Business, Ph: (510) 332-6612, smccoy@berkeley.edu.
REQUESTS FOR REPRINTS: Sandra I. McCoy, MPH, PhD, University of California, Berkeley, Institute of Business and Economic Research, F502 Haas School of Business, Ph: (510) 332-6612, smccoy@berkeley.edu

INTRODUCTION

In the absence of prospectively collected clinical data, epidemiological studies of HIV infection often rely on a variety of secondary sources for patient information, including patient report, medical records, and surveillance data. Few studies have assessed the comparability of these sources. For example, comparisons of the medical record and HIV/AIDS surveillance data found good agreement for age, race, and gender, but poorer agreement for mode of transmission, CD4 count, and the more complex category of AIDS case definition.[1, 2] Comparisons of patient report and medical records have found good agreement in reporting of CD4 lymphocyte cell counts, but not viral load.[3] Understanding the magnitude and direction of these inconsistencies is vital to correctly interpret research findings.

In this study, we focused on the date of first positive HIV test, a critical measurement for epidemiologic studies of clinical outcomes and health care access and for HIV/AIDS surveillance. The source of this information can vary, as can the various biases that threaten its validity. In interview studies, an individual is often asked to report the date they first learned their HIV-positive serostatus, subjecting the resulting information to recall bias. Thus, self-report of HIV test results may be inconsistent with serological screening.[4–6] The date of diagnosis in the medical record, a source frequently considered the gold standard for clinical information, can also be inaccurate or incomplete as the record may be based on patient report or not include testing performed at different facilities. A Massachusetts study found that 24% of cases reported to the state had different years of AIDS diagnosis than that recorded in the medical record.[1] HIV/AIDS surveillance data is another common source, given recommendations for confidential name-based reporting in all 50 states.[7] However, the completeness and accuracy of surveillance data can vary over time as surveillance practices evolve. A national study found that although 56% of self-reported years of first positive HIV test agreed with that reported to the U.S. HIV/AIDS Reporting System, 30% of participants reported an earlier diagnosis year.[8]

While the aforementioned studies present useful pairwise comparisons, no study has examined all three data sources in the same patient population. Our goal was therefore to describe and quantify the extent of differences of the date of the first positive HIV test from patient report, the medical record, and HIV/AIDS surveillance data.

METHODS

The source population for the study was the University of North Carolina at Chapel Hill Infectious Disease (UNC-ID) Clinic, a large, university-based medical center providing comprehensive HIV primary care services. For these analyses the study population included all HIV-infected patients participating in the ongoing Center for AIDS Research (CFAR) Clinical Cohort.[9] In a subset of the Clinical Cohort, we conducted an in-person interview, the UNC Clinical and Socio-Demographic Survey (CSDS). For this study, we matched participants in the Clinical Cohort to the North Carolina (NC) HIV/AIDS Reporting System (HARS) to compare the dates of the first positive HIV test. To be eligible for the study, all

patients must have provided written informed consent to participate in the CFAR Clinical Cohort and the CSDS (if applicable).

Patient Reported Date of First Positive HIV Test

The self-reported date of HIV diagnosis was collected in the CSDS survey, an in-person interview conducted with 336 HIV-infected patients to collect information on social and behavioral characteristics. Patients receiving care from July 2000 to June 2006 who were 18 years of age, English speaking, and able to provide written informed consent were eligible for the study. The self-reported date of diagnosis was the answer to the question "*When were you first told that you were HIV-positive?*" In the first version of the CSDS (2000–2002), patients were asked to identify the year of diagnosis. In subsequent versions of the questionnaire (2002–2006), patients were asked to identify the full date (month, day, year) of diagnosis. Due to this inconsistency, we present comparisons of year only.

HIV Diagnosis Date in the UNC Medical Record

The date of HIV diagnosis in the medical record is collected as part of the UNC CFAR HIV Clinical Cohort study. The Clinical Cohort includes over 2,400 HIV-infected individuals receiving primary HIV care and contains information on clinical care, medications and illnesses. The study population was comprised of HIV positive patients attending the clinic from January 2000 through October 2008 who provided written informed consent to participate, which includes authorization to acquire medical information from the UNC medical record as well as other providers and facilities, including state health department records. The date of diagnosis was abstracted from the medical record with standardized record reviews by trained personnel and was defined as the earliest date of positive laboratory tests (e.g., enzyme immuno-assay or Western Blot) or a clinic note if testing was done at another facility. In some cases, the earliest date of diagnosis recorded in the medical record was based on self-report of the patient, although the sources for this information could not be distinguished.

