HIV and cancer registry linkage identifies a substantial burden of cancers in persons with HIV in India

Sheela V. Godbole, MD,a, Karabi Nandy, PhD,b, Mansi Gauniyal, PhD,c, Pallavi Nalawade, MCM, MCA,d, Suvarna Sane, MSC, MPH,e, Shravani Koyande, MCom,f, Joy Toyama, MS,g, Asha Hegde, MBBS, MBA, Dip. in HIV Epidemiology,h Phil Virgo, BS,i Kishor Bhatia, PhD, FRCPATH,j Ramesh S. Paranjape, PhD,k Arun R. Risbud, MD, MPH,l Sam M. Mbulaiteye, MBChB, MPhil, MMEd,m Ronald T. Mitsuyasu, MDn

Abstract
We utilized computerized record-linkage methods to link HIV and cancer databases with limited unique identifiers in Pune, India, to determine feasibility of linkage and obtain preliminary estimates of cancer risk in persons living with HIV (PLHIV) as compared with the general population.

Records of 35,755 PLHIV were linked to 31,754 Pune Cancer Registry records (1996–2008) using a probabilistic-matching algorithm. Cancer risk was estimated by calculating standardized incidence ratios (SIRs) in the early (4–27 months after HIV registration), late (28–60 months), and overall (4–60 months) incidence periods. Cancers diagnosed prior to or within 3 months of HIV registration were considered prevalent.

Of 613 linked cancers to PLHIV, 188 were prevalent, 106 early incident, and 319 late incident. Incident cancers comprised 11.5% AIDS-defining cancers (ADCs), including cervical cancer and non-Hodgkin lymphoma (NHL), but not Kaposi sarcoma (KS), and 88.5% non-AIDS-defining cancers (NADCs). Risk for any incident cancer diagnosis in early, late, and combined periods was significantly elevated among PLHIV (SIRs: 5.6 [95% CI 4.6–6.8], 17.7 [95% CI 15.8–19.8], and 11.5 [95% CI 10–12.6], respectively). Cervical cancer risk was elevated in both incidence periods (SIRs: 9.6 [95% CI 4.8–17.2] and 22.6 [95% CI 14.3–33.9], respectively), while NHL risk was elevated only in the late incidence period (SIR: 18.0 [95% CI 9.8–30.2]). Risks for NADCs were dramatically elevated (SIR > 100) for eye-orbit, substantially (SIR > 20) for all-mouth, esophagus, breast, unspecified leukemia, colon-rectum-anus, and other/unspecified cancers; moderately elevated (SIR > 10) for salivary gland, penis, nasopharynx, and brain-nervous system, and mildly elevated (SIR > 5) for stomach. Risks for 6 NADCs (small intestine, testis, lymphocytic leukemia, prostate, ovary, and melanoma) were not elevated and 5 cancers, including multiple myeloma not seen.

Our study demonstrates the feasibility of using probabilistic record-linkage to study cancer/other comorbidities among PLHIV in India and provides preliminary population-based estimates of cancer risks in PLHIV in India. Our results, suggesting a potentially substantial burden and slightly different spectrum of cancers among PLHIV in India, support efforts to conduct multicenter linkage studies to obtain precise estimates and to monitor cancer risk in PLHIV in India.

Abbreviations: ADC = AIDS-defining cancers, cART = combination antiretroviral therapy, CIP = combined incidence period (4–60 months period after HIV registration), EI = early incident (cancers occurring in the 4–27 months period after HIV registration),

Editor: Akhilanand Chaurasia.

Meetings where this work was presented: 20th Conference on Retroviruses and Opportunistic Infections Seattle USA; 6th Mar 13; Indo-US Workshop on HIV Database and Cancer Registry Match, Pune, India 2013.

Ethics Committee Approvals: Ethics review committees at the National AIDS Research Institute-ICMR (NARI) Pune, the University of California, Los Angeles, and the Office of Human Subject Research at the National Institutes of Health approved the study.

Author’s contributions: SVG, SMM, RM, and KB conceived the idea, designed, and supervised execution of the study and data analysis. RM was the PI of the supporting grant and SVG and RM were the India and US PIs, respectively. SVG, MG, RN, SMM, PV, SK, KN, AR, RP, AH, RM acquired and edited the databases for match implementation. SMM and PV designed the match strategy and implemented the match with the NARI and PCR teams. KN, JT, and SS conducted statistical analyses. SVG and SMM drafted the manuscript. All authors interpreted data, read, and approved the final version of the manuscript.

Funding/support: The study was supported by a seed grant from UCLA CFAR (NIH), as well as intramural funds from ICMR and NCI. The sponsors had no role in the study design, data collection, data analysis, data interpretation, writing of the report, or the decision to submit the paper for publication.

None of the authors has any conflicts of interest to declare for this manuscript.

a National AIDS Research Institute (ICMR), Pune, India. b School of Nursing and Department Of Biostatistics University of California, Los Angeles (UCLA), USA. c Mumbai (Bombay) Cancer Registry, Indian Cancer Society, Mumbai, India. d Department of Biostatistics, School of Public Health, University of California, Los Angeles (UCLA), USA. e Maharashtra State AIDS Control Society, Mumbai, India. f Computer Services Corporation, Rockville, MD. g Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland. h University of California, Centre for AIDS Research and Education, David Geffen School of Medicine, Los Angeles (UCLA), USA.