First Positive HIV Test Date in the NC HIV/AIDS Reporting System

Name-based reporting of HIV to the NC Department of Health and Human Services (NC DHHS) has been mandated in NC since 1990 and AIDS reporting has been required since 1984. Anonymous HIV testing at publicly-funded counseling and testing sites was discontinued in NC in 1997. Laboratory reporting of HIV testing information has been actively integrated into NC HIV surveillance operations since 2002. Case reports are entered into HARS, patient identifiers removed, and cases are reported to the Centers for Disease Control and Prevention. Out-of-state diagnoses are verified with the corresponding surveillance office in the diagnosing state, and potential duplicate reports of the same person by different states are investigated and resolved. Case reports typically include the first documented HIV-positive test for the patient.

Clinical Cohort participants were matched electronically to HARS using a 4-step algorithm. First, patients were matched deterministically using the first four letters of the last name, first 3 letters of the first name, month and year of birth, and sex. In the second stage, single matches were removed and the remaining non-matched subjects were matched

deterministically by social security number, if available. The remaining non-matched subjects were then manually matched by record lookup using an inexact matching algorithm and rotating the first name, last name, date of birth, and sex though the lookup system to identify changed names and/or gender errors. Finally, multiple matches for a single patient were investigated and resolved and all matches were manually reviewed for errors. We considered the earliest date of HIV or AIDS diagnosis to represent the date of diagnosis recorded in HARS.

Statistical Analysis

We merged the Clinical Cohort and the CSDS by a unique patient code to create the analysis dataset with three years of diagnosis from the patient, medical record, and HARS. Only years of 1985 or later, when commercial testing became available, were considered valid. [10] We did not have an external gold standard and thus were not able to assess which date was the true year of diagnosis.

We first describe demographics of the patients from the three data sources to assess comparability. We then conduct three separate pairwise comparisons of the self-reported, medical record, and HARS year of diagnosis. For each comparison, we present the percent agreement, the mean, median and range of discordant results, and the weighted kappa statistic (κ) with Cicchetti-Allison weights to measure the agreement of two measures while taking into account the degree of disagreement.[11, 12] We also present Bland-Altman plots and the 95% limits of agreement, calculated as the mean difference plus or minus 1.96 standard deviations (SD).[13, 14] Ninety-five percent of differences between measurements would be expected to lie between these limits. This method is an alternative to correlation coefficients, which measure the relationship between two measurements, not agreement, and depend on the range, distribution, and scale of data and ignore any systematic bias.[13, 14] Bland-Altman methods allow for examination of agreement and the determination if two measurement methods can be used interchangeably.

To determine factors associated with the number of years discrepant between two data sources, we considered regression models using the Poisson, zero-inflated Poisson, negative binomial, and zero-inflated negative binomial distributions. Negative binomial regression provided the best fit to model the absolute difference between two years as the outcome based on likelihood ratio tests of the dispersion parameters and the Vuong test to distinguish between the zero-inflated negative binomial and the negative binomial model. Using this method, exponentiated parameter estimates represent the ratio of years different for two covariate levels. For example, an exponentiated parameter estimate of 2.0 comparing women to men means that women, on average, have twice as many years discrepant as men in the sample.

We constructed unadjusted bivariable models to assess the ratio of years discordant by selected covariates. We then constructed a multivariable model for each of the three comparisons containing all of the covariates. We did not evaluate further for confounding, as the main effects of all covariates were retained. Since people who were diagnosed longer ago had the opportunity for greater discrepancies between years, the year of diagnosis was expected be a strong predictor of the absolute difference between two years. We therefore

present the multivariable analyses only. As these models were hypothesis-generating and did not include a main effect of interest, we did not examine effect measure modification. Data management and analysis was conducted with SAS software (version 9.2, Cary, NC).

Human Subjects Protection

All participants provided written informed consent to participate in the Clinical Cohort and the CSDS (if applicable) and HIPAA authorization to access medical information. This study was approved by the UNC Institutional Review Board and the NC DHHS.

RESULTS

Two thousand forty-seven participants were included from the CFAR Clinical Cohort and 322 from the CSDS. Of the 2,047 included in the Clinical Cohort, 1,652 (81%) matched to the NC HARS system. The number of participants with data available for each comparison was 299 for patient report vs. medical record, 252 for patient report vs. HARS, and 1,624 for medical record vs. HARS. Patient demographics from each subset were similar, although the subset of Clinical Cohort participants who completed the CSDS were more likely to be Black (non-Hispanic), report a history of injection drug use (IDU), and be enrolled in public health insurance (Table 1). The subset of participants in the Clinical Cohort who matched to HARS was slightly less likely to have been diagnosed early in the epidemic.