Correspondence: Sheela V. Godbole, National AIDS Research Institute-ICMR, 73, G Block, MDC Bhosari, Pune 411026, India (e-mail: sgodbole@nariindia.org, sgodbole@icmr.org.in).

Copyright © 2016 the Author(s). Published by Wolters Kluwer Health, Inc. All rights reserved.

This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially.

Medicine (2016) 95:37(e4850)

Received: 17 February 2016 / Received in final form: 4 August 2016 / Accepted: 21 August 2016

https://dx.doi.org/10.1097/MD.0000000000004850
1. Introduction

An estimated 2.61 million persons were living with HIV (PLHIV) in India in 2004 and while this figure was revised down to 2.11 in 2015,[22] it still represents about 6% of the global HIV epidemic. The Government of India has responded to this substantial HIV epidemic by implementing free combination antiretroviral therapy (cART), starting in 2004. Today, over 800,000 PLHIV are currently receiving antiretroviral treatment in India[3] and longevity among PLHIV has increased as HIV-related mortality among cART recipients has dramatically declined.[2] However, the increased longevity of PLHIV has uncovered new threats to life from chronic complications, including cancer.[4–8] The relative contribution of cancer to chronic HIV-related morbidity in India is poorly understood.[9] Studies conducted in the United States, Europe, and Australia reported dramatic increases in the risk for Kaposi sarcoma (KS), aggressive non-Hodgkin lymphoma (NHL) and, to a lesser extent, cervical cancer. These cancers are designated AIDS-defining cancers (ADCs).[5–8] The risks for several other cancers, including lung, anal, and liver cancer, were shown to be consistently, albeit modestly, elevated in PLHIV. Thus, these cancers are designated non-AIDS-defining cancer (NADCs). Not surprisingly, the use of cART in those countries has resulted in dramatic reduction of KS and NHL and concomitant increase in the risk for NADCs in PLHIV living longer.[10] Although ADCs remain important, NADCs are becoming more important causes of morbidity and mortality among PLHIV in those countries.[11]

Whether Indian PLHIV experience similar cancer risks is currently unknown. An Indian tertiary cancer hospital study[12] confirmed that cancer is seen in Indian PLHIV,[13] but suggested that the cancer spectrum in Indian PLHIV may be different from that reported elsewhere.[9] For example, KS, the classical ADC, was not seen in that study nor has it been reported as a common cancer in Indian men who have sex with men (MSM).[14] Primary central nervous system lymphoma, whose risk is elevated several fold in more developed countries, appears not to be common in Indian PLHIV.[9,14,15] The lack of quantitative data about cancer risk in Indian PLHIV hampers efforts to address the cancer burden among PLHIV.[9]

To address this issue, we used computerized probabilistic record-linkage methods to link records from one HIV/AIDS registry to records from the Pune Cancer Registry (PCR) in India. Record-linkage methods are now widely used in developed countries to rapidly and efficiently generate cancer statistics.[5] While the record-linkage studies conducted in the United States rely on the availability of unique personal identifiers (social security numbers) in the health databases, which are not available in India, record-linkages have been successfully conducted without unique personal identifiers in developed countries, including Italy and[16] Australia.[17] Moreover, they have recently been successfully used in Uganda, Nigeria and, most recently, in South Africa.[18–21] The objective of this first computerized record-linkage study of HIV and cancer in India was to determine the feasibility of this approach in India, to obtain preliminary cancer data in PLHIV, and to explore cancer patterns in a representative sample of PLHIV.

2. Methods

Ethics review boards at National AIDS Research Institute (NARI), the University of California, Los Angeles, and the Office of Human Subject Research at the National Institutes of Health approved the study.

2.1. Study population

The study was conducted in 2011 in Pune, a major city in the western Indian state of Maharashtra, with a population of 3.12 million in 2011.[22] Pune city was one of the Indian regions where HIV was recognized early and became a site for many of the early HIV studies. In addition, its population-based cancer registry, established in 1972 as an affiliate of the Mumbai Cancer Registry, is one of the oldest cancer registries in India. In 2008, a study conducted in Pune reported a high HIV prevalence among high risk populations and pregnant women varying from > 5% to 74%.[23] PLHIV in Pune have been estimated as 40,000 in 2000[24] and 27,000 in 2008 using different methods.[23] The National AIDS Research Institute (NARI), which leads national HIV research programs of the Indian Council of Medical Research, has provided free HIV testing services in Pune city since the outset of the epidemic. In addition, HIV Integrated Testing and Counseling Centers (ICTC), established by the Government of India, were rapidly expanded since 2005 to implement free cART in Pune for persons with AIDS/CD4+ cell counts <200/mm³. These changes increased the percentage of PLHIV registered at government ICTCs who were on ART from negligible in 2004 to 37% in 2012.[25] Thus, while this increase in access to cART is significant, the duration of cART use and availability is still relatively short to demonstrate temporal trends due to provision of cART in our linkage study.

The HIV registry of Pune city was constructed using the NARI data from 1991 to 2008 and the ICTC register data until 2008. NARI HIV data, which includes unique personal identification numbers but not the patient names, was computerized to create the “HIV Serology Database.” Patient information (names, sex, birthdates/age, residence, and address), unique identification number, but not the HIV results, which was recorded on paper, were compiled and computerized to create an electronic demographic database. This database was linked to the HIV Serology Database using the unique patient numbers to create the “NARI HIV-Seropositive Database” (Fig. 1). To ensure completeness of our HIV database for Pune residents, data on HIV positive persons from Pune ICTCs were added to the “NARI HIV-Seropositive Database.” The resulting HIV database, called the “Pune HIV Database,” was formatted to construct a match compatible file for the linkage. Because individuals were registered after being diagnosed as HIV positive, their date of
sero-conversion is not known. Thus, the date of registration was considered as the date when follow-up of PLHIV started.

Cancer outcomes in PLHIV were based on being linked to a record in the PCR. PCR uses active and passive registration of cases. PCR registrars regularly visit cancer hospitals (~35) and radiotherapy centers (~9) in Pune city to identify and record new cancer cases. Cancer cases from Pune who seek treatment in Mumbai are registered by the Mumbai Cancer Registry and their data repatriated to PCR for statistical analysis. PCR is considered a high quality cancer registry based on being included in 5 volumes (IV, V, VIII, IX, and X) of the Cancer Incidence in Five Continents Monograph series\[26\] and a survey in 1999 suggested it was 80% complete.\[27\] For this analysis, we used PCR data from 1996 to 2008, which were

Figure 1. The flowchart depicts the methodology and outcomes of the computerized HIV database and Cancer Registry Match. Steps for pre-match database review, cleaning, and preparation resulting in match-compatible files are shown. Details of the 4 match passes for the computerized linkage are outlined. It ends with details of the HIV records linked to cancer by time since registration in HIV database including the early incident (EI) and late incident (LI) periods. Cancers diagnosed before or within 3 months of NARI registration were considered prevalent and were excluded from incidence analysis.
Baseline characteristics of HIV-infected persons registered with the HIV/AIDS Database in Pune, India.

| Characteristic            | HIV-infected persons, N = 32,575 | HIV-infected persons linked to cancer, N = 613 |
|--------------------------|-----------------------------------|-----------------------------------------------|
| Sex                      | Male                              | Female                                        |
|                          | 20,865 (64.1%)                    | 335 (54.6%)                                   |
|                          | Female                            |                                               |
|                          | 11,710 (35.9%)                    | 278 (45.4%)                                   |
| Age group (years)        | <14                               |                                               |
|                          | 1,298 (4%)                        | 9 (1.5%)                                      |
|                          | 14–24                             |                                               |
|                          | 5,463 (16.8%)                     | 59 (9.6%)                                     |
|                          | 24–34                             |                                               |
|                          | 14,916 (45.8%)                    | 237 (38.7%)                                   |
|                          | 34–44                             |                                               |
|                          | 7,033 (21.4%)                     | 182 (29.7%)                                   |
|                          | 44–54                             |                                               |
|                          | 2,178 (6.7%)                      | 84 (13.7%)                                    |
|                          | 55 and above                       |                                               |
|                          | 787 (2.4%)                        | 42 (6.9%)                                     |
| Marital status           | Unmarried                         |                                               |
|                          | 549 (1.7%)                        | 4 (0.7%)                                      |
|                          | Married                           |                                               |
|                          | 4,280 (13.1%)                     | 77 (12.6%)                                    |
|                          | Widowed/divorced/separated         |                                               |
|                          | 620 (1.9%)                        | 17 (2.8%)                                     |
| Calendar year            | 1991–1995                         |                                               |
|                          | 2,078 (6.4%)                      | 58 (9.5%)                                     |
|                          | 1996–2000                         |                                               |
|                          | 10,706 (32.9%)                    | 306 (49.9%)                                   |
|                          | 2001–2005                         |                                               |
|                          | 11,380 (34.9%)                    | 151 (21.4%)                                   |
|                          | 2006–2009                         |                                               |
|                          | 8,411 (25.8%)                     | 118 (16.3%)                                   |
| Total person time by calendar year |                       |                                               |
| 1991–1995                | 9,850.63 (8.25%)                  | 255.6 (17.2%)                                 |
| 1996–2000                | 50,413.77 (42.2%)                 | 1,013.8 (68.2%)                               |
| 2001–2005                | 47,351.8 (39.66%)                 | 188.2 (12.7%)                                 |
| 2006–2009                | 11,763.71 (9.85%)                 | 28.1 (1.19%)                                  |

The baseline characteristics of HIV-infected persons registered with the HIV/AIDS Database in Pune, India (1991–2008) are described along with the distribution of these characteristics among patients linked to cancer after the match.

coded using ICD-9 (1996–2005) and ICD-10 codes (2006–2008). ICD-10 codes were recoded to ICD-9, the databases were cleaned to construct match-compatible files for linkage (Fig. 1).