Comparison of Patient Report and the Medical Record

Of the participants with valid self-reported diagnosis years and medical record values, 199 of 299 (67%) had the same year of diagnosis, 19% differed by one year and 6% differed by two. (Table 2, Figure 1A). On average, the self-reported year of diagnosis was earlier than that in the medical record by 0.9 months (−0.07 years). Although the weighted kappa statistic was 0.85, indicating good agreement, the 95% limits of agreement were −4.0 to 3.9 years, suggesting that for 95% of people, the self-reported year of diagnosis may be as much as 4 years before or after the medical record date (Figure 2A). However, for several patients, the differences were well outside the 95% limits of agreement, represented by values above or below the horizontal lines. In the multivariable model, which represents the ratio of years different for two covariate levels, no factors were significantly associated with differences between the two dates (Table 3). Being uninsured or having public or other forms of insurance was moderately associated with larger differences between the two sources compared to having private insurance (p=0.15). Seventy-two percent of participants with private insurance matched exactly, compared to 67% of public insurance recipients, 62% of uninsured participants, and 64% of those with other types of insurance (p=0.64).

Comparison of Patient Report and HARS

Overall, 128 (51%) of 252 participants reported the same year of diagnosis as was reported to HARS and 54 (21%) were different by one year (Figure 1B). Fifty-seven (23%) were discordant by three or more years, with the patient report on average being 9.0 months earlier than the year reported to HARS (κ=0.64). 95% of differences were between −7.5 and 6.0 years (Figure 2B). No factors were significantly associated with differences between the two dates. Only health insurance status was moderately associated with differences between

the patient reported diagnosis year and that in HARS; being uninsured or having public or other types of insurance were associated with fewer discrepancies than private insurance.

Comparison of Medical Record and HARS

Of the 1,624 participants with medical record and HARS dates, 1,132 (70%) matched exactly (Table 2, Figure 1C). Although agreement was high (κ=0.76), 10% of medical record values were five or more years earlier than the corresponding HARS year, and 3% of HARS values were five or more years earlier than the medical record. On average, the medical record was 9.1 months earlier than HARS. Ninety-five percent of the differences were contained within the bounds −7.5 to 6.0 years (Figure 2C). In the subset of subjects that had data from all three sources (n=250) the comparison between the medical record and HARS was similar. Only gender and sexual orientation were significantly associated with discrepancies between years in the multivariable model. Seventy-four percent of women had the same year of diagnosis (ratio=0.66, 95% CI: 0.49, 0.90), as did 71% of men who have sex with men (MSM) (ratio=0.78, 95% CI: 0.57, 1.07), versus to 64% of heterosexual men.

DISCUSSION

Many epidemiologic studies use the date of the first positive HIV test as the starting point for evaluation of HIV-related clinical and psychological outcomes.[15–17] This study demonstrates that among patients attending a busy HIV outpatient clinic, the dates of diagnosis reported by the patient, recorded in the medical record, and reported to the state are highly variable. While we did not know the true diagnosis date and were unable to compare more specific date elements such and month and day, our study quantifies the direction and magnitude of these differences and highlights the importance of understanding the biases inherent in HIV testing histories.

We found that the year of first positive HIV test reported by the patient was either the same as in the medical record (67%) or within 2 years of the medical record in 92% of individuals. However, the limits of agreement suggest that these two sources could not be reliably used interchangeably, especially considering that in a few cases the year of first positive HIV test differed by 10 years or more. Differences between these two measures may be due to inaccurate or inadequate medical records from previous providers or poor patient recall. For example, patient report of "episodic" medical information (e.g., specific events or facts) has been found to be poor, especially when the diagnosis of a life threatening condition incites stress and anxiety.[18, 19] A limitation of this comparison is that the CSDS was a small survey of under 300 participants, and differences in key demographic and risk factors such as insurance status and injection drug use may limit comparability to the larger Clinical Cohort. However, participants in both datasets were selected from the same underlying source population.

In some cases, the first positive HIV test date in the medical record may be based on patient report, if no other sources are available (such as a laboratory test documenting HIV infection). Patient report is frequently used in the medical record when patients move from out of state or were diagnosed years prior to receiving care. Although we could not discern the source of the date in the medical record, a useful future project could identify exactly

what the diagnosis date in the medical record represents, especially in the common situation where the HIV diagnosis was made in a different venue than where the patient currently receives care or where no previous negative test is documented. Understanding under what circumstances the medical record is reliable and accurate would benefit researchers who regularly use these data, considering that the medical record is often considered the gold standard.