The HIV and cancer match-compatible files had a similar data structure to facilitate record-linkage. For example, the names in both were edited to remove spelling errors, titles, and to use common spelling. In addition, we developed a dictionary of short forms of the names, based on commonly adopted abbreviations for long names, for example, Ramakant, Ramesh, Ramu, or Ramchandra, for use in the match-compatible files. Because birthdates were often not complete or reported accurately, that is, day, month, and year, a birth-year group variable called “birthyear-5” was created with the birth year as the midpoint for 5-year birth interval.

Computerized probabilistic linkage of records in the match-compatible HIV and cancer files was done using a matching algorithm similar to that used in the US and in Uganda. Briefly, the similarity of records was calculated using harmonized values or characters of variable names: family and other names, sex, city PIN code, birth year and/or age at registration as was done in previous matches. Variables with an exact match received a perfect score. Variables with inexact matching values (or characters) were penalized so that the scores were reduced in proportion to the inexactness, until it was such that the variables were not similar at all and they made no contribution to the overall match score. To ensure the record linkage was as complete as possible, 4 separate match passes were run. To ensure efficiency of match passes, each pass was implemented within a block of variables for which the records being matched had the same exact values. These records were then compared for the non-blocking variables (Fig. 1). Records that were linked in a particular match pass were excluded from the subsequent pass. All potential matches were reviewed by NARI staff to select valid matches, based on the similarity of the original files.

2.2. Statistical analysis

Cancers were considered prevalent (present at or before diagnosis) if they were diagnosed before or within 3 months of NARI registration. Because of concerns related to surveillance bias around the time of registration, these cancers were excluded from incidence analysis (SIR). Cancers were defined as “incident” cancers if they were diagnosed 4 to 60 months after NARI registration. This 56-month period is referred to as the “combined incidence” period (CIP). Cancers in the CIP were further subdivided into early incident (EI) cancers diagnosed in the 4 to 27 month period after registration and late incident (LI) cancers diagnosed 28 to 60 months after registration. The risk of specific cancers in the PLHIV for the early, late, and combined incidence periods relative to the general population was estimated by calculating standardized incidence ratios (SIRs) for each cancer type by dividing the observed number of cancers linked to PLHIV by the expected number of cancers in the period. The expected number of cancers in PLHIV was calculated as the sum of the products of the person-time of PLHIV at risk and the age-group, sex, and period (1996–2000, 2001–2005, 2006–2008) specific cancer rate from the general population in Pune city. The person-time of PLHIV at risk of cancer was calculated from the date of registration at NARI until cancer diagnosis, date of censoring (60 months from date of registration), or December 31, 2008 (whichever occurred earliest). The significance of SIRs was assessed by calculating 95% confidence intervals (95% CIs) assuming that cancer incidence in PLHIV follows a Poisson distribution. We considered cancers in PLHIV to be statistically associated with HIV when their SIRs were significantly elevated both in the early and in the late incident periods. Cancers whose SIRs were significantly elevated in only one incident period and in the combined periods were considered to be possibly associated with HIV, while the cancers whose SIRs were not increased in any incidence period, were considered not associated with HIV. We also examined for a statistical trend in the SIRs from the early to the late incident period for the HIV-associated or possibly associated
| Cancer type | ICD-9 code | Observed cancers by time of onset, n (%) | Standardized incidence ratios (95% CI) | P (trend) |
|-------------|------------|----------------------------------------|--------------------------------------|-----------|
| **Cancer type** | Observed cancers by time of onset, n (%) | 4-27 m (E) | 28-60 m (L) | 4-27 m (E) | 28-60 m (L) | 4-27 m (E) | 28-60 m (L) | P (trend) |
| AIDS-defining cancers (ADC) | Breast | 34 (18.1) | 12 (11.3) | 37 (11.5) | 6.7 (3.5, 11.8) | 20.6 (14.5, 28.4) | 13.7 (10.1, 18.1) | <0.001 |
| | | X | X | X | X | X | X | X |
| | Non-Hodgkin lymphoma | 176.1-176.8 | 12 (6.3) | 1 (1.9) | 14 (4.3) | 1.6 (0.8, 3.7) | 18 (9.8, 30.2) | 10.6 (5.9, 17.5) | 0.003 |
| | | 180-180.9 | 22 (11.7) | 11 (10.4) | 23 (7.20) | 9.6 (4.8, 17.2) | 22.6 (14.3, 33.9) | 15.7 (11.2, 22) | 0.026 |
| | Non AIDS-defining cancers (NACD) | All Non AIDS-defining cancers | 200-200.8 | 202-202.9 | 27 m (EI) 28 | 60 m (LI) 4 | | | |
| | | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
The table shows the "observed cancers by type and ICD-9 code among PLHIV registered in Pune. It also shows the corresponding standardized incidence ratios by the early (EI: 4 – 60 months), and combined (CIP: 4 to 60 months) incidence periods along with the 95% Poisson confidence intervals in parenthesis. Bold emphasis represents significantly elevated SIR values.

Cancers associated with HIV: Significantly elevated SIRs both in the early and in the late incident periods

Cancers possibly associated with HIV: Significantly elevated SIRs in only one incident period and in the combined period.

Cancers not associated with HIV: SIRs not significantly elevated.

The last column depicts the \( P \)-value for statistical trend in the SIRs (\( P_{\text{trend}} \)) from the early to the late incident period to determine possible contribution of depth or duration of immunosuppression using a binomial distribution to compare SIRs in early and late incident periods.

The results show that the risk for any incident cancer diagnosis for PLHIV in Pune was significantly elevated as compared with the general population between 4 and 60 months after HIV registration (SIR: 11.5 [95% CI 10, 12.6]).

### 3.1. AIDS-defining cancers

Consistent with earlier reports from India,\(^ {12,29} \) Kaposi sarcoma was not observed in PCR. Cervical cancer was observed in 10.4% and 7.2% of the cancers in the early and late incident periods, respectively (Table 2, Fig. 2). The risk for cervical cancer was elevated in both periods and it increased 2.4 times from the early to the late period (SIR: 9.6 [95% CI 4.8–17.2] and 22.6 [95% CI 14.3–33.9], \( P_{\text{trend}} = 0.026 \)) (Table 2). NHL was observed in 1% and 4.4% in the early and late incident periods (Table 2). NHL risk was elevated in the late, but not early incident period (SIR: 18.0 [95% CI 9.8–30.2] vs 1.6 [95% CI 0–8.7]). The 11-fold increase in SIR for NHL from the early to the late incident period was statistically significant (\( P_{\text{trend}} = 0.003 \)) (Table 2, Fig. 2) and was observed both among males (13.7, 95% CI 6.3–26.1) and females (40.9, 95% CI 13.3–95.5), suggesting that this observation may be valid. The morphologic diagnosis of NHL was not specified in 15 cases, but was follicular lymphoma in one case. No case of Burkitt lymphoma was identified.

### 3.2. Non-AIDS-defining cancers

A wide range of NADCs (13 of 38 evaluated) showed significant elevations in risk compared with the general population in both
the early and late incidence periods (Fig. 3, Table 2). The risk was dramatically (SIR > 100) elevated for eye-orbit, substantially elevated (SIR > 20) for all-mouth, esophagus, breast, unspecified leukemia, colon, rectum and anus, and other/unspecified cancers; and moderately elevated (SIR > 10) for salivary gland, penis, nasopharynx, and brain-nervous system. We observed a small but statistically significant elevation for stomach (SIR > 5). We observed a statistical trend in the SIR from the early to the late incident periods for 5 of 13 NADCs that were statistically associated with HIV infection (colon-rectum-anus, esophagus, all mouth, breast, and unspecified cancer) (Fig. 3A), but not for the remaining 6 NADCs (lung-trachea-bronchus; nasopharynx, stomach, brain-nervous system, salivary gland, penis, unspecified leukemia, and eye-orbit) (Fig. 3B). Interestingly, the risk for female breast cancer was elevated in the early and late incident periods and also showed a statistical trend (SIR: 3.6, [95% CI 1.4–7.3] and 23.6 [95% CI 16.7–32.4] P\text{trend} < 0.001), but these results are based on 7 and 38 cases, respectively. The mean age at breast cancer diagnosis was 38.6 years in PLHIV, which was significantly lower than 54.3 years in women in the general population (P < .001). Histologically, the linked breast cancers were duct adenocarcinoma in 71% and 82% of the early and late incident cases. The risk for all-mouth cancer also showed notable elevation in the early and late incident periods and strong statistical trend between the periods (SIR: 4.3 [95% CI 1.9–8.5] and 27.0 [95% CI 19.7–36.1] P\text{trend} < 0.001), based on 8 and 45 cases, respectively.

There were 12 of 38 NADCs whose risks were significantly elevated in the late incident and combined periods (Fig. 3-C). Out of these, risk was substantially elevated for peritoneum and retroperitoneum (42.8, 95% CI 5.2–154.6) and uterus (63.5, 95% CI 20.6–148.1). The results for uterine cancer were based on 6 cases, including 3 endometrial adenocarcinoma, 2 squamous-cell type and 1 unspecified uterine cancer, probably misclassified squamous cell carcinoma of the cervix. SIRs were modestly elevated for gall bladder and extra hepatic biliary ducts (20.1, 95% CI 4.1–58.7), pancreas (14.7, 95% CI 4.0–37.7), kidney, renal pelvis, ureter, and urinary bladder (15.7, 95% CI 7.2–29.7), myeloid leukemia (17.6, 95% CI 8.0–33.4), and other non-epithelial skin (31.2, 95% CI 11.5–68.0). SIRs were slightly elevated for other pharynx (9.9, 95% CI 3.2–23.2), Hodgkin lymphoma (10.5, 95% CI 2.2–30.6), connective and soft tissue including heart (10.0, 95% CI 3.2–23.2), and liver (14.9, 95% CI 6.0–30.7). The liver cases included 5 with hepatocellular carcinoma, 1 each of adenocarcinoma, cholangiocarcinoma, and unspecified neoplasm. The results for Hodgkin lymphoma were based on 4 cases whose subtype was not specified.

Pleura/mesothelioma, peripheral nerves, vulva, placenta, and multiple myeloma were not seen in PLHIV. SIR for thyroid (14.1, 95% CI 3.8–36.1) and vaginal (49.0, 95% CI 5.9–177.0) cancers were elevated only in the early incidence period.