Both patient report of HIV diagnosis year and the year recorded in the medical record were approximately 9 months earlier than the year reported to HARS. Although this finding is consistent with national reports and with our expectations given the challenges of documenting HIV test results, it highlights an important bias of HIV surveillance data.[8] Most of the differences fell in a wide 13.5 year window and some were more than 15 years different. Reporting may be completed years after the initial diagnosis by a current physician with inaccurate or incomplete diagnosis information either from patient self-report or other providers. Further, although the NC DHHS works with other states to identify people diagnosed out of state (and determine residency), complete diagnostic information may not have been recorded in NC HARS. Timelier surveillance reports (e.g., those made closer to diagnosis) may be more accurate. Women tended to have fewer discrepancies between the medical record and HARS, possibly because routine tests during pregnancy are reported rapidly to the state or because increased screenings provide previous negative test results, which can lead to a more accurate diagnosis date. This preliminary hypothesis warrants consideration in future research.

Importantly, surveillance completeness must be differentiated from accuracy. Reporting completeness can be defined as the proportion of HIV positive persons who had been reported to the HARS system. While this study was not designed to determine HIV reporting completeness, only 81% of cases in this study had been reported to the state or could be adequately matched despite the mandate for name-based HIV reporting since 1990 and AIDS reporting since 1984.[20] All patients not found in surveillance records were reported to the state at the end of the study. Half of these patients were diagnosed before 1995 (data not shown). As the source for the most comprehensive data on patterns of new diagnoses, morbidity reporting is essential for effective targeting of health resources. In addition to epidemiological monitoring of the epidemic, HIV/AIDS reporting is used for Ryan White funding distribution to pay for medical care and treatment of un- and underinsured people living with HIV/AIDS. Nationally, completeness of HIV/AIDS reporting is high.[2, 21] It is likely that many patients not identified in HARS were diagnosed elsewhere and providers may have mistakenly believed that reporting was already complete when the patient entered care.

Although we have focused on HIV diagnosis, our findings can be viewed through a larger lens of measurement. Collection of clinical data from patient report or existing sources is convenient, cost-effective, efficient, and often the only feasible way to obtain necessary data. However, each method has inherent limitations. These results should encourage other investigators to identify and describe the sources of bias that influence the data used in their research endeavors.

Acknowledgments

LIST OF ABBREVIATIONS

|	
AIDS	Acquired Immune Deficiency Syndrome
CFAR	Center for AIDS Research
CSDS	Clinical and Socio-Demographic Survey
HARS	HIV/AIDS Reporting System
HIPAA	Health Insurance Portability and Accountability Act of 1996
HIV	Human immunodeficiency virus
IDU	Injection drug use
MSM	Men who have sex with men
NC	North Carolina
NC DHHS	North Carolina Department of Health and Human Services
SD	Standard deviation
UNC	University of North Carolina
UNC-ID	University of North Carolina at Chapel Hill Infectious Disease Clinic

References

1. Gallagher KM, Jara M, Demaria A Jr, Seage GR, Heeren T. The reliability of passively collected AIDS surveillance data in Massachusetts. Ann Epidemiol. 2003; 13:100–104. [PubMed: 12559668]

2. Klevens RM, Fleming PL, Li J, Gaines CG, Gallagher K, Schwarcz S, et al. The completeness, validity, and timeliness of AIDS surveillance data. Ann Epidemiol. 2001; 11:443–449. [PubMed: 11557175]

3. Kalichman SC, Rompa D, Cage M. Reliability and validity of self-reported CD4 lymphocyte count and viral load test results in people living with HIV/AIDS. Int J STD AIDS. 2000; 11:579–585. [PubMed: 10997499]

4. Phillips KA, Catania JA. Consistency in self-reports of HIV testing: longitudinal findings from the National AIDS Behavioral Surveys. Public Health Rep. 1995; 110:749–753. [PubMed: 8570830]

5. McCusker J, Stoddard AM, McCarthy E. The validity of self-reported HIV antibody test results. Am J Public Health. 1992; 82:567–569. [PubMed: 1546774]

6. Fisher DG, Reynolds GL, Jaffe A, Johnson ME. Reliability, sensitivity and specificity of self-report of HIV test results. AIDS Care. 2007; 19:692–696. [PubMed: 17505932]

7. Centers for Disease Control and Prevention. Department of Health and Human Services. Dear Colleague letter re name-based HIV reporting, from Julie Gerberding. Atlanta, Georgia: 2005.

8. Hall HI, Li J, Campsmith M, Sweeny P, Lee LM. Date of first positive HIV test: reliability of information collected for HIV/AIDS surveillance in the United States. Public Health Rep. 2005; 120:89–95. [PubMed: 15736337]

9. Napravnik S, Eron JJ Jr, McKaig RG, Heine AD, Menezes P, Quinlivan E. Factors associated with fewer visits for HIV primary care at a tertiary care center in the Southeastern U. S. AIDS Care. 2006; 18 (Suppl 1):S45–50. [PubMed: 16938674]

10. Henry, J. Kaiser Family Foundation. HIV/AIDS Policy Fact Sheet. Washington, D.C: 2006. HIV Testing in the United States.

11. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull. 1968; 70:213–220. [PubMed: 19673146]

12. Cicchetti DV, Allison T. A New Procedure for Assessing Reliability of Scoring EEG Sleep Recordings. American Journal of EEG Technology. 1971; 11:101–109.

13. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986; 1:307–310. [PubMed: 2868172]

14. Bland JM, Altman DG. Applying the right statistics: analyses of measurement studies. Ultrasound Obstet Gynecol. 2003; 22:85–93. [PubMed: 12858311]

15. Hall HI, McDavid K, Ling Q, Sloggett A. Determinants of progression to AIDS or death after HIV diagnosis, United States, 1996 to 2001. Ann Epidemiol. 2006; 16:824–833. [PubMed: 17067817]