4. Discussion

We tested the feasibility of applying computerized probabilistic record-linkage methods to study the burden and spectrum of cancer in Indian PLHIV as a proof-of-concept. Our results demonstrate the feasibility of this approach. We also provide detailed preliminary estimates of the impact of HIV on cancers seen in a well-defined population of PLHIV in India. Our findings, based on data from one city in India, point to a substantial burden of cancer in Indian PLHIV. They suggest that PLHIV may contribute to about 1.9% of cancers in the general population and that the SIR of cancer in PLHIV is elevated about 11.5-fold compared with the general population. In contrast to findings in linkage studies conducted in Africa, which have shown ADCs as the major cancers,[18,19,21] NADCs appear to predominate in Indian PLHIV.
Our results also suggest some differences in the spectrum of cancer in Indian PLHIV compared with PLHIV from developed countries.[5,7,10,16–19] For example, KS was absent and the risk for cervical, conjunctival, and breast cancer were substantially elevated. The absence of KS in Indian PLHIV mirrors results from anecdotal and case series reports and it is consistent with the absence of KS in the general population in Pune city.

Interestingly, 15% to 25% HIV-infected males in Northern India were Kaposi sarcoma-associated herpes virus (KSHV) positive,[30] but local data for Pune are not available. NHL (1–4.4%) and cervical cancer (7–10.5%) risks were elevated in Indian PLHIV, in accord with Dhir et al.[12] Although comparing with data from other countries is risky, NHL risk in Indian PLHIV appears to be lower than the risk observed among PLHIV in the United States before widespread use of HAART,[5] and cervical cancer risk was higher than has been reported in Uganda and Nigeria.[18,19] The risks for cervical cancer are concerning given reports of high prevalence of cervical high-risk HPV and precancerous abnormalities among women living with HIV (WLHIV) in India.[31–33] At the time when the study was conducted, routine cervical cancer screening was not an ingrained standard of care for either PLHIV or women in general population. Since India has one of the highest cervical cancer burdens worldwide,[34] and about 40% of the estimated 2.1 million PLHIV in India are women,[2] our results highlight a potential cervical cancer burden in patients living longer on cART and an opportunity to intensify or strengthen cervical cancer screening for WLHIV in India.

The reasons for increased cancer risk among PLHIV may include high prevalence of oncogenic viruses in PLHIV, relatively low use of cART, chronic inflammation/immune activation, and lifestyle factors, such as smoking and alcohol consumption.[35] Our estimates may be biased because we lacked data on confounders for proper adjustment, but they provide motivation to conduct hypothesis-driven studies to delineate causal factors, as is being conducted in the United States and elsewhere.

We found that SIRs for 13 NADCs, which are linked to infectious etiology (mouth, salivary glands, nasopharyngeal carcinoma, esophagus, stomach, anus, penis, and eye or orbit), or to lifestyle risk factors, such as smoking or chewing betel nut, (lung, bronchus, and mouth), were elevated. If confirmed in larger series, these findings may support development of public health messages to stress the prevention of cancer among PLHIV using cost-effective screening programs to reduce cancer risk.
morbidity and mortality.\textsuperscript{[16,37]} Similarly, our findings that 47% of cancers occurring during the first 3 months after registration are ADC suggest that evaluation of cancer at the time of HIV registration could lead to improved clinical outcomes among PLHIV.

Our finding of dramatically elevated risk for eye cancers is based on small numbers, but it is notable because similar results have been reported from south India\textsuperscript{[18]} and in PLHIV in Africa\textsuperscript{[21,39,40]} and in the United States.\textsuperscript{[41]} Exposure to ultraviolet light is implicated, but the role of infections is controversial.\textsuperscript{[44–43]} Our finding of increased risk of breast cancer in Indian PLHIV was unexpected, but is similar to one report from Uganda\textsuperscript{[18]} and in West Africa,\textsuperscript{[19]} but different from reports in developed countries.

The strengths of our study include using a high quality population-based cancer registry and a well-defined HIV population to study the spectrum of cancers in PLHIV in India using a computerized probabilistic-record linkage. This method has become a standard in developed countries including Italy, Australia, the United States, and European countries and is being vigorously applied in several countries in Africa.

The limitations of our study include a relatively small sample size, which resulted in relatively imprecise estimates, and focusing on PLHIV from one region of India.

Surveillance bias may be a factor in our results, but stigma and poverty could limit access to cancer diagnosis and lead to under ascertainment of cancers. We believe surveillance bias might not be substantial because of less awareness of cancer complications among PLHIV in India. As a previous Indian study suggests,\textsuperscript{[12]} as many as 50% or more of people with HIV and cancer were not aware of HIV status, while about a third of all cancers linked in our study were diagnosed before HIV diagnosis.

We acknowledge the possibility of imperfect/missed linkages, undiagnosed cases, or improperly completed records, and our results should be interpreted with some caution. However, our results should be comparable to those performed in other countries without unique personal identity numbers.\textsuperscript{[18]}

In conclusion, our study highlights the feasibility of using computerized record linkage to improve the quality and accuracy of cancer registration and the potential for leveraging the wide network of free government-funded antiretroviral treatment centers and integrate systematic cancer screening among PLHIV in India.