16. Samet JH, Freedberg KA, Stein MD, Lewis R, Savetsky J, Sullivan L, et al. Trillion virion delay: time from testing positive for HIV to presentation for primary care. Arch Intern Med. 1998; 158:734–740. [PubMed: 9554679]

17. Cooperman NA, Simoni JM. Suicidal ideation and attempted suicide among women living with HIV/AIDS. J Behav Med. 2005; 28:149–156. [PubMed: 15957570]

18. Kessels RP. Patients' memory for medical information. J R Soc Med. 2003; 96:219–222. [PubMed: 12724430]

19. Jansen J, Butow PN, van Weert JC, van Dulmen S, Devine RJ, Heeren TJ, et al. Does age really matter? Recall of information presented to newly referred patients with cancer. J Clin Oncol. 2008; 26:5450–5457. [PubMed: 18936478]

20. N.C. Department of Health and Human Services. Scope of the HIV/AIDS Epidemic in North Carolina. North Carolina Epidemiologic Profile for HIV/STD Prevention & Care Planning. Jul. 2007

21. Hall HI, Song R, Gerstle JE 3rd, Lee LM. Assessing the completeness of reporting of human immunodeficiency virus diagnoses in 2002–2003: capture-recapture methods. Am J Epidemiol. 2006; 164:391–397. [PubMed: 16772373]
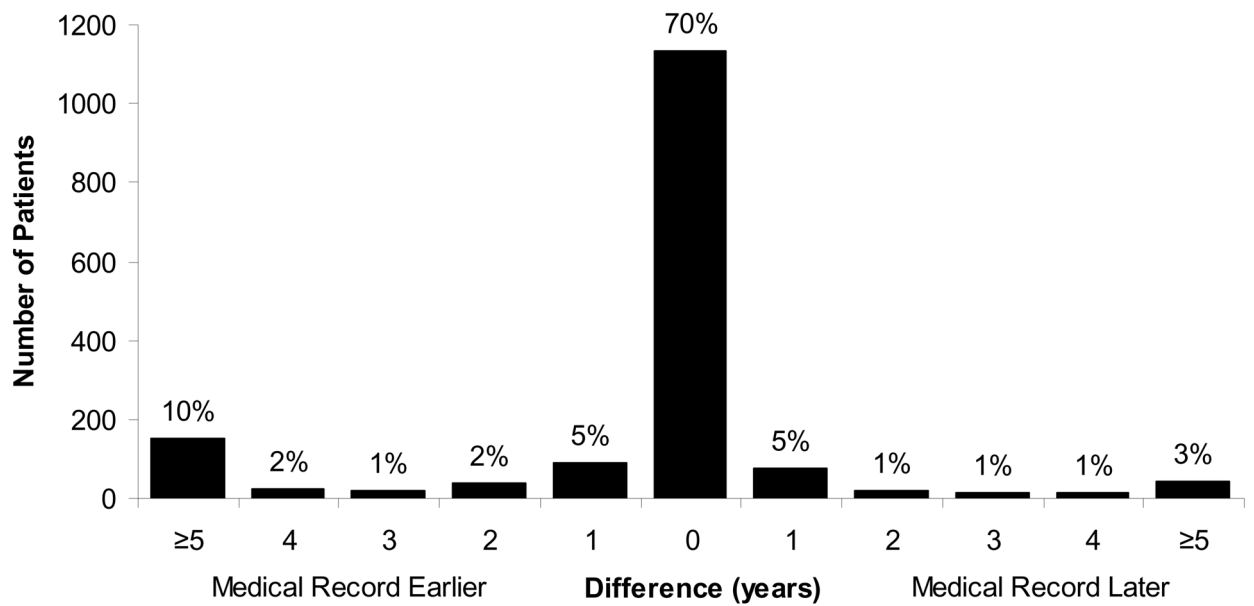
A. Patient Report vs. Medical Record
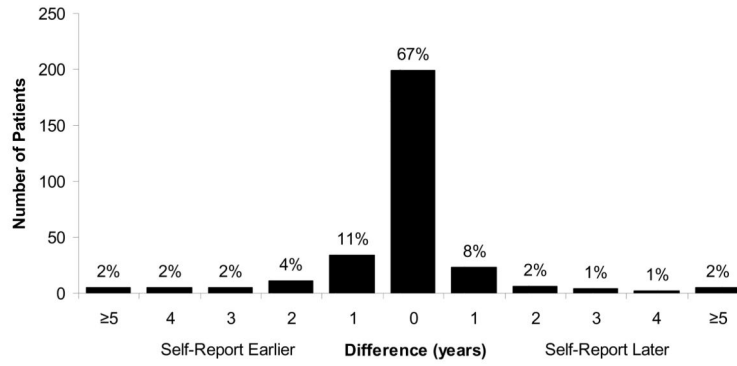
B. Patient Report vs. HIV/AIDS Reporting System
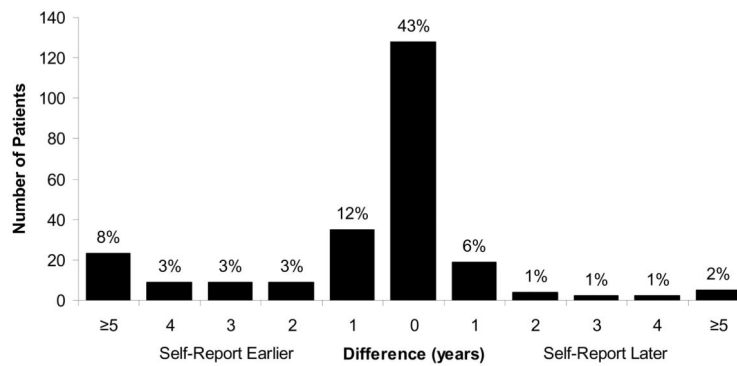
C. Medical Record vs. HIV/AIDS Reporting System

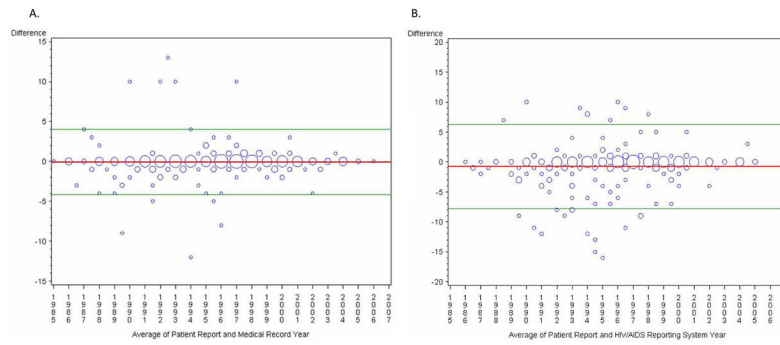**FIGURE 1.**

Comparison of sources of the year of HIV diagnosis among patients attending the University of North Carolina Infectious Disease Clinic, 2000–2008. Panel A: comparison of patient report and the medical record (n=299); Panel B: comparison of patient report and the HIV/AIDS Reporting System (n=252); Panel C: comparison of the medical record and the HIV/AIDS Reporting System (n=1,624).
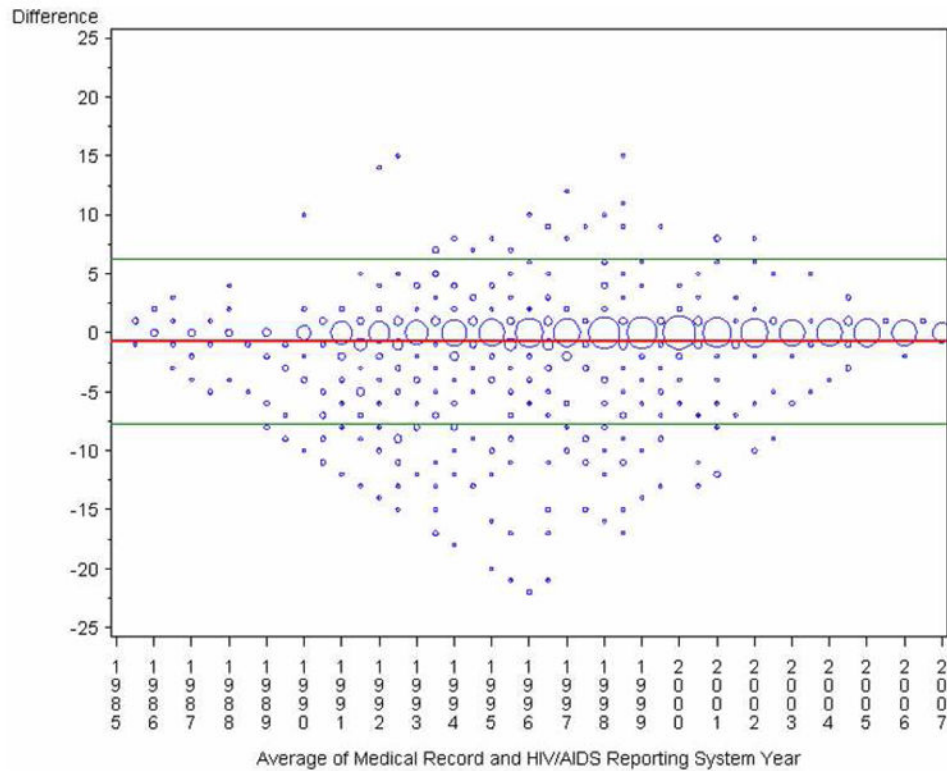
FIGURE 2.

Comparison of sources of the year of HIV diagnosis presented as a plot of the difference in years against the mean with mean difference and 95% limits of agreement indicated.[1] Sizes of the data points are proportional to the number of observations. Panel A: Comparison of patient report and the medical record; Panel B: Patient report and the HIV/AIDS Reporting System; Panel C: Medical Record and the HIV/AIDS Reporting System.

1 Method of Bland & Altman [13, 14]

TABLE 1

Demographics of participants in the University of North Carolina Clinical Cohort Study and the Clinical and Socio-Demographic Survey, North Carolina, 2000–2008. Values are based on the medical record unless otherwise indicated.[1]

Characteristic	Clinical Research Database N=2,047	NC HIV/AIDS Reporting System N=1,652	CSDS N=322
	N %	N %	N %
Gender and sexual preference			
Female	648 (31.7)	556 (33.7)	109 (36.2)
Heterosexual male	632 (30.9)	525 (31.8)	86 (28.6)
MSM	767 (37.5)	571 (34.6)	106 (35.2)
Race			
Black, non-Hispanic	1206 (59.0)	1038 (62.9)	217 (72.1)
White, non-Hispanic	663 (32.4)	469 (28.4)	73 (24.3)
Hispanic	98 (4.8)	83 (5.0)	4 (1.3)
Native American	37 (1.8)	33 (2.0)	6 (2.0)
Asian	2 (0.1)	0 (0.0)	0 (0.0)
Other	38 (1.9)	28 (1.7)	1 (0.3)
Year of diagnosis[2]			
1985–1989	194 (9.7)	53 (3.2)	40 (12.4)
1990–1994	477 (23.8)	358 (21.7)	98 (30.4)
1995–1999	632 (31.5)	556 (33.7)	125 (38.8)
2000–2004	513 (25.6)	487 (29.5)	56 (17.4)
2005–2007	188 (9.4)	198 (12.0)	3 (0.9)
Age at diagnosis[2]			
<18	24 (1.2)	13 (0.8)	3 (0.9)
18–24	366 (18.3)	272 (16.5)	35 (10.9)
25–34	669 (33.4)	515 (31.2)	130 (40.4)
35–44	568 (28.3)	500 (30.3)	109 (33.9)
45–54	317 (15.8)	295 (17.9)	35 (10.9)
55	60 (3.0)	57 (3.5)	10 (3.1)
Age at study consent[2]			
18–24	143 (7.0)		7 (2.2)
25–34	397 (19.4)	N/A	53 (16.5)
35–44	753 (36.8)		133 (41.3)
45–54	626 (30.6)		109 (33.9)
55	128 (6.3)		20 (6.2)
History of IDU			
Yes	298 (14.6)	253 (15.3)	73 (24.3)
No/Unknown	1,749 (85.4)	1399 (84.7)	228 (75.7)
Health Insurance			
None	663 (33.0)	564 (34.8)	81 (27.0)

Characteristic	Clinical Research Database N=2,047	NC HIV/AIDS Reporting System N=1,652	CSDS N=322
	N %	N %	N %
Public	664 (33.1)	539 (33.3)	133 (44.3)
Private	539 (26.8)	405 (25.0)	58 (19.3)
Other	142 (7.1)	113 (7.0)	28 (9.3)

[1] Numbers may not sum to column totals due to missing data.