Acknowledgments

The authors wish to thank the Maharashtra State AIDS Control Society (MSACS) and the National AIDS Control Organization (NACO) for sharing HIV testing records for the match. We also wish to acknowledge the support of Dr A. Nandakumar, Director-in-Charge, National Centre for Disease Informatics and Research (ICMR) and National Cancer Registry Programme. We acknowledge India’s National Cancer Registry Programme, and the in-charge, registrar, and social workers of the Pune cancer registry for the excellent and reliable registration which was valuable for this registry match study and the contributions of the late Dr B.B. Yeole, Mumbai Cancer Registry, and late Dr A. Kurkure, Indian Cancer Society, to the PCR. Additionally, we wish to thank all the clinical and laboratory staff at NARI for their efforts in HIV diagnosis and registration, Mrs. Radhika Brahme, Mr. Rajendra Yelegate, and Chitra Kadu for participating as independent manual reviewers during the match and other departmental staff for helping create the list of names to match.

References

\textsuperscript{[1]} NACO. Technical Report on HIV Estimation. 2006

\textsuperscript{[2]} NACO. India HIV Estimations 2015. 2015

\textsuperscript{[3]} NACO. Annual Report 2014-15. 2015

\textsuperscript{[4]} Deeks SG, Lewin SR, Havlir DV. The end of AIDS: HIV infection as a chronic disease. Lancet 2013;382:1523–53.

\textsuperscript{[5]} Engels EA, Pfeiffer RM, Goedert JJ, et al. Trends in cancer risk among people with AIDS in the United States 1980-2002. AIDS 2006;20:1645–54.

\textsuperscript{[6]} Fisch M, Biggar RJ, Engels EA, et al. Association of cancer with AIDS-related immunosuppression in adults. JAMA 2001;285:1736–45.

\textsuperscript{[7]} Goedert JJ, Cote TR, Virgo P, et al. Spectrum of AIDS-associated malignant disorders. Lancet 1998;351:1833–9.

\textsuperscript{[8]} Grulich AE, Cancer: the effects of HIV and antiretroviral therapy, and implications for early antiretroviral therapy initiation. Curr Opin HIV AIDS 2009;4:183–7.

\textsuperscript{[9]} Biggar RJ, Chaturvedi AK, Bhatia K, et al. Cancer risk in persons with HIV/AIDS in India: a review and future directions for research. Infect Agent Cancer 2009;4:8.

\textsuperscript{[10]} Shuls MS, Pfeiffer RM, Hall HI, et al. Proportions of Kaposi sarcoma, selected non-Hodgkin lymphomas, and cervical cancer in the United States occurring in persons with AIDS, 1980-2007. JAMA 2011;305:1450–9.

\textsuperscript{[11]} Colette J, Smith LR, Rainer Weber, et al. D:A:D Study GroupTrends in underlying causes of death in people with HIV from 1999 to 2011 (D:A: D): a multicohort collaboration. Lancet 2014;384.

\textsuperscript{[12]} Dhir AA, Sawant S, Dikshit RP, et al. Spectrum of HIV/AIDS related cancers in India. Cancer Causes Control 2008;19:147–53.

\textsuperscript{[13]} Dhir AA, Sawant SP. Malignancies in HIV: the Indian scenario. Curr Opin Oncol 2008;20:517–21.

\textsuperscript{[14]} Joshi U, Cetna DE, Ongole R, et al. AIDS related Kaposi’s sarcoma presenting with palatal and eyelid nodule. J Assoc Physicians India 2012;60:30–3.

\textsuperscript{[15]} Dhir AA. HIV-associated non-Hodgkin’s lymphoma: how much do we know? Indian J Cancer 2010;47:6–7.

\textsuperscript{[16]} Bolesel J, Franceschi S, Solberg R, et al. Cancer incidence in people with AIDS in India. Int J Cancer 2010;127:1437–45.

\textsuperscript{[17]} Grulich AE, Li Y, McDonald A, et al. Rates of non-AIDS-defining cancers in people with HIV infection before and after AIDS diagnosis. AIDS 2002;16:1155–61.

\textsuperscript{[18]} Mubaleetey SM, Katabira ET, Wablinga H, et al. Spectrum of cancers among HIV-infected persons in Africa: the Uganda AIDS-Cancer Registry Match Study. Int J Cancer 2006;118:985–90.

\textsuperscript{[19]} Akaroloo-Anthony SN, Maso LD, Ighinoba F, et al. Cancer burden among HIV-positive persons in Nigeria: preliminary findings from the Nigerian AIDS-cancer match study. Infect Agent Cancer 2014;9:1.

\textsuperscript{[20]} Bohlus F MN, Spoori A, Wainwright R, et al. for IeDEA-Southern AfricaIncidence of AIDS-defining and other cancers in HIV-positive children in South Africa: record linkage study. Pediatr Infect Dis J 2016;35:e164–70.

\textsuperscript{[21]} Sengayi M, Spoori A, Egger M, et al. Record linkage to correct under ascertainment of cancers in HIV cohorts: the Sinikithemba HIV clinic linkage project. Int J Cancer 2016;139:1209–16.