[2] Based on date of diagnosis recorded in each data source.

TABLE 2

Comparison of the first positive HIV test from patient report, the medical record, and the North Carolina HIV/AIDS Reporting System.

Characteristic	Year of Diagnosis Comparison		
	Patient Report vs. Medical Record	**Patient Report vs. HIV/AIDS Reporting System**	**Medical Record vs. HIV/AIDS Reporting System**
Number of date pairs	299	252	1,624
Date pair comparison			
Agree *(same year)*	199 (66.6)	128 (50.8)	1,132 (69.7)
Disagree	100 (33.4)	124 (49.2)	492 (30.3)
Level of Disagreement			
0 (match)	199 (66.6)	128 (50.8)	1,132 (69.7)
1 year	57 (19.1)	54 (21.4)	164 (10.1)
2 years	17 (5.7)	13 (5.2)	60 (3.7)
3 years	9 (3.0)	11 (4.4)	33 (2.0)
4 years	7 (2.3)	11 (4.4)	38 (2.3)
5 or more years	10 (3.3)	35 (13.9)	197 (12.1)
Mean difference (SD)	−0.07 (SD 2.03)	−0.75 (SD 3.47)	−0.76 (SD 3.46)
Median difference (range)	0 (−12 – 13)	0 (−16 – 10)	0 (−22 – 15)
Kappa statistic[1]	0.85	0.64	0.76

[1] Weighted kappa statistic

TABLE 3

Multivariable negative binomial models assessing factors associated with the number of years different between the HIV diagnosis year collected by patient report, recorded in the medical record and the North Carolina HIV/AIDS Reporting System. Exponentiated parameter estimates represent the ratio of years different for two covariate levels.

Characteristic	Year of Diagnosis Comparison, Ratio (95% CI)		
	Patient Report vs. Medical Record	**Patient Report vs. HIV/AIDS Reporting System**	**Medical Record vs. HIV/AIDS Reporting System**
Gender and sexual preference			
Female	0.82 (0.43, 1.56)	0.76 (0.42, 1.42)	0.66 (0.49, 0.90)[*]
Heterosexual male	Referent	Referent	Referent
MSM	1.13 (0.61, 2.10)	0.96 (0.53, 1.72)	0.78 (0.57, 1.07)
Race			
Black, non-Hispanic	Referent	Referent	Referent
White, non-Hispanic	0.76 (0.42, 1.38)	1.61 (0.96, 2.71)	1.15 (0.87, 1.51)
Hispanic	--	0.46 (0.07, 3.12)	0.66 (0.36, 1.20)
Other race	0.77 (0.15, 3.92)	0.73 (0.15, 3.64)	0.77 (0.38, 1.54)
Year of Diagnosis			
1985–1989	3.31 (1.40, 7.83)	2.84 (1.32, 6.11)	6.91 (4.42, 10.80)[*]
1990–1994	1.56 (0.85, 2.85)	1.36 (0.79, 2.34)	1.89 (1.37, 2.61)
1995–1999	Referent	Referent	Referent
2000–2004	1.27 (0.59, 2.70)	1.09 (0.56, 2.12)	0.57 (0.41, 0.78)
2005–2007	--	0.68 (0.08, 5.73)	0.40 (0.25, 0.65)
Age at diagnosis			
<25	1.04 (0.50, 2.15)	0.86 (0.43, 1.72)	0.92 (0.66, 1.28)
25–34	Referent	Referent	Referent
35–44	0.70 (0.37, 1.31)	1.19 (0.69, 2.06)	1.27 (0.94, 1.72)
45–54	1.31 (0.56, 3.11)	0.86 (0.38, 1.94)	1.10 (0.74, 1.62)
55	0.95 (0.21, 4.36)	1.61 (0.42, 6.16)	0.86 (0.39, 1.88)
IDU			
Yes	1.03 (0.56, 1.91)	1.39 (0.81, 2.39)	1.25 (0.89, 1.75)
No/Unknown	Referent	Referent	Referent
Health Insurance			
None	1.82 (0.84, 3.95)	0.75 (0.39, 1.43)	1.15 (0.84, 1.57)
Public	1.64 (0.76, 3.55)	0.49 (0.25, 0.96)	1.03 (0.75, 1.41)
Private	Referent	Referent	Referent
Other	3.09 (1.16, 8.27)	0.66 (0.26, 1.65)	1.08 (0.65, 1.80)

CI: Confidence Interval

[*] P<0.05 for likelihood ratio test of this variable (all levels)