\textsuperscript{[22]} 2011 RGoCol. Pune City Census 2011 data. 2011. http://www.census2011.co.in/census/city/375/pune.html. Accessed December 23, 2015.

\textsuperscript{[23]} India Health Action TrustHIV/AIDS situation and response in Pune district: epidemiological appraisal using data triangulation. 2011; Bangalore, India:India Health Action Trust, http://www.that.in/Annual%20report%20IHA%20D%20Triangulation/Maharashtra%20all%20District%20Inner%20Pages/Pune%20report%20Final%202011.pdf. Accessed December 23, 2015.

\textsuperscript{[24]} Godbole SV, Sane S, Mehandale SM. Use of locally generated data for estimating HIV disease burden in Pune district in Maharashtra, a high prevalent state of India. XV International AIDS Conference Bangkok Thailand 2004. Abstract no: MoPeC3640, 2004. http://www.abstractarchive.org/.

\textsuperscript{[25]} National AIDS Control OrganisationDistrict HIV/AIDS epidemiological profiles developed through data triangulation, fact sheets. Maharashtra. 2014-New Delhi:National AIDS Control Organisation, http://www.naco.gov.in/upload/NACP%20-%20HIVFACT%20HE%20NACP%20%20LAUNCH%20Maharashtra%20DEP.pdf. Accessed December 23, 2015.

\textsuperscript{[26]} International Agency for Research on Cancer (IARC) CIS-I-X - Cancer Incidence in 5 continents Volumes I-X: Cancer Registry List. http://c5.iarc.fr/CISI-X/Pages/database.aspx. Accessed May 7, 2015.
[27] Nandakumar A, Gupta PC, Gangadharan P, et al. Geographic pathology revisited: development of an atlas of cancer in India. Int J Cancer 2005;116:740–54.

[28] Jaro MA. Probabilistic linkage of large public health data files. Stat Med 1995;14:491–8.

[29] Sharma SK, Soneja M, Ranjan S. Malignancies in human immunodeficiency virus infected patients in India: Initial experience in the HAART era. Indian J Med Res 2015;142:563–7.

[30] Munawwar A, Sharma SK, Gupta S, et al. Seroprevalence and determinants of Kaposi sarcoma-associated human herpesvirus 8 in Indian HIV-infected males. AIDS Res Hum Retroviruses 2014;30:1192–6.

[31] Mane A, Nirmalkar A, Rishud AR, et al. HPV genotype distribution in cervical intraepithelial neoplasia among HIV-infected women in Pune, India. PLoS One 2012;7:e38731.

[32] Sahasrabuddhe VV, Bhosale RA, Joshi SN, et al. Prevalence and predictors of colloproscopic-biopathologically confirmed cervical intraepithelial neoplasia in HIV-infected women in India. PLoS One 2010;5:e8634.

[33] Godbole SV, Mane AK, Chidrawar SK, et al. Prevalence of anal human papillomavirus infection among HIV-infected women from India. J Acquir Immune Defic Syndr 2014;67:e111–14.

[34] Forman D, de Martel C, Lacey CJ, et al. Global burden of human papillomavirus and related diseases. Vaccine 2012;30(Suppl 5):F12–23.

[35] Grulich AE, van Leeuwen MT, Falster MO, et al. Incidence of cancers in people with HIV/AIDS compared with immunosuppressed transplant recipients: a meta-analysis. Lancet 2007;370:59–67.

[36] Sankaranarayanan R, Ramadas K, Thomas G, et al. Effect of screening on oral cancer mortality in Kerala, India: a cluster-randomised controlled trial. Lancet 2005;365:1927–33.

[37] Sankaranarayanan R, Ramadas K, Thara S, et al. Clinical breast examination: preliminary results from a cluster randomized controlled trial in India. J Natl Cancer Inst 2011;103:1476–80.

[38] Pradeep TC, Gangasagara SK, Subbaramaiah GB, et al. Prevalence of undiagnosed HIV infection in patients with ocular surface squamous neoplasia in a tertiary center in Karnataka, South India. Cornea 2012;31:1282–4.

[39] Waddell KM, Lewallen S, Lucas SB, et al. Carcinoma of the conjunctiva and HIV infection in Uganda and Malawi. Br J Ophthalmol 1996;80:503–8.

[40] Newton R, Ziegler J, Ateenyi-Agaba C, et al. The epidemiology of conjunctival squamous cell carcinoma in Uganda. Br J Cancer 2002;87:301–8.

[41] Guech-Ongey M, Engels EA, Goedert JJ, et al. Elevated risk for squamous cell carcinoma of the conjunctiva among adults with AIDS in the United States. Int J Cancer 2008;122:2590–3.

[42] Ateenyi-Agaba C. Conjunctival squamous-cell carcinoma associated with HIV infection in Kampala, Uganda. Lancet 1995;345:695–6.

[43] Ateenyi-Agaba C, Weiderpass E, Tommasino M, et al. Papillomavirus infection in the conjunctiva of individuals with and without AIDS: an autopsy series from Uganda. Cancer Lett 2006;239:98–102.